



ФОТОНИКА И ОПТОИНФОРМАТИКА

Извлечение материальных параметров плазмонного мультислоя из коэффициентов отражения и прохождения
Плазмонные солитоны, кинки и волны Фарадея в двумерной решетке металлических наночастиц
Исследование характеристик сигналов спектральной интерференции в ближней ИК области спектра
Логика с исключением на алгебре Фурье-дуальных операций: нейросетевой механизм редуцирования когнитивного диссонанса

Орлов А.А., Янковская Е.А. 1
Жуковский С.В., Белов П.А.
Носков Р.Е., Смирнова Д.А., Лапина Н.С. 5
Гуров И.П. 11
Павлов А.В. 17

ОПТИЧЕСКИЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Метод исследования зависимости h -параметра анизотропного световода от радиуса изгиба
Технологические методы снижения уровня оптических потерь в микроструктурированных волоконных световодах
Многослойное просветляющее покрытие на подложке из оптического сульфида цинка
Контроль структуры различных видов бумаги методом атомно-силовой микроскопии
Энергосберегающая технология расплавления химических веществ световым излучением
О технологических несовершенствах геометрических параметров силового стержня для заготовки оптического волокна PANDA

Шрамко О.А., Рупасов А.В., 26
Новиков Р.Л., Аксарин С.М.
Коробейников А.Г., Гатчин Ю.А., 31
Дукельский К.В., Тер-Нерсесянц Е.В.
Тан Тай До, Губанова Л.А. 38
Жуков М.В. 44
Черепанов А.Н. 50
Семенов Н.В., Труфанов Н.А., 54
Адамов А.А.

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Верификация параллельных автоматных программ
Анализ данных на основе платформы SQL-MAPREDUCE
О влиянии адаптивных пользовательских интерфейсов на надежность и эффективность функционирования автоматизированных систем
Метод отображения задач на крупногранулярные реконфигурируемые вычислительные системы
Оценка применимости модели IRI-212 для автоматизированной обработки ионограмм вертикального зондирования
Оценка точности визуализации местоположения объекта в геоинформационных системах и системах индикации навигационных комплексов пилотируемых летательных аппаратов
Применение методов машинного обучения для обнаружения бактерий в продуктах питания
Представление документов в задаче кластеризации аннотаций научных текстов
Использование контейнера BC7 для хранения текстур с глубиной цвета 10 бит
Предшествующая и последующая фильтрация шумов в алгоритмах восстановления изображений
Оценка возможности экранной репродукции насыщенных пигментов
Эксплуатационные характеристики риска нарушений безопасности информационной системы
Идентификация анонимных пользователей Интернет- порталов на основании технических и лингвистических характеристик пользователя

Лукин М.А. 60
Дергачев А.А. 66
Фуртат Ю.О. 71
Румянцев А.С. 76
Николаева В.Д., Рыбаков М.В., Котиков А.Л., 82
Кошелевский В.К.
Костишин М.О., Жаринов И.О., Жаринов О.О., 87
Нечаев В.А., Сулов В.Д.
Саенко А.П., Мусалимов В.М., Лерм Ш., Линц Г. 93
Попова С.В., Данилова В.В. 99
Перминов И.В., Палташев Т.Т. 107
Сизиков В.С., Экземляров Р.А. 112
Черевань Л.В., Тозик В.Т. 123
Щеглов К.А., Щеглов А.Ю. 129
Воробьева А.А., Гвоздев А.В. 139

ТЕХНИЧЕСКИЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Методика расчета коэффициентов облученности цилиндрического космического объекта подсветкой Земли
Автономизация нелинейных динамических систем
Современное состояние и перспективы развития основных понятий в области мехатроники
Опыт применения и перспективы технологии алмазного микроточения
Особенности формирования прямых зубьев цилиндрических колес ступенчатым долбяком

Дзитолев А.М., Ханков С.И. 145
Иванов С.Е., Мельников Г.И. 151
Шалобаев Е.В., Толочка Р.-Т. 156
Медунецкий В.М., Солк С.В. 165
Расулов Н.М., Гусейнов Г.Р., Надиров У.М. 170

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Проблема учета зависимости коэффициента объемной теплоемкости от температуры при моделировании лазерно-дуговой наплавки
Методы моделирования температурного поля при бесконтактной лазерной деформации пластины
Численное моделирование турбулентного потока воздуха с использованием метода отсоединенных вихрей
Стратегия маркетинга в системе высокочастотной алгоритмической торговли

Лукьяненко С.А., Третьяк В.А. 177
Лукьяненко С.А., Михайлова И.Ю. 182
Воронин А.А., Лукьянов Г.Н., Фролов Е.В. 187
Торопов А.В. 193

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

От традиционного дистанционного обучения к массовым открытым онлайн-курсам

Васильев В.Н., Стафеев С.К., Лисицына Л.С., 199
Ольшевская А.В.

КРАТКИЕ СООБЩЕНИЯ

Обратимая фотодеструкция наночастиц серебра в фото-термо-рефрактивных стеклах
Уравнения переноса излучения в инфракрасной томографии в случае активно-пассивной диагностики и веерного сканирования
Получение спектров электрофизиологических сигналов в режиме реального времени

Игнатьев Д.А., Игнатьев А.И., 206
Никонов Н.В., Стародубов Д.С.
Макарова А.А. 207
Толкович Д.В., Андрианова Е.С. 209

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

МЕЖДУНАРОДНЫЙ РЕДАКЦИОННЫЙ СОВЕТ

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 Slood, 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. Бобцов¹, д.т.н., профессор

А.В. Бухановский¹, д.т.н.

Ю.А. Гатчин¹, д.т.н., профессор

Н.Ф. Гусарова¹, к.т.н., с.н.с. (редактор)

Н.С. Кармановский¹, к.т.н., доцент

(заместитель главного редактора)

Ю.Л. Колесников¹, д. ф.-м.н., профессор

С.А. Козлов¹, д.ф.-м.н., профессор

А.Г. Коробейников², д.т.н., профессор

Л.С. Лисицына¹, д.т.н., доцент

В.Г. Мельников¹, д.т.н., доцент

Н.В. Никоноров¹, д.т.н., профессор

А.А. Ожиганов¹, д.т.н., профессор

Н.Н. Розанов³, д.ф.-м.н., профессор, чл.-корр. РАН

И.Г. Сидоркина⁴, д.т.н., профессор

О.А. Степанов⁶, д.т.н., профессор

А.А. Шальто¹, д.т.н., профессор

Ю.Г. Якушенков⁵, д.т.н., профессор

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

¹Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (НИУ ИТМО), Санкт-Петербург, Россия

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

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

⁴Поволжский государственный технологический университет, Йошкар-Ола, Россия

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

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

Журнал издается Санкт-Петербургским национальным исследовательским университетом информационных технологий, механики и оптики 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

Англоязычное название: "Scientific and technical journal of information technologies, mechanics and optics".

Транслитерированное название: "Nauchno-tehnicheskiy vestnik informacionnyh tehnologij, mehaniki i optiki".

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

Подписано к печати 15.12.2013 Тираж 350 экз. Заказ № 37(89)

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

РЕДАКЦИОННАЯ ПОЛИТИКА

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 по направлениям:

- фотоника и оптоинформатика (fotonics and optoinformatics),
- оптические системы и технологии (optical engineering),
- автоматическое управление и робототехника (automatic control and robotics),
- новые материалы и нанотехнологии (material science and nanotechnologies),
- компьютерные системы, информационные технологии (computer science),
- технические системы и технологии (engineering and technologies),
- математическое и компьютерное моделирование (mathematical and computer simulation),
- информационные технологии в образовании (information technologies in education).

Представленные для публикации авторские материалы должны соответствовать следующим требованиям: актуальность темы, научная новизна исследования, оригинальность исследовательской проблемы и примененных методов исследования, высокая научная и практическая значимость полученных результатов.

Редакция в своей текущей деятельности руководствуется принципами журнальной этики (подробно см. <http://ntv.ifmo.ru/>). Редакция самостоятельно принимает решение о порядке и сроках опубликования статей, исходя из их научной значимости, тематики каждого выпуска, качества предварительной подготовки рукописей.

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

Редакция вправе отказать в публикации статьи на основании результатов рецензирования, несоответствия материала статьи тематике журнала, а также при неудовлетворительном качестве оформления представленных материалов.

Журнал публикует как полнотекстовые статьи, так и краткие сообщения.

Полнотекстовая статья должна иметь четкую структуру, включающую в себя аннотацию, ключевые слова, введение, несколько содержательных разделов и заключение.

В аннотации, рассчитанной на самый широкий круг читателей, необходимо в объеме **150-250** слов в форме краткого реферата изложить научное содержание статьи: предмет, цель работы, метод или методология проведения работы, краткое описание эксперимента, полученные результаты, рекомендации по их применению.

Во введении необходимо представить содержательную постановку рассматриваемого вопроса, провести краткий анализ известных из научной литературы решений (со ссылками на источники), дать критику их недостатков, показать научную новизну и преимущество (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 должен быть разбит на содержательные разделы.

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

Максимальный объем полнотекстовой статьи – 8 страниц машинописного текста (с рисунками и таблицами), шрифт 12 pt, один интервал.

Статья в форме краткого сообщения предназначена для того, чтобы в сжатые сроки и кратко информировать научное сообщество о новых научных результатах, полученных автором. Объем краткого сообщения – до 3 страниц, шрифт 12 pt, один интервал. Объем аннотации 50-100 слов. Рубрикация текста не требуется.

Пристатейный список литературы должен составлять, как правило, для обзорной статьи – не менее 50, для полнотекстовой статьи – не менее 15, для краткого сообщения - не менее 8 литературных источников (не менее половины из них должны составлять ссылки на издания, включенные в международные базы цитирования Scopus и Web of Science).



FOTONICS AND OPTOINFORMATICS

Extraction of material parameters for plasmon multilayer from reflection and transmission coefficients	<i>A. Orlov, E. Yankovskaya, S. Zhukovsky, P. Belov</i>	1
Plasmon solitons, kinks and Faraday waves in two-dimensional lattice of metal nanoparticles	<i>R.E. Noskov, D.A. Smirnova, N.S. Lapshina</i>	5
Study of characteristics of spectral interference signals in near infrared spectral range	<i>I. Gurov</i>	11
Logic with exception on the algebra of Fourier-dual operations: neural net mechanism of cognitive dissonance reducing	<i>A. Pavlov</i>	17

OPTICAL ENGINEERING

Analysis method of anisotropic lightguide h -parameter dependence on its bending radius	<i>O. Shramko, A. Rupasov, R. Novikov, S. Aksarin</i>	26
Process methods with low level of optical losses for the microstructured fiber light guides	<i>A. Korobeynikov, Yu. Gatchin, K. Dukel'skiy, E. Ter-Nersesyantz</i>	31
Multi-zone antireflection coating on a substrate made of optical zinc sulphide	<i>Tan Tai Do, L. Gubanova</i>	38
Structure control for different types of paper by atomic force microscopy	<i>M. Zhukov</i>	44
Energy-saving technology of chemical agents melting by light radiation	<i>A. Cherepanov</i>	50
Technological imperfections of force rod geometrical parameters for PANDA optical fibers production	<i>N. Semenov, N. Trufanov, A. Adamov</i>	54

COMPUTER SCIENCE

Verification of parallel automata-based programs	<i>M. Lukin</i>	60
Data analysis by SQL-MAPREDUCE platform	<i>A. Dergachev</i>	66
On the effect of adaptive user interfaces on reliability and efficiency of the automated systems	<i>Yu. Furtat</i>	71
Tasks mapping method for coarse grain reconfigurable systems	<i>A. Rumyantsev</i>	76
IRI-2012 model adaptability estimation for automated processing of vertical sounding ionograms	<i>V. Nikolaeva, M. Rybakov, A. Kotikov, V. Koshelevsky</i>	82
Accuracy evaluation of the object location visualization for geo-information and display systems of manned aircrafts navigation complexes	<i>M. Kostishin, I. Zharinov, O. Zharinov, V. Suslov, V. Nechaev</i>	87
Detection of bacteria in foodstuff by machine learning methods	<i>A. Saenko, V. Musalimov, S. Lerm, G. Linss</i>	93
Document representation for clustering of scientific abstracts	<i>S. Popova, V. Danilova</i>	99
Usage of BC7 container for storing textures with 10-bit color depth	<i>I. Perminov, T. Paltashev</i>	107
Preliminary and subsequent filtering of noise in image restoration algorithms	<i>V. Sizikov, R. Ekzemplyarov</i>	112
Reliability estimation for screen reproduction of saturated pigments	<i>L. Cherevan, V. Tozik</i>	123
Operational characteristics of information system security threats risk	<i>K. Shcheglov, A. Shcheglov</i>	129
Anonymous website user identification based on combined feature set (writing style and technical features)	<i>A. Vorob'yeva, A. Gvozdev</i>	139

ENGINEERING AND TECHNOLOGIES

Calculation methods for irradiance coefficients of cylindrical space object by the Earth radiation	<i>A. Dzitoev, S. Khankov</i>	145
Off-line interaction of the nonlinear dynamic systems	<i>S. Ivanov, G. Melnikov</i>	151
Modern state and development prospects of the basic concepts in the field of mechatronics	<i>E. Shalobaev, R.-T. Tolochka</i>	156
Application experience and prospects of diamond micro-turning technology	<i>V. Medunetskiy, S. Solk</i>	165
Straight cogs formation features for cylindrical spur gears by stepped gear-shaped cutter	<i>N. Rasulov, G. Guseynov, U. Nadirov</i>	170

MATHEMATICAL AND COMPUTER SIMULATION

Temperature dependence consideration issue for coefficient of volumetric heat capacity in simulation of laser-arc pad weld process	<i>S. Luk'yanenko, V. Tret'yak</i>	177
Methods of temperature field modeling for contactless laser deformation of a plate	<i>S. Luk'yanenko, I. Mykhailova</i>	182
Detached-eddy simulation of turbulent airflow	<i>A. Voronin, G. Luk'yanov, E. Frolov</i>	187
Market-making strategy in the system of algorithmic high-frequency trading	<i>A. Toropov</i>	193

INFORMATION TECHNOLOGIES IN EDUCATION

From traditional distance learning to mass online open courses	<i>V. Vasiliev, S. Stafeev, L. Lisitsyna, A. Ol'shevskaya</i>	199
--	---	-----

BRIEF PAPERS

Reversible photo destruction of silver nanoparticles in photo-thermo-refractive glass	<i>D. Ignatiev, A. Ignatiev, N. Nikonorov, D. Starodubov</i>	206
Equations of radiation transfer in infrared tomography in the case of active-passive diagnosis and sweeping scanning	<i>A. Makarova</i>	207
Real time registration of the electrophysiological signals spectra	<i>D. Tolkovich, E. Andrianova</i>	209

УДК 537.6, 535.58, 538.958

ИЗВЛЕЧЕНИЕ МАТЕРИАЛЬНЫХ ПАРАМЕТРОВ ПЛАЗМОННОГО
МУЛЬТИСЛОЯ ИЗ КОЭФФИЦИЕНТОВ ОТРАЖЕНИЯ И ПРОХОЖДЕНИЯА.А. Орлов^а, Е.А. Янковская^а, С.В. Жуковский^б, П.А. Белов^а^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Alexey.orlov@phoi.ifmo.ru^б Технический университет Дании, Конгенс Люнгбю, Дания, sezh@fotonik.dtu.dk

Получены и исследованы диэлектрическая и магнитная проницаемости конечного образца многослойной металлodielectricкой наноструктуры – плазмонного мультислоя, относящегося к классу электромагнитных метаматериалов. Метаматериалы представляют собой искусственно созданные структуры, как правило, периодические, размеры элементарной ячейки которых много меньше длины волны, обладающие необычными электромагнитными свойствами, не наблюдаемыми в природе. Так, метаматериалы открывают путь к созданию оптических материалов с магнитной проницаемостью, существенно отличной от единицы, что долгое время казалось трудновыполнимой задачей. Для извлечения из коэффициентов отражения и прохождения материальных параметров, описывающих электромагнитное поведение плазмонного мультислоя, был применен классический метод Николсона–Росса–Веира. В рассматриваемом слое метаматериала наблюдается сильный магнетизм резонансного типа в оптическом диапазоне частот. Положение резонанса соответствует точке предельно малых значений диэлектрической проницаемости. Показана возможность перестройки резонансной частоты путем изменения соотношения толщины слоев в элементарной ячейке, образующей плазмонный мультислой. Наблюдаемая магнитная активность входит в установившийся режим, начиная с толщины мультислоя, составляющей порядка нескольких десятков слоев. Предложено использование таких многослойных наноструктур как простого и эффективного материала, способного демонстрировать магнитную активность во всем оптическом диапазоне.

Ключевые слова: метаматериалы, плазмоны, магнетизм, многослойные структуры.EXTRACTION OF MATERIAL PARAMETERS FOR PLASMON MULTILAYER
FROM REFLECTION AND TRANSMISSION COEFFICIENTSA. Orlov^с, E. Yankovskaya^с, S. Zhukovsky^д, P. Belov^с^с Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Alexey.orlov@phoi.ifmo.ru^д Technical University of Denmark, Kongens Lyngby, Denmark, sezh@fotonik.dtu.dk

The paper deals with acquisition and analysis of permittivity and permeability for a finite sample made of multi-layered metal-dielectric nanostructure – plasmon multilayer referred to a class of electromagnetic metamaterials. Metamaterials are artificial structures, periodical as a rule, with characteristic unit cell sizes much smaller than the wavelength in vacuum, having unusual properties not met in nature. For example, metamaterials open the way to fabrication of optical materials with permeability substantially differing from unity - the task considered as unrealizable for a long time. The classical Nicolson-Ross-Weir method has been applied for extraction of material parameters describing an electromagnetic behavior of the plasmon multilayer from reflection and transmission coefficients. Strong resonance-type magnetic activity in the optical frequency domain is observed in the metamaterial under consideration. Magnetism appears due to strong spatial dispersion inherent to the plasmon multilayers. Position of the permeability resonance is located exactly in the epsilon-near-zero region. It is shown how the resonance can be repositioned by means of the filling factor changing. Observed magnetic activity reaches the steady state with multilayer thickness equal to a few dozens of layers. Plasmon multilayers are suggested as robust and effective optical materials with a strong magnetic response in the whole optical domain.

Keywords: metamaterials, plasmons, magnetism, multilayered structures.

Введение

Многослойные структуры в оптике являются предметом рассмотрения уже достаточно длительное время [1–3]. Однако периодические многослойные структуры, включающие в себя слои металла, стали рассматриваться значительно позже. Неожиданно оказалось, что металлodielectricкие многослойные наноструктуры демонстрируют высокий коэффициент прохождения в оптическом диапазоне длин волн даже в режиме фотонного кристалла, где характерный размер периода структуры D составляет около четверти длины волны в вакууме λ [4, 5]. Данное явление, именуемое эффектом прозрачности металла, возникает вследствие того, что в периодических металлodielectricких многослойных наноструктурах на границах между слоями могут возбуждаться поверхностные плазмон-поляритоны (ППП), обеспечивая механизм туннелирования оптического излучения через такую структуру [6–8]. Вследствие присутствия ППП на границах между слоями данные структуры будут именоваться в дальнейшем плазмонными мультислоями.

Извлечение ϵ и μ

В настоящей работе рассматриваются плазмонные мультислои, образованные периодически повторяющейся элементарной ячейкой, состоящей из диэлектрика толщиной d_1 и металла толщиной d_2 . Диэлектрическая проницаемость диэлектрика ϵ_1 принимается постоянной и равной 4,6. Для металла диэлектрическая проницаемость задается моделью Друде:

$$\epsilon_2 = 1 - \omega_p / \omega (\omega + i\Gamma),$$

где плазменная частота $\omega_p = 2\pi c / \lambda_p$, $\lambda_p = 250$ нм (c – скорость света в вакууме), коэффициент затухания $\Gamma = 1,734 \text{ c}^{-1}$, что соответствует серебру с учетом омических потерь. Параметры взяты из работы [9].

Насколько нам известно, магнитная активность плазмонных мультислоев, составленных из немагнитных материалов, не была исследована до настоящего времени. Далее нами будут изучены диэлектрическая и магнитная проницаемости конечного образца мультислоя, составленного из N элементарных ячеек, освещаемого ТМ-поляризованным излучением нормально к слоям. Падение по нормали выбрано с целью извлечения продольной компоненты тензора ϵ диэлектрической проницаемости плазмонного мультислоя ввиду того, что в данном случае электромагнитное излучение нечувствительно к имеющимся в ϵ недиагональным компонентам [10, 11].

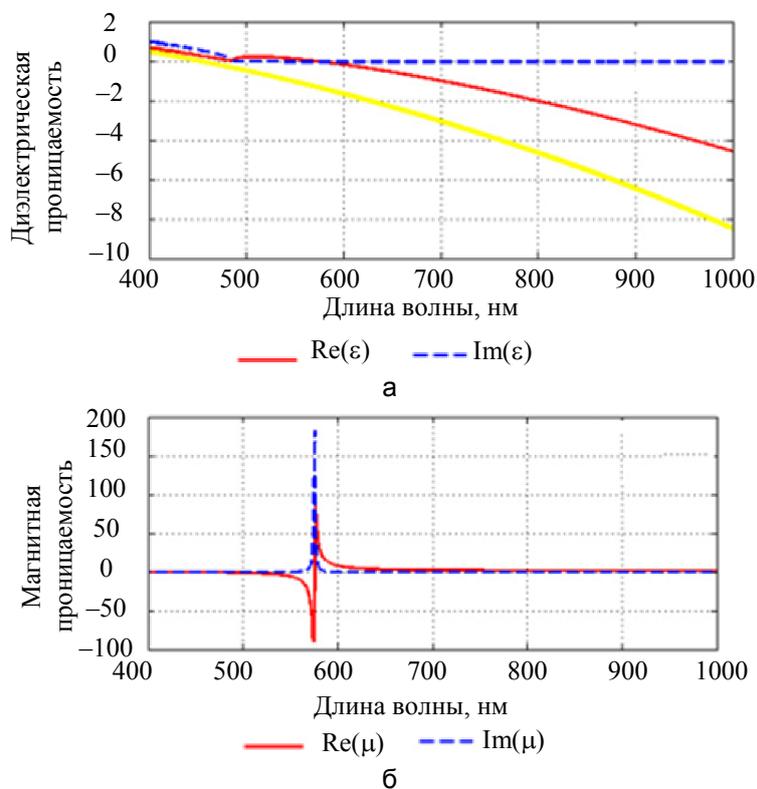


Рис. 1. Эффективная диэлектрическая (а) и магнитная (б) проницаемости (вещественные – Re и мнимые – Im части) образца плазмонной многослойной структуры, состоящей из $N = 200$ периодов, образованных слоем диэлектрика и металла с толщинами 30 и 60 нм соответственно. Желтой линией показана аналитически рассчитанная диэлектрическая проницаемость для бесконечного мультислоя

Для получения эффективных диэлектрической и магнитной проницаемостей из конечного образца плазмонного мультислоя был применен метод экстракции материальных параметров из рассчитанных при помощи матриц передачи коэффициентов отражения R и прохождения T конечного образца мультислоя согласно процедуре Николсона–Росса–Веира [12, 13]:

$$Z = \sqrt{\frac{\mu}{\epsilon}} = \pm \sqrt{\frac{(1 + R)^2 - T^2}{(1 - R)^2 - T^2}}, \quad n = \sqrt{\epsilon\mu} = \pm \frac{\arccos \left[\frac{1}{2T} (1 - R^2 + T^2) \right] + 2\pi m}{k_0 D N},$$

где $D = d_1 + d_2$, N – число периодов в конечном мультислое. Правильный выбор целого числа m был подробно обсужден в [14, 15]. Рассчитав характеристический импеданс Z , а также показатель преломления n , нетрудно установить материальные параметры ϵ и μ образца. Полученные диэлектрическая и магнитная проницаемости для мультислоя из 200 элементарных ячеек диэлектрика и металла с толщинами $d_1 = 30$ нм, $d_2 = 60$ нм соответственно представлены на рис. 1. Число периодов выбрано большим с целью создания объемной структуры во избежание проблем плохой сходимости ϵ и μ при малом числе элементарных ячеек, составляющих наноструктуру. Как видно из рис. 1, плазмонный мультислой в такой конфи-

гурации демонстрирует сильную магнитную активность в оптическом диапазоне длин волн в точке около нулевого ϵ . Обычно такое поведение обуславливается сильной пространственной дисперсией плазмонного мультислоя [16–18]. Полученная магнитная активность является устойчивой к изменению числа периодов и сходится к стабильным значениям уже при около 50 периодах, что видно из рис. 2. Также сходимость магнитной проницаемости, очевидно, зависит от толщины периода D : чем больше толщина, тем быстрее будет наблюдаться сходимость.

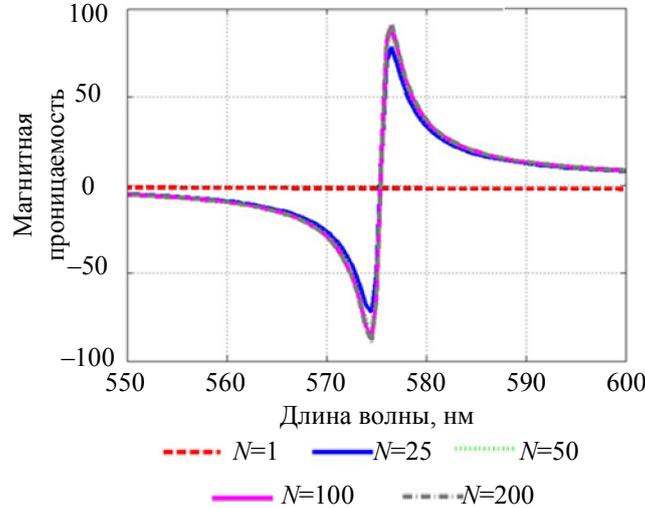


Рис. 2. Сходимость магнитной проницаемости конечного плазмонного мультислоя при увеличении числа N составляющих его элементарных ячеек

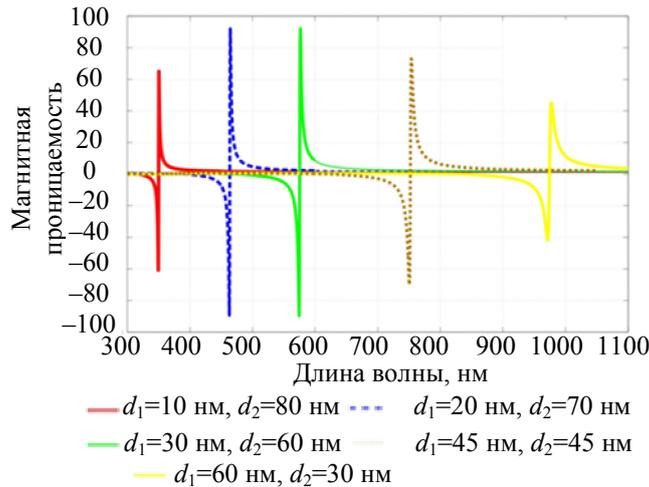


Рис. 3. Изменение положения резонанса магнитной проницаемости при изменении соотношения толщин слоев d_1/d_2 . Толщина периода фиксирована. Для всех кривых количество периодов равно 200

Положение резонансной частоты μ зависит от соотношения толщин металла и диэлектрика d_1/d_2 (рис. 3). При определенном соотношении толщин слоев наблюдается максимум раскачки резонанса. В этом максимуме толщина металла в 2–3 раза больше, чем толщина диэлектрика (точное значение зависит от толщины периода). Более сильное утолщение слоя металла приводит к уменьшению раскачки резонанса.

Заключение

Была продемонстрирована сильная магнитная активность плазмонного мультислоя во всем видимом диапазоне длин волн. Резонанс извлеченной магнитной проницаемости соответствует частотам предельно малых значений диэлектрической проницаемости. Положение резонанса зависит от соотношения толщин в периоде, обеспечивая перестраиваемость метаматериала [19], и сходится к устойчивым значениям уже от нескольких десятков слоев. В настоящее время предложенный способ получения магнитной активности в видимом диапазоне длин волн является самым простым и доступным для изготовления. К примеру, по сравнению с двумерными массивами последовательно расположенных пар наночастиц [20], где магнитный отклик наблюдается вплоть до голубой линии спектра, плазмонные мультислои не требуют применения глубокой ультрафиолетовой интерференционной литографии и являются объемным материалом, а не всего лишь тонким слоем упорядоченных наночастиц.

Литература

1. Рытов С.М. Электромагнитные свойства мелкослоистой среды // ЖЭТФ. 1955. Т. 29. № 5. С. 605–616.
2. Brekhovskikh L. Waves in Layered Media. NY: Academic Press, 1960. 574 p.
3. Борн М., Вольф Э. Основы оптики. М.: Наука, 1973. 721 с.
4. Scalora M., Bloemer M.J., Manka A.S., Pethel S.D., Dowling J.P., Bowden C.M. Transparent, metallo-dielectric one dimensional photonic band gap structures // J. Appl. Phys. 1998. V. 83. N 4. P. 2377–2383.
5. de Ceglia D., Vincenti M.A., Cappeddu M.G., Centini M., Akozbek N., D'Orazio A., Haus J.W., Bloemer M.J., Scalora M. Tailoring metallodielectric structures for superresolution and superguiding applications in the visible and near-IR ranges // Phys. Rev. A. 2008. V. 77. N 3. P. 033848-1–033848-12.
6. Tomita S., Yokoyama T., Yanagi H., Wood B., Pendry J.B., Fujii M., Hayashi S. Resonant photon tunneling via surface plasmon polaritons through one-dimensional metal-dielectric metamaterials // Optics Express. 2008. V. 16. N 13. P. 9942–9950.
7. Allen T.W., DeCorby R.G. Assessing the maximum transmittance of eriodic metal-dielectric multilayers // J. Opt. Soc. Am. B. 2011. V. 28. N 10. P. 2529–2536.
8. Feng S., Elson J.M., Overfelt P.L. Optical properties of multilayer metal-dielectric nanofilms with all-evanescent modes // Optics Express. 2005. V. 13. N 11. P. 4113–4124.
9. Zhang J., Jiang H., Galak B., Enoch S., Tayeb G., Lequime M. Towards-1 effective index with one-dimensional metal-dielectric metamaterial: a quantitative analysis of the role of absorption losses // Optics Express. 2007. V. 15. N 12. P. 7720–7729.
10. Chebykin A.V., Orlov A.A., Vozianova A.V., Maslovski S.I., Kivshar Y.S., Belov P.A. Nonlocal effective medium model for multilayered metal-dielectric metamaterials // Phys. Rev. B. 2011. V. 84. N 11. P. 115438-1–115438-11.
11. Chebykin A.V., Orlov A.A., Simovski C.R., Kivshar Y.S., Belov P.A. Nonlocal effective parameters of multi-layered metal-dielectric metamaterials // Phys. Rev. B. 2012. V. 86. N 11. P. 115420-1–115420-8.
12. Nicolson A.M., Ross G.F. Measurement of the Intrinsic Properties of Materials by Time-Domain Techniques // IEEE Trans. Instrum. Meas. 1970. V. 19. N 4. P. 377–382.
13. Weir W.B. Automatic Measurement of Complex Dielectric Constant and Permeability at Microwave Frequencies // Proc. of IEEE. 1974. V. 62. N 1. P. 33–36.
14. Симовский К.Р. О материальных параметрах метаматериалов // Оптика и спектроскопия. 2009. Т. 107. № 5. С. 766–793.
15. Chen X., Grzegorzczak T.M., Wu B.I., Pacheco J., Jr, Kong J.A. Robust method to retrieve the constitutive effective parameters of metamaterials // Phys. Rev. E. 2004. V. 70. N 1. P. 016608-1–016608-7.
16. Orlov A.A., Voroshilov P.M., Belov P.A., Kivshar Y.S. Engineered optical nonlocality in nanostructured metamaterials // Phys. Rev. B. 2011. V. 84. N 4. P. 045424-1–045424-24.
17. Elser J., Podolksiy V.A., Salakhutdinov I., Avrutsky I. Nonlocal effects in effective-medium response of nanolayered metamaterials // Appl. Phys. Lett. 2007. V. 90. N 19. P. 191109-1–191109-3.
18. Pollard R.J., Murphy A., Hendren W.R., Evans P.R., Atkinson R., Wurtz G.A., Zayats A.V., Podolksiy V.A. Optical nonlocalities and additional waves in epsilon-near-zero metamaterial // Phys. Rev. Lett. 2009. V. 102. N 12. P. 127405-1–127405-4.
19. Кившарь Ю.С., Орлов А.А. Перестраиваемые и нелинейные метаматериалы // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 3 (79). С. 1–10.
20. Jeyaram Y., Jha S.K., Agio M., Löffler J.F., Ekinici Y. Magnetic metamaterials in the blue range using aluminum nanostructures // Optics Lett. 2010. V. 35. N 10. P. 1656–1658.

Орлов Алексей Анатольевич

– младший научный сотрудник, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Alexey.orlov@phoi.ifmo.ru

Янковская Елизавета Алексеевна

– аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, adfors@gmail.com

Жуковский Сергей

– старший научный сотрудник, факультет фотонной техники, Технический университет Дании, Конгенс Люнбю, Дания, sez@fotonik.dtu.dk

Белов Павел Александрович

– доктор физ.-мат. наук, главный научный сотрудник, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, belov@phoi.ifmo.ru

- Alexei Orlov* – junior research scientist, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Alexey.orlov@phoi.ifmo.ru
- Elizaveta Yankovskaya* – postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, adfors@gmail.com
- Sergei Zhukovsky* – senior research scientist, Department of Photonics Engineering, Technical University of Denmark, Kongens Lyngby, Denmark, sezh@fotonik.dtu.dk
- Pavel Belov* – D.Sc., principal research fellow, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, belov@phoi.ifmo.ru

УДК 535.012, 530.182

ПЛАЗМОННЫЕ СОЛИТОНЫ, КИНКИ И ВОЛНЫ ФАРАДЕЯ В ДВУМЕРНОЙ РЕШЕТКЕ МЕТАЛЛИЧЕСКИХ НАНОЧАСТИЦ

Р.Е. Носков^a, Д.А. Смирнова^b, Н.С. Лапшина^a

^a Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, nanometa@gmail.com

^b Центр нелинейной физики, Австралийский национальный университет, Канберра, Австралия, namaste89@mail.ru

Рассматриваются нелинейные дискретные моды в двумерной решетке металлических наночастиц, возбуждаемой оптическим излучением на частоте, близкой к частоте поверхностного плазмонного резонанса уединенной частицы. Предполагается, что размер частиц много меньше оптической длины волны, а межчастичное расстояние достаточно велико, чтобы отклик частиц можно было рассматривать в рамках дипольного приближения. Мы также считаем, что наночастицы сделаны из серебра и обладают нелинейным откликом керровского типа.

В силу того, что каждая частица представляет собой резонансно возбуждаемый нелинейный осциллятор с относительно медленным инерционным откликом по сравнению с периодом колебаний света, динамический отклик системы описывается в терминах медленных амплитуд поляризации каждого шарика. Стандартная процедура линеаризации дает возможность получить зоны модуляционной неустойчивости и бистабильности однородного стационарного решения соответствующих динамических уравнений на плоскости параметров внешнего поля «интенсивность–частота». Используя эти данные, мы представляем и анализируем примеры генерации плазмонного аналога волн Фарадея, устойчивых двумерных солитонов, осциллонов и кинков (волн переключения), фронт которых представляет собой переход при изменении номера частицы от одного однородного распределения поляризации частиц к другому при однородном внешнем возбуждении. Также обсуждается реалистичная длительность лазерного импульса, которая должна быть достаточно большой для формирования рассмотренных нелинейных дискретных мод, но в то же время достаточно малой для предотвращения теплового разрушения наночастиц.

Ключевые слова: плазмоника, нанофотоника, металлическая наночастица, кубическая восприимчивость серебряной наночастицы, поверхностный плазменный резонанс, дискретная локализованная мода, модуляционная неустойчивость, волны Фарадея, солитон, осциллонов, кинк.

PLASMON SOLITONS, KINKS AND FARADAY WAVES IN TWO-DIMENSIONAL LATTICE OF METAL NANOPARTICLES

R.E. Noskov^c, D.A. Smirnova^d, N. S. Lapshina^c

^c Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, nanometa@gmail.com

^d Nonlinear Physics Center, Australian National University, Canberra, Australia, namaste89@mail.ru

We consider nonlinear discrete modes in a two-dimensional lattice of metallic nanoparticles driven by optical radiation at a frequency close to the frequency of the surface plasmon resonance of an individual nanoparticle. We suppose that the particles are small enough and the interparticle distance is large enough to treat nanoparticle within point-dipole approximation. We also assume that nanoparticles are made of silver and possess an intrinsic nonlinear Kerr-type response.

Since each particle acts as a resonantly excited oscillator with slow (in comparison with the light period) inertial response, we employ a slowly varying amplitude approach to describe dynamical behavior of particle polarizations. Following a standard linear stability analysis, we obtain areas of bistability and modulation instability for the homogeneous stationary solution of the corresponding dynamical system in the plane ‘intensity-frequency’. Based on these data, we present and analyze examples of generation of plasmonic Faraday waves, stable two-dimensional solitons, oscillons, and kinks (switching waves), which separate two different homogeneous states of particle polarizations. We also discuss realistic duration of the laser pulse which should be large enough to cause the formation of the considered nonlinear modes and small enough to prevent particle ablation.

Keywords: plasmonics, nanophotonics, metal nanoparticle, cubic susceptibility of silver nanoparticle, surface plasmon resonance, discrete localized mode, modulation instability, Faraday waves, soliton, oscillon, kink.

Введение

Растущий интерес к исследованию нелинейных свойств плазмонных наноструктур, таких как массивы серебряных и золотых наночастиц, объясняется их большим потенциалом для разнообразных нано-

фотонных приложений [1–3]. Благодаря усилению внутренней нелинейности металла в условиях поверхностного плазменного резонанса такие структуры позволяют уменьшить как размер, так и рабочую мощность нелинейных оптических компонент [4–6]. Исследование новых нелинейных эффектов в плазмонных системах открывает большие перспективы для дальнейших разработок нанофотонных устройств [7, 8]. В настоящей работе рассматриваются нелинейные дискретные структуры в двумерной решетке металлических наночастиц, возбуждаемой оптическим излучением. Показана возможность генерации плазмонного аналога волн Фарадея, устойчивых двумерных солитонов, осциллонов и кинков (волн переключения), фронт которых представляет собой переход при изменении номера частицы от одного однородного распределения поляризации частиц к другому при однородном внешнем возбуждении.

Теоретическая модель и основные уравнения

Рассмотрим двумерную решетку идентичных металлических наночастиц сферической формы, которая возбуждается внешним лазерным лучом с частотой, близкой к частоте поверхностного плазменного резонанса отдельной частицы, ω_0 , как показано на рис. 1. Предположим, что размер частиц много меньше оптической длины волны, а междоузельное расстояние достаточно велико, чтобы отклик частиц можно было рассматривать в рамках дипольного приближения. Ограничимся также случаем, когда внешнее электрическое поле ориентировано перпендикулярно плоскости решетки, что может быть реализовано при скользком падении света, а также посредством бинарной оптической маски [9].

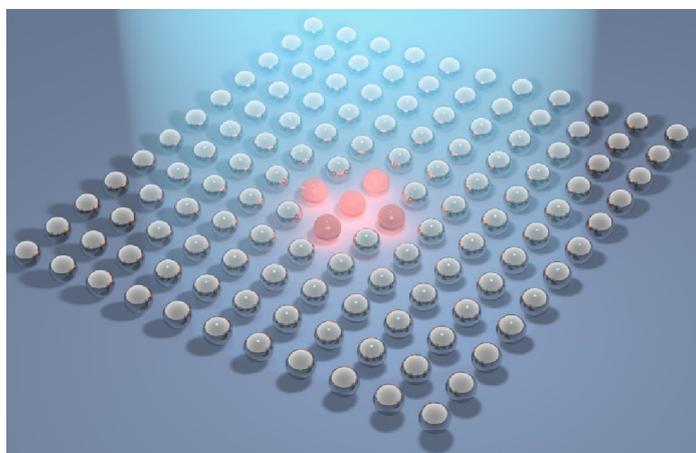


Рис. 1. Схематическое изображение квадратной решетки металлических наночастиц, возбуждаемой лазерным пучком. Потенциальная структура дискретного плазмон-солитона обозначена красными частицами

Полагая, что наночастицы сделаны из серебра и обладают нелинейным откликом керровского типа, мы представляем их диэлектрическую проницаемость в виде

$$\varepsilon_{Ag}^{NL} = \varepsilon_{\infty} - \frac{\omega_p^2}{\omega(\omega - i\nu)} + \chi^{(3)} |\mathbf{E}_{n,m}^{(in)}|^2,$$

где $\varepsilon_{\infty} = 4,96$; $\hbar\omega_p = 9,54$ эВ; $\hbar\nu = 0,055$ эВ; \hbar – постоянная Планка; $\chi^{(3)}$ – кубическая восприимчивость серебряной наночастицы; $\mathbf{E}_{n,m}^{(in)}$ – комплексная амплитуда локального поля внутри (n,m) -ой частицы. Здесь и далее принята зависимость от времени $\sim \exp(i\omega t)$.

Вообще говоря, оптическая кубическая восприимчивость металлических наночастиц зависит от типа металла, размера частиц, частоты и длительности лазерного облучения, а также некоторых других факторов [10]. В частности, аналитическая квантовая модель, развитая в работах [11,12] и позже подтвержденная численным моделированием [13], показала, что для серебряных шариков с радиусом 10 нм при возбуждении на частотах вблизи частоты поверхностного плазменного резонанса кубическая восприимчивость имеет чисто действительный характер и равна $\chi^{(3)} = 4,25 \times 10^{-18}$ В²/М² (что соответствует 3×10^{-9} ед. СГС). В силу того, что каждая частица представляет собой резонансно возбуждаемый нелинейный осциллятор с относительно медленным инерционным откликом по сравнению с периодом колебаний света, динамический отклик системы можно характеризовать в терминах медленных амплитуд поляризации каждого шарика.

Рассматриваемая модель аналогична модели, изученной ранее в работах [14–16] для одномерных цепочек наночастиц. Ее главное отличие заключается в дополнительной размерности. Соответствующие уравнения для поперечных компонент (по отношению к плоскости, в которой лежит решетка) поляризации частиц могут быть представлены в следующем виде:

$$-i \frac{dP_{n,m}^\perp}{dt} + \left(\Omega - i\gamma + |P_{n,m}^\perp|^2 \right) P_{n,m}^\perp + \sum_{\substack{n' \neq n \\ m' \neq m}} G_{n-n', m-m'}^\perp P_{n',m'}^\perp = E_{n,m}^\perp, \quad (1)$$

где

$$G_{n-n', m-m'}^\perp = \frac{\eta}{2} \left[(k_0 d)^2 - \frac{ik_0 d}{\Delta r} - \frac{1}{\Delta r^2} \right] \frac{\exp(-ik_0 d \Delta r)}{\Delta r}$$

описывает взаимодействие между частицами в решетке через полные дипольные поля, $\Delta r = \left(|n - n'|^2 + |m - m'|^2 \right)^{1/2}$, $\eta \sim (a/d)^3$, a – радиус частиц, d – период решетки, $k_0 = \frac{\omega_0}{c} \sqrt{\epsilon_h}$, ϵ_h – диэлектрическая проницаемость внешнего окружения, $P_{n,m}^\perp$ и $E_{n,m}^\perp$ – безразмерные медленно меняющиеся амплитуды поляризации (n,m) -ой частицы и внешнего электрического поля соответственно, γ описывает тепловые и радиационные потери частиц, $\Omega = (\omega - \omega_0)/\omega_0$ – относительная частотная расстройка, $\tau = \omega_0 t$ – безразмерное время. Мы используем такую же нормировку, как в работах [14–16].

Рассмотрим вначале бесконечную решетку, которая возбуждается однородным электрическим полем, т.е. $E_{n,m}^\perp = E_0^\perp$. В этом случае все дипольные моменты частиц равны $P_{n,m}^\perp = P_0^\perp$ и стационарное решение системы (1) можно представить в виде

$$\left[\Omega - i\gamma + |P_0^\perp|^2 + 2 \sum_{j=1}^{\infty} G_{j,l}^\perp \right] P_0^\perp = E_0^\perp, \quad (2)$$

где $j = |n - n'|$ и $l = |m - m'|$. В этом выражении мы воспользовались симметрией суммы относительно положительных и отрицательных значений индексов суммирования. Исследуем теперь устойчивость решения (2) по отношению к малым пространственно-временным возмущениям. Следуя стандартной процедуре линеаризации [16], представим поляризации частиц в виде суперпозиции однородного решения (2) и малого возмущения, взятого в виде собственной моды решетки:

$$P_{n,m}^\perp = P_0^\perp + \delta P \exp[-i(K_x n + K_y m)d + \lambda \tau] + \delta P^* \exp[i(K_x n + K_y m)d + \lambda^* \tau], \quad (3)$$

где δP – амплитуда малого возмущения, $K_{x,y}$ – компоненты волнового вектора собственной моды, λ – инкремент (коэффициент усиления) модуляционной неустойчивости (МН), «*» означает комплексное сопряжение. Подставляя (3) в (1) с учетом (2) и условия существования нетривиального решения, получаем следующее выражение для инкремента:

$$\lambda = \text{Im}\Gamma - \gamma + \left[|P_0^\perp|^4 - (2|P_0^\perp|^2 + \Omega + \text{Re}\Gamma)^2 \right]^{1/2},$$

где

$$\Gamma = 2 \sum_{\substack{j=1 \\ l=1}}^{\infty} G_{j,l}^\perp \cos[(K_x j + K_y l)d].$$

Очевидно, что положительные значения инкремента отвечают условию развития МН. Заметим также, что условие $\lambda > 0$ при $K_{x,y} = 0$ отвечает области существования бистабильности в решении (2).

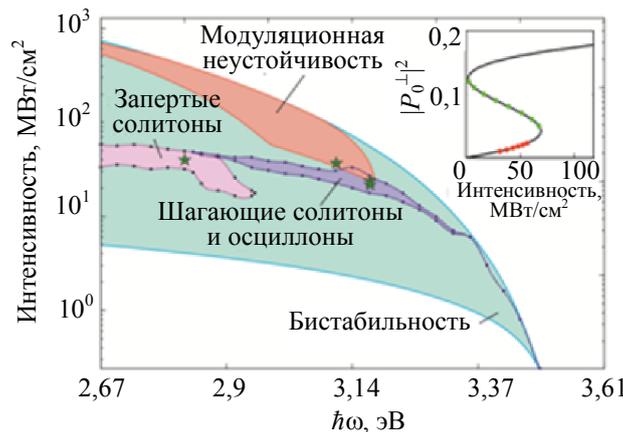


Рис. 2. Бифуркационная диаграмма, показывающая зоны бистабильности и МН, в координатах «интенсивность – энергия фотонов ($\hbar\omega$)» внешнего поля. Зелеными звездочками обозначены параметры, для которых получены рис. 3, а, в, г, и рис. 4, а. Вкладка: зависимость поляризации $|P_0^\perp|^2$ от интенсивности однородного внешнего поля. Красными и зелеными точками обозначены зоны МН, соответствующие возбуждению медленных и быстрых собственных мод решетки

На рис. 2 представлены соответствующие области модуляционной неустойчивости и бистабильности на плоскости параметров «интенсивность–частота». Для простоты мы разделили все собственные моды решетки, по отношению к которым может быть достигнута МН, на локализованные (с волновым числом модуляции $K > k_0$) и излучаемые (с $K \leq k_0$) и сложили зоны МН всех мод для каждого из двух

видов. Красным обозначена область параметров внешнего поля, при которых МН может быть достигнута по отношению только к локализованным собственным модам решетки, в то время как генерация МН по отношению к излучаемым собственным модам не может быть осуществлена в силу того, что такой сценарий развития МН соответствует неустойчивой ветке стационарного решения (см. вкладку на рис. 2). Следовательно, можно ожидать генерацию нелинейных локализованных состояний для параметров внешнего поля, принадлежащих этой области. В свою очередь, для параметров, лежащих за пределами этой области, бистабильность может привести к формированию дискретных плазмон-солитонов и волн переключения (кинков).

Результаты численного моделирования

Для анализа динамического поведения нелинейных мод было проведено численное моделирование конечной решетки размером 101×101 наночастица с $a = 10$ нм и $d = 30$ нм. Предполагалось, что решетка, погруженная в кварцевое стекло, возбуждается однородным стационарным полем. Профиль солитонов и кинков задавался через начальные условия, соответствующие различным веткам бистабильного стационарного решения бесконечной решетки в разных областях массива, в то время как краевые эффекты выступали в роли малого возмущения, необходимого для запуска МН. Характерные результаты представлены на рис. 3, 4. Для лучшей визуализации мгновенные снимки распределения поляризации частиц были увеличены, а краевые эффекты оставлены за пределами рисунков.

Если значения частоты и интенсивности внешнего поля отвечают области МН, наблюдается генерация нелинейных локализованных состояний, известных как волны Фарадея [17] (см. рис. 4, а). Симметрия таких образований определяется спектром собственных мод, возбужденных благодаря МН. На рис. 3, б, представлен спектр собственных мод нелинейного паттерна, приведенного на рис. 3, а. Интересно заметить, что этот спектр формируется двумя зонами в окрестности $Kd \sim 1,9$ и $Kd \sim 0,8$. Таким образом, волны Фарадея постепенно формируются во времени, будучи совершенно отличными от модуляций, обусловленных краевыми эффектами, которые повторяют квадратную симметрию решетки (рис. 3, а, б).

Захваченные («запертые») и шагающие солитоны представлены на рис. 3, в, г. Семейство захваченных солитонов включает в себя как симметричные, так и асимметричные солитоны с поперечным субволновым размером порядка $0,05\lambda$. Необходимо отметить, что солитоны такого типа всегда остаются неподвижными, даже когда приложенное поле является неоднородным. Это обусловлено тем, что их энергия недостаточно велика, чтобы преодолеть потенциальный барьер, создаваемый самой решеткой [18].

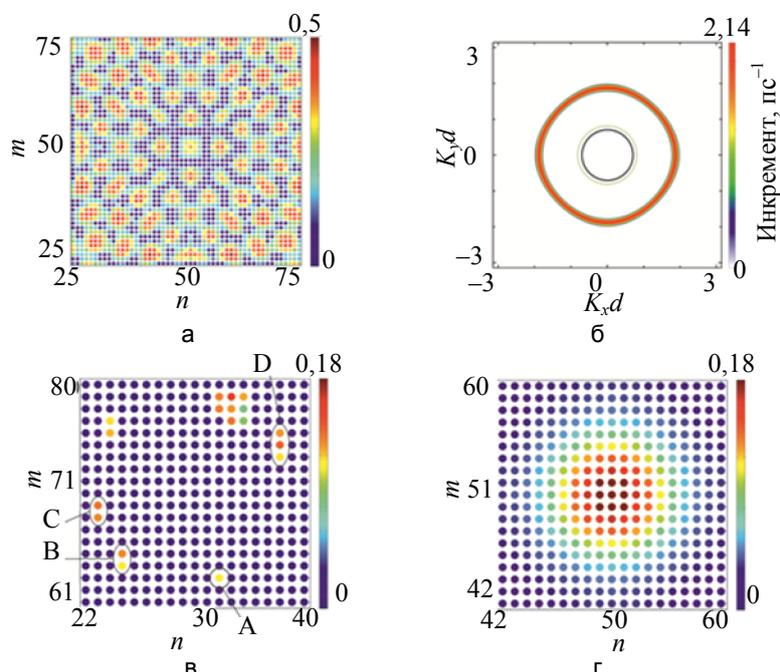


Рис. 3. Мгновенные снимки поляризации частиц $|P_{n,m}^{\perp}|^2$, соответствующие: волнам Фарадея ($\Omega = -0,01$, $|E_0^{\perp}|^2 = 1,5 \times 10^{-4}$) (а), запертым ($\Omega = -0,1$, $|E_0^{\perp}|^2 = 1,7 \times 10^{-4}$) (в) и шагающим плазмон-солитонам ($\Omega = 0,01$, $|E_0^{\perp}|^2 = 10^{-4}$) (г). На рисунке (б) показана контурная карта инкремента, соответствующая реализации, показанной на рис. (а). На рисунке (в) буквами «А», «В», «С» и «D» обозначены захваченные солитоны, локализованные на одной, двух (симметричный и асимметричный) и трех частицах соответственно

Шагающие солитоны характеризуются более широкой локализацией, которая составляет около 150 наночастиц (в силу чего их энергия достаточна для преодоления потенциала решетки), а их форма не зависит от Ω , $|E_0^\perp|^2$ и ширины профиля поляризации, задаваемого начальными условиями. Была обнаружена бифуркация шагающих солитонов при $\hbar\omega = \hbar\omega_0 = 3,14$ эВ ($\Omega = 0$) (см. рис. 2). При $\hbar\omega \geq 3,14$ эВ такие солитоны обладают устойчивым стационарным профилем, дрейфуя под влиянием краевых эффектов с относительно малой скоростью порядка $10^{-4} \times c$, где c – скорость света, и упруго сталкиваясь с границами решетки и друг с другом. Однако при $\hbar\omega < 3,14$ эВ шагающие солитоны трансформируются в осциллоны, которые перемещаются со скоростью порядка $10^{-2} \times c$, сливаются в один осциллоны и излучаются через границы решетки. Отметим, что мощным инструментом для управления солитонами являются неоднородности фазы или амплитуды внешнего поля, так как градиент неоднородности задает направление дрейфа.

Численное исследование позволило обнаружить зоны существования запертых и шагающих солитонов на плоскости параметров «интенсивность–частота» (рис. 2). Примечательно, что зоны МН и шагающих солитонов имеют перекрытие, внутри которого солитоны не разрушаются МН, а существуют в окружении волн Фарадея.

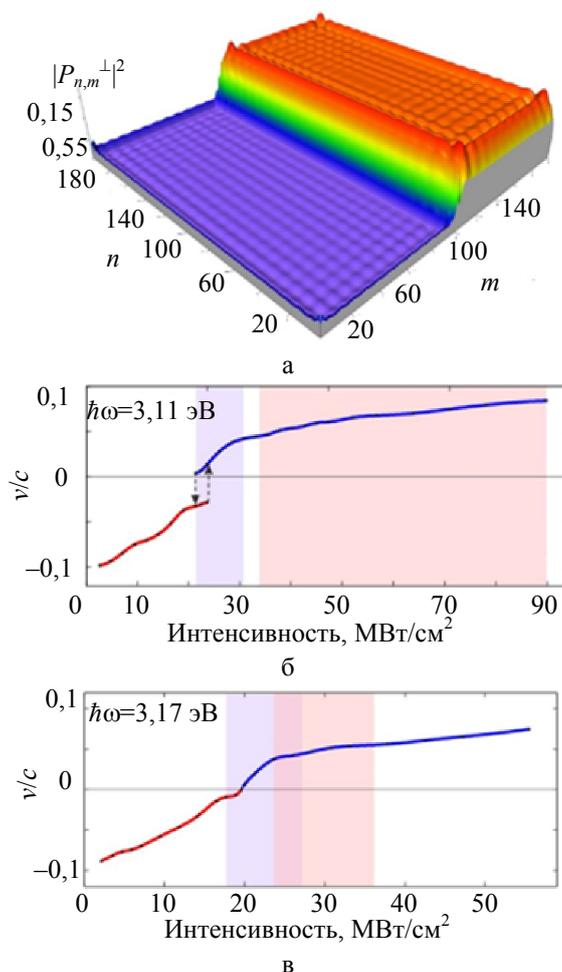


Рис. 4. Профиль поляризации $|P_{n,m}^\perp|^2$ для стационарного кинка, полученный при $\Omega = 0,01$ и $|E_0^\perp|^2 = 0,83 \times 10^{-4}$. Слабые осцилляции вызваны краевыми эффектами (а). Зависимости скорости кинка v , нормированной на скорость света c , от интенсивности приложенного поля при $\Omega = -0,01$ ($\hbar\omega = 3,11$ эВ) (б) и $\Omega = 0,01$ ($\hbar\omega = 3,17$ эВ) (в). Синим цветом отмечена область существования солитонов, красным – модуляционной неустойчивости

Интересно также сравнить исследованные солитоны с солитонами в схожей резонансной системе – двумерной решетке кольцевых резонаторов [3, 19]. Принципиальным отличием последней от решетки металлических наночастиц является отсутствие дальнего взаимодействия. В обеих этих системах были обнаружены захваченные солитоны, локализованные всего на одном или нескольких резонаторах, со схожими свойствами. Однако мобильность и форма широких солитонов оказалась совершенно различной. Дальнее взаимодействие приводит к тому, что форма широких солитонов в массиве наночастиц – всегда круглая, с фиксированным диаметром порядка 16 наночастиц вне зависимости от формы начального распределения поляризуемостей частиц. Также такие солитоны отличаются высокой степенью подвижности.

В то же время в решетке кольцевых резонаторов с ближним взаимодействием форма широких солитонов задается начальным условием (за исключением случаев разбиения одного солитона на несколько в условия развития МН), а их подвижность, по крайней мере, при однородном возбуждении, полностью отсутствует.

Рассмотрим теперь волны переключения. Для подавления влияния краевых эффектов на динамику кинков увеличим размер решетки до 201×201 частиц. В отличие от одномерных волн переключения, в цепочке частиц [16] двумерные кинки с отрицательными и положительными скоростями имеют одинаковую ширину, равную 9 наночастицам, как показано на рис. 4, а. Как следует из рис. 4, б, при $\hbar\omega < 3,14$ эВ отрицательная и положительная ветви зависимости скорости кинков от интенсивности формируют гистерезисную петлю, которая исчезает при $\hbar\omega \geq 3,14$ эВ, что приводит к появлению стационарных кинков с нулевой скоростью (рис. 4, в). Важно отметить, что бифуркационные частоты для кинков и шагающих солитонов совпадают. Более того, шагающие солитоны всегда существуют при тех же параметрах, что и кинки с минимальными скоростями, как показано на рис. 4, б. Это является типичной ситуацией для пространственно распределенных бистабильных систем, в которых солитоны могут трактоваться как два самозахватенных кинка с противоположными полярностями. В заключение необходимо отметить, что кинки могут существовать одновременно с МН, если МН генерирует только малоамплитудные волны Фарадея, что соответствует малой величине инкремента.

Реалистичные параметры

Заметим, что интенсивность внешнего поля, необходимого для наблюдения описанных выше эффектов, составляет порядка 50 МВт/см². Такие большие поля неизбежно приведут к тепловому разрушению наночастиц, если не ограничивать длительность облучения. Чтобы оценить максимально возможную длительность импульса накачки, мы воспользуемся экспериментально полученной величиной пороговой плотности энергии абляции серебряных наночастиц в кварцевой матрице, которая составляет $3,96$ Дж/см² при пикосекундном режиме облучения [20]. Принимая во внимание усиление электрического поля внутри серебряной наночастицы благодаря поверхностному плазменному резонансу, мы получаем максимальную длительность импульса порядка $0,1$ нс, что намного больше характерного времени, необходимого для генерации нелинейных диссипативных мод, которое равно примерно 500 фс. Таким образом, все рассмотренные в настоящей работе эффекты доступны для наблюдения в экспериментальных условиях.

Заключение

В работе изучены поперечные нелинейные моды в двумерной решетке нелинейных металлических наночастиц, возбуждаемой лазерным излучением. Продемонстрирована генерация плазмонного аналога волн Фарадея, захваченных и шагающих солитонов, а также кинков. Полученные результаты открывают широкие перспективы для дальнейших экспериментальных исследований нелинейных плазмонных наноструктур и метаматериалов, которые могут найти применение в нанофотонных устройствах.

Литература

1. Kauranen M., Zayats A.V. Nonlinear plasmonics // *Nature Photonics*. 2012. V. 6. N 11. P. 737–748.
2. Кившарь Ю.С., Орлов А.А. Перестраиваемые и нелинейные метаматериалы // *Научно-технический вестник информационных технологий, механики и оптики*. 2012. № 3 (79). С. 1–10.
3. Розанов Н.Н., Высотина Н.В., Шацев А.Н., Десятников А.С., Шадривов И.В., Носков Р.Е., Кившарь Ю.С. Дискретные волны переключения и диссипативные солитоны в когерентно возбуждаемых наноструктурах и метаматериалах // *Научно-технический вестник информационных технологий, механики и оптики*. 2012. № 4 (80). С. 1–12.
4. Utikal T., Hentschel M., Giessen H. Nonlinear photonics with metallic nanostructures on top of dielectrics and waveguides // *Appl. Phys. B*. 2011. V. 105. N 1. P. 51–65.
5. Schumacher T., Kratzer K., Molnar D., Hentschel M., Giessen H., Lippitz M. Nanoantenna-enhanced ultrafast nonlinear spectroscopy of a single gold nanoparticle // *Nature Commun.* 2011. V. 2. P. 333–1–333–6.
6. Ginzburg P., Krasavin A.V., Zayats A.V. Cascaded second-order surface plasmon solitons due to intrinsic metal nonlinearity // *New J. Phys.* 2013. V. 15. P. 013031–1–013031–13.
7. Zharov A.A., Noskov R.E., Tsarev M.V. Plasmon-induced terahertz radiation generation due to symmetry breaking in a nonlinear metallic nanodimer // *J. Appl. Phys.* 2009. V. 106. N 7. P. 073104–1–073104–5.
8. Noskov R.E., Zharov A.A., Tsarev M.V. Generation of widely tunable continuous-wave terahertz radiation using a two-dimensional lattice of nonlinear metallic nanodimers // *Phys. Rev. B*. 2010. V. 82. N 7. P. 073404–1–073404–4.
9. Wang H., Shi L., Lukyanchuk B., Sheppard C.J.R., Chong C.T. Creation of a needle of longitudinally polarized light in vacuum using binary optics // *Nature Photonics*. 2008. V. 2. N 8. P. 501–505.

10. Palpant B. Third-order nonlinear optical response of metal nanoparticles // *Non-Linear Optical Properties of Matter* / Eds M. G. Papadopoulos, A. J. Sadlej, J. Leszczynski. Dordrecht: Springer, 2006. V. 1. P. 461–508.
11. Раутян С.Г. Нелинейная спектроскопия насыщения вырожденного электронного газа в сферических металлических наночастицах // *ЖЭТФ*. 1997. Т. 112. № 3 (9). С. 836–855.
12. Drachev P., Buin A.K., Nakotte H., Shalaev V.M. Size dependent $\chi^{(3)}$ for conduction electrons in Ag nanoparticles // *Nano Lett.* 2004. V. 4. N 8. P. 1535–1539.
13. Govyadinov A.A., Panasyuk G.Y., Schotland J.C., Markel V.A. Theoretical and numerical investigation of the size-dependent optical effects in metal nanoparticles // *Phys. Rev. B*. 2011. V. 84. N 15. P. 155461-1–155461-12.
14. Noskov R.E., Belov P.A., Kivshar Yu.S. Subwavelength modulational instability and plasmon oscillons in nanoparticle arrays // *Phys. Rev. Lett.* 2012. V. 108. N 9. P. 093901-1–093901-5.
15. Noskov R.E., Belov P.A., Kivshar Yu.S. Subwavelength plasmonic kinks in arrays of metallic nanoparticles // *Optics Express*. 2012. V. 20. N 3. P. 2733–2739.
16. Noskov R.E., Belov P.A., Kivshar Yu.S. Oscillons, solitons, and domain walls in arrays of nonlinear plasmonic nanoparticles // *Scientific Reports*. 2012. V. 2. P. 873-1–873-8.
17. Miles J. On Faraday waves // *J. Fluid Mech.* 1993. V. 248. P. 671–683.
18. Braun O.M., Kivshar Yu.S. *The Frenkel-Kontorova Model: Concepts, Methods, and Applications*. Heidelberg: Springer, 2004. 498 p.
19. Rosanov N.N., Shadrivov I.V., Kivshar Y.S., Vysotina N.V., Shatsev A.N., Powell D.A. Discrete dissipative localized modes in nonlinear magnetic metamaterials // *Optics Express*. 2011. V. 19. N 27. P. 26500–26506.
20. Torres-Torres C., Peréa-López N., Reyes-Esqueda J.A., Rodríguez-Fernández L., Crespo-Sosa A., Cheang-Wong J.C., Oliver A. Ablation and optical third-order nonlinearities in Ag nanoparticles // *Int. J. Nanomedicine*. 2010. V. 5. P. 925–932.

- | | |
|-------------------------------------|--|
| <i>Носков Роман Евгеньевич</i> | – кандидат физ.-мат. наук, старший научный сотрудник, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, nanometa@gmail.com |
| <i>Смирнова Дарья Александровна</i> | – аспирант, Центр нелинейной физики, Австралийский национальный университет, Канберра, Австралия, namaste89@mail.ru |
| <i>Лапшина Надежда Сергеевна</i> | – аспирант, инженер-исследователь, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, n.lapshina@phoi.ifmo.ru |
| <i>Roman Noskov</i> | – PhD, senior scientific researcher, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, nanometa@gmail.com |
| <i>Daria Smirnova</i> | – postgraduate, Nonlinear Physics Center, Australian National University, Canberra, Australia, namaste89@mail.ru |
| <i>Nadezhda Lapshina</i> | – postgraduate, research engineer, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, n.lapshina@phoi.ifmo.ru |

УДК 581.787

ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК СИГНАЛОВ СПЕКТРАЛЬНОЙ ИНТЕРФЕРЕНЦИИ В БЛИЖНЕЙ ИК ОБЛАСТИ СПЕКТРА¹

И.П. Гуров^a

^a Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, gurov@mail.ifmo.ru

Рассмотрены особенности формирования сигналов в спектральной интерферометрии и оптической когерентной томографии. Приведены основные соотношения, определяющие минимальное значение координаты по глубине исследуемого объекта, на которой регистрируется один период сигнала спектральной интерференции и устанавливаемое значение приращения по длине волны с учетом диапазона глубин, в котором регистрируются сигналы спектральной интерференции. Дана оценка разрешающей способности систем спектральной интерферометрии и оптической когерентной томографии с перестраиваемой длиной волны с учетом спектрального диапазона перестройки длины волны. Показано, что отношение среднего значения длины волны к диапазону перестройки по длине волны определяет разрешающую способность по глубине исследуемого объекта, тогда как максимальный диапазон по глубине, в котором возможно исследование микроструктуры объекта при помощи метода спектральной оптической когерентной томографии, не зависит от

¹ Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации.

диапазона перестройки по длине волны и определяется шагом перестройки по длине волны (волновому числу). Представлены количественные оценки указанных параметров при использовании источников излучения в ближнем ИК диапазоне спектра, а также соотношения для оценки видности интерференционных полос в зависимости от регистрируемой относительной интенсивности измерительной волны.

Ключевые слова: спектральная интерферометрия, оптическая когерентная томография, перестраиваемая длина волны.

STUDY OF CHARACTERISTICS OF SPECTRAL INTERFERENCE SIGNALS IN THE NEAR INFRARED SPECTRAL RANGE²

I. Gurov^b

^bSaint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, gurov@mail.ifmo.ru

Peculiarities of signals formation in spectral interferometry and optical coherence tomography are considered. Basic relations are given defining minimal depth coordinate value of an investigated object, where single period of spectral interference signal is acquired and a value of the wave length increment set according to the depth range, where spectral interference signals are registered. The estimate of resolving power of the spectral interferometry and optical coherence tomography systems with tunable wave length is given taking into account a spectral range of wave length tuning. It is shown that the ratio of the wave length mean value and the range of the wave length tuning defines the resolving power in depth of an investigated object, while the maximum depth range, within which investigation of an object's micro structure by the spectral optical coherence tomography is possible does not depend on the range of the wave length tuning being determined by the wave length (wave number) tuning step. Numerical estimates of the parameters mentioned above are presented when using light sources in near infrared range, as well as relations and estimates of interference fringe visibility dependent on registered relative intensity of a measuring wave.

Keywords: spectral interferometry, optical coherence tomography, reconfigurable wave length.

Введение

Методы спектральной интерферометрии основываются на формировании интерференционных полос в широкой области спектра в зависимости от длины волны (волнового числа) [1, 2] и широко используются при измерениях малых перемещений [3], в профилометрии [4], при анализе оптических свойств материалов [5, 6], исследованиях характеристик коротких оптических импульсов [7, 8] и в других областях [9–11].

В последние годы получили активное развитие методы спектральной оптической когерентной томографии (СОКТ), обеспечивающие возможность исследования с высоким быстродействием и разрешением внутренней микроструктуры неоднородных биологических объектов и сред [12–16]. Ввиду использования новых источников излучения с перестраиваемой длиной волны в современных системах СОКТ сигналы спектральной интерференции формируются во временной области, при этом актуальная задача состоит в исследовании характеристик сигналов для оптимизации режимов функционирования систем СОКТ.

В настоящей работе представлены основные соотношения, определяющие разрешающую способность систем СОКТ с перестраиваемой длиной волны с учетом спектрального диапазона перестройки длины волны и количества устанавливаемых длин волн для исследования биологических объектов на глубине проникновения оптического излучения до 2 мм. Представлены количественные оценки указанных параметров при использовании источников излучения в ближнем инфракрасном (ИК) диапазоне со средней длиной волны 1300 нм, соответствующей «окну прозрачности» многих биологических сред.

Формирование сигналов в спектральной интерферометрии

Пусть источник излучения с напряженностью электрического поля E_0 освещает интерферометр. Обозначим значения напряженности электрического поля измерительной и опорной волн в интерферометре (рис. 1) как

$$E_1(v, t) = \kappa E_0(v, t), \quad (1)$$

$$E_2(v, t, \tau) = (1 - \kappa) E_0(v, t) \exp(-j2\pi v \tau), \quad (2)$$

где v – частота оптических колебаний; κ – амплитудный коэффициент отражения в светоделителе, $E_0(v, t) = A(v) \exp(-j2\pi v t)$ – оптические колебания с амплитудным спектром $A(v)$; $\tau = \Delta/c$ – интервал времени сдвига измерительной волны по отношению к опорной при оптической разности хода (ОРХ) Δ , c – скорость света.

Интенсивность излучения, регистрируемого на выходе интерферометра, как известно, определяется выражением

² The work was done with financial support from the Ministry of education and science of Russia

$$I(\nu, \tau) = \langle EE^* \rangle = \langle |E_1|^2 \rangle + \langle |E_2|^2 \rangle + 2\text{Re} \langle E_1 E_2^* \rangle, \quad (3)$$

где $E = E_1 + E_2$, угловые скобки обозначают усреднение во времени. Поскольку при использовании практически доступных фотодетекторов интервал усреднения значительно превышает период световых колебаний, из (1)–(3) получим

$$I(\nu, \tau) = I_1(\nu) + I_2(\nu) + 2\sqrt{I_1(\nu)I_2(\nu)}\cos 2\pi\nu\tau, \quad (4)$$

где $I_1(\nu) = \kappa^2 \langle A_1(\nu)A_1^*(\nu) \rangle$, $I_2(\nu) = (1-\kappa)^2 \langle A_2(\nu)A_2^*(\nu) \rangle$, $A_1(\nu)$ и $A_2(\nu)$ – комплексные амплитуды интерферирующих волн на выходе интерферометра.

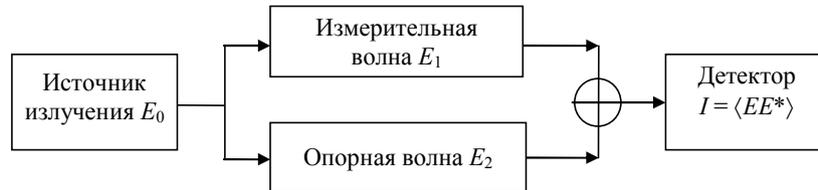


Рис. 1. Структурная схема интерферометрической системы

При $A_1(\nu) = A_2(\nu) = A(\nu)$ первые два слагаемых в (4) соответствуют спектру источника излучения, третье слагаемое представляет автокорреляционную функцию гармонической составляющей оптических колебаний с частотой ν . При таком рассмотрении излучение источника интерпретируется как состоящее из набора гармонических колебаний с амплитудами $A(\nu)$.

При фиксированной ОРХ $\Delta = c\tau$ выражение (4) можно представить в форме

$$I(\lambda) = I_1(\lambda) + I_2(\lambda) + 2\sqrt{I_1(\lambda)I_2(\lambda)} \cos(2\pi\Delta/\lambda), \quad (5)$$

где $\lambda = c/\nu$ – длина волны. При отсутствии дисперсии в плечах интерферометра и преобразовании интенсивности в сигнал $s = \mu I$, где μ – коэффициент преобразования, выражение (5) преобразуется к виду

$$s(\lambda) = s_0(\lambda)[1 + V \cos(2\pi\Delta/\lambda)],$$

где $s_0(\lambda)$ – значения, пропорциональные распределению интенсивности в спектре источника излучения, V – видность интерференционных полос.

В спектральной интерферометрии с перестраиваемой длиной волны получают последовательность отсчетов сигнала $s(k) = s(\sigma_k)$, $k = 1, \dots, K$,

$$s(k) = s_0(k)\{1 + V \cos[2\pi\Delta(\sigma_1 + (k-1)\delta\sigma)]\}, \quad (6)$$

где $\delta\sigma = \delta\lambda/[\lambda(\lambda + \delta\lambda)]$ – шаг изменения волнового числа $\sigma = 1/\lambda$ при перестройке по длине волны $\delta\lambda$. Пример сигнала вида (6) показан на рис. 2.

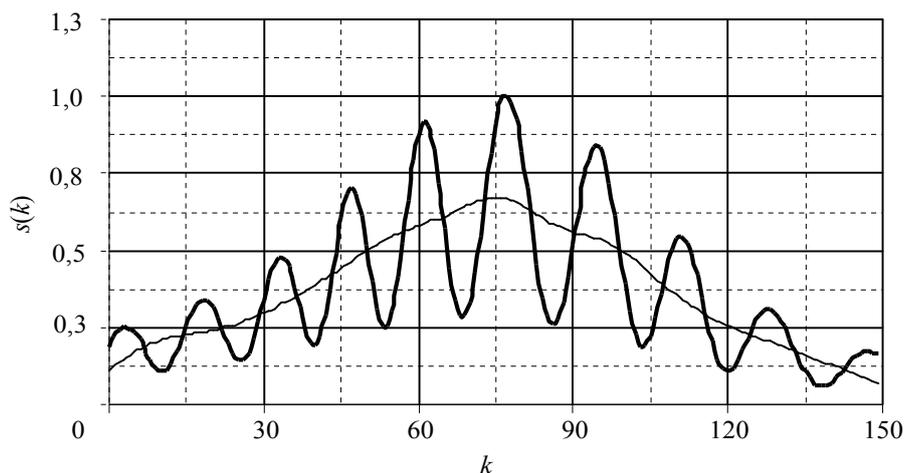


Рис. 2. Сигнал спектральной интерференции при однократном отражении, отнормированный на максимальное значение

Поскольку спектр источника априорно известен, можно соответственно отнормировать сигнал (6), при этом

$$s(k) = 1 + V \cos(2\pi\Delta\sigma_k), \quad (7)$$

где $\sigma_k = 1/\lambda_k$. Задача состоит в определении значения Δ по измеренным значениям сигнала $s(k)$ в (7).

Характеристики сигналов в СОКТ

При отражении измерительной волны от образца на глубине z (полагаем $z = 0$ на поверхности образца) ОРХ выражается как

$$\Delta = 2n_s z, \quad (8)$$

где n_s – показатель преломления материала образца. Полагая вначале $n_s \approx 1$, перепишем выражение (7), подставив в него (8):

$$s(k) = 1 + V \cos(4\pi z / \lambda_k). \quad (9)$$

При перестройке длины волны источника излучения в диапазоне от $\lambda_{\min} = \lambda_1$ до

$$\lambda_{\max} = \lambda_1 + (K - 1) \delta\lambda$$

фаза сигнала (9) для фиксированного значения координаты по глубине $z > 0$ будет изменяться. Одному периоду изменения сигнала соответствует соотношение

$$\frac{4\pi z_1}{\lambda_{\min}} - \frac{4\pi z_1}{\lambda_{\max}} = 2\pi. \quad (10)$$

При обозначении диапазона перестройки по длине волны $\lambda_{\max} - \lambda_{\min} = \Delta\lambda$ и среднего значения длины волны $\langle \lambda \rangle$ из (10) получим

$$2z_1 = \frac{\lambda_{\min} \lambda_{\max}}{\lambda_{\max} - \lambda_{\min}} \approx \frac{\langle \lambda \rangle^2}{\Delta\lambda},$$

откуда

$$z_1 = \frac{\langle \lambda \rangle \langle \lambda \rangle}{\Delta\lambda} \cdot \frac{1}{2}. \quad (11)$$

Из (11) видно, что минимальное значение глубины z_1 , на которой регистрируется один период сигнала спектральной интерференции, равно половине среднего значения длины волны, умноженного на отношение среднего значения длины волны и диапазона перестройки по длине волны.

Полагая $\langle \lambda \rangle = 1300$ нм и $\Delta\lambda = 100$ нм, из (11) находим $z_1 \approx 8,5$ мкм ($n_s \approx 1$). Таким образом, при исследовании объекта методом СОКТ положение начальной границы диапазона измерений по глубине отличается от нулевой ОРХ на весьма малую величину (в приведенном примере на 8,5 мкм).

Используя введенные выше обозначения, выразим количество шагов перестройки по длине волны как

$$\frac{\Delta\lambda}{\delta\lambda} = K.$$

Изменение фазы сигнала (9) для одного шага перестройки $\delta\lambda$ по длине волны составит

$$\delta\varphi = \frac{4\pi z}{\langle \lambda \rangle} - \frac{4\pi z}{\langle \lambda \rangle + \delta\lambda} \approx \frac{4\pi z}{\langle \lambda \rangle} \frac{\delta\lambda}{\langle \lambda \rangle}. \quad (12)$$

Из (12) видно, что приращение фазы пропорционально значению глубины z . Для проведения измерений на максимальной глубине $z = z_{\max}$ должно выполняться условие

$$\delta\varphi_{\max} = \frac{4\pi z_{\max}}{\langle \lambda \rangle} \frac{\delta\lambda}{\langle \lambda \rangle} < \pi, \quad (13)$$

выражающее критерий Найквиста и соответствующее получению не менее двух отсчетов для максимальной частоты в спектре сигнала. При этом шаг $\delta\lambda$ по длине волны определяется из (13) как

$$\delta\lambda < \frac{\langle \lambda \rangle^2}{4z_{\max}}. \quad (14)$$

Согласно (14), при $\langle \lambda \rangle = 1300$ нм и $z_{\max} = 2$ мм $\delta\lambda < 0,2$ нм. При диапазоне перестройки по длине волны $\Delta\lambda = 100$ нм количество шагов перестройки составит $K = 500$. С учетом последующего выполнения быстрого преобразования Фурье целесообразно принять $K = 512$.

В приведенных выше соотношениях (8), (11), (14) несложно учесть отличающееся от единицы значение n_s показателя преломления материала образца.

Оценим разрешающую способность системы, а именно, минимальное значение изменения координаты δz по глубине, которое возможно зарегистрировать. Предположим, что система обработки позволяет определять приращения фазы не менее $\delta\varphi_{\min}$. При этом малое изменение δz по глубине должно

обуславливать изменение фазы сигнала (9) в диапазоне перестройки по длинам волн от λ_{\min} до λ_{\max} согласно условию

$$\frac{4\pi\delta z}{\lambda_{\min}} - \frac{4\pi\delta z}{\lambda_{\max}} > \delta\varphi_{\min}$$

или

$$\delta\varphi_{\min} < 2\pi \frac{\Delta\lambda}{\langle\lambda\rangle} \frac{2\delta z}{\langle\lambda\rangle},$$

откуда (см. также (11))

$$\delta z > \frac{\delta\varphi_{\min}}{2\pi} \frac{\langle\lambda\rangle}{\Delta\lambda} \frac{\langle\lambda\rangle}{2} = \frac{\delta\varphi_{\min}}{2\pi} z_1. \quad (15)$$

Из (15) видно, что разрешающая способность по глубине пропорциональна отношению средней длины волны и диапазона перестройки по длине волны и минимальному значению глубины z_1 , на которой регистрируется один период сигнала спектральной интерференции.

При $\langle\lambda\rangle = 1300$ нм и $\Delta\lambda = 100$ нм из (11) и (15) имеем

$$\delta z > \frac{\delta\varphi_{\min}}{2\pi} \times 8,5 \text{ мкм},$$

что означает отсутствие строгих требований к точному определению приращений фазы: даже при $\delta\varphi_{\min} = 2\pi$ разрешающая способность составляет несколько микрометров.

Более важную роль играет видность интерференционных полос, поскольку при исследовании неоднородных объектов и сред в СОКТ лишь небольшая часть интенсивности падающей на образец волны после отражения регистрируется фотодетектором, что может снижать видность полос, и, следовательно, амплитуду полезного сигнала вида (9).

Видность интерференционных полос определяется известным соотношением

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} = \frac{2}{\sqrt{I_1/I_2} + \sqrt{I_2/I_1}} = \frac{2}{\sqrt{\gamma} + \sqrt{1/\gamma}}, \quad (16)$$

где для упрощения записи формул введено обозначение $\gamma = [\kappa/(1-\kappa)]^2$. Видность максимальна при условии минимизации выражения в знаменателе (16), т.е.

$$\gamma^{1/2} + \gamma^{-1/2} \rightarrow \min. \quad (17)$$

Вычисление производной по γ в левой части выражения (17) дает

$$(\gamma^{1/2} + \gamma^{-1/2})'_\gamma = \frac{\gamma - 1}{2\gamma\sqrt{\gamma}}. \quad (18)$$

Приравнивание производной (18) к нулю приводит к условию $\gamma = 1$.

Выражение (16) при $\gamma \ll 1$ имеет простой вид

$$V \approx 2\sqrt{\gamma}. \quad (19)$$

Из (19) следует, что даже при весьма малом значении $\gamma = 0,01$ величина видности составляет $V = 0,2$. Таким образом, параметр γ не является критичным, однако обеспечение его значения в пределах от 1 до 0,15 позволяет получить видность соответственно от 1 до 0,8.

Заключение

Полученные соотношения показывают, что в СОКТ отношение среднего значения длины волны к диапазону перестройки по длине волны определяет минимальное регистрируемое отклонение от нулевой ОРХ и разрешающую способность по глубине исследуемого объекта. Разрешающая способность возрастает при увеличении относительной ширины спектра перестройки по длине волны.

Максимальный диапазон по глубине, в котором возможно исследование микроструктуры объекта при помощи метода СОКТ, не зависит от диапазона перестройки по длине волны и определяется шагом перестройки по длине волны (волновому числу). Отметим, что ширина спектральной линии излучения в современных перестраиваемых лазерах достаточно мала и не оказывает заметного влияния на видность спектральных интерференционных полос.

Приведенные выше соотношения не учитывают влияния ряда факторов, в частности, дисперсии излучения в среде и шума спеклов. Для учета влияния дисперсии требуется получать оценки не только амплитуды, но и фазы спектральных интерференционных полос (см., например, [17]). Снижение влияния шума спеклов достигается при усреднении нескольких зарегистрированных сигналов с учетом статисти-

ки спеклов (см., например, [18]), что требует дополнительных вычислительных ресурсов, однако заметно повышает качество визуализации спектральных томограмм.

Литература

1. Vienot J.-C., Goedgebuer J.-P., Lacourt A. Space and time variables in optics and holography: recent experimental aspects // *Applied Optics*. 1977. V. 16. N 2. P. 454–461.
2. Emde M.F., de Boeij W.P., Pshenichnikov M.S., Wiersma D.A. Spectral interferometry as an alternative to time-domain heterodyning // *Optics Lett.* 1997. V. 22. N 17. P. 1338–1340.
3. Hlubina P., Gurov I., Chugunov V. Slightly dispersive white-light spectral interferometry to measure distances and displacements // *Optik*. 2003. V. 114. N 9. P. 389–393.
4. Reolon D., Jacquot M., Verrier I., Brun G., Veillas C. Broadband supercontinuum interferometer for high-resolution profilometry // *Optics Express*. 2006. V. 14. N 1. P. 128–137.
5. Kumar V.N., Rao D.N. Using interference in the frequency domain for precise determination of thickness and refractive indices of normal dispersive materials // *J. Opt. Soc. Am. B*. 1995. V. 12. N 9. P. 1559–1563.
6. Hlubina P., Lunacek J., Ciprian D. White-light spectral interferometry and reflectometry to measure thickness of thin films // *Proc. of SPIE*. 2009. V. 7389. P. 738926-1–738926-8.
7. Iaconis C., Walmsley I.A. Spectral phase interferometry for direct electric-field reconstruction of ultrashort optical pulses // *Optics Lett.* 1998. V. 23. N 10. P. 792–794.
8. Yetzbacher M.K., Courtney T.L., Peters W.K. Spectral restoration for femtosecond spectral interferometry with attosecond accuracy // *J. Opt. Soc. Am. B*. 2010. V. 27. N 5. P. 1104–1117.
9. Hlubina P., Gurov I., Chugunov V. White-light spectral interferometric technique to measure the wavelength dependence of the spectral bandpass of a fibre-optic spectrometer // *J. Mod. Optics*. 2003. V. 50. N 13. P. 2067–2074. Lepetit L., Chériaux G., Joffre M. Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy // *J. Opt. Soc. Am. B*. 1995. V. 12. N 12. P. 2467–2474.
10. Lepetit L., Chériaux G., Joffre M. Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy // *J. Opt. Soc. Am. B*. 1995. V. 12. N 12. P. 2467–2474.
11. Albrecht A.W., Hybl J.D., Gallagher S.M., Jonas D.M. Experimental distinction between phase shifts and time delays: implications for femtosecond spectroscopy and coherent control of chemical reactions // *J. Chem. Phys.* 1999. V. 111. P. 10934–10956.
12. Oldenburg A.L., Xu C., Boppart S.A. Spectroscopic optical coherence tomography and microscopy // *IEEE J. Select. Topics Quantum Electron.* 2007. V. 13. N 6. P. 1629–1640.
13. Verrier I., Jacquot M., Brun G., Veillas C., Ben Houcine K. Imaging through scattering medium by recording 3D «spatial-frequential» interferograms // *Optics Commun.* 2006. V. 267. N 2. P. 310–317.
14. Choma M.A., Sarunic M.V., Yang C.H., Izatt J.A. Sensitivity advantage of swept source and Fourier domain optical coherence tomography // *Optics Express*. 2003. V. 11. N 18. P. 2183–2189.
15. Васильев В.Н., Гуров И.П. Сравнительный анализ методов оптической когерентной томографии // *Изв. вузов. Приборостроение*. 2007. Т. 50. № 7. С. 30–40.
16. Гуров И.П., Вольнский М.А., Жукова Е.В., Маргарянц Н.Б. Исследование растительных тканей методом оптической когерентной микроскопии // *Научно-технический вестник информационных технологий, механики и оптики*. 2012. № 5 (81). С. 42–47.
17. Hillmann D., Bonin T., Lührs C., Franke G., Hagen-Eggert M., Koch P., Hüttmann G. Common approach for compensation of axial motion artifacts in swept-source OCT and dispersion in Fourier-domain OCT // *Optics Express*. 2012. V. 20. N 6. P. 6761–6776.
18. Wong A., Mishra A., Bizheva K., Clausi D.A. General Bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery // *Optics Express*. 2010. V. 18. N 8. P. 8338–8352.

- Гуров Игорь Петрович** – доктор технических наук, профессор, зав. кафедрой, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, gurov@mail.ifmo.ru
- Igor Gurov** – D.Sc., Professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, gurov@mail.ifmo.ru

УДК 004.387; 535.417

ЛОГИКА С ИСКЛЮЧЕНИЕМ НА АЛГЕБРЕ ФУРЬЕ-ДУАЛЬНЫХ ОПЕРАЦИЙ: НЕЙРОСЕТЕВОЙ МЕХАНИЗМ РЕДУЦИРОВАНИЯ КОГНИТИВНОГО ДИССОНАНСА¹

А.В. Павлов^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Pavlov@phoi.ifmo.ru

В развитие подхода к реализации методом голографии Фурье немонотонных логик из класса нечетко-значимых показан механизм редуцирования когнитивного диссонанса. Когнитивный диссонанс возникает при восприятии новой информации, противоречащей ранее сложившейся субъективной картине мира, представленной в виде двойного каскада преобразования Фурье с голограммой Фурье-матрицы связей нейронных слоев репрезентации входной информации и заключения. Голограмма реализует монотонную логику по правилу вывода «Обобщенный Modus Ponens». Новая информация представлена голограммой исключения – матрицей связей слоев заключения и исключений, связанных преобразованием Фурье-операцией, задающей дуальность определяющих алгебру логики операций конъюнкции и дизъюнкции. Голограмма исключения формирует вывод, дуальный выводу по правилу «Обобщенный Modus Ponens». Показано, что обученная основному правилу и исключению система может быть представлена моделью двухслойной нейронной сети с разными матрицами связей для прямой и обратной итераций. Введена функция энергии сети, определяющая циклический характер динамики, и дан анализ диссипативного члена, обуславливающего конвергентность динамики. Показано влияние условий записи голограмм правила «Обобщенный Modus Ponens» и исключения на характер динамики и скорость схождения системы к устойчивому состоянию, характеризующемуся независимостью значения логического заключения от значений входных переменных. Такой тип динамики, ведущий к формированию толерантности, характерен для обыденной формы мышления, нацеленной на устойчивость внутренней картины мира. Для реализации научной формы мышления, ориентированной на адекватность внутренней картины мира реальности, необходим механизм, останавливающий релаксацию сети; этот механизм должен быть внешним по отношению к модели логики. Приведены результаты численного моделирования для условий обучения, адекватных условиям записи реальных голограмм Фурье.

Ключевые слова: информационные технологии, когнитивные механизмы, когнитивный диссонанс, голография, нейронные сети, диссипативный фактор.

LOGIC WITH EXCEPTION ON THE ALGEBRA OF FOURIER-DUAL OPERATIONS: NEURAL NET MECHANISM OF COGNITIVE DISSONANCE REDUCING²

A. Pavlov^b

^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Pavlov@phoi.ifmo.ru

A mechanism of cognitive dissonance reducing is demonstrated with approach for non-monotonic fuzzy-valued logics by Fourier-holography technique implementation developing. Cognitive dissonance occurs under perceiving of new information that contradicts to the existing subjective pattern of the outside world, represented by double Fourier-transform cascade with a hologram – neural layers interconnections matrix of inner information representation and logical conclusion. The hologram implements monotonic logic according to “General Modus Ponens” rule. New information is represented by a hologram of exclusion that implements interconnections of logical conclusion and exclusion for neural layers. The latter are linked by Fourier transform that determines duality of the algebra forming operations of conjunction and disjunction. Hologram of exclusion forms conclusion that is dual to the “General Modus Ponens” conclusion. It is shown, that trained for the main rule and exclusion system can be represented by two-layered neural network with separate interconnection matrixes for direct and inverse iterations. The network energy function is involved determining the cyclic dynamics character; dissipative factor causing convergence type of the dynamics is analyzed. Both “General Modus Ponens” and exclusion holograms recording conditions on the dynamics and convergence of the system are demonstrated. The system converges to a stable status, in which logical conclusion doesn't depend on the inner information. Such kind of dynamics, leading to tolerance forming, is typical for ordinary kind of thinking, aimed at inner pattern of outside world stability. For scientific kind of thinking, aimed at adequacy of the inner pattern of the world, a mechanism is needed to stop the net relaxation; the mechanism has to be external relative to the model of logic. Computer simulation results for the learning conditions adequate to real holograms recording are presented.

Keywords: information technologies, cognitive systems, cognitive dissonance, holography, neural networks, dissipative factor.

¹ Работа выполнена при поддержке РФФИ, грант 12-01-00418-а, и НИР СПб НИУ ИТМО «Исследование эффектов нелинейной голографической записи информации и поиск возможностей реализации биологически мотивированных механизмов креативного мышления».

² The work was done with supporting from the Russian Foundation for Basic Research, grant 12-01-00418-a and SPb NRU ITMO S&R work «Research of nonlinear holographic information record effects and retrieval of implementation facilities for biologically motivated procedures of creative thinking»

Введение

Одно из актуальных направлений развития информационных технологий – реализация моделей и механизмов, обуславливающих проявление феноменов, отчасти аналогичных тем, что присущи биологическому мозгу [1]. В рамках когнитивного подхода к обработке информации на первый план выходят механизмы восприятия и включения новой информации в структуру существующего индивидуального знания [1] – когнитивные механизмы и их внешние проявления – когнитивные феномены.

Один из таких феноменов, когнитивный диссонанс (КД), проявляет конфликт между двумя когнициями – уже наличествующим знанием, представленным в виде внутренней картины мира, и знанием новым, ей противоречащим [2]. С КД неразрывно связан защитный механизм его редуцирования, направленный на обеспечение субъективной непротиворечивости внутренней картины мира [3]. Можно выделить две парадигмы редуцирования КД, соответствующие двум когнитивным стилям – научному и обыденному [4]. Для первого, ориентированного на адекватность картины мира реальности, характерно повышение когнитивной активности для получения новой информации, позволяющей согласовать диссонирующие когниции их дополнением. В этом случае КД может формально рассматриваться как ошибка обучения, устранимая также методами обучения. Для обыденного когнитивного стиля характерен примат устойчивости картины мира, даже вопреки ее адекватности реальности, и КД относится уже более к психологической, чем формально-логической противоречивости [3]. Здесь только дообучение не даст эффекта в силу иных приоритетов носителя данного когнитивного стиля [4].

В качестве модели, описывающей механизм включения новой информации во внутреннюю картину мира, интерес представляют немонотонные логики, в которых заданное при начальном обучении значение логического заключения может меняться при восприятии новой информации [5]. Применительно к феномену КД актуален частный случай немонотонных логик – логики с исключением [6]. Но логика – не только абстрактная модель (алгебра), но в первую очередь – атрибут психической сферы [7]. Соответственно, чтобы претендовать на адекватность, алгебра логики должна моделировать феномены, присущие человеческой психике, в том числе феномен КД и его редуцирования в форме, характерной для обыденного когнитивного стиля.

В числе информационных технологий, соответствующих основным постулатам когнитивного подхода [1], интерес, среди прочих, представляет и голография – как в силу естественного параллелизма оптической обработки двумерных массивов информации, так и в силу ряда аналогий между свойствами голограмм и мозга [8, 9]. В работах [10–15] развит подход к реализации нечетко-значимых логик методом голографии Фурье. Этот подход объединяет формальную модель логико-лингвистического моделирования по Л. Заде [16] с биологически мотивированным методом представления информации паттернами внутренней репрезентации и их обработки [12, 17, 18]. В рамках подхода в [13] дана модель логики с исключением (ЛИ), а в [15] показана реализация на ЛИ феномена, аналогичного КД. В [15] остался открытым вопрос о внутреннем механизме и закономерностях редуцирования КД, их зависимости от свойств материального носителя интеллекта и условий его обучения.

В настоящей работе, в развитие [10–15], показан обусловленный диссипативным фактором имманентный данной модели механизм редуцирования КД, характерный для обыденного когнитивного стиля, и выявлены закономерности динамики в связи с условиями обучения и свойствами голографических регистрирующих сред.

Подход

Для удобства читателя, с учетом междисциплинарного характера работы, кратко изложим основные моменты развиваемого подхода и иллюстрируем формальные построения простым примером. Детали заинтересованный читатель найдет в [10–15].

В живых системах информация из внешнего мира воспринимается сенсорами, преобразуется сенсорами и сенсорными трактами, и на коре головного мозга активируются картины нейронной активности – паттерны внутренней репрезентации информации, которые и обрабатываются мозгом – биологической нейронной сетью [17, 18]. Применение голографии в искусственном интеллекте мотивировано аналогией между обработкой паттернов мозгом и изображений голограммой [4, 7–9]. В частности, $4f$ -схема голографии Фурье (на рис. 1, а, ограничена пунктиром) реализует двухслойную нейронную сеть (на рис. 1, б, ограничена пунктиром) с матрицей связей по модели внешнего произведения векторов, описывающих изображения. Запись голограммы Фурье реализует правило обучения Хебба. Метод наложенных голограмм позволяет связать набор пар векторов без использования статистических методов.

В работах [10, 11, 13] показано, что схема голографии Фурье порождает при обработке изображений алгебру Фурье-дуальных определяющих модель операций. Преобразование Фурье, удовлетворяющее аксиоматическому определению операции, задающей дуальность, обуславливает нечетко-значимость адекватной схеме логики как отображение в модели реального физического явления дифракции [13].

Задача объединения двух подходов – нейросетевого, основанного на обработке паттернов, и формально-алгебраического, основанного на представлении смысла значений лингвистических переменных

нечеткими числами и их обработке по правилам арифметики нечетких чисел [16] – решена в [12] в пространстве Фурье приравниванием спектров амплитуд нечетких чисел (абстрактная модель) и изображений как аналогов паттернов внутренней репрезентации. Этот подход позволил применить аппарат нечетких чисел к описанию обработки произвольных изображений и реализации вывода «Обобщенный Modus Ponens» на универсальной шкале. Применительно к логике в рамках развиваемого подхода формальной операции изменения моды нечеткого числа, описывающего изменение смысла значения лингвистической переменной в [16], соответствует физически реальная операция изменения спектра амплитуд изображения, представляющего значение лингвистической переменной. Поскольку в рамках голографической реализации [11] непосредственная оценка моды нечеткого числа невозможна, она проводится измерением радиуса его корреляции, совпадающего с радиусом корреляции описываемого нечетким числом изображения.

В алгебре логики [10, 11, 13] дизъюнкция реализуется сверткой. Поскольку при обработке изображений их свертка в общем случае – многомодальное распределение, т.е. не удовлетворяет критерию однозначности интерпретации, предъявляемому к логическому заключению, то оператор импликации определен через вычитание. Вычитание (абстрактное) в алгебре [10, 11, 13] реализуется операцией корреляции, результат которой формируется в +1-м порядке дифракции схемы голографии Фурье.

ЛИ обычно строится добавлением к монотонной логике по правилу вывода «Modus Ponens», оператора исключения «если не ...» [7]. Пример ЛИ из повседневной жизни: «Если яблоко большое, то оно хорошее, **если не** перекармлено нитратами».

Первая часть вывода – монотонная логика «Если яблоко большое, то оно хорошее» в [10–12] реализована 4f-схемой голографии Фурье (рис. 1, а, ограничена пунктиром). Связи слоев репрезентаций In и заключений C реализованы каскадом двойного преобразования Фурье L_1, L_2 с голограммой H^{GMP} , записанной паттернами: в слое In – $Im_{Immin}(x,y)$, представляющим минимальное значение лингвистической переменной «размер», и в C – $Im_{Cmin}(\zeta,\eta)$, представляющим связанное с ним значение заключения – лингвистической переменной «качество»:

$$H^{GMP}(v_x, v_y) = \eta^{GMP} \left(F^* \left(Im_{Inmax}(x, y) \right) F \left(Im_{Cmin}(\zeta, \eta) \right) \right), \quad (1)$$

где v_x, v_y – частоты; ζ, η – координаты в слое C ; η^{GMP} – оператор, учитывающий обуславливающую дополнительную фильтрацию на голограмме (1) нелинейность голографической регистрирующей среды, астериск – символ комплексного сопряжения.

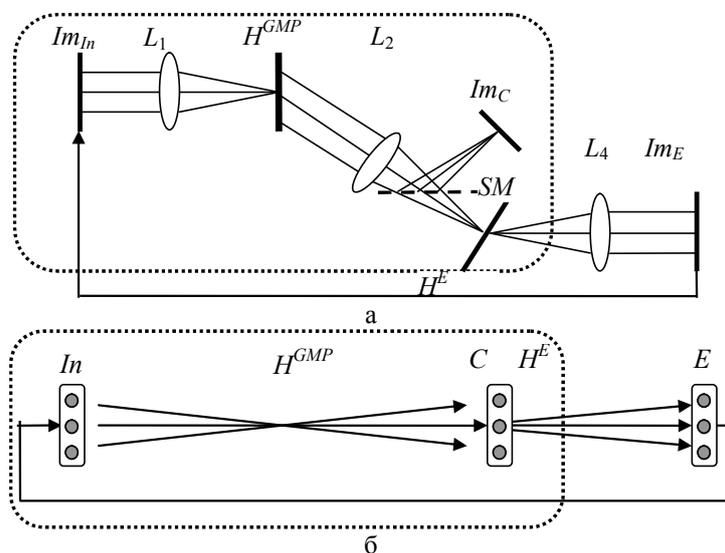


Рис. 1. 6f-схема голографии Фурье (а) и модель трехслойной нейронной сети (б)

ЛИ построена в [14] добавлением к двухслойной нейронной сети, реализующей монотонную логику (рис. 1, б, ограничена пунктиром), слоя исключений E , связанного со слоем C преобразованием Фурье (рис. 1, б). Таким образом, исключение связано с заключением монотонной логики операцией, задающей дуальность определяющих модель операций. Оптическая схема, реализующая данную трехслойную сеть на этапе обучения исключению, представляет собой 6f-схему голографии Фурье (рис. 1, а).

После обучения сети монотонной логике – записи голограммы (1), записывается голограмма (матрица связей) исключения H^E : в слое исключения E предъявляется паттерн исключения $Im_{RE}(\chi_x, \chi_y)$, а в

слое C – паттерн заключения $Im_{CRE}^{GMP}(\zeta, \eta)$, символ R в нижних индексах означает «эталон». Этот паттерн формируется голограммой (1) при предъявлении в слое In паттерна $Im_{In}(x, y)$, с которым связано исключение (в нашем примере – большое яблоко, подозрительное на нитраты):

$$H^E(\zeta, \eta) = \eta^E \left(F \left(Im_{RE}(\chi_x, \chi_y) \right) Im_{CRE}^{GMP}(\zeta, \eta) \right). \quad (2)$$

В данной модели на вход сети поступает только паттерн, представляющий посылку монотонной логики – значение переменной «размер». Значение исключения ассоциативно вызывается из памяти заключением, сформированным по правилу монотонной логики, и, в свою очередь, модифицирует его. Такая модель ЛИ адекватна реальным ситуациям – покупатель может оценить значение переменной «размер», но не может непосредственно оценить значение переменной исключения (уровень нитратов), так как у него нет лабораторного оборудования. Однако ранее полученная информация об опасности нитратов связана в памяти с размером – именно большое яблоко подозрительно на повышенный уровень нитратов. Следовательно, оценка размера вызывает из памяти ассоциированное с ней значение исключения.

Для модификации исключением начального вывода, слой E связан со слоем In так, что паттерн из слоя E , подается в слой без изменений In . При предъявлении в слое In паттерна $Im_{In}(x, y)$, представляющего текущее значение переменной «размер», в слое C формируется паттерн $Im_C^{GMP}(\zeta, \eta)$, представляющий значение лингвистической переменной «качество», полученное по правилу монотонной логики:

$$Im_C^{GMP}(\zeta, \eta) = Im_{In}(x, y) \otimes \eta^{GMP} \left(Im_{In\min}(x, y) \right), \quad (3)$$

где \otimes – символ операции корреляции. Этот паттерн, проходя матрицу связей слоев C и E (дифрагируя на голограмме (2)), формирует в слое E паттерн

$$\begin{aligned} Im_E(\chi_x, \chi_y) &= F \left\{ Im_C^{GMP}(\zeta, \eta) \eta^E \left[F \left(Im_{RE}(\chi_x, \chi_y) \right) Im_{CRE}^{*GMP}(\zeta, \eta) \right] \right\} = \\ &= Im_{RE\eta^E}(\chi_x, \chi_y) * \left(F \left(Im_C^{GMP}(\zeta, \eta) \right) \otimes F \left(Im_{CRE}^{GMP}(\zeta, \eta) \right) \right) \end{aligned} \quad (4)$$

где нижний индекс η^E означает дополнительную фильтрацию вследствие нелинейности голографических регистрирующих сред, а $*$ – символ операции свертки. В выражении (3) значение заключения в слое E связано со значением заключения, формируемым по правилу монотонной логики в слое C , преобразованием Фурье, удовлетворяющим аксиоме невозрастания, входящей в определение задающей дуальность операции [13]. В результате шкала, формируемая голограммой H^E в слое E , инверсна шкале в слое C .

Модификация первоначального заключения осуществляется подачей паттерна (4) на входной слой In . Тогда в слое C формируется новый паттерн – заключение ЛИ:

$$Im_C^E(\zeta, \eta) = Im_E(\chi_x, \chi_y) \otimes Im_{In\min}^{GMP}(x, y). \quad (5)$$

Если при обучении исключению выполнено условие

$$Im_{RE}(\chi_x, \chi_y) \in \{ Im_{In}(x, y) \},$$

то паттерн (5) в области глобального максимума автокорреляционной функции является унимодальным, т.е. удовлетворяет критерию однозначности интерпретации логического заключения и пригоден для подачи на исполнительные органы.

Выражение (5) описывает паттерн, формируемый в одном цикле $Im_{In} \rightarrow Im^{GMP} \rightarrow Im_C^E$. Вместе с тем, в модели нет факторов, останавливающих процесс – паттерн (5) вновь поступает на голограмму (2), и процесс развивается далее. Ниже покажем, что диссипативные свойства данной системы определяют ее динамику, аналогичную редуцированию КД в форме, характерной для обыденного когнитивного стиля – имеет место не изменение внутренней картины мира с учетом новой информации, а формирование толерантности [19], и определим связь характеристик процесса со свойствами элементов схемы.

Динамика системы

Как видно из приведенного выше рассмотрения, третий слой E необходим только для дообучения сети исключению – записи голограммы (2). Полностью обученная система может быть представлена моделью двухслойной сети с разными матрицами связей: (1) – для $In \rightarrow C$ и (2) – для обратной $C \rightarrow In$ итераций. Тогда для обученной сети координаты (χ_x, χ_y) в слое исключения E , совпадающего со слоем In , можно заменить на координаты (x, y) слоя In .

Поскольку матрица связей – голограмма (1) (при описании паттернов как векторов $\mathbf{Im}_{In\min}$ и $\mathbf{Im}_{C\min}$) определена как их внешнее произведение, то для итерации $In \rightarrow C$ определим энергию как частный случай функции Ляпунова посредством скалярного произведения векторов \mathbf{Im}_{In} и \mathbf{Im}_C , описывающих текущее состояние слоев In и C [20]:

$$E_{InC} \propto - \left(\mathbf{Im}_{In} \mathbf{H}^{GMP} \mathbf{Im}_C^* \right). \quad (6)$$

Для итерации $C \rightarrow In$, поскольку голограмма (2) также может быть представлена как внешнее произведение $F(\mathbf{Im}_{CRE}^{GMP})$ и \mathbf{Im}_E , энергию сети определим аналогично:

$$\mathcal{E}_{CIn} \propto -\left(F(\mathbf{Im}_C^{GMP}) \right) \mathbf{H}^E \mathbf{Im}_{In}^* = -\left(F(\mathbf{Im}_{In}) \eta \left(F^* \left(\mathbf{Im}_{In \min \eta}^{GMP} \right) \right) \right) \mathbf{H}^E \mathbf{Im}_{In}^*. \quad (7)$$

Полная энергия системы определяется как сумма (6) и (7):

$$\mathcal{E} = \mathcal{E}_{InC} + \mathcal{E}_{CIn} = -c_{InC} \left(\mathbf{Im}_{In} \mathbf{H}^{GMP} \mathbf{Im}_C^* \right) - c_{CIn} \left(\left(F(\mathbf{Im}_{In}) \eta \left(F^* \left(\mathbf{Im}_{In \min \eta}^{GMP} \right) \right) \right) \mathbf{H}^E \mathbf{Im}_{In}^* \right),$$

где c_{InC} и c_{CIn} – нормировочные коэффициенты. Компонент \mathcal{E}_{InC} определяет конвергентность динамики сети [20] как схождение к аттрактору

$$\begin{cases} \mathbf{Im}_{In} \rightarrow \mathbf{Im}_{In \min}, \\ \mathbf{Im}_C \rightarrow \delta, \end{cases}$$

где δ – дельта-функция, описывающая точечный источник как вектор размерностью 1. Компонент \mathcal{E}_{CIn} , в свою очередь, определяет динамику сети

$$\begin{cases} F(\mathbf{Im}_C^{GMP}) \rightarrow F(\mathbf{Im}_{C \max}^{GMP}), \\ \mathbf{Im}_{In} \rightarrow \mathbf{Im}_{In \max}. \end{cases}$$

Таким образом, имеем противоположное направление динамики системы на встречных итерациях $In \rightarrow C$ и $C \rightarrow In$.

Для анализа динамики примем допущение, что обрабатываемые паттерны описываются как реализации однородных изотропных случайных полей и отношение площадей паттернов и корреляции S_{Im}/S_{Corr} достаточно велико для того, чтобы применительно к глобальным максимумам их автокорреляционных функций можно было пренебречь свойством некоммутативности операции корреляции. Тогда паттерн – заключение ЛИ (5) на n -ой итерации описывается выражением

$$Im_{C(n)}^E(\zeta, \eta) = F \left(Im_{C(n-1)}^E(\zeta, \eta) \right)^* \left\{ \left[Im_{E \eta^E}(x, y) \otimes F \left(Im_{CRE}^{GMP}(\zeta, \eta) \right) \right] \otimes Im_{In \min \eta}^{GMP}(x, y) \right\}, \quad (8)$$

где $Im_C^{GMP}(\zeta, \eta)$ из (4), начиная со второй итерации, мы обозначаем $Im_{C(n-1)}^E(\zeta, \eta)$. В (8) от номера итерации n зависит $F \left(Im_{C(n-1)}^E(\zeta, \eta) \right)$, а член в фигурных скобках является диссипативным фактором, учитывающим дифракцию на голограммах H^{GMP} и H^E .

Для консервативной системы динамика сети в устоявшемся режиме имела бы колебательный характер (6)–(7). Наличие в (8) диссипативного члена обуславливает затухание колебаний – схождение процесса к устойчивому состоянию

$$Im_{C(n)}^E(\zeta, \eta) = F \left(Im_{C(n)}^E(\zeta, \eta) \right)^* \left\{ \left[Im_{E \eta^E}(x, y) \otimes F \left(Im_{CRE}^{GMP}(\zeta, \eta) \right) \right] \otimes Im_{In \min \eta}^{GMP}(x, y) \right\}. \quad (9)$$

Из (9) следует зависимость скорости схождения от характеристик минимального элемента последовательности входных паттернов $Im_{In \min}(x, y)$ и условий обучения системы – условий записи голограмм.

При радиусе корреляции минимального элемента последовательности $Im_{In \min}(x, y)$, сравнимом с дифракционным пределом разрешения, т.е. $r_{In \min \eta}^{GMP}(\zeta, \eta) \approx \delta(\zeta, \eta)$, голограмма H^{GMP} будет иметь постоянную во всем частотном диапазоне дифракционную эффективность. Если голограмма H^E записана паттерном $Im_{RE} = Im_{In \min}$, то также $r_E(x, y) \approx \delta(x, y)$, и в части, обусловленной выбором паттерна исключения при обучении системы, диссипативный член минимален, поскольку

$$Im_{E \eta^E}(x, y) \otimes Im_{In \min \eta}^{GMP}(x, y) = \delta(\zeta, \eta).$$

В [14] показано, что для обеспечения максимальной чувствительности логики целесообразно проводить обучение системы исключению паттерном $Im_{CRE}^{GMP} = Im_{C \max}^{GMP}$. Поскольку $Im_{C \max}^{GMP}$ – максимальный элемент последовательности $\{Im_C\}$, то

$$\left| F \left(Im_{CRE}^{GMP}(\zeta, \eta) \right) \right| = \delta(v_x, v_y).$$

При таком выборе условий обучения монотонной логике исключению диссипативный член в (9) минимален:

$$\left\{ \left[Im_{E \eta^E}(x, y) \otimes F \left(Im_{CRE}^{GMP}(\zeta, \eta) \right) \right] \otimes Im_{In \min \eta}^{GMP}(x, y) \right\} \approx \delta(x, y).$$

В пределе этот член теряет диссипативный характер – система стремится к консервативности, а ее аттрактор – к циклическому. Однако в силу внутренней коррелированности как атрибута информации,

отличающего ее от шума, при обучении системы должны выполняться условия, исключающие разрушение внутренней коррелированности, в первом приближении эти условия могут быть определены [21] как

$$r_{In\min}^{GMP}(\zeta, \eta) \geq 3\delta(\zeta, \eta), \quad r_E(x, y) \geq 3\delta(x, y).$$

По мере увеличения радиусов корреляций паттернов расти будет также и член (9) – скорость затухания колебаний будет возрастать. При обучении сети исключению паттернами $Im_{RE}(x, y) = Im_{In\max}(x, y)$ и $F(Im_{CRE}^{GMP}(\zeta, \eta)) = F(Im_{Cmin}^{GMP}(\zeta, \eta))$ оба паттерна являются максимальными элементами соответствующих последовательностей, и член (9) будет иметь максимальное значение, а скорость конвергенции – максимальна.

Радиусы корреляций паттернов, входящих в диссипативный член, определяются как радиусами корреляции исходных паттернов, так и условиями записи голограмм вкупе с экспозиционными характеристиками голографических регистрирующих сред. В силу ограниченности динамического диапазона регистрирующих сред голограммы (1) и (2) всегда записываются в ограниченном частотном диапазоне [22–24]. Высокочастотная фильтрация ведет к убыванию радиуса корреляции и, тем самым, к убыванию значения диссипативного члена, низкочастотная – к его росту.

Из (9) видно, что устойчивое состояние не зависит от начальной точки процесса релаксации – для всех значений входных переменных система релаксирует к одному аттрактору, определяемому диссипативным членом в (9). Иными словами, формируется толерантность – одинаковое значение Im_C^E для всех значений набора $\{Im_{In}\}$ [19].

Численное моделирование

Моделировалась динамика системы (рис. 1) при обработке паттернов – реализаций случайного процесса с гауссовым спектром амплитуд и случайным спектром фаз, равномерно распределенных на интервале $[0, 2\pi]$. Размер реализаций – 1024 пикселей. Голограмма (1) записывалась паттернами $Im_{In\min}$, голограмма (2) – $Im_{RE} = Im_{In\min}$ и Im_{Cmax}^{GMP} – заключением монотонной логики, полученным при максимальном значении входной лингвистической переменной (ЛПП) $Im_{In\max}$. В рамках примера предьявлялась последовательность паттернов $\{Im_{In}\} = \{Im_{Size}\} = \{Im_{Small}, Im_{Little}, Im_{Medium}, Im_{Big}\}$, отличающихся, согласно [10, 12], спектрами амплитуд с одинаковыми спектрами фаз. Условия записи голограмм и свойства регистрирующих сред моделировались введением дополнительных высокочастотных и низкочастотных фильтров Гаусса.

Условия моделирования применительно к голограмме H^{GMP} даны в таблице. Для двух серий наборов входных ЛП с радиусами корреляции $\{r_{0,606}^a\} = \{14, 27, 53, 105\}$ (а) и $\{r_{0,606}^b\} = \{9, 18, 35, 70\}$ (б) введены значения радиуса корреляции минимального элемента $Im_{In\min}$ – импульсного отклика системы $\delta_{0,606}$, описывающего фильтрацию на голограмме (1), в пикселях по уровню 0,606 и нормированные на длину реализации параметры фильтра Гаусса – центральная частота ν_0 и ширина по уровню 0,606 – $\nu_{0,606}$.

На рис. 2, а, приведены графические представления логик, порождаемых разными голограммами H^{GMP} (строки 1, 4 и 7 в таблице) при линейных H^E – три семейства зависимостей, связывающих входную (ось абсцисс) и выходную (ось ординат) лингвистические шкалы для следующих логик: монотонной (возрастающие зависимости), с исключением на первой итерации (убывающие зависимости) и на 10-й итерации. Справа и сверху даны метрические шкалы, соответствующие лингвистическим. Градуировка метрических шкал согласно субъективным лингвистическим шкалам эксперта задается при обучении [12, 13]. Видно, что к 10-й итерации система при линейной 4 и низкочастотной 7 голограммах H^{GMP} приходит в состояние толерантности (9) – для всех значений входной переменной формируется одно значение выходной переменной. При высокочастотной голограмме H^{GMP} диссипативный член имеет меньшее значение – скорость конвергенции ниже, и к 10-й итерации система еще не достигла состояния (9).

На рис. 2, б, приведены примеры динамики для этих моделей голограммы H^{GMP} . Для наглядности приведена динамика для максимального элемента последовательности значений входных лингвистических переменных «большое». Для других значений динамика имела такой же характер, сходилась, как и следует из (8), к устойчивому состоянию, отличаясь, в соответствии со значением входной переменной, меньшим размахом колебаний. Для оценки скорости схождения системы зависимости значения (8) по уровню 0,606 от номера итерации n аппроксимировались экспонентой $Im_{C0,606}^E(n) = y_0 + \exp(-An)$, значения коэффициента A и количество итераций N_m , за которое система достигает устойчивого состояния, также даны в таблице.

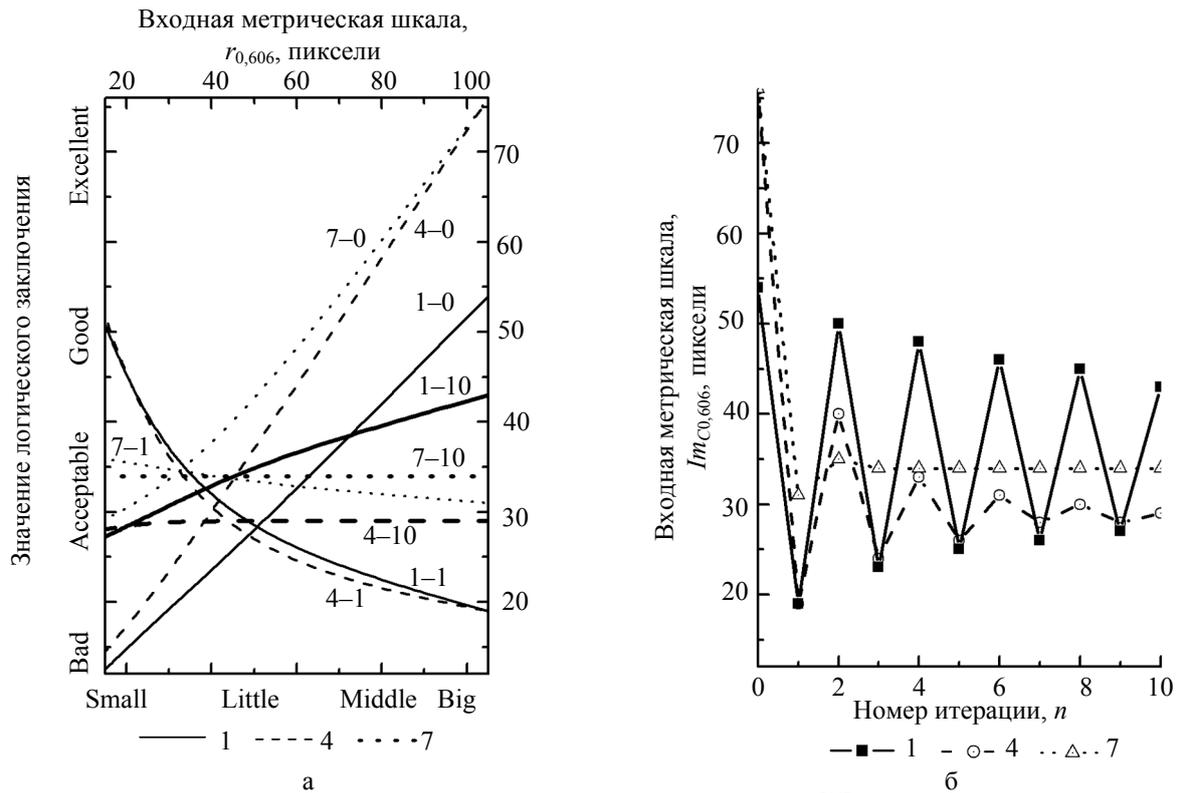


Рис. 2. Динамика системы для ряда условий записи голограммы H^{GMP} (табл.): а – графические представления логик зависимостями значений на выходных шкалах (лингвистической – слева и метрической – справа) от значений на шкалах входных (лингвистической – снизу и метрической – сверху); первая цифра в обозначениях кривых – номер строки в таблице, вторая – номер итерации: 1-0, 4-0, 7-0 – логики монотонные, 1-1, 4-1, 7-1 – ЛИ, 1-ая итерация, 1-10, 4-10, 7-10 – ЛИ, 10-я итерация; б – динамика системы как зависимость заключения, формируемого при предъявлении на входе максимального значения, от номера итерации; номер кривой соответствует строке таблицы

Номер	$\delta_{0,606}$, пиксели		Параметры фильтра			A		N_m	
	а	б	Тип фильтра	ν_0	$\nu_{0,606}$	а	б	а	б
1	12	8	ВЧ	30	20	0,14	0,2	>10	>10
2	13	9	ВЧ	20	20	0,52	0,37	>10	>10
3	13	9	ВЧ	10	20	0,66	0,56	>10	>10
4	14	9	нет	–	–	0,74	0,57	10	>10
5	19	17	НЧ	30	40	0,88	0,71	8	9
6	24	19	НЧ	20	40	1,1	0,94	6	6
7	29	28	НЧ	20	20	1,85	1,45	3	3

Таблица. Параметры динамики в зависимости от характеристик голограммы H^{GMP} (ВЧ – высокочастотный фильтр, НЧ – низкочастотный фильтр)

Из таблицы видно, что с ростом радиуса корреляции минимального элемента, используемого для записи голограммы H^{GMP} – импульсного отклика системы, растет и значение параметра A , описывающего скорость схождения системы к устойчивому состоянию (9). Радиус корреляции, в свою очередь, определяется как собственно входным паттерном (серия а или б), так и условиями записи голограммы H^{GMP} , обуславливающими тот или иной вид дополнительной фильтрации.

Заключение

Таким образом, динамический характер нейронной сети, реализующей логику с исключением на алгебре Фурье-дуальных определяющих модель операций, обуславливает колебательный тип динамики как частный случай когнитивного диссонанса, возникающего при восприятии информации, противоречащей уже сложившейся внутренней картине мира, представленной монотонной логикой. Редуцирование когнитивного диссонанса определяется диссипативным характером системы – при оптической реализации диссипативным фактором является фундаментальное физическое явление дифракции на голограммах мо-

нотонной логики и исключения. Поскольку реализуемая алгебра логики основана на Фурье-дуальности определяющих модель операций, а преобразование Фурье описывает дифракцию, то данный диссипативный фактор и соответственно механизм редуцирования когнитивного диссонанса, имманентен модели. Скорость редуцирования как схождения к устойчивому состоянию определяется свойствами голографических регистрирующих сред и условиями записи голограмм, интегральной оценкой здесь может служить импульсный отклик системы, описывающий степень диссипативности системы в терминах дифракционного уширения точки.

Этот механизм может быть соотнесен с характерным для обыденного когнитивного стиля конфликтом между стремлением к устойчивости внутренней модели мира и необходимостью учета нового знания – отказа от сложившихся стереотипов. Если система релаксирует, то она приходит в устойчивое состояние безразличия или толерантности – значение заключения логики с исключением не зависит от значения входных лингвистических переменных. Уровень этой толерантности определяется операторами η^{GMP} и η^E , описывающими свойства регистрирующих сред и условиями обучения.

Этот сценарий, характерный для обыденного когнитивного стиля, неприемлем в рамках научного стиля познания – здесь необходим механизм, останавливающий итерационную процедуру. Скорее всего, он должен быть внешним по отношению к модели и может быть соотнесен с присущим интеллекту высокого уровня механизмом саморефлексии. Отметим, что остановка итерационного процесса – лишь первый шаг формирования новой внутренней картины мира. Эта картина должна быть не только адекватна реальности, но и внутренне непротиворечива. Последнее обуславливает, вероятно, реализацию новой картины мира не в виде двух голограмм, поскольку они связаны оператором, задающим дуальность, т.е. противоречат друг другу, но одной голограммы. Иными словами, если это предположение верно, то логика с исключением – закономерный и необходимый, но лишь промежуточный этап процесса познания как процесса формирования и постоянной корректировки внутренней картины мира, адекватной реальности.

Литература

1. Кузнецов О.П. Когнитивная семантика и искусственный интеллект // Искусственный интеллект и принятие решений. 2012. № 4. С. 32–42.
2. Психологическая энциклопедия / Под ред. Р. Корсини, А. Ауэрбаха. 2-е изд. СПб: Питер, 2006. 1096 с.
3. Хекхаузен Х. Мотивация и деятельность. 2-е изд. СПб: Питер; М.: Смысл, 2003. 860 с.
4. Кузнецов О.П. Быстрые процессы мозга и обработка образов // Новости искусственного интеллекта. 1998. № 2 [Электронный ресурс]. Режим доступа: <http://raai.org/library/ainews/1998/2/DISTR.ZIP>, свободный. Яз. рус. (дата обращения 5.12.2013).
5. Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах. 2-е изд. М.: ФИЗМАТЛИТ, 2008. 712 с.
6. Reiter R. A logic for default reasoning // Artificial Intelligence. 1980. V. 13. N 1–2. P. 81–132.
7. Кузнецов О.П. Неклассические парадигмы в искусственном интеллекте // Изв. РАН. Сер. Теория и системы управления. 1995. № 5. С. 3–23.
8. Прибрам К. Нелокальность и локализация: голографическая гипотеза о функционировании мозга в процессе восприятия и памяти // Синергетика и психология. В.1. Методологические вопросы. М.: МГСУ «Союз», 1997. С. 156–183.
9. Кузнецов О.П., Марковский А.В., Шипилина Л.Б. Голографические механизмы обработки образной информации. М.: Институт проблем управления им. В.А.Трапезникова РАН, 2007. 82 с.
10. Павлов А.В. Математические модели оптических методов обработки информации // Изв. РАН. Сер. Теория и системы управления. 2000. № 3. С. 111–118.
11. Павлов А.В. Об алгебраических основаниях Фурье-голографии // Оптика и спектроскопия. 2001. Т. 90. № 3. С. 515–520.
12. Алексеев А.М., Константинов А.М., Павлов А.В. Использование метода Фурье-голографии для моделирования принципа образности мышления // Оптический журнал. 2006. Т. 73. № 9. С. 77–82.
13. Павлов А.В. Об алгебраических основаниях голографической парадигмы в искусственном интеллекте: алгебра Фурье-дуальных операторов // Сборник научных трудов V Международной научно-практической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте». М.: ФИЗМАТЛИТ, 2009. Т. 1. С. 140–148.
14. Павлов А.В. Алгебра Фурье-дуальных операций: логика с исключением // Искусственный интеллект и принятие решений. 2012. № 3. С. 26–38.
15. Павлов А.В. Логика с исключением на алгебре Фурье-дуальных операций: феномен сомнений и колебаний // Сборник научных трудов VII Международной научно-практической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте». М.: ФИЗМАТЛИТ, 2013. Т. 3. С. 1035–1045.

16. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. Сер. Математика. Новое в зарубежной науке. М.: Мир, 1976. В. 3. 167 с.
17. Борисюк Г.Н., Борисюк Р.М., Казанович Я.Б., Иваницкий Г.Р. Модели динамики нейронной активности при обработке информации мозгом – итоги «десятилетия» // Успехи физических наук. 2002. Т. 172. № 10. С. 1189–1214.
18. Тушканов Н.Б., Тушканова О.Н. К построению мультисенсорных систем: принципы работы неокортекса головного мозга при распознавании объектов внешнего мира // Труды конгресса по интеллектуальным системам и информационным технологиям (IS-IT'12). М.: ФИЗМАТЛИТ, 2012. Т. 2. С. 373–378.
19. Общая психология: Словарь / Под ред. А.В. Петровского. М.: Пер Сэ, 2005. 251 с.
20. Магницкий Н.А. Распознавание образов распределенными динамическими системами // Доклады академии наук. 1994. Т. 338. № 3. С. 320–321.
21. Шубников Е.И. Отношение сигнал/помеха при корреляционном сравнении изображений // Оптика и спектроскопия. 1987. Т. 62. № 2. С. 450–456.
22. Кулешов А.М., Шубников Е.И., Смаева С.А. Об оптимальности голографического согласованного фильтра // Оптика и спектроскопия. 1986. Т. 60. № 6. С. 1273–1276.
23. Александрина С.А., Кулешов А.М. Влияние режекции низких пространственных частот спектра на параметры сигнала в голографическом корреляторе // Оптика и спектроскопия. 1990. Т. 68. № 3. С. 652–655.
24. Кулешов А.М., Шубников Е.И. Влияние нелинейности среды и пространственных ограничений фильтра на параметры сигнала в голографическом корреляторе // Оптика и спектроскопия. 1986. Т. 60. № 3. С. 606–609.

Павлов Александр Владимирович – кандидат технических наук, доцент, старший научный сотрудник, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Pavlov@phoi.ifmo.ru

Alexander Pavlov – PhD, Associate professor, Senior scientific researcher, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Pavlov@phoi.ifmo.ru

УДК 535.55

МЕТОД ИССЛЕДОВАНИЯ ЗАВИСИМОСТИ
 h -ПАРАМЕТРА АНИЗОТРОПНОГО СВЕТОВОДА ОТ РАДИУСА ИЗГИБА¹О.А. Шрамко^а, А.В. Рупасов^а, Р.Л. Новиков^а, С.М. Аксарин^а^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, shramko_oa@mail.ru

Рассмотрены известные методы измерения h -параметра волокна, разработан новый интерферометрический метод. На его основе предложен метод исследования зависимости h -параметра анизотропного оптического волокна от радиуса изгиба. Микроизгибы волокна, являющиеся дефектами намотки чувствительного элемента волоконно-оптического гироскопа, приводят к ухудшению ряда его оптических характеристик, в частности, коэффициента экстинкции и коэффициента перекрестной связи мод, что ухудшает точностные параметры волоконно-оптического гироскопа. Предложен способ повышения точности измерения и адаптации процесса измерения h -параметра к существующей технологии производства волоконно-оптических гироскопов. Механизм влияния поляризационных свойств на точностные параметры волоконно-оптического гироскопа описывается через анализ вторичных волн, индуцируемых преобразованием поляризации в волоконном контуре. Дана оценка влияния измеряемого параметра на точностные характеристики волоконно-оптического гироскопа. Для гироскопа с коэффициентом экстинкции поляризатора 27 дБ, h -параметром волокна $2 \cdot 10^{-5} \text{ м}^{-1}$, длиной волоконного контура 1500 м и длиной деполаризации волокна 4 мм амплитудная фазовая ошибка не превысит $3,2 \cdot 10^{-5}$ рад, а фазовая ошибка интенсивности составит $2 \cdot 10^{-10}$ рад.

Ключевые слова: оптическое волокно, коэффициент экстинкции, h -параметр.ANALYSIS METHOD OF ANISOTROPIC LIGHTGUIDE h -PARAMETER
DEPENDENCE ON ITS BENDING RADIUS²O. Shramko^b, A. Rupasov^b, R. Novikov^b, S. Aksarin^b^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, shramko_oa@mail.ru

The paper deals with the known methods for h -parameter measuring of the fiber and the new developed interferometry method. Analysis method of the anisotropic lightguide h -parameter dependence on its bending radius for the optical fiber is proposed. Micro bends of a fiber, being the defects of a sensitive element winding in a fiber-optic gyroscope (FOG), lead to the deterioration of a number of optical characteristics, in particular, the extinction coefficient and the cross-coupling coefficient. As a result, it worsens the FOG accuracy. The way of increasing the measurement accuracy and adaptation of the h -parameter measuring process to the existing technology of fiber-optic gyroscopes production is proposed. The mechanism of the polarization properties influence on the FOG accuracy is described through the analysis of secondary waves induced by the polarization transformation in the optical circuit. The influence estimation of the measured parameter on the FOG accuracy characteristics is given. For the gyro with the polarizer extinction coefficient equal to 27 dB, h -parameter $2 \cdot 10^{-5} \text{ м}^{-1}$, the length of the fiber loop 1500 m and the length of depolarization 4 mm the amplitude phase error does not exceed $3.2 \cdot 10^{-5}$ rad and the intensity phase error is $2 \cdot 10^{-10}$ rad.

Keywords: optical fiber, interference, extinction coefficient, h -parameter, lightguide.

Введение

Одной из основных задач, стоящих перед разработчиками волоконно-оптического гироскопа (ВОГ), является изготовление чувствительного элемента. Его основная часть – волоконный контур, представляющий собой многослойную и многovitковую катушку оптического волокна, создаваемую путем квадрупольной намотки световода на каркас. При этом радиус изгиба волокна равен [1]

$$r_N = \frac{D_K}{2} + \frac{D_B}{2} + (N-1) \cdot \frac{\sqrt{3}}{2} \cdot D_B. \quad (1)$$

В выражении (1) r_N – радиус изгиба волокна; N – номер слоя контура; D_K – диаметр каркаса; D_B – диаметр волокна.

Традиционно в ВОГ высокой точности используется анизотропное оптическое волокно [2], в котором благодаря наличию напрягающей эллиптической оболочки существуют две ортогональные оптические оси с различными друг от друга показателями преломления, их называют быстрой осью и медленной осью. Структура волокна приведена на рис. 1. При вводе линейно-поляризованного излучения в оптиче-

¹ Работа выполнена в НИУ ИТМО и ОАО «Концерн «ЦНИИ «Электроприбор» при финансовой поддержке Министерства образования и науки Российской Федерации (проект № 02.G25.31.0044).

² The work was done in NRU ITMO and «Concern «CSRI «Elektroprigor» PLC with financial support from the Ministry of education and science of the Russian Federation (project № 02.G25.31.0044).

скую ось такое волокно способно сохранять состояние поляризации света. В том случае, когда линейно-поляризованное излучение вводится в световод под углом к оптическим осям, возбуждаются две ортогональные моды.



Рис. 1. Структура волоконного световода, сохраняющего поляризацию излучения

Изгиб волокна приводит к ухудшению ряда его оптических характеристик, в частности, коэффициента экстинкции и коэффициента перекрестной связи мод, используемых для оценки степени сохранения поляризации распространяющегося в световоде излучения. Другими словами, по значениям этих параметров можно судить о том, какая часть оптической мощности перетекает из основной моды в паразитную ортогональную. Подробнее этот процесс описан ниже.

Механизм случайной связи поляризационных мод в анизотропном оптическом волокне может быть представлен с помощью простой модели. Перекачка оптической мощности из одной ортогональной моды в другую происходит в дискретных точках M_i , расположенных в пределах одной длины деполаризации L_d на протяжении всего волновода. При длине волокна L количество таких точек будет равно L/L_d . Каждая пара точек преобразования, расположенных симметрично относительно середины волокна, возбуждает две пары вторичных волн (в каждом направлении), которые вносят одинаковую фазовую ошибку $\Delta\varphi_{ei}$. Среднее значение фазовой ошибки близко к нулю, а ее среднее квадратичное отклонение зависит от ширины спектра источника света, коэффициента экстинкции поляризатора и h -параметра волокна. Шум, создаваемый фазовой ошибкой, приводит к нестабильности смещения и увеличению минимальной измеряемой угловой скорости [3, 4].

В настоящее время для измерения h -параметра используется несколько способов [5, 6], каждый из которых имеет свои существенные недостатки, такие как невозможность использовать световоды малой длины, жесткие требования к характеристикам поляризаторов и источников излучения, сложности, связанные с внедрением в налаженный и отработанный технологический процесс производства волоконно-оптических гироскопов, и др. Описанный в работе метод лишен ряда этих недостатков.

В рамках настоящей работы был проведен анализ литературы, в ходе которого было рассмотрено несколько методов измерения степени сохранения поляризации. Для измерения степени сохранения поляризации обычно используется методика скрещенных поляризаторов: с помощью входного поляризатора обеспечивается возбуждение в световоде только одной поляризационной моды, а путем вращения выходного поляризатора измеряется соотношение мощностей на выходе световода. В работе [5] описан другой метод, основанный на применении когерентного источника и измерении интерференционных колебаний, вызванных дополнительной фазовой модуляцией на входе световода.

Указанные методы либо не применимы к коротким длинам волокон, либо требуют дополнительного оборудования, т.е. трудно встраиваются в отработанный технологический процесс производства ВОГ. Суть предложенного метода изложена ниже.

Представленный метод исследования

Исследуемое волокно наматывается на катушки различных диаметров с постоянным натяжением в один слой. Линейно-поляризованный свет вводится строго в оптическую ось исследуемого волокна. После прохождения всей длины световода излучение попадает в поляризационный интерферометр Майкельсона. Зная параметры получившейся интерференционной картины, а именно – наибольшее и наименьшее значение оптической мощности, можно вычислить отношение интенсивностей паразитной моды и изначально возбужденной моды.

На рис. 2 представлена разработанная схема установки для измерения h -параметра.

При распространении одной поляризационной моды в волокне происходит перекачка мощности в ортогональную моду, а также возникает набег разности фаз между ними. С помощью микропозиционера с закрепленным на нем зеркалом в одно из плеч интерферометра вносится соответствующая фазовая задержка, что позволяет наблюдать интерференционную картину ортогональных поляризационных мод (рис. 3).

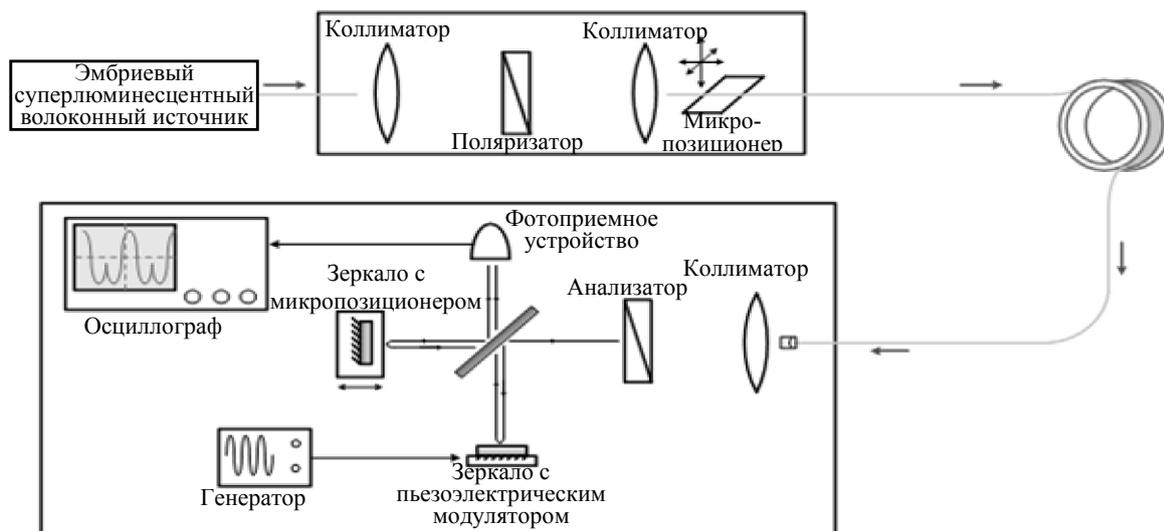


Рис. 2. Схема установки для измерения h -параметра

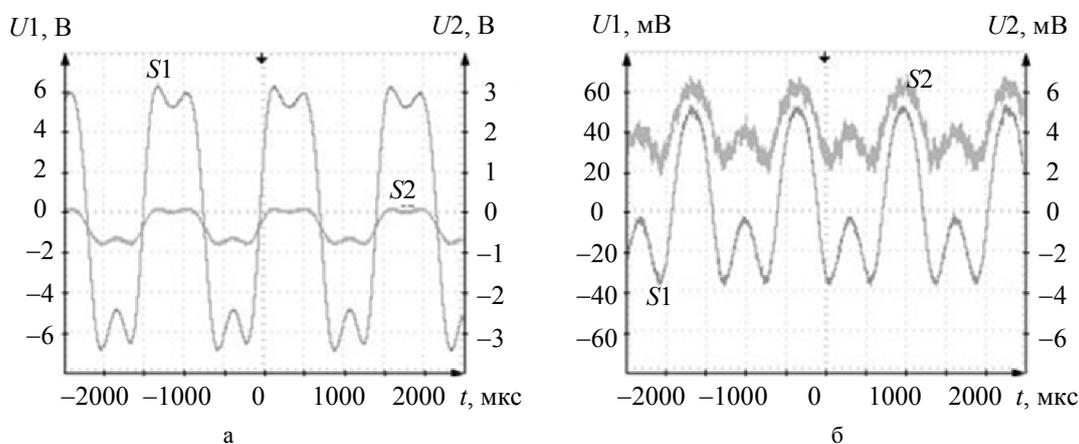


Рис. 3. Осциллограммы интерференционного сигнала: при вводе линейно-поляризованного излучения в световод под углом 45° к оптическим осям (а); при вводе линейно-поляризованного излучения строго в оптическую ось (S1 – интерференционный сигнал в режиме переменного тока, S2 – интерференционный сигнал в режиме постоянного тока) (б)

Параметры полученной интерференционной картины зависят от интенсивностей мод на выходе исследуемого волокна следующим образом:

$$I_{\min} = \frac{1}{2}I_1 + \frac{1}{2}I_2 - 2\sqrt{\frac{I_1}{4} \times \frac{I_2}{4}}, \quad (2)$$

$$I_{\max} = \frac{1}{2}I_1 + \frac{1}{2}I_2 + 2\sqrt{\frac{I_1}{4} \times \frac{I_2}{4}}, \quad (3)$$

где I_{\min} , I_{\max} – интенсивность интерференционного сигнала при деструктивной и конструктивной интерференции соответственно; I_1 – интенсивность первоначально возбужденной моды; I_2 – интенсивность паразитной моды.

Из формул (2) и (3) получаем выражение для искомого отношения:

$$\frac{I_2}{I_1} = \frac{(I_{\max} + I_{\min}) - 2\sqrt{I_{\max} \times I_{\min} - \frac{3}{4}(I_{\max} - I_{\min})^2}}{(I_{\max} + I_{\min}) + 2\sqrt{I_{\max} \times I_{\min} - \frac{3}{4}(I_{\max} - I_{\min})^2}}. \quad (4)$$

Таким образом, зная I_{\min} и I_{\max} и используя выражение (4), можно определить h -параметр световода.

Для проведения исследования зависимости степени сохранения поляризации от радиуса контура предлагается наматывать оптическое волокно на каркасы различных диаметров и проводить измерения согласно описанному выше методу.

Оценка влияния h -параметра на точностные характеристики ВОГ

Механизм влияния поляризационных свойств на точностные параметры ВОГ описывается через анализ вторичных волн, индуцируемых преобразованием поляризации в волоконном контуре. Традиционно искажения оптического сигнала в волоконной катушке, вызванные поляризационной невязанностью, разделяют на амплитудные ошибки и ошибки интенсивности [7]. Амплитудная ошибка возникает, когда свет, поляризованный ортогонально оси пропускания поляризатора при распространении по волоконному контуру, переходит в основную поляризационную моду на выходе контура. Амплитудная фазовая ошибка характеризуется следующим выражением:

$$|\Delta\varphi_e| < 2\varepsilon \times \sqrt{\rho_{in} \times \rho_{cr}}. \quad (5)$$

В выражении (5) $\rho_{in} = \frac{|A_2|^2}{|A_1|^2}$, $\rho_{cr} = \frac{|A_{12}|^2}{|A_{11}|^2}$, A_1 и A_2 – амплитуда оптического сигнала, поляризованного вдоль и поперек оси пропускания поляризатора, при входе в контур соответственно, A_{11} и A_{12} – амплитуда оптического сигнала, поляризованного вдоль и поперек оси пропускания поляризатора, при выходе из контура соответственно. Для оценки примем, что в волокно вводится оптическая мощность P , тогда отношение мощности оптических сигналов $|A_1|^2$ и $|A_2|^2$ соответствует (6), а отношение мощности оптических сигналов $|A_{11}|^2$ и $|A_{12}|^2$ рассчитывается как результат преобразования основной поляризационной моды в неосновную в соответствии со значением h -параметра волокна. Таким образом,

$$\frac{|A_2|^2}{|A_1|^2} = \frac{P \cdot \varepsilon}{P \cdot (1 - \varepsilon)} = \frac{\varepsilon}{(1 - \varepsilon)}, \quad (6)$$

$$\frac{|A_{12}|^2}{|A_{11}|^2} = \frac{A_2^2 \cdot (1 - h \cdot L) + A_1^2 \cdot h \cdot L}{A_1^2 \cdot (1 - h \cdot L) + A_2^2 \cdot h \cdot L} = \frac{\varepsilon + h \cdot L - 2 \cdot \varepsilon \cdot h \cdot L}{1 - \varepsilon - h \cdot L + 2 \cdot \varepsilon \cdot h \cdot L}, \quad (7)$$

где h – h -параметр оптического анизотропного волокна чувствительного элемента; ε – коэффициент экстинкции поляризатора; L – длина волоконного контура. Предположим, что коэффициент экстинкции поляризатора составляет 27 дБ ($\varepsilon = 0,002$), h -параметр волокна – $2 \cdot 10^{-5} \text{ м}^{-1}$, длина волоконного контура – $L = 1500$, тогда с учетом (5)–(7) получим, что максимальная амплитудная фазовая ошибка составляет

$$|\Delta\varphi_e|_{\max} = 2\varepsilon \cdot \sqrt{\frac{\varepsilon}{(1 - \varepsilon)} \cdot \frac{\varepsilon + h \cdot L - 2 \cdot \varepsilon \cdot h \cdot L}{1 - \varepsilon - h \cdot L + 2 \cdot \varepsilon \cdot h \cdot L}} \approx 3,2 \cdot 10^{-5} \text{ рад}.$$

Для контура диаметром 200 мм и длины волны 1550 нм указанное значение соответствует погрешности по скорости вращения, приблизительно равной $8 \cdot 10^{-6} \text{ рад/с}$.

Ошибка интенсивности возникает, когда свет, поляризованный вдоль оси пропускания поляризатора на входе в контур, при распространении по волоконному контуру переходит в неосновную поляризационную моду на выходе контура. Среднее квадратичное отклонение ошибки интенсивности характеризуется следующим выражением:

$$\sigma_{\Delta\varphi_e} = \varepsilon^2 \cdot h \cdot L_d \cdot \sqrt{N} = \frac{\varepsilon^2 \cdot h \cdot L}{\sqrt{N}}. \quad (8)$$

В выражении (8) L_d – длина деполяризации анизотропного волокна, N – отношение длины волоконного контура к длине деполяризации. Для $\varepsilon = 0,002$, $h = 2 \cdot 10^{-5} \text{ м}^{-1}$, $L = 1500$ м, $L_d = 4$ мм получим

$$\sigma_{\Delta\varphi_e} = \frac{\varepsilon^2 \cdot h \cdot L}{\sqrt{N}} = \frac{\varepsilon^2 \cdot h \cdot L}{\sqrt{L/L_d}} \approx 2 \cdot 10^{-10} \text{ рад}.$$

Для контура диаметром 200 мм и длины волны 1550 нм указанное значение соответствует погрешности по скорости вращения, приблизительно равной $5 \cdot 10^{-11} \text{ рад/с}$.

Также выделяют ошибки первого рода как разновидность амплитудной ошибки, зависящие от коэффициента экстинкции поляризатора по амплитуде ε , и ошибки второго рода как разновидность ошибки интенсивности, зависящие от коэффициента экстинкции поляризатора по интенсивности ε^2 [8].

Во время распространения в чувствительной катушке ортогональные компоненты A и B производят множество вторичных волн из-за перехлестов волокна в чувствительной катушке, при этом накапливается ошибка, являющаяся результатом интерференции между волнами нулевого, второго, четвертого порядка (вторичными волнами компоненты A) и волнами первого, третьего, пятого порядка (вторичными волнами компоненты B). Наибольшее влияние на выходной сигнал ВОГ оказывает интерференция между волной нулевого порядка компоненты A и волной первого порядка компоненты B . Эта ошибка может быть компенсирована при помощи методов [9, 10], однако существуют более сложные комбинации вторичных волн, обусловленные преобразованием поляризации в чувствительной катушке, которые также

вызывают смещение рабочей точки гироскопа. Ошибкой первого рода называется результат интерференции между вторичной волной второго порядка компоненты A и вторичной волной первого порядка компоненты B . Влияние интерференции остальных комбинаций волн принимается пренебрежимо малым и не учитывается. Ошибка второго порядка является результатом интерференции между вторичными волнами первого, третьего и пятого порядков компонент A и B .

Согласно работе [8], моделирование для рассмотренных эффектов показывает, что ошибка второго рода имеет меньшее значение по сравнению с ошибкой первого рода, обе ошибки снижаются с улучшением поляризационных свойств контура и могут быть полностью нивелированы, если коэффициент преобразования поляризации достаточно мал.

Заключение

На основе обзора методов измерения степени сохранения поляризации [5, 6] разработан интерференционный метод измерения h -параметра. На его основе предложен метод исследования зависимости h -параметра анизотропного оптического волокна от радиуса изгиба. Представленный метод имеет ряд преимуществ по сравнению с уже существующими: позволяет снизить требования к разрешающей способности фотодетектора, работоспособен при низком коэффициенте экстинкции анализатора, адаптирован к существующей технологии производства интерферометров волоконно-оптического гироскопа. Оценка влияния измеряемого параметра на точностные характеристики ВОГ с коэффициентом экстинкции поляризатора 27 дБ, h -параметром волокна $2 \cdot 10^{-5} \text{ м}^{-1}$, длиной волоконного контура 1500 м и длиной деполяризации волокна 4 мм показала, что амплитудная фазовая ошибка не превысит $3,2 \cdot 10^{-5}$ рад, а фазовая ошибка интенсивности составит $2 \cdot 10^{-10}$ рад.

Литература

1. Мешковский И.К., Киселев С.С., Куликов А.В., Новиков Р.Л. Дефекты намотки оптического волокна при изготовлении чувствительного элемента волоконно-оптического интерферометра // Изв. вузов. Приборостроение. 2010. Т. 53. № 2. С. 47–51.
2. Li Y., Chen X. Effect of highly birefringence fibers on fiber optic gyroscope // Proc. of SPIE. 2007. V. 6423. 64233S-1–64233S-6.
3. Gu H., Yang G., Yang Y., Weng H., Zhao Q. Analysis and simulation of optical polarization fluctuation of interferometric fiber optic gyroscope // Proc. of SPIE. 2007. V. 6595. 65953N-1–65953N-6.
4. Wang X., He Z., Hotate K. Reduction of polarization-fluctuation induced drift in resonator fiber optic gyro by a resonator with twin 90° polarization-axis rotated splices // Optics Express. 2010. V. 18. N 2. P. 1677–1683.
5. Котов О.И., Ликумович Л.Б., Медведев А.В. Интерференционный метод измерения коэффициента экстинкции двулучепреломляющих волоконных световодов // ЖТФ. 2007. Т. 77. № 9. С. 102–107.
6. Азам Р., Башара Н. Эллипсометрия и поляризованный свет. М.: Мир, 1981. 584 с.
7. Lefevre H. The Fiber Optic Gyroscope. Boston–London: Artech House, 1993. 313 p.
8. Xu X., Pan X., Song J. Analysis of sensing coil polarization properties' effect on performance of fiber optical gyroscope // Applied Optics. 2012. V. 51. N 5. P. 621–625.
9. Cordova A., Patterson R.A., Goldner E.L., Rozelle D.M. Interferometric fiber optic gyroscope with inertial navigation performance over extended dynamic environments // Proc. of SPIE. 1994. V. 2070. P. 164–180.
10. Малькин Г.Б. Модуляционный метод устранения вызванного поляризационной невязимостью сдвига нуля волоконного кольцевого интерферометра // Письма в ЖТФ. 1999. Т. 25. № 16. С. 78–82.

Шрамко Олег Анатольевич	– инженер-исследователь, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, shramko_oa@mail.ru
Рупасов Андрей Викторович	– аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, sadbender@yandex.ru
Новиков Роман Леонидович	– аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, novikov_r_l@mail.ru
Аксарин Станислав Михайлович	– аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, oskar2002@list.ru
Oleg Shramko	– research engineer, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, shramko_oa@mail.ru
Andrei Rupasov	– postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, sadbender@yandex.ru

Roman Novikov

– postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, novikov_r_l@mail.ru

Stanislav Aksarin

– postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, oskar2002@list.ru

УДК 681.7

ТЕХНОЛОГИЧЕСКИЕ МЕТОДЫ СНИЖЕНИЯ УРОВНЯ ОПТИЧЕСКИХ ПОТЕРЬ В МИКРОСТРУКТУРИРОВАННЫХ ВОЛОКОННЫХ СВЕТОВОДАХ

А.Г. Коробейников^а, Ю.А. Гатчин^б, К.В. Дукельский^б, Е.В. Тер-Нерсесянц^б

^а Санкт-Петербургский филиал учреждения Российской академии наук «Институт Земного магнетизма, ионосферы и распространения радиоволн им. Н.В.Пушкова РАН», Санкт-Петербург, Россия, Korobeynikov_A_G@mail.ru

^б Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, KDukel@GOI.ru

Представлены результаты разработки технологии получения микроструктурированных волоконных световодов. В ходе проводимых работ были созданы световоды с одним или двумя циклами отверстий, получаемые по двухстадийной технологии. Использование двух циклов позволяют снизить на длине волны накачки затухание до 0,1 дБ/м. В ходе экспериментов было установлено, что эффективность вытекания коррелирует с толщиной депрессированной оболочки – чем больше толщина оболочки, тем потери меньше. В ходе работ была разработана базовая технология получения дырчатых волоконных световодов из кварцевого стекла с четырьмя циклами отверстий, базирующаяся на двухстадийном процессе. В отличие от дырчатых световодов с одноэлементной сердцевинной, в разработанных системах при равных диаметрах сердцевинной шаг структур меньше в два и три раза. Это позволило получить большую устойчивость излучения к возмущениям световодного тракта. При постановке задачи получения микроструктурированного волокна с малым значением отношения диаметра отверстия конструкционной оболочки к шагу структуры и (или) с относительно малым шагом структуры возможен одностадийный процесс, так как в этом случае в исходной сборке можно использовать капилляры с небольшим по величине внутренним диаметром. Были проведены исследования по поиску технологических условий, выполнение которых делает затухание излучения определенного типа в микроструктурированных световодах минимальным. Основным способом, позволившим выявлять влияние технологической схемы на затухание излучения, являлось сравнение одно- и двухстадийных методов создания микроструктурированных световодов. Была определена зависимость потерь излучения в световодах от температуры вытяжки капилляров. Проведено сравнение методов очистки капилляров в дырчатых световодах. Предложены методы снижения затухания излучения.

Ключевые слова: оптическое волокно, микроструктурированные волоконные световоды, затухание излучения.

PROCESS METHODS WITH LOW LEVEL OF OPTICAL LOSSES FOR THE MICROSTRUCTURED FIBER LIGHT GUIDES

A. Korobeynikov^a, Yu. Gatchin^b, K. Dukel'skiy^b, E. Ter-Nersesyantz^b

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

^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, KDukel@GOI.ru

The paper deals with results of the technology process development for the microstructured fiber light guides. Light guides with one or two cycles of the openings, processed by two-phase technology were created during carried out research. The use of two cycles makes it possible to decrease attenuation to 0.5 dB/m on excitation wavelength. The experiments show that effluence efficiency correlates with thickness of depressed cover – the more is thickness of a cover, the less are losses. The basic technology of perforated fiber light guides creation from quartz glass with four cycles of the openings, based on two-phase process was developed. Unlike perforated light guides with a single-element core in the developed systems with equal diameters of a core the step of structures is two and three times less. It gives the possibility to receive big resistance of radiation to indignations of a light-guide path. At a problem definition of receiving the microstructured fiber with small value of the relation of an opening diameter of a constructional cover to a step of structure and/or with rather small step of structure single-stage process as in this case in initial assembly it is possible to use capillaries with small size internal diameter. Researches were conducted to search technological conditions for minimum attenuation of a certain type radiation in the microstructured light guides. The main way to receive answers about influence of the technological scheme on radiation attenuation is comparison of one - and two-phase methods of microstructured light guides creation. Dependence of radiation losses in light guides from temperature of capillaries extract was defined. Comparison of capillaries cleaning methods in perforated light guides is carried out. Methods of radiation attenuation decrease are offered.

Keyword: optical fiber, microstructured fiber light guides, radiation attenuation.

Введение

Большим достижением последних лет в оптических технологиях является создание фотонно-кристаллических (микроструктурированных, «дырчатых») волоконных световодов. Фотонно-кристал-

лическое волокно (ФКВ) (Photonic-crystal fiber) представляет собой оптическое волокно, у которого оболочка имеет структуру двумерного фотонного кристалла. ФКВ сегодня рассматривается как один из перспективнейших оптических элементов для использования в лазерной технике в качестве активной среды. Кроме того, ФКВ можно использовать для передачи и нелинейно-оптического преобразования мощных световых потоков с заданным распределением светового излучения в поперечном сечении.

В настоящее время наиболее хорошо известны два типа ФКВ – со сплошной и полый световедущими жилами (рис. 1). На территории России их принято называть дырчатыми волокнами. Но необходимо знать, что между этими типами ФКВ имеется существенное отличие в механизмах, отвечающих за волноводные свойства ФКВ.

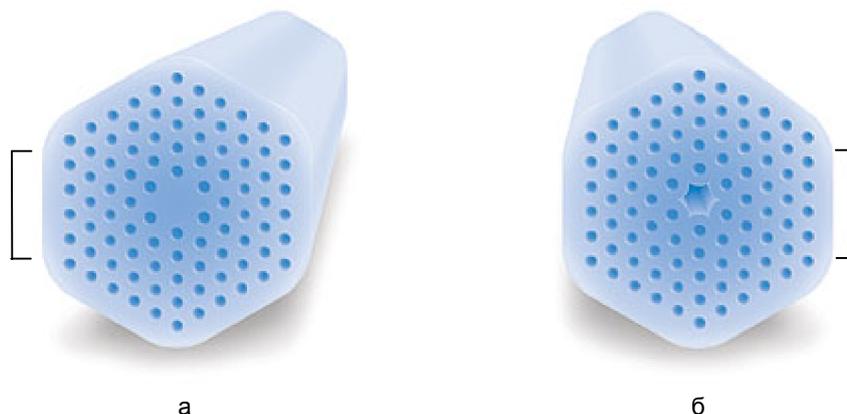


Рис. 1. Поперечное сечение дырчатого волокна: со сплошной световедущей жилой в центре и пятью циклами отверстий (а); с полый световедущей жилой в центре (б)

Дырчатое волокно со сплошной световедущей жилой (рис. 1, а) в центре представляет собой сердцевину из кварцевого стекла в оболочке из фотонного кристалла (кварцевое стекло с воздушными полостями-каналами), которая имеет по отношению к жиле в среднем более низкий коэффициент преломления. Отсюда следует, что волноводные свойства этих световодов гарантируются двумя одновременными эффектами – зонными свойствами фотонного кристалла и полным внутренним отражением (как в обычных световодах). Оболочка в виде фотонного кристалла значительно отличает ФКВ от обычных волоконных световодов.

Проведенный анализ литературных источников показал, что технология дырчатых световодов базируется в основном на двух методах: сборка системы капилляров, которые спекаются при последующей температурной обработке, и высверливание определенного множества отверстий в штабике из синтетического кварцевого стекла, который содержит минимальное количество примесей [1–7].

Первый метод характеризуется технологичностью, но имеет серьезный недостаток, заключающийся в проблеме очистки исходной сборки от механических примесей. Для второго метода необходимо специальное дорогостоящее оборудование, чтобы получать отверстия требуемой глубины. Но его применение позволяет получать лучшие характеристики по минимальным потерям излучения.

Если сравнить существующие методы газофазного осаждения диоксида кремния при изготовлении заготовок для традиционных волокон из сплошных сред, которые обязательно предусматривают эффективные меры для создания стерильной атмосферы в зонах осаждения и последующего остекловывания порошкообразных слоев диоксида кремния, с существующими методами приготовления заготовок для микроструктурированных световодов, то последние не позволяют удалять полностью механические примеси. Отсюда следует, что задача минимизации содержания примесей в заготовке является одной из основных, решаемой при разработке технологии микроструктурированных оптических волокон. Но, тем не менее, множество технологических задач, решая которые, можно в той или иной степени понизить затухание излучения, достаточно велико. Оно содержит следующие задачи: получение требуемой системы отверстий с заданными размерами и формой; создание гарантированной стабильной структуры в длиномерных отрезках световодов (даже не говоря о конкуренции традиционным связным световодам); выбор температурного режима вытяжки, который совмещает получение необходимой структуры с наличием стандартных величин механической прочности кварцевых волокон.

Технологии получения микроструктурированных световодов

В ходе проводимых работ сначала были созданы световоды гексагональной структуры с одним и двумя циклами отверстий вокруг сердцевины. Волокна были получены по двухстадийной технологии. На первой стадии производилось вытягивание предволокна диаметром 1–2 мм из поликапиллярной сборки (уложенных в необходимом порядке внутри кварцевой опорной трубы капилляров и штабиков малого диаметра), а на второй стадии предволокно помещалось в центр сборки, составленной из микроштаби-

ков, которые при перетяжке этой сборки образовывали сплошную конструкционную оболочку. Использование двух циклов отверстий (по сравнению со структурой, содержащей один цикл) позволяет снизить затухание излучения на длине волны накачки на порядок. Достигнутый минимум составил 0,5 дБ/м.

При получении световодов с двумя или одним циклами отверстий применялись элементы из кварцевого стекла, базой для которых является сырье из природных материалов – один из основных первоисточников повышенного уровня оптических потерь. В [8] было показано, что дырчатые световоды с большим уровнем оптических потерь удовлетворяют требованиям к элементу, применяемому в качестве преобразователя фемтосекундных импульсов, так как величины наиболее значимых параметров, а именно, нелинейная и дисперсионная длины, составляют не больше нескольких миллиметров и 1,5 м соответственно. В указанных пределах длин доля поглощенного излучения в световоде не превысила 20%. Последующее улучшение структуры дырчатых световодов было необходимо для понижения уровня потерь излучения до значения, которое обеспечит рациональность доставки излучения потребителю на расстоянии не больше 40 м. Такие расстояния характерны для задач, в которых волоконные световоды используют не для передачи информации.

Для решения вышеназванной задачи была разработана новая структура, состоящая из четырех циклов отверстий вокруг сердцевин, а также технология создания упорядоченной системы отверстий для получения микроструктурированных оптических волокон. Увеличение числа циклов создания отверстий продиктовано тем, что показатель преломления сердцевин и показатель преломления внешней конструкционной оболочки выше эффективного показателя преломления светоотражающей оболочки. В таких световодах (по общепринятой в волоконной оптике терминологии – световоде с депрессированной оболочкой) идет процесс вытекания направленных по сердцевине мод во внешнюю оболочку. Этот процесс обусловлен туннелированием излучения через депрессированную оболочку. Экспериментально установлено, что эффективность вытекания коррелирует с толщиной депрессированной оболочки – потери меньше, когда толщина больше. Причем это справедливо как для дырчатых световодов, так и для волоконных световодов со сплошными средами [9, 10].

В световодах, изготовленных из сплошных сред, спектральная зависимость потерь на вытекание излучения коррелирует с длиной волны. Это связано с расширением модовых полей и увеличением периферийной доли мощности, которая попадает во внешнюю (конструкционную) оболочку с последующим поглощением части излучения в полимерном покрытии. Кроме того, известно, что микро- и макроизгибы в световодах с депрессированной оболочкой содействуют более активному вытеканию излучения [11–14].

В работе [9] показано, что, увеличив число циклов отверстий до четырех, можно практически довести до нуля потери на вытекание излучения в конструкционную оболочку не только при $k \sim 0,9$ ($k = d/\Lambda$, где d – диаметр отверстия, Λ – шаг структуры), но и при значениях k , равных 0,3–0,4, и, следовательно, переходе распространения излучения в одномодовый режим.

В ходе работ была разработана базовая технология получения дырчатых волоконных световодов из кварцевого стекла с четырьмя циклами отверстий, базирующаяся на двухстадийном процессе.

Сначала в опорной кварцевой трубе (диаметр 20 мм) собирается структура, базирующаяся на центральном кварцевом штабике и объединенная в плотную гексагональную структуру с периферийными капиллярами. Выдерживается соотношение внутреннего и внешнего диаметров капилляров $\sim 0,8$ (порядок окружения равен четырем, величина диаметра структурных элементов находится в пределах 0,8–2 мм). Фотография исходной поликапиллярной сборки представлена на рис. 2.

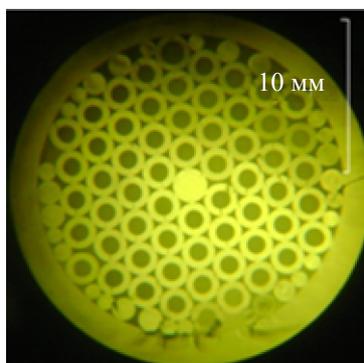


Рис. 2. Исходная поликапиллярная сборка

Из полученной таким образом заготовки вытягивается первичное волокно (предволокно) диаметром 2–3 мм (рис. 3). Существующие в начальный момент зазоры между капиллярами в заготовке необходимо сколлапсировать в ходе вытяжки. Для этой цели концы капилляров заплывают. Кроме того,

внутри опорной трубы в процессе вытяжки необходимо существование отрицательного давления. Для более эффективного коллапсирования зазоров необходимо подобрать температуру вытяжки и разрежение в опорной трубе. Диаметр сердцевины в предволокне получается 100–200 мкм. После этого полученное предволокно помещается внутрь другой кварцевой опорной трубы диаметром 20 мм, как сердцевина в центр сборки, и окружается микроштабиками, сложенными в гексагональную структуру (так называемое «шубирование»). Затем эта сборка перетягивается в новое предволокно, а потом это предволокно перетягивается в волокно (т.е. снова используется двухстадийный метод вытяжки). На выходе получается дырчатый световод, изображенный на рис. 4.

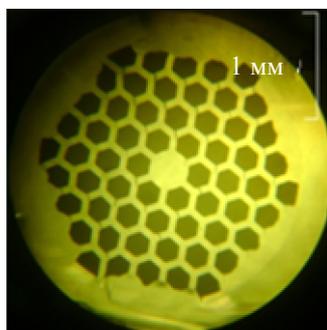


Рис. 3. Предволокно

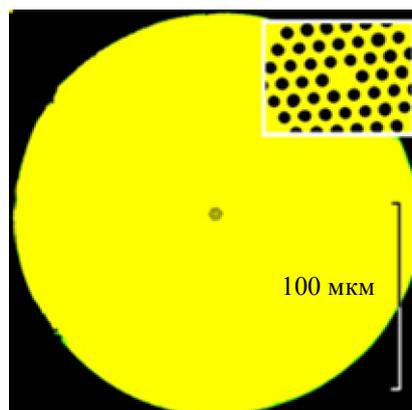


Рис. 4. Микроструктурированный световод
(на вставке показано увеличенное изображение центральной области)

В тех случаях, когда ставилась цель получения структуры с величиной шага $\Lambda > 5$ мкм, необходимость в создании конструкционной оболочки отпала. В этой связи вторая стадия технологического процесса ограничивалась перетягиванием предволокна без его «шубирования» в волокно диаметром 50–200 мкм. На самом деле получить микроструктурированный световод можно и по одностадийной технологии, т.е. прямо из исходной сборки. Но в этом случае придется выполнять два противоречащих друг другу условия. Первое – температура нагревателя обязана быть достаточно высокой, так как это необходимо для коллапсирования межкапиллярных промежутков за счет низкой вязкости стекла и действия сил поверхностного натяжения в области образования луковицы заготовки. А второе условие гласит, что повышение температуры приводит к коллапсированию внутренних капиллярных полостей и снижению параметра содержания воздуха в микроструктурированном волокне. Таким образом, практически невозможно подобрать температуру вытяжки для получения заданной структуры, так как стремление сколлапсировать межкапиллярные промежутки приводит к получению монолитного волокна. В такой ситуации получить микроструктурированное волокно в одну стадию практически невозможно, если внутренний диаметр капилляров составляет около 1,5 мм (соотношение внутреннего и внешнего диаметров в исходной трубе 0,8). При получении волокна со значением k порядка 0,9 луковица раздувается под воздействием избыточного давления в капиллярах при всех возможных значениях температуры нагревателя.

Если же ставится задача получения волокна с малым значением k и (или) с относительно малым шагом структуры Λ , то решение задачи возможно благодаря тому, что в исходной сборке можно использовать капилляры с небольшим по величине внутренним диаметром. В этом случае, вследствие малого

радиуса кривизны поверхности, избыточное давление внутри капилляра уравнивается силами поверхностного натяжения.

Методы снижения затухания излучения

Одной из задач исследований являлась задача поиска технологических условий, выполнение которых делает затухание излучения определенного типа в микроструктурированных световодах минимальным [15]. Одним из методов, позволяющим изучить влияние технологической схемы на затухание излучения, является сравнение одно- и двухстадийного методов создания микроструктурированных световодов.

Как уже говорилось, одностадийный метод можно применять, когда капилляры в исходной сборке имеют небольшой внутренний диаметр. Исходя из этого, был выбран тип световода при проведении запланированного эксперимента – световод с сердцевинной, в которой было произведено замещение в исходной сборке 7 центральных капилляров. Кроме того, выбор вида световода вызван задачей проектирования такого передающего элемента, в котором излучение распространялось бы в одной поперечной моде по «большой сердцевине». Применение термина «большая сердцевина» к одномодовым оптическим волокнам на основе кварцевого стекла означает, что ее диаметр больше 10 мкм.

Представленные в [16] положения о ведущем значении Λ при формировании оптических потерь, которые обусловлены вытеканием излучения в микроструктурированных световодах, подвели к идее разработки световодов с сердцевинами, сформированными заменой в исходной гексагональной сборке 7 или 19 центральных капилляров одним стержнем, который образует затем сердцевину световода.

В отличие от дырчатых световодов с одноэлементной сердцевиной [17, 18] в созданных нами системах при равных диаметрах сердцевины шаг структур меньше в два и три раза. Это позволило получить большую устойчивость излучения к возмущениям световодного тракта. Кроме того, снижение шага структуры ведет к уменьшению размеров капилляров в исходной сборке. А из этого следует, что одностадийный процесс получения световода становится вполне реальным.

Для проведения эксперимента капилляры диаметром 1,3 мм, вытянутые из трубы марки F300, были уложены внутри опорной трубы из кварцевого стекла (соотношение диаметров 17 мм : 22 мм) в плотную гексагональную структуру. Кроме того, микроштабиками из синтетического кварцевого стекла были заполнены сегментарные промежутки между крайними капиллярами и внутренней окружностью опорной трубы. 7 центральных капилляров, как говорилось выше, были заменены одним стержнем из кварцевого стекла с низким содержанием гидроксильных групп.

В предварительных опытах [19] было определено, как влияет на величину потерь излучения в световодах температура вытягивания капилляров. Механизм таков: чем выше температура вытягивания, тем ниже в зоне луковицы заготовки вязкость стекла. Под воздействием сил поверхностного натяжения происходит более интенсивное «выглаживание» поверхностей капилляров. А это влечет снижение затухания.

В ходе экспериментов было замечено, что регистрируемое в процессе вытягивания свечение поверхностей капилляров (вызываемое распространением излучения вдоль капилляра из высокотемпературной области луковицы) сильно различается возле печи разогрева исходной трубы, до и после механизма вытягивания. Вытягивание капилляров происходило при помощи двух обрезиненных роликов, которые вращались в противоположных направлениях.

Свечение практически отсутствует на первом из перечисленных участков. После же прохода вытяжного механизма образуется множество светящихся точек на поверхности капилляра. Причина их появления, скорее всего, вызвана повреждением поверхности капилляра при контакте с вытяжным устройством, а также оседанием пылевидных частиц.

Отсюда было сделано заключение о целесообразности вытяжки капилляров бесконтактным методом. Например, это можно делать при помощи захватов, которые двигаются по направляющим от печи разогрева трубы до основания вытяжной башни. В этом случае, если расстояние между захватами превышает их размеры, то значительная величина поверхности капилляра не будет травмирована вытяжным механизмом.

При получении образцов с варьируемым параметром k в пределах 0,15–0,5 вытяжка световодов производилась двухстадийным методом. Этот метод был выбран потому, что при небольших значениях k (0,15–0,25), которые обеспечивают преимущества световодов с семиэлементной сердцевиной, двухстадийный метод оказался наиболее подходящим, так как получение небольших величин воздушности с заданной точностью может быть обеспечено лишь на второй стадии вытяжки волоконного световода.

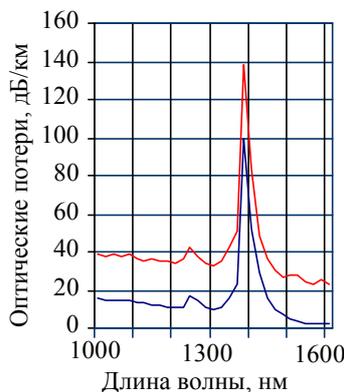
Возможность осуществлять одностадийный процесс существует для образцов с $k \sim 0,4–0,5$ и не очень жесткими требованиями к соблюдению величины этого параметра. В этом случае изготавливаются капилляры с начальным соотношением внутреннего и внешнего диаметров в пределах 0,3–0,4.

После вытяжки капилляров происходила их нарезка на куски длиной 150–200 мм. После этого концы капилляров запаивались, и различными способами производилась очистка от механических примесей их внешней поверхности. Например, капилляры промывались изопропиловым спиртом, помещались в раствор фтористоводородной кислоты, после этого промывались дистиллированной водой, изопропиловым спиртом и осушались в потоке химически чистого азота. При другом методе капилляры промывались последовательно органическими растворителями с различной полярностью (четырёххлористый углерод и изопропиловый спирт) и осушались в потоке азота. Так как в ходе экспериментов при использовании перечисленных способов очистки капилляров не было замечено существенных отличий в величине оптических потерь в микроструктурированных световодах, то в дальнейшем как наиболее технологичный применялся метод, не требующий кислоты.

Главным следствием очистки капилляров в дырчатом световоде с высокой степенью воздушности структурированной оболочки ($k \sim 0,9$) явилось понижение затухания излучения в области $\lambda = 1550$ нм с нескольких десятков дБ/км до величины порядка 10 дБ/км [20].

На рис. 5 представлены результаты измерения спектров затухания излучения в образцах световодов, полученных по одностадийному и двухстадийному методам. Необходимо пояснить, что распространение излучения по световоду с параметром $k \sim 0,5$ происходит в режиме нескольких мод. Величина диаметра сердцевины световодов составляла 20 мкм. Как показывает рис. 5, при вытяжке световода по одностадийной схеме происходит снижение потерь излучения на порядок. Основная причина такой большой разницы в величине оптических потерь, по-видимому, заключается в разной степени спекаемости исходных элементов в процессе вытягивания световода.

При использовании двухстадийного метода вытяжка как предволокна, так и волоконного световода происходит при температуре нагревателя на 50–100°C ниже, чем при использовании одностадийного процесса. Кроме того, в одностадийном процессе величина скорости подачи заготовки в печь разогрева меньше на порядок и равна 1 мм/мин. Такие параметры, по всей видимости, являются хорошими начальными условиями для того, чтобы в конечном продукте (световоде) однородность стекломассы в промежутках между воздушными отверстиями, обусловленными высокотемпературными диффузионными процессами в области границы раздела капилляров, оказывалась лучше. А это является причиной более низкого уровня потерь излучения в световоде.



— БС-12-1 двухстадийный процесс вытяжки волокна
— БС-12-2 одностадийный процесс вытяжки волокна

Рис. 5. Зависимость уровня оптических потерь излучения от технологической схемы получения световода с семиэлементной сердцевинной и значением параметра $k \sim 0,5$

Заключение

В результате исследований по разработке технологий получения микроструктурированных световодов были получены световоды с низкими оптическими потерями. Исследования оптических параметров микроструктурированных световодов показали, что эти объекты обладают большими возможностями, с одной стороны, как среды с выраженными нелинейно-оптическими свойствами, а с другой стороны – как средства передачи излучения с минимальными нелинейно-оптическими искажениями.

Разрабатываемые технологии производства микроструктурированных световодов из соображений конкурентоспособности должны выгодно отличаться малозатратностью и производительностью с максимально возможным использованием материалов отечественного производства. Кроме того, разработка конкурентоспособных микроструктурированных световодов диктует необходимость дальнейшего изучения технологических процессов для уменьшения затухания излучения и создания новых технических решений для улучшения этого параметра.

Литература

1. Яценко Ю.П., Левченко А.Е., Пряников А.Д., Косолапов А.Ф., Семенов С.Л., Дианов Е.М. Четырех-волновое смешение в двухслойных микроструктурированных световодах // Квантовая электроника. 2005. Т. 35. № 8. С. 715–719.
2. Wang S.-Y. Microstructured optical fiber with improved transmission efficiency and durability. Patent US N 6.418.258 B1; заявл. 09.06.00; опубл. 09.07.02.
3. White C.A. Microstructured multimode fiber. Patent US N 6.594.429 B1; заявл. 20.10.00; опубл. 15.07.03.
4. Russel P.S.T., Birks T.A., Knight J.C. Photonic crystal fibers. Patent US N 6.888.992 B2; заявл. 07.11.03; опубл. 03.05.05.
5. Forbes L., Geusic J.E. Hollow core photonic bandgap optical fiber. Patent US N 6.829.421 B2; заявл. 13.03.02; опубл. 07.12.04.
6. Коробейников А.Г., Гатчин Ю.А., Дукельский К.В., Ероньян М.А., Тер-Нерсесянц Е.В., Нестерова Н.А. Анализ совместимости фторсиликатных и боросиликатных слоев стекла для изготовления оптического волокна // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 3 (79). С. 15–18.
7. Коробейников А.Г., Гатчин Ю.А., Дукельский К.В., Тер-Нерсесянц Е.В. Проблемы производства высокопрочного оптического волокна // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 2 (84). С. 18–23.
8. Петровский Г.Т., Дукельский К.В., Кондратьев Ю.Н., Хохлов А.В., Шевандин В.С., Желтиков А.М. Дырчатые световоды с кварцевой сердцевинной для нелинейно-оптического преобразования импульсов лазерного излучения // Оптический журнал. 2006. Т. 73. № 9. С. 42–47.
9. Yan M., Shum P. Antiguiding in microstructured optical fibers // Optics Express. 2004. V. 12. N 1. P. 104–116.
10. Ferrarini D., Vincetti L., Zoboli M., Cucinotta A., Selleri S. Leakage properties of photonic crystal fibers // Optics Express. 2002. V. 10. N 23. P. 1314–1319.
11. Моршнев С.К., Францессон А.В. Пропускание светового излучения крутыми изгибами волоконных световодов // Квантовая электроника. 1982. Т. 9. № 2. С. 284–290.
12. Иванов С.И. Дополнительные потери, обусловленные нерегулярностями многомодовых световодов // Электросвязь. 1982. № 1. С. 41–44.
13. Jeunhomme L.B. Single-mode fiber optics principles and applications. NY: Marcel Dekker, 1983. 275 p.
14. Андреев А.Ц., Белов А.В., Власов А.В., Гурьянов А.Н., Дианов Е.М., Жиц И.Г., Иноземцев В.П., Хопин В.Ф. Потери на микроизгибах в волоконных световодах и волоконно-оптических кабелях // Квантовая электроника. 1980. Т. 7. № 1. С. 217–219.
15. Дукельский К.В., Коробейников А.Г., Тер-Нерсесянц Е.В. Методы уменьшения оптических потерь в фотонно-кристаллическом оптическом волокне // Научно-технический вестник СПбГУ ИТМО. 2010. № 3 (67). С. 5–11.
16. Дукельский К.В., Кондратьев Ю.Н., Комаров А.В., Тер-Нерсесянц Е.В., Хохлов А.В., Шевандин В.С. Влияние шага структуры дырчатого оптического волокна на его световодные свойства // Оптический журнал. 2006. Т. 73. № 11. С. 80–85.
17. Nielsen M.D., Folkenberg J.R., Mortensen N., Bjarklev A. Bandwidth comparison of photonic crystal fibers and conventional single-mode fibers // Optics Express. 2004. V. 12. N 3. P. 430–435.
18. Nielsen M.D., Mortensen N.A., Albertsen A., Folkenberg J.R., Bjarklev A., Bonaccini D. Predicting macrobending loss for large-mode area photonic crystal fiber // Optics Express. 2004. V. 12. N 8. P. 1775–1779.
19. Дукельский К.В., Комаров А.В., Тер-Нерсесянц Е.В., Хохлов А.В., Шевандин В.С. Работы ГОИ им. С.И. Вавилова по снижению затухания сигнала в микроструктурированных оптических волокнах // Труды Международных научно-технических конференций «Интеллектуальные системы» (AIS'07) и «Интеллектуальные САПР» (CAD-2007). М.: ФИЗМАТЛИТ, 2007. Т. 3. С. 160–166.
20. Дукельский К.В., Комаров А.В., Кондратьев Ю.Н., Тер-Нерсесянц Е.В., Хохлов А.В., Шевандин В.С. Дырчатые световоды из поликапиллярной сборки с затуханием излучения в 10 дБ/км // Сборник трудов VII Международной конференции «Прикладная оптика». СПб, 2006. Т. 2. С. 216–220.

Коробейников Анатолий Григорьевич

– доктор технических наук, профессор, заместитель директора, Санкт-Петербургский филиал учреждения Российской академии наук «Институт Земного магнетизма, ионосферы и распространения радиоволн им. Н.В.Пушкова РАН» Санкт-Петербург, Россия, Korobeynikov_A_G@mail.ru

Гатчин Юрий Арменакович

– доктор технических наук, профессор, зав. кафедрой, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Gatchin@mail.ifmo.ru

- Дукельский Константин Владимирович* – кандидат технических наук, доцент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, KDukel@GOI.ru
- Тер-Нерсесянц Егизе Вавикович* – кандидат технических наук, доцент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, volokno@goi.ru
- Anatoly Korobeynikov* – D.Sc., Professor, Deputy director for science, Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences St. Petersburg Branch (SPbF IZMIRAN), Saint Petersburg, Russia, Korobeynikov_A_G@mail.ru
- Yury Gatchin* – D.Sc., Professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Gatchin@mail.ifmo.ru
- Konstantin Dukel'skiy* – PhD, Associate professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, KDukel@GOI.ru
- Egizhe Ter-Nersesyantz* – PhD, Associate professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, volokno@goi.ru

УДК 681.7.064.454

МНОГОЗОННОЕ ПРОСВЕТЛЯЮЩЕЕ ПОКРЫТИЕ НА ПОДЛОЖКЕ ИЗ ОПТИЧЕСКОГО СУЛЬФИДА ЦИНКА

Тан Тай До^а, Л.А. Губанова^а

^аСанкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, La7777@mail.ru

Представлена методика создания многозонного просветляющего покрытия на подложке из оптического сульфида цинка. Покрытие одновременно эффективно работает в нескольких спектральных диапазонах: в видимой области (450–700 нм), в ближней инфракрасной области (1000–1100 нм), на длине волны 1,55 мкм, а также в средней инфракрасной области спектра (3–5 мкм). Коэффициент отражения в диапазоне 450–700 нм не превышает 2%, в диапазоне 1000–1100 нм – менее 0,5%, в диапазоне 1500–1700 нм близок к 1,5% и в диапазоне 3–5 мкм – к 0,6%. В работе представлены результаты анализа влияния отклонений в толщинах слоев, формирующих разработанное покрытие, на изменение значения энергетического коэффициента отражения в рассматриваемых областях. Получено, что отклонение в толщине слоя, граничащего с воздухом, оказывает максимальное влияние на спектральную характеристику полученного покрытия. Отклонение показателя преломления этого слоя влияет на величину остаточного отражения.

Ключевые слова: просветляющие покрытия, градиентные покрытия, многозонные покрытия, сульфид цинка.

MULTI-ZONE ANTIREFLECTION COATING ON A SUBSTRATE MADE OF OPTICAL ZINC SULPHIDE

Тан Тай До^б, L. Gubanova^б

^бSaint Petersburg National Research University of Information Technologies, Mechanics and Optics, La7777@mail.ru

The paper deals with creation technique for a multi-zone antireflection coating on a substrate made of the optical zinc sulphide ZnS. The coating effectively operates simultaneously in the following spectral ranges: visible region of 450 - 700 nm, in the near infrared region of 1000 - 1100 nm, at the wavelength of 1.55 μm , and in the mid-infrared (IR) spectrum of 3 - 5 microns. Reflection coefficient in the range of 450 - 700 nm is not more than 2%, in the range of 1000 - 1100 nm is less than 0.5%, in the range of 1500 - 1700 nm is close to 1.5% and in the range of 3 - 5 μm is equal to 0.6%. Analysis results of the deviation impact in the thickness of layers on the value changing of the energy reflection coefficient in the considered areas are given. Deviation in the thickness of the layer, contiguous with the air, is shown to have the greatest effect on the spectral characteristics of the obtained coating. Refractive index deviation for this layer influences the magnitude of the residual reflection.

Keywords: antireflection coatings, gradient coatings, multi-zone coatings, zinc sulphide.

Введение

Потребность в оптических приборах, способных работать в нескольких спектральных (видимой, ближней и средней инфракрасной (ИК)) областях, постоянно возрастает [1–7]. Такие системы оснащены датчиками, которые работают в определенной области спектра, и имеют ряд преимуществ, в том числе

позволяют получать более полную информацию об объекте, использовать общий точный механизм стабилизации, компактны, имеют минимальный вес и габаритные размеры [8–11].

Среди материалов, используемых в оптических схемах при создании многоспектральных оптических устройств, сульфид цинка (ZnS) является одним из наиболее часто используемых. Это можно объяснить следующими причинами: он обладает широким диапазоном прозрачности (0,4–14 мкм), не ядовит, нерастворим в воде [6]. ZnS имеет показатель преломления $n_m = 2,2$, поэтому оптический элемент, изготовленный из ZnS без просветляющего покрытия, пропускает не более 73% падающего излучения. Формирование просветляющих покрытий на оптических элементах, изготовленных из сульфида цинка, существенно увеличивает эффективность его использования.

Просветление оптического ZnS было изучено разными авторами [12–14]. Покрытия, рассмотренные в работе [12], состоят из трех и пяти слоев. Эти покрытия снижают энергетический коэффициент отражения в диапазоне длин волн 0,5–0,9 мкм, на длине волны 1,064 мкм и в диапазоне 8–12 мкм. Среднее значение энергетического коэффициента отражения в этих областях спектра составляет 2–3%. Конструкция разработана так, что входящие в его состав слои имеют дискретно меняющиеся показатели преломления. Рассматриваемое покрытие не обеспечивает снижение отражения в средней ИК области. В работе [13] описывается просветляющее покрытие, рассчитанное численным методом. Покрытие снижает отражение до 1–2% в трех диапазонах: 8–12 мкм, 3–5 мкм и на длине волны 1,064 мкм. Недостатками этого покрытия является отсутствие просветления в видимом диапазоне и узкая зона просветления в ближней ИК области. В работе [14] рассматривается покрытие, у которого в формирующих слоях изменение показателя преломления является градиентным. Полученное покрытие имеет среднее отражение не меньше 5% в видимом (0,45–0,65 мкм) и среднем ИК (3–5 мкм) диапазонах.

Часто встречается случай, когда в состав покрытия входят слои с «нереальным показателем преломления», т.е. рассматривается пленкообразующий материал, которого нет в природе. Эту проблему можно решить путем совместного напыления нескольких материалов из одного испарителя или из отдельных испарителей. Применение такого метода в настоящее время ограничивается из-за отсутствия точных технологий.

Представляемая авторами работа расширяет ранее полученные результаты и посвящена методике конструирования просветляющего покрытия на оптическом ZnS с использованием реальных и устойчивых пленкообразующих материалов. Покрытие одновременно эффективно работает в видимой (0,4–0,6 мкм) области для лазера с длиной волны генерации 1064 нм, лазерного дальномера, работающего на длине волны 1,55 мкм и в средней ИК (3–5 мкм) области. Покрытие разработано с использованием градиентного метода [15], который основан на том, что показатель преломления слоев, формирующих покрытие, является функцией общей толщины покрытия.

Методика расчета

Чтобы получить многослойное просветляющее покрытие, выбираем конструкцию покрытия, в которой оптическая толщина каждого слоя равна, причем показатель преломления слоев непрерывно увеличивается от среды (воздуха), из которой падает излучение, до подложки согласно уравнению пятой степени [15]:

$$n(x) = n_0 + (n_m - n_0) \cdot \left(10 \cdot \left(\frac{x}{D} \right)^3 - 15 \cdot \left(\frac{x}{D} \right)^4 + 6 \cdot \left(\frac{x}{D} \right)^5 \right), \quad (1)$$

где x – расстояние от границы с воздухом до границы раздела между i -м и $(i+1)$ -м слоями ($i = 0, 1, 2, 3, 4$); D – общая физическая толщина покрытия, состоящего из нескольких слоев; n_m – показатель преломления подложки; n_0 – показатель преломления окружающей среды (для воздуха $n_0 = 1$).

Синтез покрытий осуществляется в несколько этапов.

На первом этапе задаются значения D и x_i . По формуле (1) для каждого набора значений $(x_{i+1} - x_i)$ получаем значения показателей преломления $(i+1)$ -го слоя. Таким образом, получаем структуру просветляющего покрытия, состоящего из $(i-1)$ слоев, имеющих полученное значение показателя преломления. Физическая толщина каждого слоя определяется как $d_i = x_{i+1} - x_i$.

Как известно из [16], энергетический коэффициент отражения многослойной системы может быть определен из выражения

$$R(\lambda) = |r(\lambda)|^2,$$

где

$$r(\lambda) = \frac{n_0 m_{11} - n_m m_{22} + i \cdot (n_0 n_m m_{12} - m_{21})}{n_0 m_{11} + n_m m_{22} + i \cdot (n_0 n_m m_{12} + m_{21})}, \quad \begin{vmatrix} m_{11} & i m_{12} \\ i m_{21} & m_{22} \end{vmatrix} = \prod_{j=1}^k \begin{vmatrix} \cos \varphi_j(\lambda) & \frac{i \sin \varphi_j(\lambda)}{n_j} \\ i n_j \sin \varphi_j(\lambda) & \cos \varphi_j(\lambda) \end{vmatrix},$$

k – количество слоев, формирующих систему, $\varphi_j(\lambda)$ – фазовая толщина j -го слоя, n_j – показатель преломления j -го слоя.

Рассмотрим структуру, формирующую покрытие из четырех слоев, характеризующихся показателями преломления (n_1, n_2, n_3, n_4), физическими толщинами (d_1, d_2, d_3, d_4), оптическими толщинами ($n_1d_1, n_2d_2, n_3d_3, n_4d_4$) и фазовыми толщинами ($\varphi_1 = 2\pi n_1d_1(\lambda)^{-1}$, $\varphi_2 = 2\pi n_2d_2(\lambda)^{-1}$, $\varphi_3 = 2\pi n_3d_3(\lambda)^{-1}$, $\varphi_4 = 2\pi n_4d_4(\lambda)^{-1}$). Световое излучение падает из окружающей среды с показателем преломления n_0 . Покрытие сформировано на подложке с показателем преломления n_m .

Среднее значение энергетического коэффициента отражения в диапазоне длин волн (λ_1, λ_2)

$$R_{\text{средний}} = \frac{1}{\lambda_2 - \lambda_1} \int_{\lambda_1}^{\lambda_2} R(\lambda) d\lambda.$$

Если среднее арифметическое значение энергетического коэффициента отражения просветляющего покрытия в указанных диапазонах больше требуемого значения отражения, то процедура повторяется с первого этапа до тех пор, пока не получается желаемое значение энергетического коэффициента отражения в указанных диапазонах.

Практическая реализация методики

По представленной методике была получена структура покрытия, параметры которой приведены в табл. 1. С целью получения оптимального энергетического коэффициента отражения в видимом и среднем ИК диапазонах следует изменять оптическую толщину каждого слоя. Первоначальное значение толщины равняется 50 нм [17], финальное значение толщины равняется 450 нм, перебор осуществляется с шагом 10 нм. В итоге было получено покрытие с конструкцией H0,964H'V0,349V' , где H, H' – четвертьволновый слой с низким показателем преломления $n_H = 1,35$, $n_{H'} = 1,5$; В, В' – четвертьволновые слои с высоким показателем преломления $n_B = 1,8$, $n_{B'} = 2,15$; $\lambda_0 = 1,66$ мкм. Данное покрытие имеет характеристику, приведенную в табл. 2. На рис. 1 представлена спектральная характеристика энергетического коэффициента отражения разработанного покрытия.

Номер слоя	$n(x)$	Физическая толщина d , нм	Оптическая толщина nd , нм
1	1,35	307	415
2	1,5	276	415
3	1,82	228	415
4	2,15	193	415
Суммарная толщина покрытия		$D = 1004$	1660

Таблица 1. Значения показателей преломления и толщин слоев разработанного просветляющего покрытия

Диапазон спектра λ , нм	450–700	1000–1100	1500–1700	1000–1700	3000–5000
$R_{\text{средний}}$	1,5%	0,4%	1,4%	2%	0,6%

Таблица 2. Средний коэффициент энергетического отражения просветляющего покрытия с конструкцией H0,964H'V0,349V'

Используемые в разработанной конструкции пленкообразующие материалы должны удовлетворять следующим условиям: иметь полученный при расчетах показатель преломления и быть прозрачными в рабочих диапазонах спектра. В качестве слоев с низким показателем преломления $n_H = 1,35$, $n_{H'} = 1,5$ могут быть использованы пленкообразующие материалы MgF_2 ($n = 1,35$, область прозрачности 0,11–11 мкм) и YF_3 ($n = 1,5$, область прозрачности 0,3–14 мкм). В качестве слоев с высоким показателем преломления $n_B = 1,82$, $n_{B'} = 2,15$ могут быть использованы пленкообразующие материалы Y_2O_3 ($n = 1,82$, область прозрачности 0,3–12 мкм) и CeO_2 ($n = 2,15$, область прозрачности 0,4–12 мкм).

При изготовлении покрытия существенную роль играет устойчивость спектральной характеристики коэффициента отражения покрытия к ошибкам в слоях, которые могут возникнуть при его формировании. Далее приводится спектральная характеристика просветляющего покрытия при наличии ошибок в толщинах. В расчете для отдельного j -го слоя с оптимальной оптической толщиной $n_j d_j$ ($j = 0, 1, 2, 3, 4$) прибавляем значение ошибки в толщинах Δj . Рассмотрим, как при этом изменится спектральная характеристика отражения просветляющего покрытия.

Изменения спектральной характеристики отражения просветляющего покрытия приведены на рис. 2, 3. Авторами были рассмотрены ситуации, когда оптическая толщина каждого слоя имеет отрицательные и положительные ошибки, и, таким образом, ошибка была лишь в рассматриваемом слое. Анализ показал, что наиболее существенное влияние на спектральную характеристику оказывает слой, гра-

ничающий с воздухом. Слой, лежащий на подложке, оказывает минимальное влияние на спектральную характеристику просветляющего покрытия.

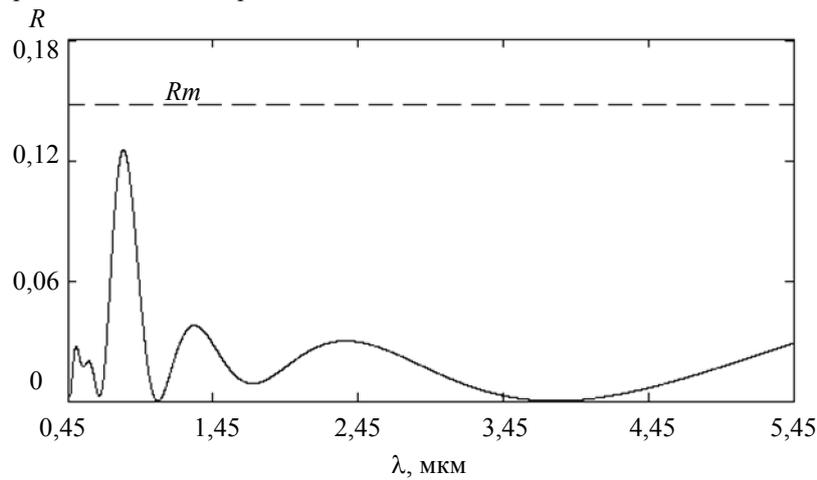


Рис. 1. Спектральная характеристика отражения просветляющего покрытия на подложке из оптического ZnS. Покрытие имеет конструкцию $H0,964H^1V0,349V^1$, $\lambda_0 = 1,66$ мкм.

R_m – коэффициент отражения подложки из оптического сульфида цинка без просветляющего покрытия

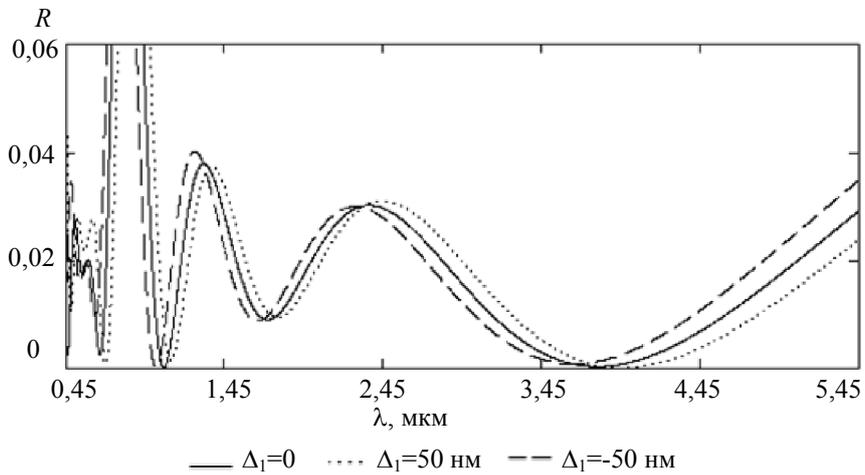


Рис. 2. Спектральная характеристика просветляющего покрытия $H0,964H^1V0,349V^1$, $\lambda_0 = 1,66$ мкм. $n_H = 1,35$, $n_H^1 = 1,5$, $n_V = 1,8$, $n_V^1 = 2,15$, $n_m = 2,2$. Первый слой принимает значения оптической толщины больше и меньше оптимальной толщины

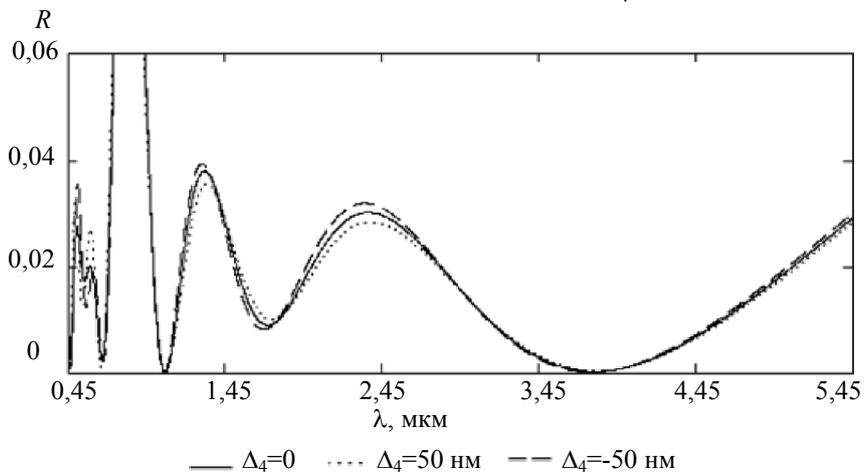


Рис. 3. Спектральная характеристика просветляющего покрытия $H0,964H^1V0,349V^1$, $\lambda_0 = 1,66$ мкм. $n_H = 1,35$, $n_H^1 = 1,5$, $n_V = 1,8$, $n_V^1 = 2,15$, $n_m = 2,2$. Четвертый слой принимает значения оптической толщины больше и меньше оптимальной толщины

Аналогично рассмотрим, как будет изменяться спектральная характеристика просветляющего покрытия при ошибках в показателе преломления отдельного слоя.

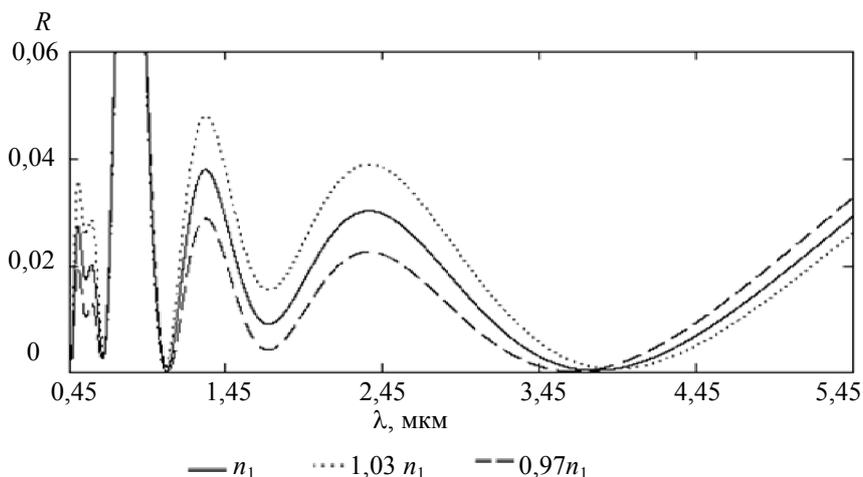


Рис. 4. Спектральная характеристика просветляющего покрытия $H\ 0,964H'V0,349V'$, $\lambda_0 = 1,66$ мкм. $n_H = 1,35$, $n_H' = 1,5$, $n_B = 1,8$, $n_B' = 2,15$, $n_m = 2,2$. Первый слой принимает значения показателя преломления больше и меньше 3% от исходного значения

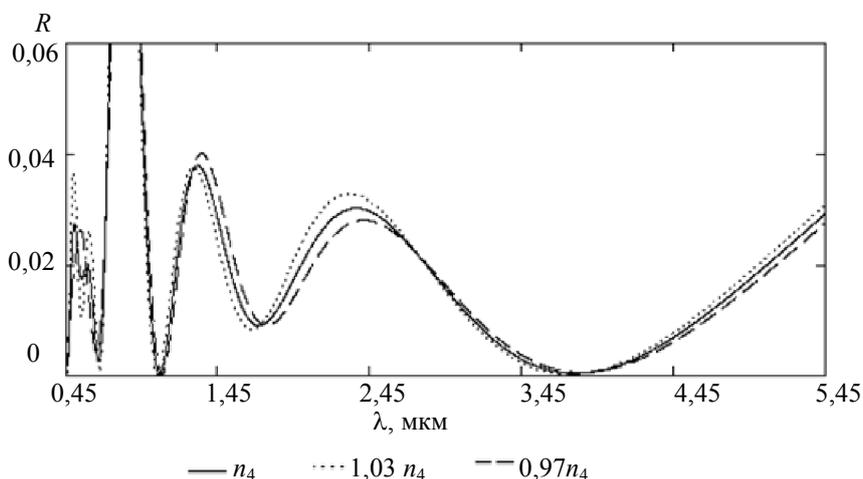


Рис. 5. Спектральная характеристика просветляющего покрытия $H\ 0,964H'V0,349V'$, $\lambda_0 = 1,66$ мкм. $n_H = 1,35$, $n_H' = 1,5$, $n_B = 1,8$, $n_B' = 2,15$, $n_m = 2,2$. Четвертый слой принимает значения показателя преломления больше и меньше 3% от исходного значения

На рис. 4, 5 приведены изменения спектральной характеристики просветляющего покрытия при ошибках в показателе преломления слоя, входящего в состав полученного покрытия. Ошибки принимают положительное и отрицательное значение, которое составляет 3% от исходного значения показателя преломления отдельного слоя. Проведенный анализ показал, что изменение показателя преломления слоя, граничащего с воздухом, оказывает наиболее существенное влияние на спектральную характеристику просветляющего покрытия по сравнению со слоем, лежащим на подложке.

Заключение

В работе была показана возможность получения структуры покрытия, обеспечивающая низкое отражение в нескольких диапазонах спектра. Структура покрытия была получена с использованием градиентных слоев, в которых показатель преломления является функцией общей толщины покрытия.

В ходе выполнения работы получены конструкции просветляющих интерференционных покрытий на основе четырех слоев. Полученное покрытие способно снизить энергетический коэффициент отражения подложки из сульфида цинка одновременно в видимой области (450–700 нм), в диапазонах длины волн лазеров 1000–1100 нм, 1500–1700 нм и в средней ИК области 3–5 мкм.

В работе исследовано изменение спектральной характеристики просветляющего покрытия при отклонении значения оптических толщин и показателя преломления от расчетных. Получено, что наиболее существенное влияние на спектральную характеристику оказывает слой, граничащий с воздухом. С уменьшением показателя преломления уменьшается энергетический коэффициент отражения, при увели-

чении показателя преломления энергетический коэффициент отражения увеличивается. Отклонение показателя преломления слоя, лежащего на подложке, практически не оказывает влияния на спектральную характеристику. При изменении оптической толщины слоя, находящегося на границе с воздухом, изменяется положение минимального энергетического коэффициента отражения в шкале длин волн. С уменьшением толщины минимумы отражения смещаются в коротковолновую область, с увеличением – в длинноволновую область.

Литература

1. Вауа S., Sanghera J., Kim W., Gibson D., Fleet E., Shaw B., Hunt M., Aggarwal I. New multiband IR imaging optics // Proc. of SPIE. 2013. V. 8704. P. 870428-1–870428-6.
2. Фатеев В.Ф., Лагуткин В.Н., Лукьянов А.П., Подкорытов Е.Н., Слынько Ю.В., Старостенко А.М. Моделирование входной информации оптической системы малых космических аппаратов // Изв. вузов. Приборостроение. 2009. Т. 52. № 4. С. 18–21.
3. Vaillancourt J., Vasinajindakaw P., Lu X., Qian X., Vangala S.R., Goodhue W.D. Thermo-electrically cooled shortwave infrared and longwave infrared dual band quantum-dot photodetector // Proc. of SPIE. 2011. V. 8012. P. 801200-1–801200-7.
4. Мордвин Н.Н., Попов Г.Н. Концепция построения оптико-электронных приборов наблюдения универсального назначения // Изв. вузов. Приборостроение. 2009. Т. 52. № 6. С. 34–39.
5. Пулов Й.Д., Чуйкин В.М., Ширококов А.М. Многоспектральный тепловизор «Терма-2» и его метрологические возможности // Научно-технический вестник СПб ГИТМО (ТУ). 2002. № 1 (5). С. 105–106.
6. Голицын А.В., Журавлев П.В., Журов Г.Е., Корякин А.В., Чихонадских А.П., Шлишевский В.Б., Яшина Т.В. Псевдобинокулярный двухканальный прибор обнаружения потенциальных угроз // Изв. вузов. Приборостроение. 2009. Т. 52. № 6. С. 27–34.
7. Zadavec D., Franks J.W., Rogers K.A., Hendry A.F., Drach P. A Multi-Spectral Optical System (1.55 μ m and 8–12 μ m) of GASIR ® 1 Design and Coating Aspects // Proc. of SPIE. 2009. V. 7298. P. 72982L-1–72982L-9.
8. Мальцев Г.Н., Кунгурцев В.В., Козинев И.А. Дистанционное зондирование Земли на основе кластера малых космических аппаратов многоспектрального наблюдения // Изв. вузов. Приборостроение. 2009. Т. 52. № 4. С. 22–28.
9. Горбулин В.И., Панченко В.В., Радионов Н.В. Техничко-экономический подход к выбору системы малых космических аппаратов // Изв. вузов. Приборостроение. 2009. Т. 52. № 4. С. 46–51.
10. Павлов Н.И., Ясинский Г.И. Малогабаритный оптико-электронный сканирующий прибор для аэросъемки в видимом и инфракрасном диапазонах // Оптический журнал. 2003. Т. 70. № 4. С. 11–14.
11. Тарасов В.В., Якушенков Ю.Г. Оптическое считывание - один из путей совершенствования микроболометрических приемников излучения (по материалам зарубежной печати) // Изв. вузов. Приборостроение. 2008. Т. 51. № 9. С. 41–46.
12. Соколова Р.С., Муранова Г.А., Михайлов А.В., Горелкина Ж.Н. Мультиспектральные просветляющие покрытия для видимой, ближней и ИК областей спектра // Оптический журнал. 2005. Т. 72. № 10. С. 60–62.
13. Cohen D., Stolov Y., Azran A., Gilo M. Dual- and triple-band AR coatings for IR systems // Proc. of SPIE. 2013. V. 8704. P. 870422-1–870422-6.
14. Awasthi S., Nautiyal B.B., Kumar R., Bandyopadhyay P.K. Multi-spectral antireflection coating on zinc sulphide simultaneously effective in visible, eye safe laser wave length and MWIR region // Infrared Physics & Technology. 2012. V. 55. N 5. P. 395–398.
15. Southwell W.H. Coating design using very thin high- and low-index layers // Applied Optics. 1985. V. 24. N 4. P. 457–460.
16. Бергнинг П.Х. Теория и методы расчета оптических свойств тонких пленок. Физика тонких пленок / Под ред. Г. Хасса. М.: Мир, 1967. Т. 1. С. 91–151.
17. Губанова Л.А., Путилин Э.С. Принципы формирования структуры интерференционных спектроредителей // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 1 (77). С. 1–6.

До Тан Тай

– аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, tantaido@yahoo.com

Губанова Людмила Александровна

– доктор технических наук, профессор, доцент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, La7777@mail.ru

Tan Tai Do

– postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, tantaiddo@yahoo.com

Lyudmila Gubanova

– D.Sc., Professor, Associate professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, La7777@mail.ru

УДК 676.22.017; 676.262; 539.211

КОНТРОЛЬ СТРУКТУРЫ РАЗЛИЧНЫХ ВИДОВ БУМАГИ МЕТОДОМ АТОМНО-СИЛОВОЙ МИКРОСКОПИИ

М.В. Жуков^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, cloudjyk@yandex.ru, zerocomb@bk.ru

Рассматриваются вопросы высокоточного контроля параметров выпускаемой бумажной продукции, а именно, различных видов бумаги и фотобумаги для печати. В рамках данной работы методом атомно-силовой микроскопии проведены исследования необработанной, матовой, глянцевой и ламинированной бумаги. Для исследований использовался учебно-научный сканирующий зондовый микроскоп NanoEducator LE. Проведена визуализация характерной структуры каждого вида исследуемой бумаги, получены гистограммы шероховатости, определена средняя шероховатость перепадов высот для каждого типа бумаги. Наименьшей шероховатостью обладает ламинированная бумага (R_a около 70 нм), шероховатость глянцевой бумаги составляет около 170 нм, шероховатость необработанной бумаги с волокнами целлюлозы – 530–540 нм. Самыми высокими параметрами шероховатости характеризуется матовая бумага (R_a порядка 670–680 нм). Продемонстрировано, что применение сканирующей зондовой микроскопии позволяет проводить экспресс-контроль параметров целлюлозной бумажной продукции, что может быть использовано при ее производстве.

Ключевые слова: бумага, атомно-силовая микроскопия, шероховатость, фотобумага, сканирующий зондовый микроскоп.

STRUCTURE CONTROL FOR DIFFERENT TYPES OF PAPER BY ATOMIC FORCE MICROSCOPY

M. Zhukov^b

^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, cloudjyk@yandex.ru, zerocomb@bk.ru

The paper deals with the precision control for the parameters of manufactured paper production, such as various kinds of paper and photo paper for printing. Research of untreated, matte, glossy and laminated paper is conducted by atomic force microscopy by means of educational and scientific scanning probe microscope NanoEducator LE in the framework of this paper. Visualization of characteristic structure for each type of studied paper was conducted, histogram of roughness was obtained, and average roughness of height differences was defined. A laminated paper has got the lowest roughness (R_a of about 70 nm) and glossy paper has got R_a of about 170 nm; roughness of untreated paper with cellulose fibers is about 530-540 nm, and matte paper has got the highest roughness parameters (R_a about 670-680 nm). Scanning probe microscopy application for parameters monitoring of cellulosic paper production is shown to give the possibility of such microscopy type application in the production of paper products and high-precision control of its parameters.

Keywords: paper, atomic force microscopy, roughness, photo paper, scanning probe microscope.

Введение

Вопросы высокоточного контроля параметров выпускаемой бумажной продукции имеют все возрастающее значение, особенно при изготовлении бумаги и фотобумаги для печати, бумаги для денежных знаков. Структура и свойства этой продукции напрямую влияют на качество печатных документов, отображение мелких особенностей изображений. Многие компании-производители проводят контроль выпускаемой бумажной продукции методами сканирующей электронной микроскопии [1, 2] и конфокальной оптической микроскопии [3–5]. Наиболее актуальными в последнее время являются исследования, направленные на высокоточное отображение параметров бумаги в нанометровом диапазоне методами атомно-силовой микроскопии (АСМ) [6, 7], визуализацию ее структуры [8, 9], изучение показателей шероховатости и адгезии [10]. Особое значение приобретают работы по неразрушающему исследованию манускриптов, пергаментов и древней бумаги, сохранившейся до наших дней [11, 12], для чего также используются методы АСМ [13].

При этом перечисленные исследования довольно дорогостоящи и требуют присутствия высококвалифицированного, специально обученного персонала. Предложенный в настоящей работе метод визуализации предполагает использование доступного учебно-исследовательского сканирующего зондового микроскопа (СЗМ) NanoEducator LE (НТ-СПб, Россия) [14], не требующего для работы на нем специализированной подготовки в области методов силовой зондовой микроскопии. В качестве рабочего зондового датчика в NanoEducator LE используется вольфрамовая игла, которую можно легко изготовить методом электрохимического травления вольфрамовой проволоки [15]. Стоимость вольфрамовой про-

локи существенно ниже кантилеверных зондов, используемых в традиционной зондовой микроскопии, что позволяет существенно снизить стоимость единичного анализа.

Целью настоящей работы являлось изучение возможности контроля параметров выпускаемой бумажной целлюлозной продукции с помощью метода АСМ с использованием доступного и простого в использовании сканирующего зондового микроскопа NanoEducator LE, подходящего для применения на производстве.

Описание экспериментальной установки

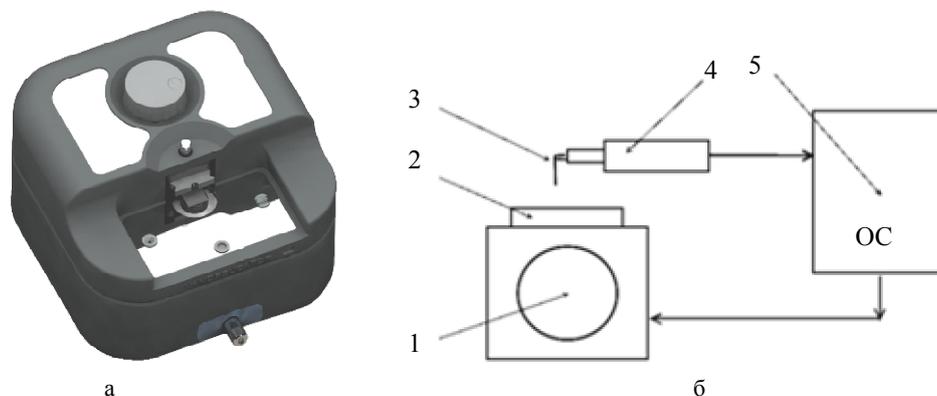


Рис. 1. Внешний вид конструкции прибора NanoEducator LE [14] (а) и принципиальная схема микроскопа (б): 1 – сканер, 2 – образец, 3 – W-зонд, 4 – пьезорезонансный зондовый датчик, 5 – цифровая система обратной связи (ОС)

Для исследования структуры бумаги использовался СЗМ NanoEducator LE (НТ-СПб, Россия) [14], внешний вид которого представлен на рис. 1, а. Прибор позволяет проводить измерения в полуконтактной силовой и туннельной модах. Область взаимодействия зонда и образца контролируется съемным оптическим микроскопом. Основные параметры зондового микроскопа NanoEducator LE приведены в табл. 1. Для неразрушающего контроля поверхности и получения информации о параметрах изучаемой структуры бумаги применялась полуконтактная силовая мода. Данный метод основан на контроле взаимодействия иглы (зонда) с поверхностью изучаемого объекта. В схеме микроскопа имеется сканер, перемещающий образец относительно зонда, система обратной связи, позволяющая контролировать взаимодействие между образцом и зондом на заданном уровне, и система детектирования, позволяющая получать информацию о положении зонда (рис. 1, б). Полученный сигнал обрабатывается специализированным программным обеспечением, представляющим рельеф исследуемой поверхности в виде трехмерного массива точек.

Атомно-силовая микроскопия является методом высокоточного изучения поверхностных свойств объектов различной природы вплоть до атомарного уровня, что позволяет исследовать наноразмерные структуры на поверхности объектов, а также изучать их параметры (шероховатость, силу адгезии, поверхностные заряды и т.д.). Подобным разрешением обладают и другие методы высокоточного анализа поверхности, например, сканирующая электронная микроскопия. При этом данный метод обладает своими недостатками. В частности, исследуемые образцы для сканирующей электронной микроскопии должны быть проводящими (в случае диэлектрических образцов их поверхность покрывается тонкими слоями металлов, что искажает поверхность объекта исследования). С помощью методов сканирующей электронной микроскопии невозможно точное определение профиля высоты, а следовательно, информации о шероховатости поверхности материала.

Сканирование	образцом
Диапазон измерений линейных размеров в плоскости XY, мкм	не менее 70
Диапазон измерений линейных размеров по оси Z, мкм	не менее 10
Среднеквадратичное отклонение (СКО) результатов измерений линейных размеров в плоскости XY и по оси Z	не более 5%
Разрешение в плоскости XY	не более 50 нм
Разрешение по оси Z	не более 1 нм
Нелинейность сканирования в плоскости XY	не более 30 нм

Таблица 1. Основные параметры прибора NanoEducator LE [14]

Предлагаемый в работе метод, реализуемый в приборе NanoEducator LE, позволяет избежать данных затруднений и позволяет проводить экспресс-диагностику наноразмерных особенностей структуры

бумаги и ее параметров, таких как шероховатость поверхности и общий уровень локальных перепадов высот.

Экспериментальные данные

Для исследования структуры и свойств параметров целлюлозной бумажной продукции были отобраны четыре основных вида бумаги: стандартная необработанная печатная бумага с волокнами целлюлозы [16], обработанная матовая, глянцевая и ламинированная фотобумага [17]. Плотность необработанной бумаги составила около 80 г/м², плотность обработанной фотобумаги составила около 230 г/м². Разрешающая способность метода АСМ зависит от размеров кончика зонда, чувствительности пьезокерамики и параметров аппаратуры. При СЗМ-исследованиях параметров бумаги использовался полуконтактный силовой режим СЗМ NanoEducator LE с радиусом закругления вершины *W* зонда около 20–50 нм. На рис. 2, 3 представлены полученные СЗМ-изображения поверхности исследованных видов бумаги. На СЗМ-изображениях светлые области соответствуют более высоким точкам, а темные – более низким точкам относительно среднего уровня, равного половине максимального перепада высот исследуемой области. Следует отметить, что перепад высот достигает максимального значения на необработанной и матовой бумаге и существенно снижается на глянцевой и ламинированной бумаге. На основании приведенных данных были получены гистограммы распределения высот (рис. 4).

Из анализа представленных гистограмм можно сделать вывод, что характерные особенности рельефа (неровности) необработанной бумаги составляют около 1,5–1,9 мкм, на матовой бумаге значения выше и составляют около 2,5–2,7 мкм. На обработанных глянцевой и ламинированной бумаге диапазон наиболее вероятных значений перепадов высот резко снижается и составляет около 1,4–1,6 мкм и 0,2–0,35 мкм соответственно.

Для контроля гладкости поверхности в данной работе проведено СЗМ-исследование поверхности образцов бумаги. С помощью специализированного программного обеспечения обработки СЗМ-изображений для каждого сорта бумаги проведен расчет средней арифметической шероховатости *Ra* [18], основанный на вычислении средних перепадов высот на поверхности заданного размера (табл. 2).

Необработанная бумага	Матовая бумага	Глянцевая бумага	Ламинированная бумага
535	677	173	71

Таблица 2. Значения *Ra* исходных образцов бумаги, нм

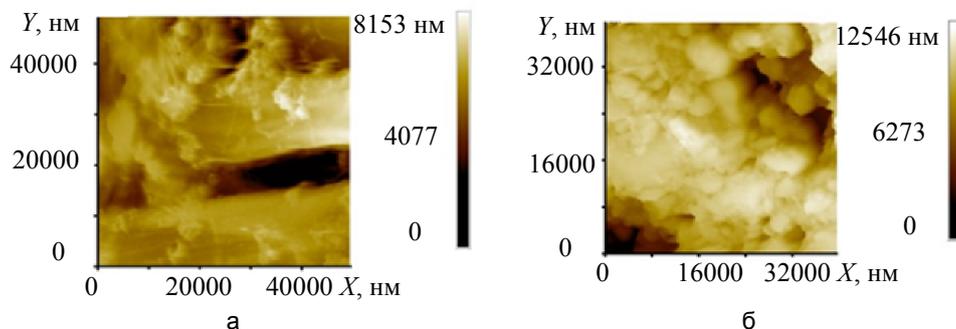


Рис. 2. Характерные АСМ-изображения поверхности необработанной (а) и матовой (б) бумаги

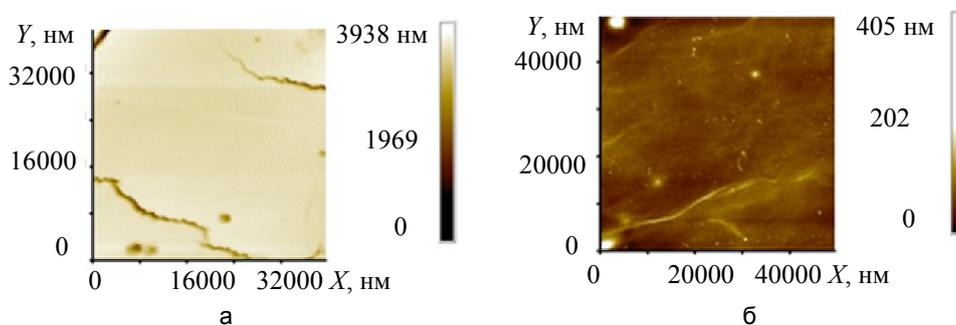


Рис. 3. Характерные АСМ-изображения поверхности глянцевой (а) и ламинированной (б) бумаги

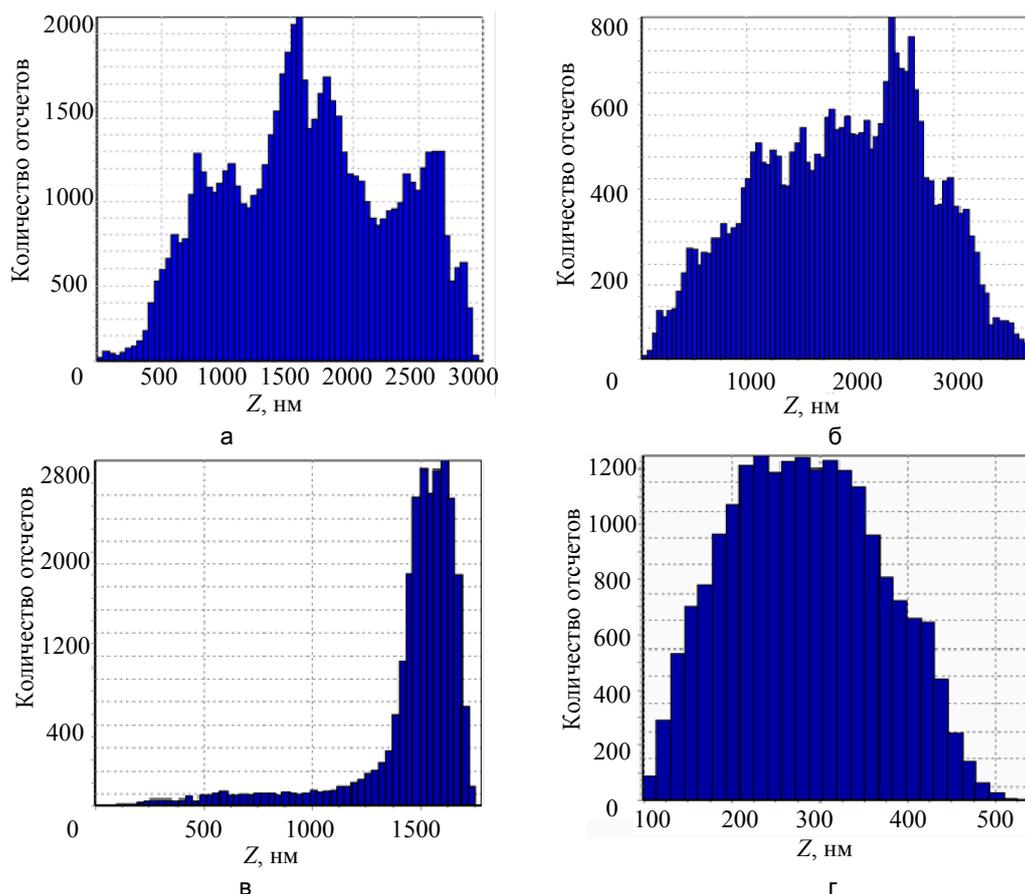


Рис. 4. Гистограммы шероховатости для АСМ-изображений поверхности необработанной (а), матовой (б), глянцевой (в) и ламинированной (г) бумаги

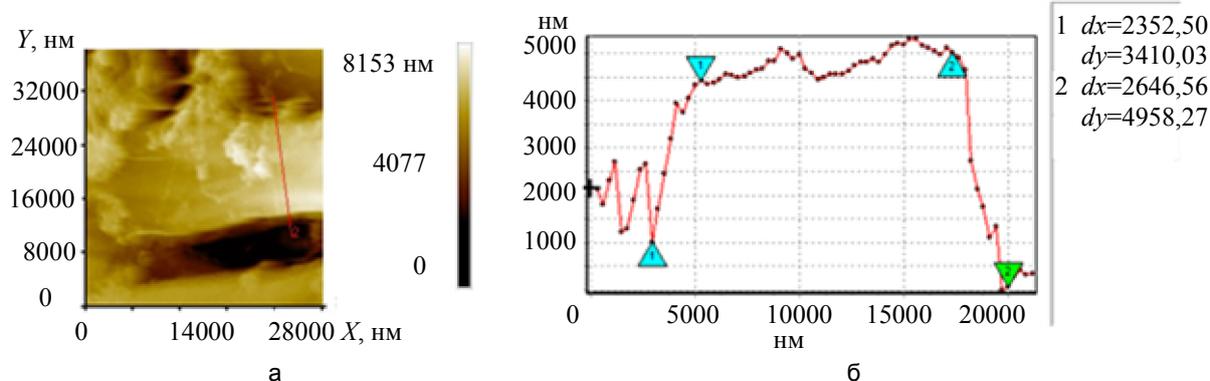


Рис. 5. Характерное АСМ-изображение поверхности необработанной бумаги (а) и сечение одного волокна целлюлозы (б). Метки обозначают крайние положения для измерения приращений по осям x и y

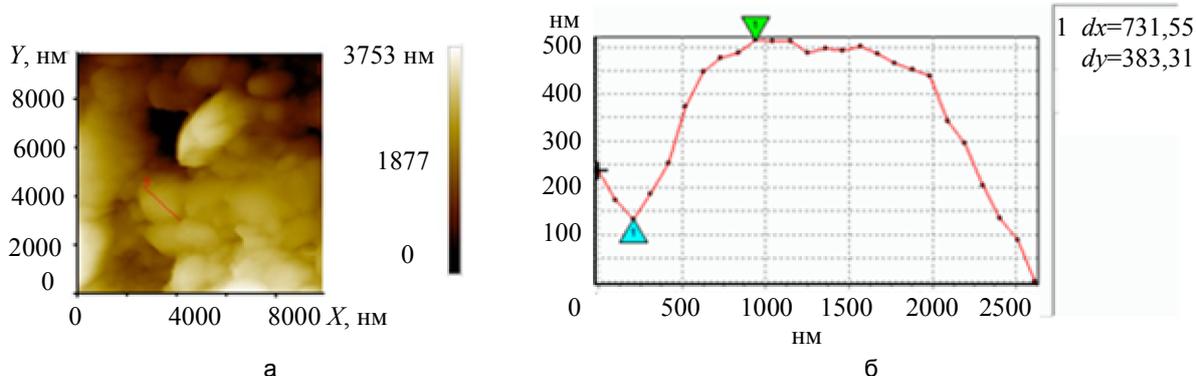


Рис. 6. Характерное АСМ-изображение поверхности матовой бумаги (а) и сечение одного зерна (б)

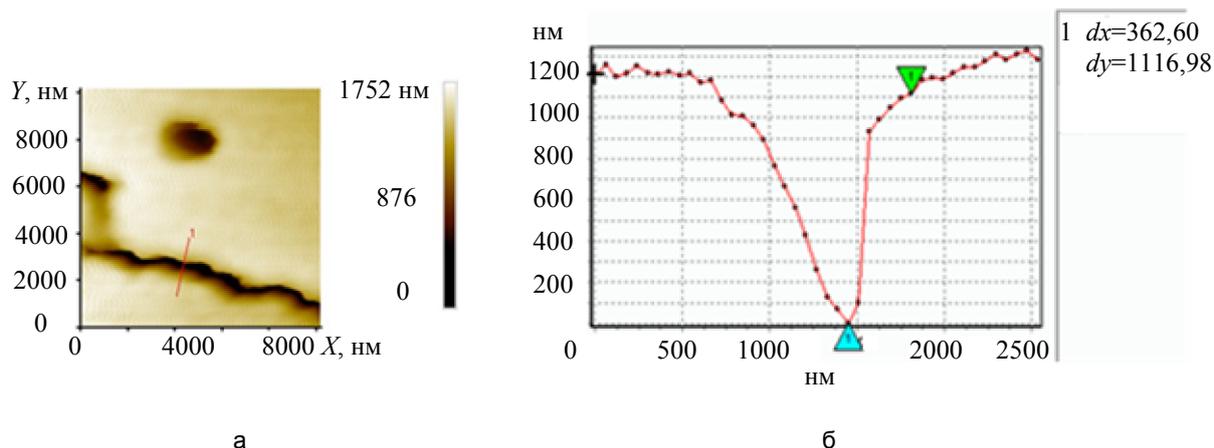


Рис. 7. Характерное АСМ-изображение поверхности глянцевой бумаги (а) и сечение трещины (б)

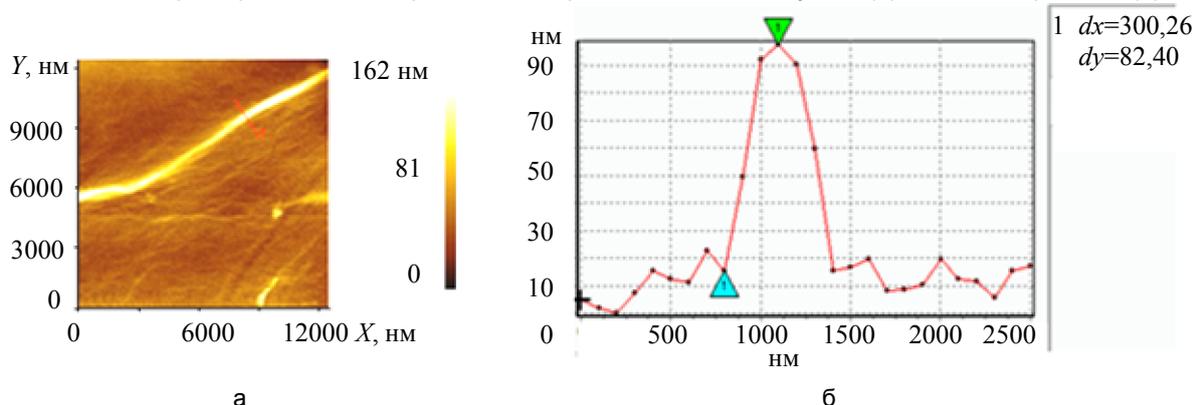


Рис. 8. Характерное АСМ-изображение поверхности ламинированной бумаги (а) и сечение одной микроскладки (б)

Обсуждение результатов

При исследовании необработанной бумаги были выявлены отдельные перекрывающиеся волокна целлюлозы, хаотично расположенные на всей поверхности. Высота волокон составила 3,5–5 мкм, ширина около 12–15 мкм (рис. 5, б). При исследовании матовой бумаги на ее поверхности были выявлены гранулы размером около 2 мкм в диаметре и 400–500 нм в высоту (рис. 6, б). На глянцевой бумаге были отмечены трещины и углубления на поверхности шириной около 1 мкм и глубиной 1–1,2 мкм (рис. 7, б). Поверхность ламинированной бумаги обладает относительно низкой шероховатостью, при этом на ней наблюдаются складки ламинарного покрытия, составляющие в высоту около 80 нм и в ширину около 500 нм (рис. 8, б).

Таким образом, была исследована структура и особенности наноразмерных объектов на поверхности бумаги разного вида, а также получены данные о перепадах высот и общей шероховатости образцов. Метод АСМ позволяет проводить высокоточную диагностику структуры и контроль параметров образцов в наноразмерном масштабе, что невозможно при использовании стандартных методов диагностики поверхности, таких как оптическая микроскопия или использование профилометров. Показана возможность применения доступного и простого в использовании на производстве прибора, позволяющего проводить высокоточный экспресс-контроль параметров, качества и типа выпускаемой бумаги.

Заключение

На основании выполненных исследований показана возможность проведения высокоточной диагностики поверхности наноразмерных структур различных видов бумаги методом атомно-силовой микроскопии. Проведены исследования параметров бумаги различных сортов, измерены основные структурные элементы и поверхностные особенности в нанометровом диапазоне. Проанализированы гистограммы перепадов высот, получены значения средней шероховатости бумаги различного типа.

Выявлено, что наименьшей шероховатостью обладают ламинированная и глянцевая бумага, а более высокими параметрами шероховатости обладают необработанная целлюлозная и матовая бумаги. Показана возможность использования учебно-научного сканирующего зондового микроскопа NanoEducator LE для анализа на производстве получаемых структур бумаги на конечной стадии выпуска продукции (контроль структуры и параметров бумаги).

Литература

1. Список продукции материалов для струйной печати Lomond [Электронный ресурс]. Режим доступа: <http://printer.ucoz.ua/Lomond-2005.pdf>, свободный. Яз. рус. (дата обращения 05.12.2013).
2. Chinga G. Structural studies of LWC paper coating layers using SEM and image analysis techniques. Ph D. Thesis. Norwegian University of Science and Technology Department of Chemical Engineering. February 2002. 122 p. [Электронный ресурс]. Режим доступа: <http://www.divaportal.org/smash/get/diva2:121743/FULLTEXT01.pdf>, свободный. Яз. англ. (дата обращения 05.12.2013).
3. Yang D., Parlange J.-Y., Walker L.P., Moran-Mirabal J.M. Investigation of the porous structure of cellulosic substrates through confocal laser scanning microscopy // *Biotechnology and Bioengineering*. 2013. V. 110. N 11. P. 2836–2845.
4. Witec. Focus innovations. Confocal Raman & AFM Imaging of Paper [Электронный ресурс]. Режим доступа: <http://www.witec.de/assets/Downloads/afm/Paper.pdf>, свободный. Яз. англ. (дата обращения 05.12.2013).
5. CD-Laboratory for Surface Chemical and Physical Fundamentals of Paper Strength [Электронный ресурс]. Режим доступа: <https://www.ipz.tugraz.at/index.php/wbPage/wbShow/cd?lang=en>, свободный. Яз. англ. (дата обращения 05.12.2013).
6. Shimadzu – AFM Observation of Coated Paper and Printed Surface [Электронный ресурс]. Режим доступа: <http://www.shimadzu.com/an/surface/spm/data/appli/paper.html>, свободный. Яз. англ. (дата обращения 05.12.2013).
7. Maattanen A., Fors D., Wang S., Valtakari D., Ihalainen P., Peltonen J. Paper-based planar reaction arrays for printed diagnostics // *Sensors and Actuators B*. 2011. V. 160. N 1. P. 1404–1412.
8. JPK Instruments. Application note. Studying texture of paper with AFM and Upright optical microscopy using the BioMAT Workstation [Электронный ресурс]. Режим доступа: https://www.google.ru/url?sa=t&rct=j&q=&esrc=s&source=web&cd=7&ved=0CGoQFjAG&url=http%3A%2F%2Fwww.jpk.com%2Findex.download.2cd8ad3d77e6aa0fb7b45cc3b9d0d192&ei=GxSfUtlfGMrd4QTN34HIBA&usg=AFQjCNEUFL_TpQfKNzaYtg2ANUGALZSKFQ&bvm=bv.57155469,d.bGE&cad=rjt, свободный. Яз. англ. (дата обращения 05.12.2013).
9. Paiva A.T., Sequeira S.M., Evtuguin D.V., Kholkin A.L., Portugal I. Nanoscale Structure of Cellulosic Materials: Challenges and Opportunities for AFM // *Modern Research and Educational Topics in Microscopy/ Eds A. Méndez-Vilas, J. Díaz. Formatex, 2007. V. 2. P. 726–733.*
10. Persson B.N.J., Ganser C., Schmied F., Teichert C., Schennach R., Gilli E., Hirn U. Adhesion of cellulose fibers in paper // *J. Phys.: Condens. Matter*. 2013. V. 25. N 4. P. 045002-1–045002-11.
11. Pinzari F., Pasquariello G., De Mico A. Biodeterioration of Paper: A SEM Study of Fungal Spoilage Reproduced Under Controlled Conditions // *Macromol. Symp*. 2006. V. 238. N 1. P. 57–66.
12. Ion R.-M., Doncea S.M., Ion M.-L., Rădițoiu V., Amărieuți V. Surface investigations of old book paper treated with hydroxyapatite nanoparticles // *Applied Surface Science*. 2013. V. 285. Part A. P. 27–32.
13. Piantanida G., Pinzari F., Montanari M., Bicchieri M., Coluzza C. Atomic Force Microscopy Applied to the Study of Whatman Paper Surface Deteriorated by a Cellulolytic Filamentous Fungus // *Macromol. Symp*. 2006. V. 238. N 1. P. 92–97.
14. NT-SPb. Наноэдюкатор LE [Электронный ресурс]. Режим доступа: <http://nt-spb.dev.ntmdt.ru/device/view/nanoeducator-1-1>, свободный. Яз. рус. (дата обращения 05.12.2013).
15. Голубок А.О., Пинаев А.Л., Феклистов А.А., Чивилихин С.А. Об устойчивости вольфрамовых зондов при функционировании сканирующего зондового микроскопа в режимах динамической силовой литографии и наноиндентирования // *Научно-технический вестник СПбГУ ИТМО*. 2011. № 4 (74). С. 91–97.
16. Бумага для принтера – Svetocopy Paper [Электронный ресурс]. Режим доступа: <http://sveto-copy.ru/>, свободный. Яз. рус. (дата обращения 05.12.2013).
17. Материалы для струйной печати офисных форматов Lomond [Электронный ресурс]. Режим доступа: <http://www.lomond.ru/>, свободный. Яз. рус. (дата обращения 05.12.2013).
18. ISO 4287:1997. Geometrical Product Specifications (GPS) – Surface texture: Profile method – Terms, definitions and surface texture parameters. 01.04.1997. Geneva, International Organization for Standardization. 35 p.

Жуков Михаил Валерьевич

– инженер, студент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, cloudjyk@yandex.ru, zerocomb@bk.ru

Mikhail Zhukov

– engineer, student, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, cloudjyk@yandex.ru, zerocomb@bk.ru

УДК 661.185.4, 661.187, 535.345.1, 536.331

**ЭНЕРГОСБЕРЕГАЮЩАЯ ТЕХНОЛОГИЯ РАСПЛАВЛЕНИЯ
ХИМИЧЕСКИХ ВЕЩЕСТВ СВЕТОВЫМ ИЗЛУЧЕНИЕМ****А.Н. Черепанов^а**^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Leva0007@rambler.ru

Представлена новая технология расплавления химических веществ с просветляющейся жидкой фазой световым излучением. Предложен и опробован способ расплавления, основанный на воздействии световым излучением. Разработана конструкция излучателя на основе галогенных ламп, которая использовалась в качестве насадки на контактный нагреватель. Лампы помещались в индивидуальные прозрачные корпуса, что позволяет выровнять распределение температур по поверхности излучателя и избежать сильного перегрева ламп. Такие излучатели применялись для расплавления синтетических жирных кислот, используемых в производстве моющих средств. При проектировании конструкции излучателя использовались ранее разработанные методики расчета скоростей процесса плавления при комбинированном воздействии на расплавляемое вещество светового излучения и тепловой мощности, выделяемой на поверхности прозрачного корпуса. На основе расчетов и данных по световым и тепловым характеристикам излучателей определялось их необходимое количество и требуемое расстояние между ними. Экспериментальная проверка показала возможность проплавления канала глубиной 1 м за 5–6 ч. Это в 3 раза меньше, чем необходимо для проплавления канала контактным нагревателем. Энергопотребление увеличено всего на 7% относительно энергозатрат контактного нагревателя. При этом был проплавлен канал большего диаметра, чем при использовании только контактного нагревателя.

Ключевые слова: синтетические жирные кислоты, источники светового излучения, фазовый переход, галогенные лампы.

**ENERGY-SAVING TECHNOLOGY OF CHEMICAL AGENTS MELTING
BY LIGHT RADIATION****A. Cherepanov^b**^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Leva0007@rambler.ru

The paper deals with new melting technology of chemical agents with antireflection liquid phase by means of light radiation. A method of melting is proposed, based on the light radiation effect. Emitter design based on halogen lamps is developed and used as a nozzle to the contact heater. The lamps are placed into individual transparent cases which gives the possibility to align the temperature distribution on the emitter surface and avoid excessive heat of the lamps. Such emitters were used for melting of synthetic fatty acids used in the production of detergents. Previously developed calculating methods for the melting process velocity of the combined influence of light radiation and thermal power released on the surface of the transparent case on the melting agents are used in the process of emitter designing. The number of emitters and the distance between them are determined in view of calculations and data of light and thermal characteristics of emitters. Experimental verification showed the possibility of channel depth melting equal to 1 m in 5-6 hours. This is three times less than it is necessary for channel melting by contact heater. At the same time energy consumption is increased only by 7% as compared with that one of the contact heater. Furthermore, the channel melted down has got a diameter larger than by using only a contact heater.

Keywords: synthetic fatty acids, sources of light, phase transfer, halogen lamp.

Введение

Большую техническую проблему на месте потребления химических продуктов составляет их перевод из твердой в жидкую фазу. Исходный продукт транспортируется в цистернах и крупногабаритных контейнерах из полимерных материалов с тонкой стенкой. Традиционные способы их расплавления энергозатратны, неэкологичны и требуют значительных затрат времени. Они основаны на подводе тепловой энергии, что малоэффективно вследствие низкой теплопроводности расплавляемых веществ. Обычно операция расплавления основывается на использовании «острого пара» [1] или на проплавлении канала контактным нагревателем. Наиболее распространенный способ расплавления основан на проплавлении канала в твердой фазе продукта с последующей рециркуляцией, т.е. прокачиванием подогретой вне контейнера жидкой фазы через проплавленный канал. При этом из отверстия на донной части контейнера принимается жидкая фаза, подогретая в отдельном теплообменнике, и подается обратно в контейнер через горловину в его верхней части. Из анализа многочисленных публикаций с описаниями традиционных методов нагрева и плавления [1–6] очевидна необходимость минимизации энергозатрат и ускорения процесса расплавления.

Объект исследований

Описанные в настоящей работе методы светового облучения применялись для расплавления синтетических жирных кислот (СЖК) фракции C₁₆–C₂₂, выделенных ректификацией или получаемых окислением твердого парафина или смеси твердого и жидкого парафинов. Эти кислоты применяются в производстве синтетических моющих средств, в мыловаренной промышленности в качестве компонентов для получения поверхностно активных веществ и в других производствах. Поставки таких жирных кислот,

как правило, осуществляются в полиэтиленовых контейнерах с формой, близкой к кубической, объемом около 1 м^3 , а иногда – в форме параллелепипеда с небольшим различием размера сторон. Заливка СЖК в контейнер производится сверху через горловину круглой формы диаметром 150–300 мм, а откачка осуществляется через штуцер с запорным краном, расположенным в нижней части контейнера. В обычных условиях СЖК находится в контейнере в твердом состоянии.

Температура застывания СЖК разных сортов составляет для высшего сорта – 47–54°C, первого сорта – 46,5–53°C, второго сорта – 46–53°C. Не допускается длительное хранение СЖК в жидком состоянии при повышенных, свыше + 90°C, температурах. При использовании традиционных методов нагрева с подводом тепловой энергии для таких веществ нельзя применять высокотемпературные источники тепловыделений вследствие достаточно низкой температуры воспламенения (примерно 120°C). Нежелательно попадание в контейнер в процессе разогрева кислоты посторонних веществ, например, конденсированной воды.

Ранее были разработаны и сформулированы критерии эффективности способов и устройств для расплавления химических продуктов [7]. Эти критерии позволяют определять совокупность параметров, обеспечивающих наилучшие рабочие характеристики каждого конкретного устройства для нагрева и расплавления твердой фазы органического вещества.

Целью настоящей работы является:

- описание разработанных автором новых энергосберегающих технологий расплавления химических веществ (на примере СЖК) световым излучением в крупногабаритном контейнере из полиэтилена;
- описание разработанного компактного излучателя, используемого в качестве насадки на традиционно используемый контактный нагреватель;
- проведение экспериментальных исследований на реальном объекте – крупногабаритном контейнере с СЖК с целью проверки эффективности предложенного способа и разработанного устройства на основе сопоставления с традиционно используемым контактным нагревателем.

Технология расплавления химических веществ световым излучением

Для обеспечения существенной экономии энергозатрат и времени, затрачиваемого на проведение операции расплавления химических продуктов, были проведены исследования и разработки новых способов и устройств для их реализации. Исследования проводились для смеси СЖК фракций C_{16} – C_{22} . Новый способ нагрева и плавления СЖК основан на применении светового излучения, поглощаемого как поверхностью непрозрачной твердой фазы, отделенной от излучателя расширяющимся во времени слоем частично прозрачной жидкой фазы, так и частично поглощающим, прилегающим к источнику излучения слоем жидкой фазы. При использовании поступательно движущегося излучателя дополнительным источником нагрева являются тепловыделения в самом источнике излучения. В качестве комбинированного источника излучения и тепловыделений использовалась галогенная лампа, помещаемая в прозрачный стеклянный корпус в виде пробирки, что позволяет выровнять распределение температур по поверхности излучателя и избежать сильного перегрева ламп [8–15].

Эффективность способа расплавления СЖК световым излучением была обоснована ранее результатами экспериментальных исследований спектров пропускания жидкой фазы СЖК, а также спектров излучения галогенных ламп [8, 12]. Одновременно были исследованы интегральные характеристики пропускания слоев жидкой фазы, а также световые и тепловые характеристики источников излучения на основе галогенных ламп. Выбор галогенных ламп обосновывался их сопоставлением с лампами накаливания по критерию скорости расплавления СЖК. Результаты этих исследований обобщены в работе [15].

Описание технологической установки для расплавления СЖК

На рис. 1 приведена фотография технологической установки для расплавления СЖК в полиэтиленовом контейнере.

Процесс расплавления канала в объеме СЖК продолжается до 18 ч, что значительно превышает длительность одной рабочей смены. Суть предлагаемого метода заключается в применении компактного излучателя, используемого в качестве насадки на наконечник контактного нагревателя. В качестве источника излучения использованы галогенные лампы, которые в результате проведенных нами исследований показали значительное преимущество перед лампами накаливания.

Конструктивная схема пучкового излучателя на основе галогенных ламп

Для эффективного проплавления канала в твердой фазе СЖК необходимо обеспечить достаточно большой диаметр канала, заполненного жидкой фазой. Для этого требуется построить конструкцию расплавляющего излучателя, состоящего из необходимого количества источников излучения. При этом во избежание воспламенения необходимо обеспечить допустимый тепловой режим на поверхности излу-

телей. Для этого необходимо размещение галогенных ламп в прозрачных стеклянных или кварцевых корпусах, при этом должна обеспечиваться рабочая температура галогенной лампы 300°C.

Рабочий излучатель служит насадкой на оконечность контактного нагревателя и может быть построен по схеме параллельно установленных излучателей, каждый из которых представляет собой галогенную лампу в корпусе (пучковая конструкция излучателя). Конструкция подобного излучателя представлена на рис. 2.

Для представленной конструкции был проведен расчет по ранее разработанной методике [15] с целью выбора мощности галогенных ламп и расстояний между корпусами излучателей. Излучатель был изготовлен и опробован в эксперименте по проплавлению канала в твердой фазе СЖК, заполняющей пластиковый контейнер кубической формы объемом 1 м³.

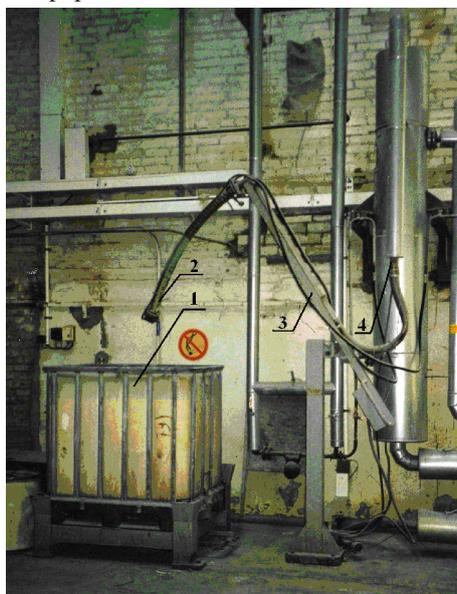


Рис. 1. Технологическая установка для расплавления синтетических жирных кислот в полиэтиленовом контейнере: 1 – контейнер с СЖК; 2 – контактный нагреватель с трубопроводом, в который подается горячая вода или пар; 3 – рычаг для подачи контактного нагревателя; 4 – узел для снабжения паром теплообменника, размещенного на контактном нагревателе

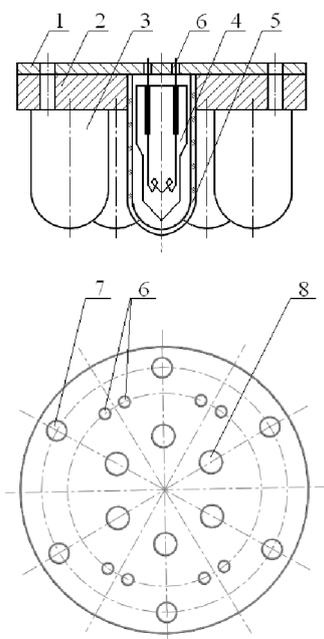


Рис. 2. Конструктивная схема оптического наконечника семилампового излучателя для проплавления канала в СЖК: 1 – текстолитовая шайба; 2 – шайба из дюралюминия с отверстиями для установки галогенных ламп; 3 – кварцевый корпус излучателя; 4 – галогенная лампа; 5 – прозрачная стенка кварцевого корпуса; 6 – отверстия для электрических проводов; 7 – отверстия для крепления шайб; 8 – отверстия для пропускания жидкой фазы

Результаты экспериментальных исследований

Экспериментальная проверка показала, что применение данного излучателя позволило проплавлять канал глубиной 1 м за 5–6 ч, что в 3 раза меньше, чем необходимо для проплавления канала контактным нагревателем. При этом мощность установки увеличена всего на 7% относительно энергозатрат контактного нагревателя, а потребляемая энергия для проплавления канала снижена в 2,8 раза. Кроме того, обеспечивался больший диаметр канала, заполненного жидкой фазой.

Основным недостатком конструкции излучателя, представленной на рис. 2, является застревание нагревателя с излучающей насадкой на краевых участках конструкции в случае перегорания галогенной лампы. А при выходе из строя центрального излучателя скорость проплавления понижается в 3 раза, что лишает пучковый излучатель преимуществ перед контактным нагревателем. Данного недостатка можно избежать за счет использования конструкции моноблочного излучателя, в котором все лампы размещаются в общем корпусе и влияние отдельно взятой лампы сглаживается. Однако это является предметом отдельного исследования.

Заключение

Разработана новая технология расплавления химических продуктов, основанная на сочетании светового и теплового воздействий. Предложена конструкция пучкового излучателя на основе галогенных ламп. По результатам экспериментальной проверки излучатель, используемый в качестве насадки на контактный нагреватель, показал высокую эффективность: достигнуто увеличение скорости проплавления канала в твердой фазе синтетической жирной кислоты в 3 раза. Контактный нагреватель необходим для поддержания химического продукта в расплавленном состоянии после прохождения светового излучателя через уже расплавленные участки канала. Приращение потребляемой установкой мощности на 7% компенсируется снижением общих затрат энергии в 2,8 раза за счет уменьшения длительности процесса. Весь процесс проплавления канала в твердой фазе синтетической жирной кислоты занимает 5–6 ч вместо 18 ч по традиционной технологии. Полученные результаты обосновывают необходимость дальнейшего совершенствования конструкции излучателя. Целесообразность развития работ в этом направлении неоспорима.

Литература

1. Бережковский М.И. Хранение и транспортирование химических продуктов. Л.: Химия, 1982. 256 с.
2. Бухштаб З.И., Мельник А.П., Ковалев В.М. Технология синтетических моющих средств: Учеб. пособие для вузов. М.: Легпромбытиздат, 1988. 320 с.
3. Бурдо Ю.И., Палевский Л.В., Бадещенков С.В., Грищенко Е.Н. Особенности процесса непрерывного дозирования жидкого сульфанола в производстве СМС // Новое в области технологии, аппаратурного оформления и разработки средств автоматизации и механизации процессов производства ТБХ: Сборник научных трудов. М.: НИИТЭХИМ, 1990. С. 23–28.
4. Herman de Groot W., Adami I., Moretti G.F. The manufacture of modern detergent powders. Wassenaar, The Netherlands: Herman de Groot Academic Publisher, 1995. 199 p.
5. Woollatt E. The manufacture of soaps, other detergents and glycerine. Chichester, UK: Ellis Horwood, 1985. 473 p.
6. Handbook of Detergents. Part F: Production/ Eds U. Zoller, P. Sosis. Boca Raton, USA: CRC Press, 2009. 593 p.
7. Черепанов А.Н. Критерии эффективности методов расплавления химических продуктов, транспортируемых в твердой фазе // Экономика и производство. 2003. № 3. С. 63–67.
8. Волинкин В.М., Ханков С.И., Черепанов А.Н. Экспериментальное исследование возможностей расплавления синтетических жирных кислот методом светового облучения // Химическая промышленность. 2004. № 6. С. 46–58.
9. Волинкин В.М., Ханков С.И., Черепанов А.Н. Энергетический баланс в технологических установках для расплавления синтетических жирных кислот // Химическая технология. 2004. № 10. С. 34–41.
10. Волинкин В.М., Ханков С.И., Черепанов А.Н. Методика для инженерных расчетов скорости проплавления канала в твердой фазе синтетических жирных кислот трубчатым контактным нагревателем // Химическая технология. 2003. № 8. С. 40–44.
11. Волинкин В.М., Ханков С.И., Черепанов А.Н. Методика расчета скорости расплавления химических продуктов класса синтетических жирных кислот оптическими излучателями // Теплоэнергетика. 2003. № 7. С. 45–51.
12. Волинкин В.М., Ханков С.И., Черепанов А.Н. Исследование эффективности методов расплавления синтетических жирных кислот, основанных на использовании источников светового излучения // Химическая технология. 2004. № 1. С. 18–21.

13. Волынкин В.М., Ханков С.И., Черепанов А.Н. Устройство для нагрева и плавления химических продуктов. Свидетельство на полезную модель № 20909; заявка № 2001118373; приоритет от 04.07.2001. Зарегистрировано в Госреестре ПМ РФ 10.12.2001. Публикация 10.12.2001. Бюл. № 34. МПК 7 В 67 D 5/00, В 65 D 88/74.
14. Волынкин В.М., Ханков С.И., Черепанов А.Н. Устройство для плавления и нагрева химических продуктов. Свидетельство на полезную модель № 21195; заявка № 2001118375; приоритет от 04.07.2001. Зарегистрировано в Госреестре ПМ РФ 27.12.2001. Публикация 27.12.2001. Бюл. № 36. МПК 7 В 67 D 5/00.
15. Черепанов А.Н., Ханков С.И. Методы расчета процессов расплавления органического химического сырья с просветляющейся жидкой фазой. СПб: НТЦ им. Л.Т. Тучкова, 2006. 230 с.

Черепанов Аркадий Николаевич

– кандидат технических наук, докторант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Leva0007@rambler.ru

Arkady Cherepanov

– PhD, doctoral candidate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Leva0007@rambler.ru

УДК 681.7.068: 519.25

О ТЕХНОЛОГИЧЕСКИХ НЕСОВЕРШЕНСТВАХ ГЕОМЕТРИЧЕСКИХ ПАРАМЕТРОВ СИЛОВОГО СТЕРЖНЯ ДЛЯ ЗАГОТОВКИ ОПТИЧЕСКОГО ВОЛОКНА PANDA¹

Н.В. Семенов^a, Н.А. Труфанов^a, А.А. Адамов^b

^a Пермский национальный исследовательский политехнический университет, Пермь, Россия, nikita@studio-gd.ru

^b Институт механики сплошных сред УрО РАН, Пермь, Россия, adamov@icmm.ru

На основе экспериментальных данных о геометрии сечений партии силовых стержней для заготовки анизотропного оптического волокна типа Panda выявлены зависимости изменения радиальных размеров заготовок от угловой координаты. Описан алгоритм статистической обработки геометрических параметров, который позволяет найти подобие в геометрии сечений разных стержней и выявить характерную форму сечения для всех образцов. Алгоритм заключается в смещении точек зависимостей радиусов от угловых координат с целью нахождения кривых с максимальным показателем корреляции, а затем на основе найденных кривых находится распределение значений радиусов для каждой угловой координаты. Показано, что при изготовлении силовых стержней с переменным содержанием легирующих добавок по радиусу модифицированным методом химического осаждения из газовой фазы не удается получить идеально круглую форму поперечного сечения стержня. При этом наиболее существенные отклонения (до 10%) от кругового профиля фиксируются для границы, разделяющей легированную сердцевину стержня и наружный слой чистого кварца, а отклонения других границ от круга незначительны (десятые доли процента). Установлено, что закономерным образом в условиях данного технологического процесса реализуется некруговая форма поперечного сечения легированного ядра, которую можно условно назвать формой типа «огранка» (с тремя «гранями» разной длины). Последующий длительный контроль формы поперечного сечения силовых стержней подтвердил повторяемость полученных результатов.

Ключевые слова: силовой стержень, оптическое волокно Panda, технологический разброс геометрических параметров, статистическая обработка результатов измерений.

TECHNOLOGICAL IMPERFECTIONS OF FORCE ROD GEOMETRICAL PARAMETERS FOR PANDA OPTICAL FIBERS PRODUCTION²

N. Semenov^c, N. Trufanov^c, A. Adamov^d

^c Perm National Research Polytechnic University, Perm, Russia, nikita@studio-gd.ru

^d Institute of Continuous Media Mechanics, Ural Branch of the Russian Academy of Sciences, Perm, Russia, adamov@icmm.ru

The paper deals with the dependence in radial dimensions changes of the workpieces on the angular coordinate based on experimental data about the geometry of the cross-sections of force rods batch for the production of anisotropic Panda-type optical fibers. An algorithm of geometrical parameters statistical processing is described, which gives the possibility to find the similarity in the geometry of the cross-sections of different force rods, and reveal the characteristic shape of the cross-section for all the samples. The algorithm contains the shift of points for dependence changes in radial dimensions of the workpieces on the angular coordinate in order to find curves with maximum correlation coefficient. Afterwards, the distribution of the radius values for each angular coordinate is found basing on the obtained curves. It is demonstrated that in case of the force rods production with variable amount of alloying dopes on the radius via the MCVD method, it is impossible to get

¹ Работа выполнена при финансовой поддержке гранта РФФИ (проект №13-08-96036 p_ural_a).

² The work was done with financial support from the Russian Foundation for Basic Research (project №13-08-96036 r_ural_a)

ideally round shape of the cross-section. At the same time, the biggest deviation from the round profile (up to 10%) is fixed on the border between the alloyed core of the force rod and the external layer of the pure quartz, whereas the deviations on other borderlines are negligible (a tenth hundredth of a percent). It is determined that within this technological process the non-circular cross-sectional shape of alloyed core is realized, which can be called “faceting” for convenience (with three “faces” of different length). The following long control of the cross-sectional shape of the force rods has proved the results recurrence.

Keywords: force rod, Panda optical fiber, technological geometry scattering, statistical processing of measurement results.

Введение

Для изготовления волоконно-оптических датчиков широко применяются анизотропные кварцевые оптические волокна [1–4], поддерживающие поляризацию сигнала за счет разности механических напряжений в светопроводящей жиле, например, волокна типа Panda. Создание необходимого уровня разности главных напряжений в жиле оптического волокна Panda обеспечивается введением в состав заготовки для вытяжки волокна специальных силовых элементов из легированных кварцевых стекол, имеющих большее значение коэффициента линейного температурного расширения, чем чистый кварц [5–7]. Волокна типа Panda имеют два круглых цилиндрических силовых стержня с переменным содержанием легирующих добавок по радиусу, изготовленных с помощью модифицированного метода химического осаждения из газовой фазы (MCVD) [8].

Конструирование волокна Panda обычно ведется с использованием предположения о том, что сечения силовых стержней имеют круговую форму [9–11], однако на практике профиль сечения зависит от множества факторов, предопределенных технологическим процессом изготовления. Технологический разброс геометрических параметров формы сечения может сказываться на прочности самих стержней и на оптических характеристиках волокон, в составе которых применяются такие силовые элементы.

В настоящей работе на основе измерений, анализа и статистической обработки [12] значений геометрических параметров большой партии изготовленных силовых стержней для заготовок анизотропного волокна Panda с помощью программного пакета MATLAB [13, 14] устанавливается математический закон, описывающий возможные отклонения радиуса легированной зоны от круговой формы.

Описание технологии получения данных

Измерение геометрических параметров исследуемых стержней производится с помощью анализатора «2600 Photon Kinetics». Прибор позволяет осуществлять поворот исследуемой заготовки на заданный угол и выбор сечения по длине стержня. Изучается профиль коэффициента преломления в сечении заготовки на основе изменения углов отклонения лазерного луча, сканирующего заготовку по диаметру, а затем определяется положение границ слоев с разными оптическими средами.

Рассмотренные силовые стержни производились на одном и том же оборудовании, сходны по конструкции и состоят из трех различных по химическому составу слоев (рис. 1). Основным интересом представляет изучение геометрии наружного кварцевого слоя силового стержня, а также границ между различными оптическими средами, образованных за счет легирования заготовок разными добавками.

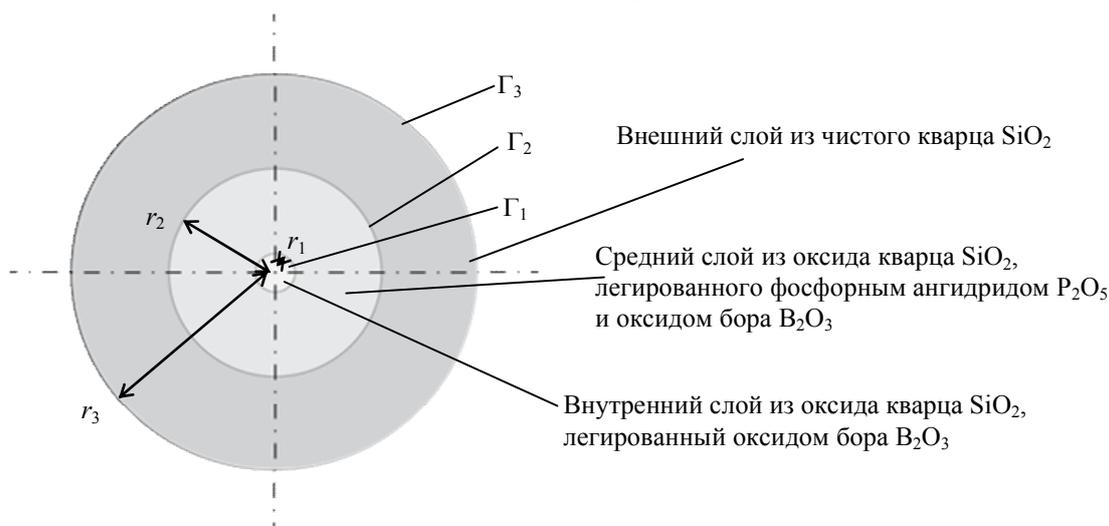


Рис. 1 – Схема осевого сечения заготовки силового стержня: Γ_1 – граница между внутренним и средним легированными слоями; Γ_2 – граница между средним и внешним слоями; Γ_3 – наружная граница заготовки

Для каждого стержня замеры положения границ Γ_1 , Γ_2 и Γ_3 снимаются в 15 радиальных сечениях (30 точек) и 10 осевых сечениях. На рис. 2 в качестве примера показаны полученные данные для одного из типовых осевых сечений. Всего было обработано 100 стержней, изготовленных друг за другом на одном и том же оборудовании за 5 месяцев. Таким образом, имеется по 30 наборов фактически замеренных зависимостей радиусов от угла в 30 точках для каждого стержня $R_{nij}(\varphi_k)$, $n = \overline{1,3}$, $i = \overline{1,100}$, $j = \overline{1,10}$, $k = \overline{1,30}$, где n – номер границы; i – номер стержня из набора; j – номер сечения по длине; k – номер углового сечения. Длина заготовок составляет $L \approx 0,5$ м. Проектные значения радиусов: сердцевин $r_1 = 0,46$ мм, двухкомпонентного слоя $r_2 = 2,8$ мм, наружного слоя $r_3 = 5,3$ мм.

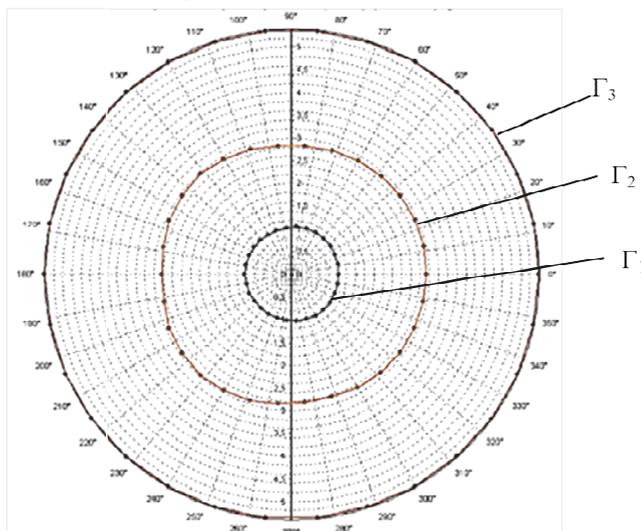


Рис. 2. Диаграмма зависимости экспериментально установленных радиусов от угловой координаты $R_{nij}(\varphi_k)$ для одного из сечений типового силового стержня

Статистическая обработка результатов

Анализ данных на границах Γ_1 и Γ_3 не позволил выявить существенных отклонений от проектных значений (разброс в значениях $R_{1ij}(\varphi_k)$ и $R_{3ij}(\varphi_k)$ составляет десятые доли процента от среднего значения в рамках конкретного стержня). Внешняя граница стержня (Γ_3) имеет очень малые случайные отклонения от круговой формы. Производитель исходных опорных трубок, из которых изготавливаются заготовки, указывает максимальное отклонение радиуса внешней границы не более чем 0,2% от проектного значения. В связи с этим далее речь идет о положении точек на границе Γ_2 между зоной чистого оксида кремния SiO_2 и легированной частью, так как разброс значений функций $R_{2ij}(\varphi_k)$ на этой границе довольно велик и достигает 10% (до 275 мкм) и более в рамках одного стержня. Для упрощения записей индекс n принимаем равным двум и опускаем $R_{ij}(\varphi_k) = R_{2ij}(\varphi_k)$.

Будем считать функцию радиуса от угловой координаты $R_{ij}(\varphi_k)$ случайной величиной, где i – номер стержня; j – номер осевого сечения; k – номер угловой координаты.

При построении функции изменения радиуса от угла $R_{ij}(\varphi_k)$ для 10 сечений по длине для каждого стержня было выявлено, что при наложении графиков друг на друга они практически полностью повторяются. Исходя из этого, для описания геометрии сечений отдельного стержня можем использовать среднее арифметическое значение радиусов соответствующих точек по всем осевым сечениям стержня:

$$R_i(\varphi_k) = \sum_{j=1}^{10} \frac{R_{ij}(\varphi_k)}{10}.$$

Анализ зависимостей $R_i(\varphi_k)$ на границе Γ_2 для разных стержней позволяет установить их качественное сходство. На всех графиках визуально определяется по три характерных «максимума» и по три «минимума» (вид зависимостей соответствует изображенному на рис. 3).

Поскольку при помещении заготовки в анализатор начальное положение отсчета по углу никак не регламентируется, заготовки, несмотря на схожесть геометрии, оказываются повернутыми друг относительно друга по угловой координате. Наложение графиков друг на друга показывает, что схожести общей картины без предварительной обработки не видно. Построение нормированной корреляционной матрицы также указывает на то, что графики в таком виде между собой не коррелируют.

Для обработки результатов измерений используется следующий алгоритм.

1. За основу выбирается любая зависимость из исследуемого набора $R_n(\varphi_k)$, $n = \overline{1,100}$.
2. Выбирается произвольная зависимость $R_m(\varphi_k)$, $m \neq n$, $m = \overline{1,100}$ из исследуемого набора и проверяется на наличие корреляции с кривой $R_n(\varphi_k)$ по формуле вычисления нормированного корреляционного момента r_{nm} [15]:

$$r_{nm} = \frac{M(R_n - M(R_n))M(R_m - M(R_m))}{\sigma_n \sigma_m},$$

где M – математическое ожидание; σ_n, σ_m – средние квадратичные отклонения для R_n и R_m соответственно.

Кривые сравниваются 30 раз (количество точек соответствует числу возможных комбинаций), при этом каждый раз происходит сдвиг: все точки смещаются влево, а самая левая становится самой правой, $R'_m(\varphi_k) = R_m(\varphi_{k+1})$, $k = \overline{1,29}$ и $R'_m(\varphi_{30}) = R_m(\varphi_1)$.

3. Среди 30 проверенных комбинаций находится та, у которой получается максимальный нормированный корреляционный момент с эталонной кривой n , и запоминается, а график сохраняется с учетом найденного сдвига $R'_i(\varphi_k) = R'_m(\varphi_k)$, где $i = m$. Таким образом, комбинация, полученная с учетом сдвига всех точек на m позиций влево, является максимально подобной эталонной среди возможных.
4. Пункты 2 и 3 повторяются до тех пор, пока не будут обработаны данные обо всех исследуемых стержнях.

Реализация описанной процедуры дает набор кривых $R'_i(\varphi_k)$, максимально коррелирующих между собой (рис. 3): между всеми кривыми нормированный корреляционный момент $r_{nm} > 0,9$. На рис. 3 и далее значения радиусов записываются в виде отношения реального радиуса к среднему.

Таким образом, отклонение формы легированной сердцевинки от круговой повторяется практически на всех изделиях данной партии. Проведем статистический анализ полученных зависимостей $R'_i(\varphi_k)$. Установлено, что распределения радиусов $f(R'_i(\varphi_k))$, $k = \overline{1,10}$ близки к нормальному закону (рис. 4).

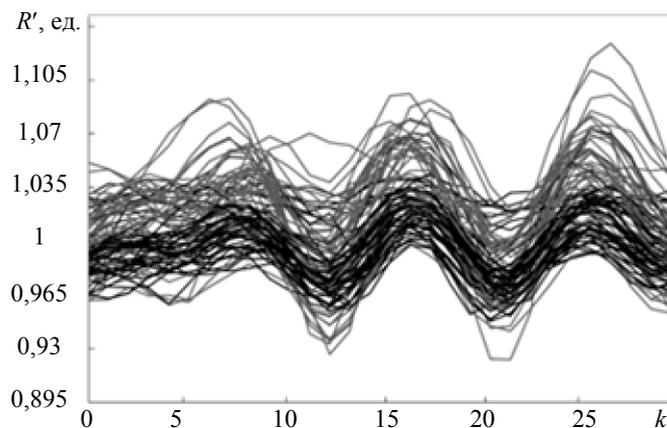


Рис. 3. Распределения $R'_i(\varphi_k)$ для всех исследуемых стержней после обработки

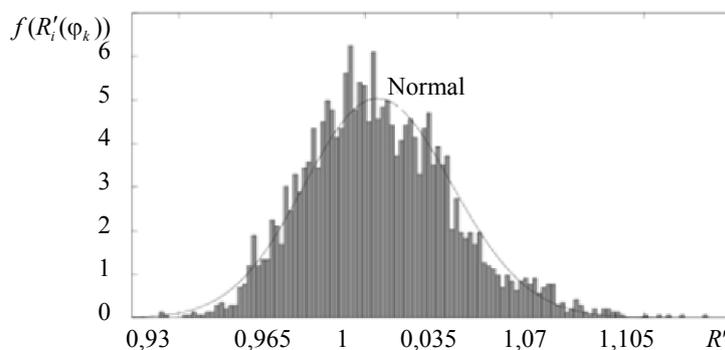


Рис. 4. Вероятность распределения радиусов $R'_i(\varphi_k)$ в одном типовом угловом положении φ_k

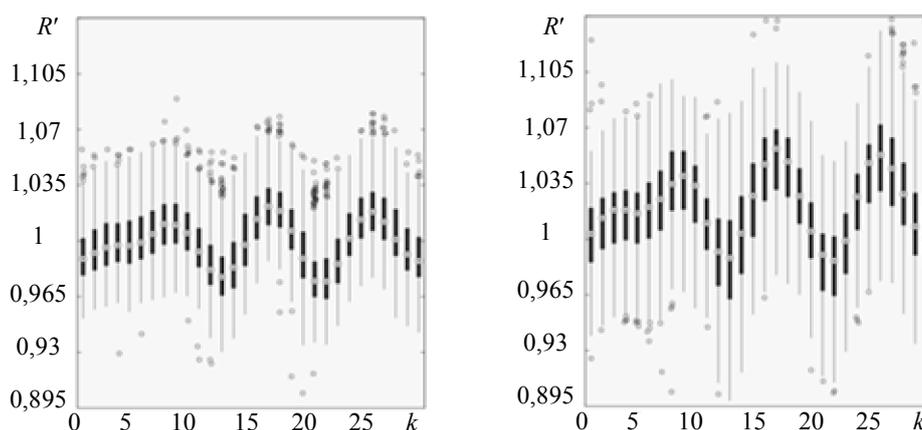


Рис. 5. Распределения вероятностей радиуса $R'_i(\varphi_k)$, $k = \overline{1,30}$: заготовки прошедшие (а) и не прошедшие (б) технологический контроль соответственно

Для наглядности распределения радиусов во всех сечениях можно представить в виде общего графика типа «ящик с усами» (рис. 5). Жирными вертикальными линиями обозначен интерквартильный размах, т.е. 50% всех значений входит в этот диапазон. Границы тонких линий («концы усов») – это границы статистически значимой выборки. Примерно посередине жирных линий указаны медианы, соответствующие серединам выборки, а за пределами линий – одиночные круги, соответствующие случайным выбросам. Видно, что у заготовок, прошедших технологический контроль (рис. 5, а), размах значений меньше, чем у заготовок, не прошедших контроль (отбракованных) (рис. 5, б). Случайных выбросов гораздо больше на графике рис. 5, б. По положению медиан можно судить о симметричности распределения значений в каждом угловом сечении φ_k .

Зависимости математических ожиданий от номера углового сечения $M[R'_i(\varphi_k)]$ также меньше у заготовок, прошедших заводской контроль. Это свидетельствует о том, что увеличение радиальных размеров легированного слоя, увеличение отклонения от круглости или величина случайных выбросов радиуса могут быть связаны с непрохождением технологического контроля. Также следует отметить, что на всех графиках $M[R'_i(\varphi_k)]$ наблюдаются три явных максимума и три минимума, т.е. особенности данного технологического процесса изготовления силовых стержней закономерно приводят к реализации некруговой формы поперечного сечения, которую можно условно назвать формой типа «огранка» (с тремя «гранями» разной длины).

Заключение

Показано, что при изготовлении методом MCVD силовых стержней с переменным содержанием легирующих добавок по радиусу, предназначенных для заготовки анизотропного оптического волокна типа Panda, не удастся получить идеально круглую форму поперечного сечения стержня. При этом наиболее существенные отклонения от кругового профиля фиксируются для границы, разделяющей легированную сердцевину стержня и наружный слой чистого кварца, а отклонения других границ от круга незначительны. В результате статистической обработки результатов измерений геометрических параметров стержней за достаточно длительный срок установлено, что закономерным образом в условиях данного технологического процесса реализуется некруговая форма поперечного сечения легированного ядра, которую можно условно назвать формой типа «огранка» (с тремя «гранями» разной длины). Имеются основания считать, что результат есть следствие реализации при изготовлении стержня цепочки технологических операций, сопровождающихся неосесимметричными циклическими термосиловыми воздействиями.

Литература

1. Guan R., Zhu F., Gan Z., Huang D., Liu S. Stress birefringence analysis of polarization maintaining optical fibers // *Optical Fiber Technology*. 2005. V. 11. N 3. P. 240–254.
2. Окоси Т., Окамото К., Оцу М., Нисихара Х., Каюма К., Хататэ К. Волоконно-оптические датчики: Пер. с яп. Л.: Энергоатомиздат, 1990. 256 с.
3. Гроднев И.И., Ларин Ю.Т., Теумин И.И. Оптические кабели: конструкции, характеристики, производство и применение. 2-е изд., перераб. и доп. М.: Энергоатомиздат, 1991. 264 с.
4. Бутусов М.М., Галкин С.Л., Орбинский С.П., Пал Б.П. Волоконная оптика и приборостроение. Л.: Машиностроение, 1987. 328 с.
5. Мазурин О.В., Стрельцина М.В., Швайко-Швайковская Т.П. Свойства стекол и стеклообразующих расплавов: Справочник. Л.: Наука, 1973. Т. 1. 444 с.

6. Лeko В.К., Мазурин О.В. Свойства кварцевого стекла. Л.: Наука, 1985. 166 с.
7. Shelby J.E. Introduction to glass science and technology. 2nd ed. Cambridge: The Royal Society of Chemistry, 2005. 291 p.
8. Бурков В.Д., Иванов Г.А. Физико-технологические основы волоконно-оптической техники: Учеб. пособие. М.: ГОУ ВПО МГУЛ, 2007. 222 с.
9. Trufanov A.N., Smetannikov O.Yu., Trufanov N.A. Numerical analysis of residual stresses in preform of stress applying part for PANDA-type polarization maintaining optical fibers // Optical Fiber Technology. 2010. V. 16. N 3. P. 156–161.
10. Сметанников О.Ю., Труфанов Н.А. Технологические и остаточные напряжения в неоднородном стеклующемся цилиндрическом стержне // Механика композиционных материалов и конструкций. 2009. № 2. С. 126–140.
11. Труфанов Н.А., Сметанников О.Ю., Труфанов А.Н. Модели формирования полей технологических и остаточных напряжений в условиях релаксационного перехода (стеклования) // Вестник Нижегородского университета им. Н.И.Лобачевского. 2011. № 4 (5). С. 2534–2536.
12. Пугачев В.С. Теория вероятностей и математическая статистика. М.: ФИЗМАТЛИТ, 2002. 91 с.
13. Чек К., Джиблин П., Ирвинг А. Matlab в математических исследованиях: Пер. с англ. М.: Мир, 2001. 317 с.
14. Мартынов Н.Н., Иванов А.П. Matlab 5.x. Вычисления, визуализация, программирование. М.: КУДИЦ-ОБРАЗ, 2000. 47 с.
15. Вентцель Е.С. Теория вероятностей: Учебник для вузов. 7-е изд. М.: Высш. шк., 2001. 575 с.

<i>Семенов Никита Владимирович</i>	– аспирант, Пермский национальный исследовательский политехнический университет, Пермь, Россия, nikita@studio-gd.ru
<i>Труфанов Николай Александрович</i>	– доктор технических наук, профессор, зав. кафедрой, Пермский национальный исследовательский политехнический университет, Пермь, Россия, nat@pstu.ru
<i>Адамов Анатолий Арсангалеевич</i>	– доктор физ.-мат. наук, ведущий научный сотрудник, Федеральное государственное учреждение науки Институт механики сплошных сред УрО РАН, Пермь, Россия, adamov@icmm.ru
<i>Nikita Semenov</i>	– postgraduate, Perm National Research Polytechnic University, Perm, Russia, nikita@studio-gd.ru
<i>Nikolai Trufanov</i>	– D.Sc., Professor, Department head, Perm National Research Polytechnic University, Perm, Russia, nat@pstu.ru
<i>Anatoly Adamov</i>	– D.Sc., Professor, leading scientific researcher, Institute of Continuous Media Mechanics, Ural Branch of the Russian Academy of Sciences, Perm, Russia, adamov@icmm.ru

УДК 004.05

ВЕРИФИКАЦИЯ ПАРАЛЛЕЛЬНЫХ АВТОМАТНЫХ ПРОГРАММ

М.А. Лукин^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, lukinma@gmail.com

Рассмотрен интерактивный метод верификации параллельных автоматных программ, в которых иерархические автоматы могут реализовываться в разных потоках и взаимодействовать друг с другом. Верификация проводится при помощи инструментального средства Spin, включает в себя автоматическое построение модели на языке Promela, приведение LTL-формулы в формат, определяемый инструментальным средством Spin и построение контрпримера в терминах автоматов. Интерактивная верификация позволяет сократить время верификации и увеличить максимально возможный размер верифицируемых программ. Рассмотренный метод позволяет верифицировать параллельную систему иерархических автоматов, которые взаимодействуют между собой через сообщения и общие переменные. Особенность автоматной модели состоит в том, что каждый автомат объявляется как новый тип данных и может иметь произвольное (но заранее заданное) число экземпляров. Каждый конечный автомат в системе может запускать другой автомат в новом потоке или иметь вложенный автомат. Была проведена апробация инструментального средства Stater, разработанного на основе данного метода. На всех примерах Stater отработал правильно.

Ключевые слова: автоматы, параллельные автоматные программы, верификация, проверка моделей, линейная темпоральная логика, Spin.

VERIFICATION OF PARALLEL AUTOMATA-BASED PROGRAMS

M. Lukin^b

^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, lukinma@gmail.com

The paper deals with an interactive method of automatic verification for parallel automata-based programs. The hierarchical state machines can be implemented in different threads and can interact with each other. Verification is done by means of Spin tool and includes automatic Promela model construction, conversion of LTL-formula to Spin format and counterexamples in terms of automata. Interactive verification gives the possibility to decrease verification time and increase the maximum size of verifiable programs. Considered method supports verification of the parallel system for hierarchical automata that interact with each other through messages and shared variables. The feature of automaton model is that each state machine is considered as a new data type and can have an arbitrary bounded number of instances. Each state machine in the system can run a different state machine in a new thread or have nested state machine. This method was implemented in the developed Stater tool. Stater shows correct operation for all test cases.

Keywords: state machines, parallel automata-based programs, verification, model checking, linear temporal logic, Spin.

Введение

Формальные методы все шире используются для обеспечения качества программного обеспечения. Эти методы не конкурируют с традиционным тестированием, а дополняют его. В данной работе рассматривается верификация методом проверки моделей (model checking) [1–3] для автоматных программ при помощи верификатора Spin [4]. Метод проверки моделей характеризуется высокой степенью автоматизации [1]. По данной теме проводятся исследования в России и за рубежом [5–30]. Большинство из этих работ верифицируют автоматные модели, которые сложно охватить взглядом. Таким образом, теряется одно из главных достоинств автоматных программ – наглядность. Например, блок-схемы и SDL-диаграммы представляют собой, по сути, одномерную структуру и почти всегда выходят за пределы одного экрана, т.е. их нельзя охватить взглядом. Автоматные модели в работах [10–12, 15–19, 25, 26] избавлены от этого недостатка, но они предназначены для однопоточных программ. Метод [28] предназначен для многопоточных программ, но в нем нет возможности задавать свою спецификацию при помощи темпоральной или другой логики. Проводится проверка только заранее заданных свойств. В настоящей работе, которая является продолжением исследований [10, 17, 22], ставится задача построения метода верификации параллельных автоматных программ, обладающего свойством наглядности.

В данной работе предлагается метод интерактивной верификации параллельных автоматных программ. На основе предложенного подхода разработано инструментальное средство Stater, которое позволяет создавать параллельную систему конечных иерархических автоматов, импортировать конечные автоматы из инструментального средства Stateflow, верифицировать созданную систему конечных автоматов при помощи верификатора Spin и генерировать программный код по созданной системе конечных автоматов.

Описание автоматной модели

Предлагаемый подход предназначен для построения распределенных систем взаимодействующих иерархических конечных автоматов [31–33]. При этом каждый такой автомат работает в отдельном потоке. Под иерархическим автоматом в настоящей работе понимается система вложенных автоматов.

В данной работе каждый граф переходов задает не конкретный автомат, а тип автоматов, по аналогии с типом данных или классом в объектно-ориентированном программировании (ООП). Назовем его автоматным типом. У каждого автоматного типа может быть несколько экземпляров (по аналогии с объектом в ООП). Назовем эти объекты автоматными объектами. Каждый автоматный объект имеет уникальное имя. В дальнейшем, если не указано иное, автоматные объекты будут называться просто автоматами.

Переходы автоматов осуществляются по событиям. На переходе могут также быть охранные условия [34]. Если встретилось событие, по которому нет перехода, то автомат может либо завершить работу и перейти в недопускающее состояние, либо игнорировать это событие. Все события – общие для всей системы автоматов. Вводится специальное событие «*», которое означает переход по любому событию, кроме тех, которые указаны на других переходах из этого состояния (аналог *default* в блоках *switch* для C-подобных языков или *else* в условных конструкциях).

Автомат может иметь конечное число переменных целочисленных типов (включая массивы). Для переменных вводятся следующие модификаторы:

- *volatile* – переменная может применяться в любом месте программы;
- *external* – переменная может использоваться другим автоматом;
- *param* – переменная является параметром автомата.

По умолчанию считается, что переменная не используется нигде, кроме как на диаграмме переходов автомата.

Выходные воздействия автомата бывают двух типов:

1. на переходах и в состояниях может быть выполнен любой код, однако верификатор и генератор кода перенесут его без изменений, поэтому код должен быть допустимым в целевом языке;
2. на переходах и в состояниях запускаются функции, определяемые пользователем на целевом языке программирования (после того, как сгенерирован код).

Автомат может иметь вложенные автоматы любого типа, кроме собственного, во избежание бесконечной рекурсии. Циклическая рекурсия также запрещена. Автомат может запускать поток с новым автоматом любого типа. Задается тип автомата `<StateMachine>` и имя `<concreteStateMachine>`. Нельзя запускать несколько автоматов с одним именем. Нельзя запускать автоматы своего типа. Автомат может взаимодействовать с другим автоматом, выступая источником событий для него. События формируются асинхронно. Автомат может использовать переменные другого автомата, отмеченные специальным модификатором. Таким образом, в системе могут существовать несколько автоматов с одинаковым графом переходов, более того, часть этих автоматов могут быть вложенными, а часть не обладать этим свойством.

Описание процесса верификации

Чтобы провести верификацию программы методом проверки моделей, требуется составить модель программы и формализовать требуемые свойства (спецификацию) на языке темпоральной логики [1]. В данной работе используется верификатор Spin, и языком темпоральной логики является LTL [1]. Так как модель строится для автоматной программы, то это может быть выполнено автоматически. Построение модели описано в разделе «Генерация кода на языке Promela».

Обозначим автоматный тип через `AType`, автоматный объект – через `aObject`. Пусть состояния `AType` называются `s0`, `s1` и т. д., в автомат поступают события `e0`, `e1` и т. д., а переменные называются `x0`, `x1` и т. д., внешние воздействия второго типа `z0`, `z1` и т. д. Пусть автоматный тип `AType` имеет вложенный автомат `nested`. Пусть `AType` запускает автомат `fork`.

Процесс верификации состоит из следующих этапов.

1. Построение модели – генерация кода на языке Promela [4]. Для автоматных программ, как отмечено выше, это выполняется автоматически.
2. Преобразование LTL-формулы (переход от нотации автоматной программы в нотацию Spin).
3. Запуск верификатора Spin.
4. Преобразование контрпримера в термины исходной системы автоматов. Это преобразование автоматных программ также выполняется автоматически, аналогично работе [22].

Этапы процесса верификации описаны ниже. Эти этапы похожи на этапы ручной верификации при помощи Spin. Основным отличием является больший уровень автоматизации и большая приближенность модели к реализации, чем при верификации неавтоматных программ. Ниже описана реализация интерактивности, а затем все четыре этапа верификации.

Интерактивность

Одна из главных проблем при верификации методом проверки моделей – это размер модели Крипке. Чтобы уменьшить модель (отсечь лишние подробности), будем строить ее интерактивно. Для этого вводится возможность выбирать, какие уровни абстракции автоматной системы входят в модель, а какие нет. Кроме того, модель структурируется понятным для человека образом, чтобы пользователь мог самостоятельно модифицировать построенную модель. Ниже описаны уровни абстракции по разным аспектам верификации – по переменным, параллелизму и источникам событий.

Переменные. Для переменных введем следующие уровни абстракции.

1. Переменные в модели не учитываются.
2. Переменные в модель включены, но модель абстрагируется от их значения. Недетерминированно выбирается, какое охранный условие будет верно.
3. Модель вычисляет значения переменных. При этом переменные могут быть следующих видов:
 - Локальные. Эти переменные могут быть изменены только самим конечным автоматом. Все изменения таких переменных находятся только в выходных воздействиях автомата;
 - Параметры. Извне изменяются только один раз при запуске автомата. В остальном они подобны локальным переменным;
 - Публичные. Такие переменные могут быть изменены в любом месте программы, в которую входит построенная автоматная система. В модели перед каждым переходом автомата таким переменным недетерминированно присваивается произвольное значение;
 - Совместно используемые. К таким переменным данного автомата имеют доступ другие автоматы, параллельно работающие с данным автоматом.

Параметры и публичные переменные могут быть также одновременно и совместно используемыми.

Параллелизм. Вводятся два уровня – параллелизм поддерживается либо нет. Если параллелизм не поддерживается, то в модель не вводятся взаимодействия параллельных автоматов, остаются только взаимодействия по вложенности.

Источники событий. В качестве источников событий для автоматов в системе могут выступать внешняя среда и другие автоматы. Внешняя среда как источник событий для каждого автомата может работать в одном из трех режимов:

- внешняя среда не взаимодействует с автоматом (события от внешней среды не приходят);
- внешняя среда отправляет только те события, которые автомат может в данный момент обработать;
- внешняя среда отправляет любые события.

Другие автоматы как источники событий можно отключить, если отключить параллелизм.

Генерация кода на языке Promela

Все состояния каждого автоматного типа перенумеровываются, и для них создаются константы. Для каждого автоматного типа состояния нумеруются отдельно. Имя константы состоит из имени автоматного типа и имени состояния, разделенных знаком подчеркивания. Это сделано для того, чтобы состояния разных автоматов с одинаковыми именами не конфликтовали друг с другом.

Пример:

```
#define AType_s0 0
#define AType_s1 1
```

Все события перенумеровываются, и для них создаются константы. Для событий применяется сквозная нумерация.

Пример:

```
#define e0 1
#define e1 1
```

Все внешние воздействия второго типа (вызываемые функции) перенумеровываются, и для них создаются константы аналогично состояниям. Все вызовы вложенных и запуски параллельных автоматов перенумеровываются аналогично состояниям. Каждый тип автоматов записывается в inline-функцию, которая моделирует один шаг автомата. Переходы записываются при помощи охранных команд Дейкстры [34]. Для каждого типа автоматов создается структура. Элементы структуры:

- byte state – номер текущего состояния;
- byte curEvent – номер последнего пришедшего события;
- byte ID – номер автомата;
- byte functionCall – номер последней запущенной функции, если такая существует;
- byte nestedMachine – номер текущего вложенного автомата, если такой существует;
- все переменные автомата.

Для каждого экземпляра автомата создается экземпляр структуры и канал, по которому происходит передача событий. Для каждого экземпляра автомата, кроме вложенных, создается процесс, который извлекает из канала событие и запускает встраиваемую (inline) функцию автомата с этим событием. Для каждого экземпляра автомата, кроме вложенных, создается процесс, который недетерминированно выбирает событие и отправляет его в канал автомата. Для публичных переменных на каждом шаге автомата вызывается специальная функция, которая их недетерминированно изменяет. Для переменных-параметров такая функция вызывается один раз – при запуске автомата.

Если по данному событию нет перехода и в текущем состоянии есть вложенный автомат, то он запускается (запускается встраиваемая функция автомата). Если в текущем состоянии автомат запускает другой автомат, то запускается заранее созданный процесс запускаемого автомата. Если автомат отправляет событие другому автомату, то он записывает его номер в канал этого автомата.

Апробация метода

Для поддержки описанного метода разработано инструментальное средство Stater, которое позволяет построить параллельную систему автоматов, импортировать автоматы из Stateflow, провести верификацию системы автоматов и сгенерировать программный код, эквивалентный этой системе автоматов. Апробация метода проводилась на нескольких программах. Несколько модулей Stater были разработаны при помощи самого инструмента Stater, а именно:

- модуль генерации программного кода;
- модуль преобразования LTL-формулы;
- модуль импорта диаграмм из Stateflow;
- модуль загрузки диаграмм из файла.

Также был разработан прототип программы управления гусеничным шасси и несколько иных программ. Продемонстрируем предложенный подход на примере прототипа программы управления гусеничным шасси для робота. В шасси два двигателя: по одному на левую и правую гусеницы.

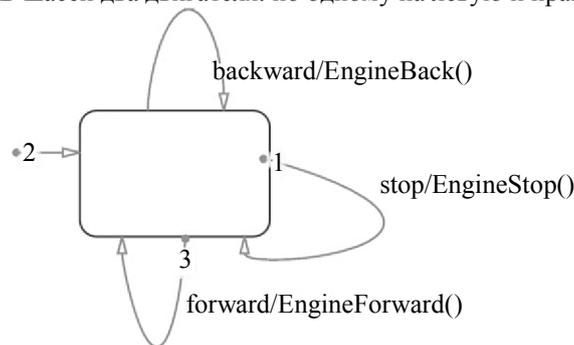


Рис. 1. Граф переходов автоматного типа AEngine

Прототип программы состоит из двух автоматных типов: AEngine и AManager. Два автомата left и right типа AEngine (рис. 1) управляют соответственно левым и правым двигателями.

Автомат типа AManager (рис. 2) отправляет команды на управление двигателями в зависимости от команд для шасси. При входе в состояния он отправляет события автоматам AEngine (слева от стрелки написано имя автомата, справа – событие):

- Stopped: left ← stop, right ← stop.
- MoveForward: left ← forward, right ← forward.
- MoveBackward: left ← backward, right ← backward.
- TurnRight: left ← backward, right ← forward.
- TurnLeft: left ← forward, right ← backward.
- ForwardRight: left ← stop, right ← forward.
- ForwardLeft: left ← forward, right ← stop.
- BackwardRight: left ← backward, right ← stop.
- BackwardLeft: left ← stop, right ← backward.

Проверим свойство: «В любой момент, если поступила команда «стоп», то будет подана команда остановки левого двигателя». Рассматриваемое свойство формализуется следующим образом: в любой момент времени в автомат manager пришло событие stop, следовательно, в будущем автомат left вызовет функцию EngineStop:

$$\mathbf{G} (\{\text{manager.stop}\} \Rightarrow (\mathbf{F} \{\text{left.EngineStop}\}))$$

Данное свойство не должно выполняться в следующих состояниях: StartState, Ready и Stopped. В первых двух шахи еще не готово к работе, а в состоянии Stopped двигатель и так остановлен. В итоге получаем следующую формулу:

$$\lceil ((\{manager.stop\} \ \&\& \ !\{manager.StartState\} \ \&\& \ !\{manager.Ready\} \ \&\& \ !\{manager.Stopped\}) \rightarrow (\langle \{left.EngineStop\} \rangle)) \rceil \quad (1)$$

Выполняем верификацию с формулой (1) и получаем ответ, который означает, что верифицируемое свойство выполняется в построенной системе:

$$0. \lceil ((\{manager.stop\} \ \&\& \ !\{manager.StartState\} \ \&\& \ !\{manager.Ready\} \ \&\& \ !\{manager.Stopped\}) \rightarrow (\langle \{left.EngineStop\} \rangle)) \rceil$$

Verification successful!

Во время апробации Stater отработал правильно на всех задачах.

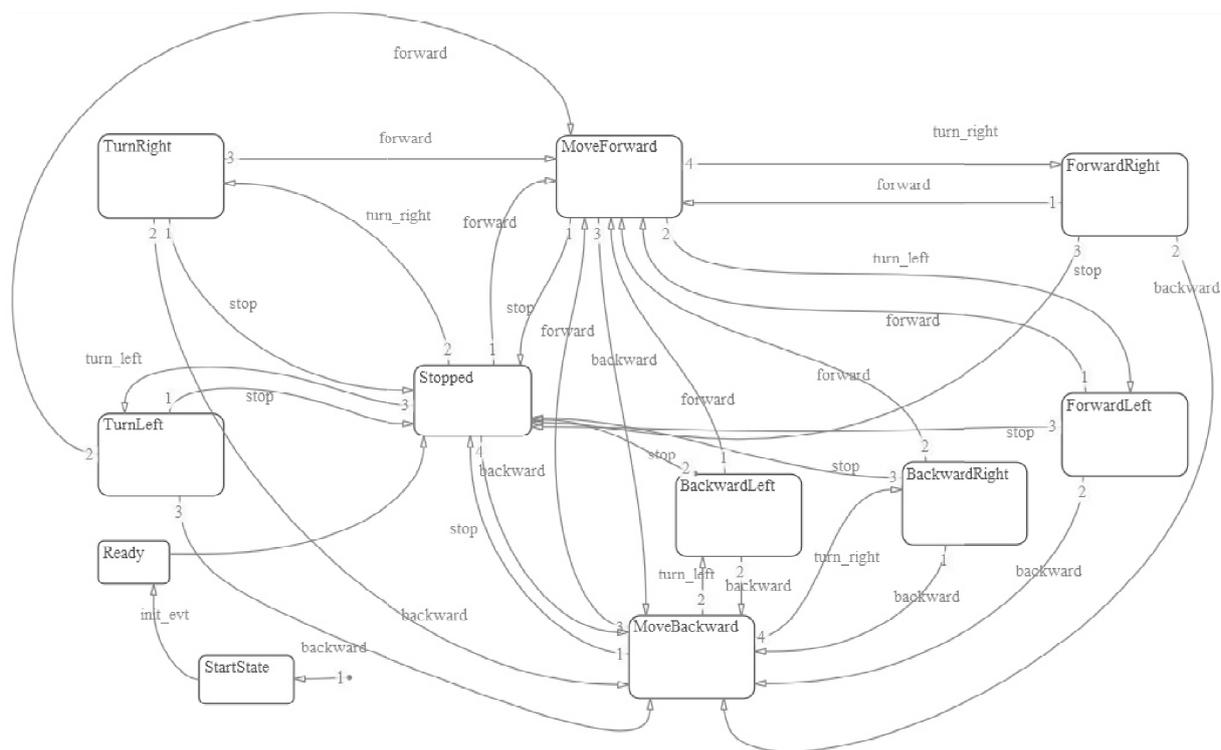


Рис. 2. Граф переходов автоматного типа AManager

Заключение

Таким образом, в работе продемонстрирован метод интерактивной верификации параллельных автоматных программ. Апробация метода показала его работоспособность. Основные результаты работы состоят в разработке метода верификации параллельных автоматных систем, которые отличаются от остальных важнейшим свойством – наглядностью. Применение интерактивности позволило сократить модель Крипке для верифицируемых программ и тем самым увеличить размер программ, которые можно верифицировать. Перспективы работы – более полная верификация программ, разработанных в Stateflow, а также статический анализ кода традиционных программ, основанный на автоматическом построении автоматных моделей по программам и верификации построенных моделей.

Литература

1. Кларк Э.М., Грамберг О., Пелед Д. Верификация моделей программ: Model Checking. М.: МЦНМО, 2002. 416 с.
2. Вельдер С.Э., Лукин М.А., Шалыто А.А., Яминов Б.Р. Верификация автоматных программ. СПб: Наука, 2011. 244 с.
3. Карпов Ю.Г. Model Checking: верификация параллельных и распределенных программных систем. СПб: БХВ-Петербург, 2010. 560 с.
4. Официальный сайт инструментального средства Spin [Электронный ресурс]. Режим доступа: <http://spinroot.com>, свободный. Яз. англ. (дата обращения 03.09.2013).

5. Gnesi S., Mazzanti F. On the fly model checking of communicating UML state machines. 2004 [Электронный ресурс]. Режим доступа: <http://fmt.isti.cnr.it/WEBPAPER/onthefly-SERA04.pdf>, свободный. Яз. англ. (дата обращения 25.11.2013).
6. Gnesi S., Mazzanti F. A model checking verification environment for UML statecharts // Proc. of XLIII Congresso Annuale AICA. 2005 [Электронный ресурс]. Режим доступа: <http://fmt.isti.cnr.it/~gnesi/matdid/aica.pdf>, свободный. Яз. англ. (дата обращения 20.07.2013).
7. Канжелев С.Ю., Шалыто А.А. Автоматическая генерация автоматного кода // Информационно-управляющие системы. 2006. № 6 (25). С. 35–42.
8. Виноградов Р.А., Кузьмин Е.В., Соколов В.А. Верификация автоматных программ средствами CPN/Tools // Моделирование и анализ информационных систем. 2006. Т. 13. № 2. С. 4–15.
9. Васильева К.А., Кузьмин Е.В. Верификация автоматных программ с использованием LTL // Моделирование и анализ информационных систем. 2007. Т.14. № 1. С. 3–14.
10. Лукин М.А. Верификация автоматных программ. Бакалаврская работа. СПбГУ ИТМО, 2007 [Электронный ресурс]. Режим доступа: http://is.ifmo.ru/papers/_lukin_bachelor.pdf, свободный. Яз. рус. (дата обращения 25.11.2013).
11. Яминов Б.Р. Автоматизация верификации автоматных UniMod-моделей на основе инструментального средства Bogog. Бакалаврская работа. СПбГУ ИТМО, 2007 [Электронный ресурс]. Режим доступа: http://is.ifmo.ru/papers/jaminov_bachelor.pdf, свободный. Яз. рус. (дата обращения 25.11.2013).
12. Вельдер С.Э., Шалыто А.А. О верификации простых автоматных систем на основе метода Model Checking // Информационно-управляющие системы. 2007. № 3 (28). С. 27–38.
13. Ma G. Model checking support for CoreASM: model checking distributed abstract state machines using Spin. 2007 [Электронный ресурс]. Режим доступа: <http://summit.sfu.ca/item/8056>, свободный. Яз. англ. (дата обращения 25.11.2013).
14. David A., Möller O., Yi W. Formal Verification of UML Statecharts with Real-time Extensions // Lecture Notes in Computer Science. 2002. V.2306. P. 218–232.
15. Егоров К.В., Шалыто А.А. Методика верификации автоматных программ // Информационно-управляющие системы. 2008. № 5 (36). С. 15–21.
16. Курбацкий Е.А. Верификация программ, построенных на основе автоматного подхода с использованием программного средства SMV // Научно-технический вестник СПбГУ ИТМО. 2008. № 8 (53). С. 137–144.
17. Лукин М.А., Шалыто А.А. Верификация автоматных программ с использованием верификатора SPIN // Научно-технический вестник СПбГУ ИТМО. 2008. № 8 (53). С. 145–162.
18. Гуров В.С., Яминов Б.Р. Верификация автоматных программ при помощи верификатора UNIMOD.VERIFIER // Научно-технический вестник СПбГУ ИТМО. 2008. № 8 (53). С. 162–176.
19. Егоров К.В., Шалыто А.А. Разработка верификатора автоматных программ // Научно-технический вестник СПбГУ ИТМО. 2008. № 8 (53). С. 177–188.
20. Гуров В.С., Мазин М.А., Шалыто А.А. Автоматическое завершение ввода условий в диаграммах состояний // Информационно-управляющие системы. 2008. № 1 (32). С. 24–33.
21. Prashanth C.M., Shet K.C. Efficient Algorithms for Verification of UML Statechart Models // Journal of Software. 2009. V. 4. N 3. P. 175–182.
22. Лукин М.А. Верификация визуальных автоматных программ с использованием инструментального средства SPIN. Магистерская диссертация. СПбГУ ИТМО, 2009 [Электронный ресурс]. Режим доступа: http://is.ifmo.ru/papers/_lukin_master.pdf, свободный. Яз. рус. (дата обращения 25.11.2013).
23. Тимофеев К.И., Астафуров А.А., Шалыто А.А. Наследование автоматных классов с использованием динамических языков программирования (на примере языка RUBY) // Информационно-управляющие системы. 2009. № 4 (41). С. 21–25.
24. Ремизов А.О., Шалыто А.А. Верификация автоматных программ // Сборник докладов научно-технической конференции «Состояние, проблемы и перспективы создания корабельных информационно-управляющих комплексов ОАО «Концерн Моринформсистема-Агат». М., 2010. С. 90–98 [Электронный ресурс]. Режим доступа: http://is.ifmo.ru/works/_2010_05_25_verific.pdf, свободный. Яз. рус. (дата обращения 25.11.2013).
25. Клебанов А.А., Степанов О.Г., Шалыто А.А. Применение шаблонов требований к формальной спецификации и верификации автоматных программ // Семантика, спецификация и верификация программ: теория и приложения: Труды семинара. 2010. С. 124–130 [Электронный ресурс]. Режим доступа: http://is.ifmo.ru/works/_2010-10-01_klebanov.pdf, свободный. Яз. рус. (дата обращения 25.11.2013).
26. Вельдер С.Э., Шалыто А.А. Верификация автоматных моделей методом редуцированного графа переходов // Научно-технический вестник СПбГУ ИТМО. 2009. № 6 (64). С. 66–77.
27. Янкин Ю.Ю., Шалыто А.А. Автоматное программирование ПЛИС в задачах управления электроприводом // Информационно-управляющие системы. 2011. № 1 (50). С. 50–56.

28. Chen C., Sun J., Liu Y., Dong J., Zheng M. Formal modeling and validation of Stateflow diagrams // International Journal on Software Tools for Technology Transfer. 2012. V. 14. N 6. P. 653–671.
29. Малаховски Я.М., Корнеев Г.А. Применение зависимых систем типов со структурной индукцией для верификации реактивных программ // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 6 (82). С. 63–67.
30. Катериненко Р.С., Бессмертный И.А. Верификация данных в системах отслеживания задач с помощью продукционных правил // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 1 (83). С. 86–90.
31. Шалыто А.А. Switch-технология. Алгоритмизация и программирование задач логического управления. СПб: Наука, 1998. 617 с.
32. Cardei I., Jha R., Cardei M., Pavan A. Hierarchical architecture for real-time adaptive resource management // Proc. of the IFIP/ACM International Conference on Distributed Systems Platforms. Secaucus, NJ, USA: Springer-Verlag, 2000. P. 415–434.
33. Поликарпова Н.И., Шалыто А.А. Автоматное программирование. СПб: Питер, 2010. 176 с.
34. Dijkstra E.W. Guarded commands, non-determinacy and formal derivation of programs // Communications of the ACM. 1975. V. 18. N 8. P. 453–457 [Электронный ресурс]. Режим доступа: <http://www.cs.virginia.edu/~weimer/615/reading/DijkstraGC.pdf>, свободный. Яз. англ. (дата обращения 25.11.2013).

Лукин Михаил Андреевич

– программист, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, lukinma@gmail.com

Michael Lukin

– programmer, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, lukinma@gmail.com

УДК 004.65

АНАЛИЗ ДАННЫХ НА ОСНОВЕ ПЛАТФОРМЫ SQL-MAPREDUCE

А.А. Дергачев^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, dam600@mail.ru

Рассмотрены проблемы, связанные с применением реляционных СУБД в области анализа больших объемов данных, в том числе данных, предоставляемых для аналитики посредством веб-сервисов в Интернет.

Возможность их решения может быть представлена веб-ориентированной распределенной системой анализа данных, исполнителем ядром которой является процессор сервисных запросов. Функции такой системы аналогичны функциям реляционных СУБД, только применительно к веб-сервисам. Процессор сервисных запросов необходим для формирования и исполнения плана вызова веб-сервисов анализа данных. Эффективность такой веб-ориентированной системы зависит от эффективности плана вызова веб-сервисов и программной реализации веб-сервисов, основным элементом которых являются средства хранения анализируемых данных – реляционные СУБД. Развитию возможностей реляционных СУБД для анализа больших объемов данных и уделено основное внимание в данной работе, а именно – оценке перспективности реализации веб-сервисов анализа данных на основе платформы SQL/MapReduce. Для достижения поставленной цели в качестве прикладной была выбрана аналитическая задача, характерная для различных социальных сетей и веб-порталов, связанная с анализом данных об их посещаемости различными пользователями. В рамках практической части исследования был реализован алгоритм формирования плана вызова веб-сервисов для решения прикладной аналитической задачи и выполнен эксперимент, подтверждающий эффективность технологии SQL/MapReduce и перспективность применения ее при реализации веб-сервисов анализа данных.

Ключевые слова: анализ данных, веб-сервисы, SQL, MapReduce, СУБД.

DATA ANALYSIS BY SQL-MAPREDUCE PLATFORM

A. Dergachev^b

^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, dam600@mail.ru

The paper deals with the problems related to the usage of relational database management system (RDBMS), mainly in the analysis of large data content, including data analysis based on web services in the Internet. A solution of these problems can be represented as a web-oriented distributed system of the data analysis with the processor of service requests as an executive kernel. The functions of such system are similar to the functions of relational DBMS, only with the usage of web services. The processor of service requests is responsible for planning of data analysis web services calls and their execution. The efficiency of such web-oriented system depends on the efficiency of web services calls plan and their program implementation where the basic element is the facilities of analyzed data storage – relational DBMS. The main attention is given to extension of functionality of relational DBMS for the analysis of large data content, in particular, the perspective estimation of web services data analysis implementation on the basis of SQL/MapReduce platform. With a view of obtaining this result, analytical task was chosen as an application-oriented part, typical for data analysis in various social networks and web portals, based on data analysis of users' attendance. In the practical part of this research the algorithm for planning of web services

calls was implemented for application-oriented task solution. SQL/MapReduce platform efficiency is confirmed by experimental results that show the opportunity of effective application for data analysis web services.

Keywords: data analysis, web services, SQL, MapReduce, DBMS.

Введение

Рынок аналитических инструментов на сегодняшний день широко представлен готовыми программными решениями. Однако программные продукты такого рода обычно сложно поддаются настройке под конкретные требования организации [1] и не имеют технической возможности доступа к данным, предоставляемым в Интернет посредством веб-сервисов. В связи с этим в последнее время возрос интерес к построению систем анализа данных на основе готовых аналитических платформ [2], позволяющих с использованием интегрированных в них средств создавать новые аналитические веб-ориентированные инструменты [3]. Такие аналитические системы представляют собой сложные программно-аппаратные комплексы, основным элементом которых являются средства распределенного хранения анализируемых данных [4–6]. Согласно статистике, в качестве таких средств на сегодняшний день наибольшее распространение получили систем управления базами данных (СУБД), среди которых лидирующие позиции, порядка 80–90%, занимают реляционные СУБД [7, 8]. Исходя из этого, в данной работе вопрос построения аналитической платформы рассматривается с позиций развития концепции реляционных СУБД как перспективной применительно к веб-сервисам [9].

Целью настоящей работы является подтверждение эффективности подхода к расширению парадигмы, заложенной в основе реляционных СУБД, моделью распределенных вычислений MapReduce, и его перспективности при реализации веб-сервисов, входящих в состав веб-ориентированной системы анализа данных.

Расширение языка SQL моделью распределенных вычислений MapReduce

Структурированный язык запросов SQL (Structured Query Language) и параллельные архитектуры СУБД на сегодняшний день уже не удовлетворяют требуемым показателям производительности по анализу данных [10]. Это связано с несколькими причинами.

Во-первых, одним из важнейших свойств языка SQL является свойство декларативности – в запросе указывается, какие данные необходимо извлечь или модифицировать, но не указывается, каким образом этот запрос должен быть обработан. Это значительно упрощает процесс формулировки SQL-запросов, но усложняет задачу оптимизации при формировании плана выполнения запроса, который позволил бы минимизировать время, необходимое для его выполнения. Одним из перспективных вариантов решения проблемы является расширение языка SQL, а именно, его интеграция с моделью распределенных вычислений MapReduce, которая используется для организации параллельной обработки больших объемов данных в компьютерных кластерах. Модель MapReduce была представлена компанией Google в 2004 г. [11]. Данная модель реализована в виде фреймворка, работающего поверх распределенной файловой системы GFS (Google File System), и широко применяется в программных продуктах самой компании Google. Однако, являясь сугубо проприетарной, она недоступна для сторонних разработчиков. Альтернативной свободно доступной реализацией стал проект сообщества Apache Software Foundation под названием Apache Hadoop. Фреймворк для реализации MapReduce-вычислений, который называется Hadoop MapReduce, работает поверх распределенной файловой системы HDFS (Hadoop Distributed File System), предназначенной для хранения файлов большого размера, поблочно распределенных между узлами вычислительного кластера. Все файловые блоки, кроме последнего, имеют одинаковый размер, и при этом каждый блок может быть размещен в нескольких узлах. Благодаря использованию механизмов репликации обеспечивается устойчивость распределенной системы к отказам отдельных узлов. Файлы в HDFS могут записываться лишь однажды, при этом не поддерживаются механизмы их модификации, и запись в файл в одно и то же время может вести только один процесс. Таким образом, реализована модель однократной записи в файл с последующим его многократным чтением, способствующая упрощению механизмов обеспечения целостности данных. Работа всей файловой системы находится под централизованным управлением узла имен, хранящего все метаданные о файлах системы, в том числе информацию об их размерах, размещении блоков и их реплик и т.д. Узел имен отвечает за обработку операций уровня файлов и каталогов, таких как открытие и закрытие файлов, манипуляцию каталогами. Сами блоки данных хранятся в серии узлов данных. Узлы данных отвечают за обработку операций по чтению и записи данных. Широкая известность в области информационных технологий модель MapReduce получила именно благодаря открытости и доступности реализации Hadoop, которая применяется в различных научных и исследовательских проектах, стимулирует тем самым разработчиков данной модели к постоянному ее совершенствованию. Однако важно заметить, что реализация Hadoop MapReduce полностью основана на спецификациях компании Google.

Во-вторых, концепция «мастер–работник», заложенная в архитектуре параллельных СУБД, также накладывает ряд ограничений [12]. Основной недостаток данных систем состоит в том, что процесс-мастер

является «узким местом» всей системы. Это связано с тем, что он осуществляет всю финальную обработку данных в последовательной форме, поскольку не все операции могут обрабатываться параллельно на стороне процессов-работников. Соответственно, чем больше данных возвращается от процессов-работников, тем дольше процесс-мастер будет осуществлять их финальную обработку, что может негативно сказаться на времени выполнения запроса в целом. При этом использование различных форм распараллеливания запросов между процессами-работниками приводит к усложнению задачи их оптимизации.

Все эти факторы указывают на необходимость поиска решений, которые позволили бы обойти ряд сложившихся ограничений по применению данных технологий в области анализа данных. Одним из таких решений является подход, предложенный компанией Aster Data Systems, приобретенной позже компанией Teradata. Он заключается в расширении возможностей языка SQL посредством хорошо распараллеливаемых табличных функций, которые можно вызывать прямо из операторов выборки. Для обеспечения работы данного механизма компанией была разработана технология под названием SQL/MapReduce, реализованная в SQL-ориентированной массивно-параллельной СУБД nCluster, которая поставляется в составе одноименной аналитической платформы SQL/MapReduce [13]. Согласно модели MapReduce, обработка данных осуществляется в две фазы. При выполнении фазы Map из набора входных данных происходит формирование промежуточных пар ключ-значение. Затем все пары с одинаковыми значениями промежуточных ключей передаются в фазу Reduce для их финальной обработки. Обработка данных организуется на основе использования двух типов функций – Map-функций и Reduce-функций. Функции могут быть вызваны из SQL-запросов в любом порядке произвольное количество раз с возможностью вложенных вызовов, в отличие от классической модели MapReduce, где исполнение этих фаз носит последовательный характер. Основным достоинством такого решения является реализация логики по обработке, сортировке и группировке данных внутри SQL/MapReduce-функций, которые исполняются на стороне работников. За счет этого значительно упрощаются сами SQL-запросы и, как следствие, задача их оптимизации. Также происходит уменьшение роли мастера, поскольку в данном случае он отвечает лишь за распределение задач между работниками, сбор от них всех готовых результатов и формирование итогового результата без какой-либо его дополнительной обработки на своей стороне. Все это позволяет значительно сократить время выполнения аналитического запроса, что наглядно демонстрируют результаты эксперимента, представленные ниже. Помимо этого, благодаря заимствованию принципов модели MapReduce, появилась возможность анализа полуструктурированных данных, таких как файлы-журналы и XML-файлы, а также неструктурированных данных, таких как простой текст, для анализа которых язык SQL и реляционные СУБД являются малоприменимыми в силу того, что они основываются на фиксированной схеме хранения данных.

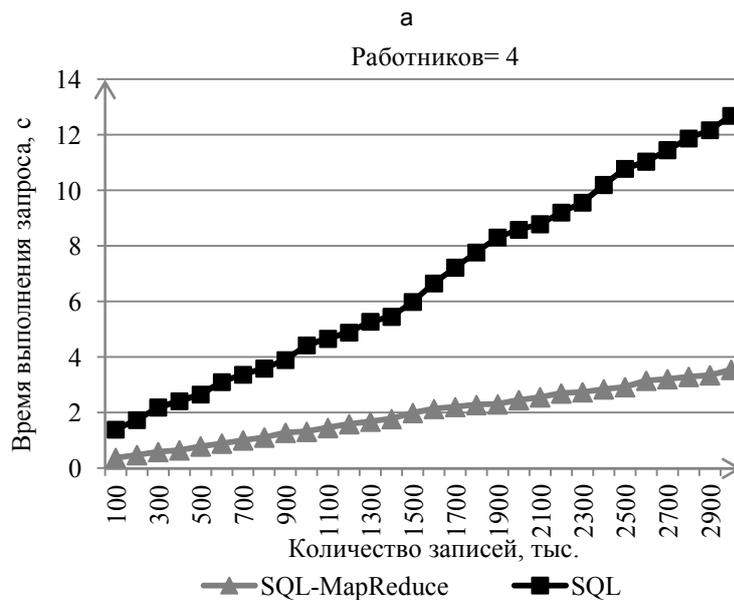
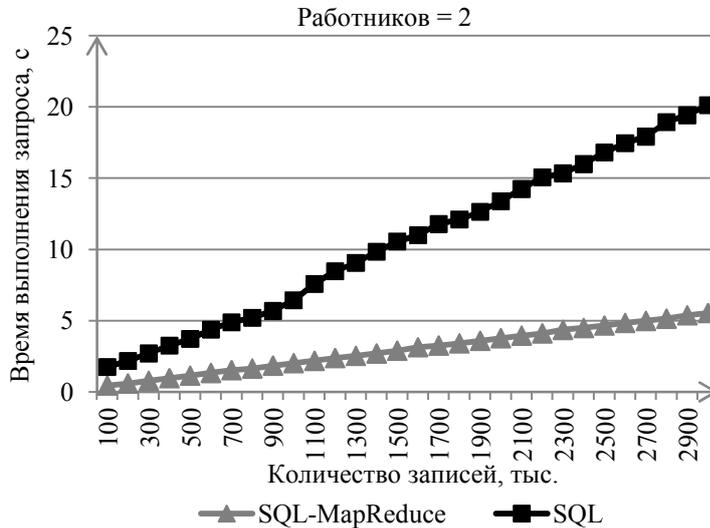
Экспериментальная реализация SQL/MapReduce-функций

Для выполнения экспериментальной части исследования был создан программно-аппаратный комплекс на базе компьютера, оснащенного процессором AMD Phenom II Six-Core 1075T с тактовой частотой 3 ГГц, 16 ГБ оперативной памяти и жестким диском емкостью 1 ТБ, под управлением 64-разрядной операционной системы Windows 7. Программная реализация системы организации доступа к веб-сервисам анализа данных развертывалась на сервере приложений Java EE GlassFish 3.1 и пробной версии СУБД nCluster компании Aster Data Systems. При создании модели были развернуты две виртуальные машины с использованием платформы VMware Workstation 9.0, которые затем были объединены в единую сеть. В качестве операционных систем на виртуальных машинах использовался 64-разрядный дистрибутив SUSE Linux Enterprise 11. Один из образов использовался для моделирования узла-распорядителя (мастера), а другой узла-исполнителя (рабочего). Такая довольно упрощенная конфигурация модели была обусловлена, во-первых, ограничением на количество моделируемых узлов кластера, накладываемым пробной версией реляционной СУБД nCluster, во-вторых, вычислительными ресурсами, располагаемыми в ходе выполнения данного эксперимента. В качестве прикладной была выбрана аналитическая задача, характерная для различных социальных сетей и веб-порталов, связанная с анализом данных об их посещаемости различными пользователями. Для проведения эксперимента был создан файл-журнал, содержащий информацию о времени посещения, идентификаторе пользователя и имени посещаемой страницы. Всего файл содержал информацию приблизительно о тридцати тысячах различных пользователей, которые посещали одиннадцать различных страниц на протяжении трех месяцев. Весь файл состоял из трех миллионов уникальных записей. На основе этих данных осуществлялся подсчет количества посещений пользователями различных страниц по часам для каждого дня каждого месяца. Для решения данной аналитической задачи было разработано два типа SQL-запросов. Один из них выполнял ее стандартными средствами языка SQL:

```
SELECT Год, Месяц, Число, Час_посещения, Страница,  
COUNT(Пользователь) AS Количество_посещений  
FROM (  
SELECT TO_CHAR(DESTAMP, 'YYYY') AS Год,  
TO_CHAR(DESTAMP, 'MM') AS Месяц,
```

```

TO_CHAR(DATESTAMP, 'DD') AS Число,
TO_CHAR(DATESTAMP, 'HH24":00") AS Час_посещения,
PAGE AS Страница,
CUSTOMER_ID Пользователь
FROM test_table) AS sub_select
GROUP BY Год, Месяц, Число, Час_посещения, Страница
ORDER BY Страница, Год, Месяц, Число, Час_посещения
    
```



б

Рисунок. Графики времени выполнения простого SQL-запроса и запроса, содержащего вызов SQL/MapReduce-функции, при использовании двух (а) и четырех (б) работников

Обработка приведенного выше SQL-запроса осуществляется следующим образом. Сначала происходит выполнение вложенного подзапроса. В нем для разбиения входного значения, содержащего временную отметку вида «2013-03-17 16:35:59», используется встроенная функция TO_CHAR, которая на основе указанного формата (YYYY, MM, DD и т.д.) осуществляет выделение в качестве отдельных полей года, месяца, числа и времени в 24-часовом формате. Для большей наглядности запроса в нем были использованы псевдонимы, задаваемые с помощью оператора AS, такие, как Год, Месяц, Число и т.д. В результате подзапрос возвращает таблицу, содержащую все полученные значения в качестве отдельных столбцов, с добавлением к ним столбцов с именем страницы и идентификатором пользователя. Далее эта таблица передается в основной запрос. В нем осуществляется подсчет количества посещений с помощью оператора COUNT() по полю Пользователь, хранящего информацию об идентификаторе пользователя, по строкам, в которых содержатся одинаковые значения в полях Год, Месяц, Число, Час_посещения, Стра-

ница, с их последующей группировкой. В конце производится сортировка результирующих строк в алфавитном порядке по значению поля Страница и в порядке естественного возрастания значений полей Год, Месяц, Число, Час_посещения.

Другой запрос сформулирован с применением SQL/MapReduce-функции, которая была разработана с использованием языка программирования Java:

```
SELECT *
FROM LogAnalyzer(
    ON test_table
    PARTITION BY PAGE)
```

Вызываемая в запросе SQL/MapReduce-функция LogAnalyzer реализована как функция над разделами или же, если говорить в рамках терминологии технологии MapReduce, как Reduce-функция. При выполнении функции над разделами каждая группа строк, образованная на основе спецификации раздела PARTITION BY вызова функции, обрабатывается ровно одним экземпляром данной функции. При этом способе обработки данных экземпляр получает всю группу строк целиком. В данном случае разбиение на разделы осуществляется по имени страницы (PAGE). Таким образом, каждый экземпляр функции занимается обработкой всей информации о посещении пользователями только для одной конкретной страницы. Теперь вся логика по обработке, группировке и сортировке входных данных, а также формат результирующих данных определяются внутри функции LogAnalyzer. После своего исполнения функция возвращает на место ее вызова в основном запросе результирующую таблицу.

При проведении эксперимента осуществлялась оценка времени выполнения простого SQL-запроса и запроса, содержащего вызов функции при изменении количества входных строк данных, что отражено на рисунке. При этом модель была сконфигурирована таким образом, чтобы можно было поставить эксперимент с имитацией работы одного мастера и двух или четырех работников.

На рисунке представлены результаты эксперимента в виде графиков, отражающих время выполнения простого SQL-запроса и запроса, содержащего вызов SQL/MapReduce-функции. Графики строились на основе усредненных значений четырех измерений времени обработки для обоих запросов с точностью до микросекунд, которые выполнялись на каждом шаге эксперимента. Из них хорошо видно, что использование SQL/MapReduce-функции позволяет существенно уменьшить время выполнения запроса по обработке исходных данных в соответствии с поставленной задачей. Так, в данном случае время, которое проходит с момента запуска пользователем запроса на исполнение до того, когда ему будет возвращен результат, удалось уменьшить в среднем в 3,5 раза.

За счет увеличения же количества работников с двух до четырех удалось сократить время выполнения запроса в среднем лишь на 30%. Здесь сказываются накладные расходы, связанные с обработкой вызова функции на стороне мастера, а также координацией и назначением на выполнение ее экземпляров работниками.

Заключение

Полученные в ходе эксперимента положительные результаты подтверждают преимущества изменения логики обработки и принципов построения аналитического запроса, которые были реализованы при расширении концепции реляционной СУБД и языка SQL моделью распределенных вычислений MapReduce. Исходя из этого, можно сделать вывод, что применение технологии SQL/MapReduce для реализации веб-сервисов, входящих в состав веб-ориентированной системы анализа данных, является перспективным.

Следующим шагом развития концепции реляционных СУБД применительно к веб-сервисам может стать разработка перспективной архитектуры веб-ориентированной системы анализа данных, исполнительным ядром которой должен стать процессор сервисных запросов, нацеленный на формирование эффективного плана вызова веб-сервисов анализа данных.

Литература

1. Курочкин Д.Э., Бураков П.В. Задачи развития IT-инфраструктуры предприятия // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 2 (78). С. 74–77.
2. Марьин С.В., Ковальчук С.В. Сервисно-ориентированная платформа исполнения композитных приложений в распределенной среде // Изв. вузов. Приборостроение. 2011. Т. 54. № 10. С. 21–28.
3. Алексеев С.А. Формирование общего информационного ресурса в корпоративной сети социальной организационно-технической системе // Изв. вузов. Приборостроение. 2009. Т. 52. № 12. С. 8–11.
4. Кириллов В.В., Лукьянов Н.М. Анализ факторов, влияющих на качественные и количественные показатели функционирования систем распределенного хранилища данных // Научно-технический вестник СПбГУ ИТМО. 2008. № 11 (56). С. 9–16.

5. Новосельский В.Б., Павловская Т.А. Выбор и обоснование критерия эффективности при проектировании распределенных баз данных // Научно-технический вестник СПбГУ ИТМО. 2009. № 2 (60). С. 76–82.
6. Лукьянов Н.М., Дергачев А.М. Организация сетевого взаимодействия узлов распределенной системы хранения данных // Научно-технический вестник СПбГУ ИТМО. 2011. № 2 (72). С. 137–140.
7. DB-Engines. Ranking the popularity of database management systems. 2012 [Электронный ресурс]. Режим доступа: http://db-engines.com/en/blog_post/1, свободный. Яз. англ. (дата обращения 09.06.2013).
8. Зализняк Е. Рынок СУБД. 2009 [Электронный ресурс]. Режим доступа: http://www.cnews.ru/reviews/index.shtml?2005/08/15/184770_1, свободный. Яз. рус. (дата обращения 09.06.2013).
9. Дергачев А.М. Проблемы эффективного использования сетевых сервисов // Научно-технический вестник СПбГУ ИТМО. 2011. № 1 (71). С. 83–86.
10. Agrawal R., Ailamaki A., Bernstein P.A., Brewer E.A., Carey M.J., Chaudhuri S., Doan A., Florescu D., Franklin M.J., Garcia-Molina H., Gehrke J., Gruenwald L., Haas L.M., Halevy A.Y., Hellerstein J.M., Ioannidis Y.E., Korth H.F., Kossmann D., Madden S., Magoulas R., Ooi B.C., O'Reilly T., Ramakrishnan R., Sarawagi S., Stonebraker M., Szalay A.S., Weikum G. The Claremont Report on Database Research // Sigmod Record. 2008. V. 37. N 3. P. 9–19.
11. Dean J., Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters // Proc. of the Sixth Symposium on Operating System Design and Implementation. San Francisco, CA, 2004. P. 137–150.
12. van der Lans R.F. Using SQL-MapReduce® for Advanced Analytical Queries [Электронный ресурс]. Режим доступа: http://www.asterdata.com/resources/assets/ar_SQL-MapReduce_for_Advanced_Analytics.pdf, свободный. Яз. англ. (дата обращения 06.06.2013).
13. Friedman E., Pawlowski P., Cieslewicz J. SQL/MapReduce: A practical approach to self-describing, polymorphic, and parallelizable userdefined functions // Proc. of the 35th VLDB Conference. Lyon, France, 2009. P. 1402–1413.

Дергачев Александр Андреевич

– аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, dam600@mail.ru

Alexander Dergachev

– postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, dam600@mail.ru

УДК 004.043, 004.5, 37.04

О ВЛИЯНИИ АДАПТИВНЫХ ПОЛЬЗОВАТЕЛЬСКИХ ИНТЕРФЕЙСОВ НА НАДЕЖНОСТЬ И ЭФФЕКТИВНОСТЬ ФУНКЦИОНИРОВАНИЯ АВТОМАТИЗИРОВАННЫХ СИСТЕМ

Ю.О. Фуртат^а

^а Институт проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины, Киев, Украина, saodhar@ukr.net

В современных автоматизированных системах пользователи часто сталкиваются с проблемой информационной перегрузки из-за постоянно возрастающих объемов информации, требующей обработки за короткое время. Работа в таких условиях отрицательно сказывается на качестве работы операторов систем и на надежности самих систем.

Одним из подходов к решению задачи информационной перегрузки является создание для автоматизированных систем персонализированных интерфейсов, учитывающих особенности работы пользователей с информацией. Характеристики оператора системы, определяющие предпочитаемые им форму и темп представления информации, формируют когнитивный портрет пользователя.

Для диагностирования характеристик применяется или профессиональное тестирование с привлечением специалистов-психологов, или оперативное тестирование на рабочем месте пользователя. Вторым вариантом представляется более предпочтительным для использования в автоматизированных системах, поскольку не возникает проблемы нехватки специалистов-психологов. Составление когнитивного портрета при этом проводится в результате взаимодействия пользователя с программными средствами диагностирования, основанными на методиках когнитивной психологии.

Эффект от применения в автоматизированной системе персонализированного пользовательского интерфейса можно оценить, установив, как уменьшение времени реакции пользователя на критические события влияет на уровень надежности и эффективности функционирования системы. Для этого используются формулы теории надежности сложных автоматизированных систем, показывающие зависимость надежности системы от времени реагирования пользователя на критическое событие.

Ключевые слова: автоматизированная система, пользовательский интерфейс, персонализация, адаптация интерфейса, когнитивный профиль.

ON THE EFFECT OF ADAPTIVE USER INTERFACES ON RELIABILITY AND EFFICIENCY OF THE AUTOMATED SYSTEMS

Yu. Furtat^b

^b Ukrainian National Academy of Sciences, Pukhov Institute for Modeling in Energy Engineering, Kiev, Ukraine, saodhar@ukr.net

In modern automated systems users often have to face the information overload problem because of ever increasing volumes of information with short time processing requirements. Working in such conditions affects the system operator's work quality and the systems reliability. One possible approach to solving the information overload problem is to create personalized interfaces that take into account the user's information management particularities. System operator's features, which determine the shape and pace of information representation preferred by him, form the user's cognitive portrait. To determine the values of portrait characteristics professional testing with the assistance of psychologists or operational testing at the user's workplace is performed. The second option is more preferable for use in automated systems, since it has no issue with lack of psychologists. Cognitive portrait is then built as a result of user interaction with the software diagnostic tools that are based on the cognitive psychology methods. The effect of personalized user interface application in an automated system can be estimated by quantifying how the reduction in user's response time to critical events affects the system reliability and efficiency. For this purpose, the formulae of reliability theory for complex automated systems are used, showing the dependence between the system reliability and user's response time to critical event.

Keywords: automated system, user interface, personalization, interface adaptation, cognitive portrait.

Введение

Надежность и эффективность функционирования современных автоматизированных систем в значительной мере зависит от действий пользователя-оператора. В наше время из-за постоянного усложнения автоматизированных систем операторы работают в ситуации информационного перегрузки, которая отрицательно влияет на сосредоточенность на рабочем процессе и время реакции на критические события. Для решения этой проблемы в современных автоматизированных системах предпринимаются попытки повысить эффективность взаимодействия пользователя системы с рабочим интерфейсом. Используются различные подходы – организационный, эргономичный и т.п., однако они не всегда дают удовлетворительный результат, поскольку не учитывают индивидуальных особенностей пользователей, ориентируясь на некоторого «среднего» пользователя системы.

Вследствие этого возникла необходимость создания системы, которая предоставляет возможность персонализации взаимодействия пользователя с системой, реализует механизм диагностирования психофизиологических и когнитивных характеристик пользователя для учета их вместе с особенностями деятельности пользователя в системе при построении адаптивного интерфейса.

Особенности обработки информации операторами автоматизированных систем

К формальным характеристикам сообщения принадлежат содержание, форма, темп и ритм поступления информации (интервалы между отдельными сообщениями, регулярность поступления и т.п.) [1]. На рис. 1 показаны этапы обработки человеком информации от внешней среды (или пользователя от системы).

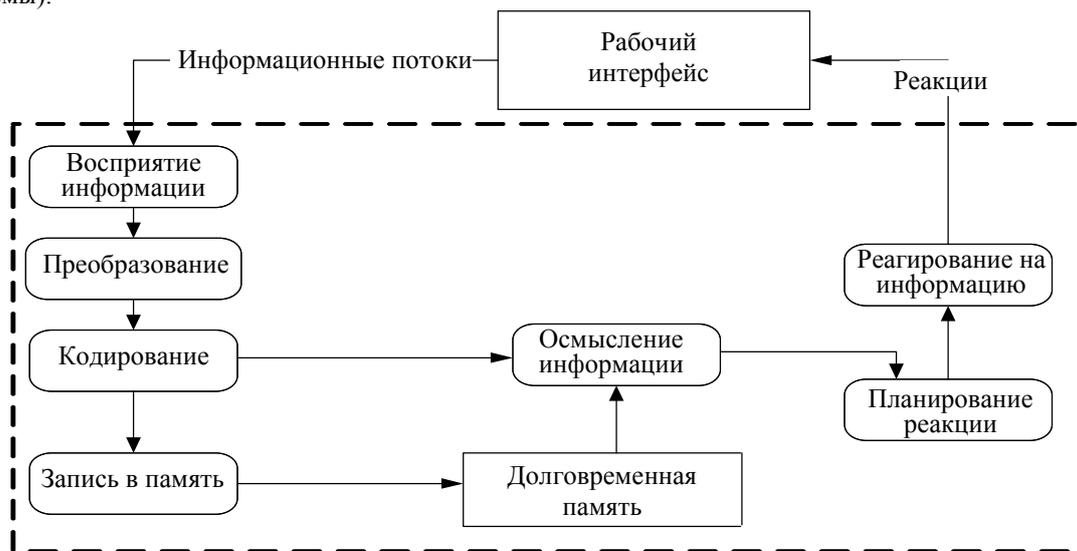


Рис. 1. Процесс восприятия человеком информации от внешней среды

Результаты экспериментальных исследований свидетельствуют о том, что этап репрезентации исходной ситуации занимает особое место [2].

Таким образом, поскольку форма представления информации является существенным фактором, который обеспечивает восприятие, понимание и усвоение информации от системы, ее можно рассматривать как средство управления деятельностью пользователя. Известно, что овладевая материалом во время обучения, человек выборочно относится к его содержанию, виду и форме [3–5].

У каждого пользователя выделяют ряд характеристик, от которых зависит предпочитаемый вид представления информации, темп и насыщенность информационного потока от системы к пользователю. Эти характеристики разделяют на когнитивные, психофизиологические и интеллектуальные.

Когнитивные характеристики, главным образом, определяют скорость, с которой пользователь ориентируется в новой информации и переключается между разными типами деятельности, предпочитаемый им для восприятия тип информации. К ним принадлежат:

- полезависимость–полнезависимость;
- узкий–широкий диапазон эквивалентности;
- импульсивность–рефлексивность;
- интеллектуальная лабильность–ригидность.

Эти характеристики определяют, насколько быстро и точно оператор автоматизированной системы способен считать информацию с большого количества элементов вывода рабочего интерфейса (индикаторов, числовых дисплеев, шкал) и насколько эффективно он способен переключаться между восприятиями различных видов информации (звук, текст, графика). Для каждой характеристики в рамках когнитивной психологии были разработаны соответствующие методики диагностирования, многие из которых формализованы в достаточной мере для реализации в рамках подсистемы автоматизированной диагностики характеристик.

Для оценки психофизиологического состояния пользователя часто используют такие параметры, как его отношение к риску, качество оперативной памяти и внимание. При этом время реакции на тестовое влияние и ошибки в ответах должны позволить оценивать собранность и координацию действий пользователя, поскольку зависят от этих личностных особенностей [6].

К интеллектуальным характеристикам принадлежат уровни IQ и EQ (коэффициенты интеллекта и эмоционального интеллекта соответственно) [7]. Хотя значение коэффициента эмоционального интеллекта связано со значением IQ, прямая зависимость отсутствует, а именно, значение EQ определяет склонность к творческому подходу при работе с информацией. Эта характеристика очень важна для пользователей, например, систем проектирования или учебных систем.

Психофизиологические и интеллектуальные характеристики определяют комфортную интенсивность потока информации. Психофизиологические характеристики при этом являются наиболее динамичными. Например, объем внимания зависит от уровня усталости пользователя и изменяется в течение всего периода работы с системой. Именно эта характеристика влияет на время реакции пользователя на события и поступающие данные. При нормальной работе пользователя на величину объема внимания следует ориентироваться для корректирования скорости вывода информации для обработки пользователем. В случае же критической ситуации эта величина может использоваться в качестве порогового значения – при повышенной усталости оператор должен быть при возможности отстранен от участия в процессе обработки критической ситуации, поскольку замедленная реакция и повышенная вероятность ошибки при принятии решения могут иметь негативные последствия для работы системы.

Интеллектуальные характеристики пользователя могут быть использованы для определения необходимости вывода дополнительной вспомогательной информации (инструкций, справочных материалов) при работе с данными от системы.

В наше время для пользователей автоматизированных систем уже составляют когнитивные портреты на основе профессионального психологического тестирования. Но этот метод имеет два существенных недостатка. Во-первых, он нуждается в участии профессионального психолога, и количество автоматизированных систем и их операторов возрастает быстрее, чем может быть подготовлено соответствующее количество специалистов–психологов. Во-вторых, такое тестирование не является оперативным, оно проходит с отрывом пользователя от рабочего процесса и рабочего места.

Оба этих недостатка могут быть устранены благодаря внедрению в систему адаптации пользовательских интерфейсов подсистемы автоматического тестирования. Схема одного из возможных вариантов реализации такой подсистемы приведена на рис. 2.

Тестирование пользователя проводится как явным образом (в процессе выполнения им тестовых заданий), так и в фоновом режиме, когда на автоматизированном рабочем месте отслеживаются такие показатели, как скорость выполнения действий или реакции на смену ситуации. В обоих случаях информация о характеристиках пользователя обрабатывается отдельным модулем системы тестирования и дополняет или изменяет изначальный когнитивный профиль, который используется при дальнейшей работе системы управления пользовательскими интерфейсами. Именно таким способом – видоизменяя интерфейс пользователя и наблюдая за изменением эффективности работы с ним (которая выражается в по-

вышении скорости реакции и уменьшении количества ошибок при работе с информацией), можно за конечное количество шагов итерационного процесса оптимизировать вид интерфейса под конкретного пользователя (в случае системы с автоматической адаптацией интерфейсов).

Чтобы оценить эффект от применения в автоматизированной системе адаптивного пользовательского интерфейса, необходимо установить, как уменьшение времени реакции пользователя на критические события влияет на уровень надежности и эффективности функционирования системы.



Рис. 2. Система диагностирования характеристик пользователя

Влияние скорости обработки критического события на надежность системы

Условимся считать, что автоматизированная система работает эффективно, если время T_1 , фактически затраченное на обработку критического события, не превышает предельно допустимого времени T_2 . Время T_1 состоит из времени появления критического события и времени собственно реакции на него. Тогда условие эффективности в рассматриваемой системе может быть записано (полагая T_1 и T_2 случайными величинами) в виде вероятности P :

$$P\{T_2 - T_1 \geq 0\} \geq a, \tag{1}$$

где a – вероятность, с которой гарантируется эффективность работы системы. Величины T_1 и T_2 определяются квалификацией и когнитивным и психофизиологическим состоянием обслуживающего персонала. На основании (1) имеем

$$P\{T_1 < T_2\} = \int_0^{\infty} F(t) dG(t) = \int_0^{\infty} F(t) g(t) dt, \tag{2}$$

где функции распределения случайной величины $F(t)$ и $G(t)$ определяются так:

$$F(t) = P\{T_1 < t\}; G(t) = P\{T_2 < t\}; g(t) = G'(t). \tag{3}$$

Формулы (2) и (3) можно использовать при любых законах распределения случайных величин T_1 и T_2 , которые принимают только положительные значения и имеют плотности распределения, равные нулю при отрицательных значениях аргументов. Вместе с тем эти формулы получены исходя из того, что за время эксплуатации или выполнения задачи возникает не более одной критической ситуации.

В реальных условиях за время выполнения задачи продолжительностью t критическая ситуация может возникать неоднократно. В этом случае эффективность функционирования автоматизированной системы на протяжении заданного времени выполнения задачи будет обеспечена при следующих возможных событиях:

- критические ситуации отсутствуют;
- возникла одна критическая ситуация, которая была обработана за допустимое время;
- возникли две критические ситуации, каждая с которых также была обработана за допустимое время.

Обозначим через $\xi_i (i=0, 1, \dots)$ интервал времени между i -м в $(i+1)$ -м событиями, а через $F_i(t) = P\{\xi_i < t\}$ – функцию распределения ξ_i . Предположим, что поток событий образует поток с ограниченным последствием (ξ_i независимы в совокупности). Тогда, согласно определению отказа системы, по формуле полной вероятности можно записать вероятность обеспечения эффективности функционирования на протяжении времени t :

$$P(t) = \overline{F}_1(t) + \sum_{k=1}^{\infty} \int_0^t \int_0^{t-x_1} \dots \int_0^{t-\sum_{s=1}^{k-1} x_s} \overline{F}_k \left(t - \sum_{s=1}^k x_s \right) \prod_{s=1}^k dF_s(x_s) W \left(t - \sum_{l=1}^s x_l \right), \tag{4}$$

где $\bar{F}_i(t) = 1 - F_i(t)$. Здесь $W(t)$ – вероятность того, что событие, которое возникло в некоторый момент времени, не приведет через время t к непоправимой ситуации. Определим эту вероятность. Очевидно,

$$W(t) = P\{T_2 > t\} + \int_0^t P\{T_1 < u\} dP\{T_2 < u\} = \bar{G}(t) + \int_0^\infty F(u) dG(u). \quad (5)$$

Выражение (5) получено в предположении, что каждое событие обрабатывается независимо. Часто предполагают, что поток критических событий системы образует простейший поток с параметром λ . В этом случае

$$F_i(t) = \begin{cases} 1 - e^{-\lambda t} & \text{при } t \geq 0, \\ 0 & \text{при } t < 0. \end{cases} \quad (6)$$

Подставляя (6) в равенство (4), получаем искомую вероятность:

$$P(t) = e^{-\lambda t} + \sum_{k=1}^{\infty} \lambda^k e^{-\lambda t} \int_0^t \int_0^{t-x_1} \dots \int_0^{t-\sum_{s=1}^{k-1} x_s} \prod_{s=1}^k W\left(t - \sum_{l=1}^s x_l\right) dx_s. \quad (7)$$

На практике иногда имеет место ситуация, когда может быть обработано только одно событие, а наличие двух критических событий в системе приводит к аварии. Очевидно, появление некоторого события в момент $\sum_{l=1}^s x_l$ не приведет к аварии: или тогда, когда $T_2 > t - \sum_{l=1}^s x_l$, если $x_{s+1} \geq t - \sum_{l=1}^s x_l$, т.е. больше критических событий не было до момента t и допустимое время T_2 достаточно большое, или тогда, когда следующее событие возникло к моменту t , т.е. $x_{s+1} < t - \sum_{l=1}^s x_l$, но предыдущее событие было до этого момента обработано прежде допустимого времени T_2 . Иначе говоря, в этом случае имеет место ситуация $\{T_1 < \min(x_{s+1}, T_2)\}$.

В первом случае вероятность отсутствия аварии равняется $W\left(t - \sum_{l=1}^s x_l\right)$, во втором случае вероятность ситуации (8) такая:

$$W^*(x_{s+1}) = \int_0^{x_{s+1}} \bar{G}(x) dF(x). \quad (9)$$

Формулы (7) и (4) принимают следующий вид:

$$P(t) = \bar{F}_1(t) + \sum_{k=1}^{\infty} \int_0^t \dots \int_0^{t-\sum_{s=1}^{k-1} x_s} \bar{F}_k\left(t - \sum_{s=1}^k x_s\right) \prod_{s=1}^k dF_s(x_s) \prod_{s=1}^{k-1} W^*(x_s) W\left(t - \sum_{s=1}^k x_s\right),$$

$$P(t) = e^{-\lambda t} + e^{-\lambda t} \sum_{k=1}^{\infty} \lambda^k \int_0^t \dots \int_0^{t-\sum_{s=1}^{k-1} x_s} \prod_{s=1}^{k-1} W^*(x_s) W\left(t - \sum_{s=1}^k x_s\right) dx_1 \dots dx_k.$$

Вероятность (9) является важным компонентом в обоих выражениях. Она зависит от величин T_1 , T_2 и значений моментов $\sum_{l=1}^s x_l$. Значения моментов – величина случайная, величина T_2 определяется организа-

ционными требованиями и особенностями рабочего процесса. Наиболее реально повлиять на показатели эффективности функционирования автоматизированных систем, меняя (снижая) значение величины T_1 . Эта величина зависит от особенностей работы пользователя с информацией, поступающей от системы, и уровня комфортности работы с информацией. Поскольку именно с помощью пользовательского интерфейса происходит взаимодействие пользователя с информацией в системе, то удобство и эффективность интерфейса влияет на важный параметр T_1 . Для оценки этого влияния можно использовать в персонализированных интерфейсах таймер для определения времени, за которое пользователь реагирует на событие в системе или поступление информации. В некритических ситуациях можно отслеживать скорость взаимодействия с элементами интерфейса (кнопками, окнами вывода информации, пунктами меню) или реакции на специальные тестовые информационные импульсы (звуковые, графические). Вместе со статистикой количества ошибок при принятии решений или взаимодействии с элементами интерфейса скорость реакции может использоваться для оценки общей эффективности пользовательского интерфейса.

Заключение

Предложенный подход к построению персонализированных пользовательских интерфейсов автоматизированных систем позволяет учитывать особенности когнитивного портрета отдельных пользователей и адаптировать взаимодействие с информацией в системе к этим особенностям. Это может повысить уровень комфорта работы пользователя системы с информацией, снизить уровень информационной перегрузки, улучшить сосредоточенность и время реакции на критические события.

Поскольку величина времени реакции пользователя влияет на общий уровень надежности и эффективности функционирования системы, можно рассчитывать, что применение адаптивных персонализированных пользовательских интерфейсов позволит повысить показатели надежности и эффективности автоматизированных систем.

Литература

1. Турзин П.С., Пономаренко В.А., Рысакова-Ромашкан С.Л. Уровни понимания информации и структура коммуникативного акта // Психологический журнал. 1992. Т. 13. № 1. С. 30–39.
2. Солсо Р. Когнитивная психология. 6-е изд. СПб: Питер, 2006. 589 с.
3. Якиманская И.С. Принципы построения образовательных программ и личностное развитие учащихся // Вопросы психологии. 1999. Т. 13. № 3. С. 39–47.
4. Гальскова Н.Д. Современная методика обучения иностранным языкам: Пособие для учителя. М.: АРКТИ, 2004. 192 с.
5. Ляховицкий М.В. Методика преподавания иностранных языков: Методическое пособие. М.: Дрофа, 2006. 369 с.
6. Верлань А.Ф., Сопель М.Ф., Фуртат Ю.О. Особенности оперативного тестирования на рабочем месте операторов систем поддержки принятия решений (СППР) // Математичне та комп'ютерне моделювання. Сер. Технічні науки. 2010. В. 3. С. 37–45.
7. Немов Р.С. Психология: Учебник для студентов высш. пед. учеб. заведений: В 3-х кн. 4-е изд. М.: Гуманитарный издательский центр «ВЛАДОС», 2003. Кн. 1. 688 с.

Фуртат Юрий Олегович

– младший научный сотрудник, Институт проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины, Киев, Украина, saodhar@ukr.net

Yurii Furtat

– junior scientific researcher, Ukrainian National Academy of Sciences, Pukhov Institute for Modeling in Energy Engineering, Kiev, Ukraine, saodhar@ukr.net

УДК 004.274

МЕТОД ОТОБРАЖЕНИЯ ЗАДАЧ НА КРУПНОГРАНУЛЯРНЫЕ РЕКОНФИГУРИРУЕМЫЕ ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

А.С. Румянцев^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, darkstreamray@gmail.com

Произведен анализ существующих подходов к отображению задач на реконфигурируемые вычислительные системы, особое внимание уделялось методам отображения на крупногранулярные реконфигурируемые вычислительные системы. На основе произведенного анализа сформированы цель и задачи создания нового эвристического метода отображения задач на крупногранулярные реконфигурируемые вычислительные системы, который базируется на методе разделения графа с выталкиванием вершин, алгоритме покрытия графов, эвристическом подходе к оптимизации и упаковке графа для конкретного варианта крупногранулярной реконфигурируемой вычислительной системы и разработанном методе отображения графа потока данных задачи на ресурсы крупногранулярной реконфигурируемой вычислительной системы. В ходе работы было осуществлено имитационное моделирование разработанного метода и существующих подходов к отображению задач на крупногранулярные реконфигурируемые вычислительные системы на модели системы с крупногранулярной реконфигурируемым аппаратным ускорителем MATRIX. Приведены экспериментальные результаты, доказывающие эффективность предлагаемого подхода по сравнению с широко используемыми методами отображения задач на крупногранулярные реконфигурируемые вычислительные системы и возможность использования динамических параметров функционирования крупногранулярной реконфигурируемой вычислительной системы для дальнейшего улучшения получаемого отображения задачи.

Ключевые слова: крупногранулярные реконфигурируемые вычислительные системы, отображение задач на вычислительные системы, покрытие графов.

TASKS MAPPING METHOD FOR COARSE GRAIN RECONFIGURABLE SYSTEMS

A. Romyantsev^b

^b Saint Petersburg State University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, darkstreamray@gmail.com

This paper deals with analysis of existing approaches to tasks mapping on reconfigurable computing systems with special attention paid to mapping methods for coarse grained reconfigurable computing systems. The purpose and objectives of a new heuristic method for tasks mapping on coarse grained reconfigurable computing systems are produced on the base of the carried out analysis. This novel method for tasks mapping on coarse grained reconfigurable computing systems is based on the method of graph partitioning with pushing vertices, graph covering algorithm, a heuristic approach to optimizing and packaging for a particular graph on coarse grained reconfigurable computing system and displaying methods of data flow graph on resources of coarse grained reconfigurable computing system. The simulation was conducted on system model with coarse grained reconfigurable hardware accelerator MATRIX. Experimental results are given. They prove the effectiveness of the proposed approach as compared with the widely used methods for tasks mapping on coarse grained reconfigurable computing systems and the ability to use dynamic functional parameters of coarse grained reconfigurable computing system to further improvement of the mapping results.

Keywords: coarse grain reconfigurable computing systems, tasks mapping on computing systems, graph covering.

Введение

Для того чтобы крупногранулярные реконфигурируемые вычислительные системы (КРВС) [1] заняли свою нишу между цифровыми сигнальными процессорами и программируемыми логическими интегральными схемами (ПЛИС), они должны быть просты в использовании. Целевое приложение должно быть легко компилируемым и загружаемым во встроенную систему, в которой используется КРВС. Отображение задачи на КРВС подразумевает выделение в задачи вычислительно-интенсивных последовательностей операций, которые могут быть отображены на КРВС, назначение временных интервалов исполнения операций задачи, отображение данных операций на конкретную архитектуру вычислительных элементов КРВС и маршрутизацию входных и выходных потоков данных и управления вычислительных элементов КРВС [1]. В процессе отображения задачи на КРВС необходимо учитывать следующие проектные ограничения из-за фиксированного набора ресурсов КРВС: организацию вычислительных элементов, способ доступа к памяти, способ организации межсоединения, порты ввода-вывода, синхронизацию и механизм реконфигурации [2]. Стоит отметить, что в своей общей трактовке задача отображения приложения на КРВС входит в класс комбинаторных задач с неполиномиальной оценкой числа итераций (является NP-сложной), так же как и проблема планирования задач на многопроцессорных системах [3].

В настоящей работе предлагается новый эвристический метод отображения задачи на встроенную систему с КРВС, который поддерживает различные типы КРВС и их конфигурации, обеспечивая наименьшее время поиска решения из всех рассмотренных эвристических методов отображения. В первом разделе представлен обзор существующих методов отображения задачи на КРВС. Во втором разделе приведено описание разработанного эвристического метода отображения. В третьем разделе приведено сравнение предлагаемого метода отображения с существующими решениями.

Обзор существующих решений

Подход в [4] применяет отображение на основе модульного планирования для максимального использования параллелизма в циклах. Модульное планирование позволяет конвейеризировать выполнение циклов [5]. Данный подход создает расписание для одной итерации заданного цикла таким образом, чтобы это же расписание можно было выполнять через определенные промежутки времени с учетом всех внутренних и внешних зависимостей и ограничений по ресурсам. В работе [6] представлен метод отображения задачи на основе модульного планирования с приоритизацией по выходным потокам данных. Метод призван ускорить процесс отображения задачи по сравнению с подходом на основе модульного планирования. Таким образом, вместо того чтобы до маршрутизации по потокам данных произвести отображение операций на вычислительные элементы КРВС, данный метод откладывает задачу отображения операций на вычислительные элементы до тех пор, пока не будет собрана необходимая информация о маршрутизации входных и выходных поток данных.

Подход в работе [7] использует высокоуровневый синтез для отображения задач на КРВС FloRA. Подход использует параллелизм в циклах для их конвейеризации. Для решения задачи отображения сначала применяется алгоритм планирования на основе списка для поиска начального решения. В дальнейшем для улучшения решения применяется эволюционный алгоритм QEA (Quantum-inspired Evolutionary Algorithm) [8].

В разработанном методе отображения используется подход, базирующийся на подходах к эффективному покрытию графа с помощью заданных или генерируемых шаблонов [9, 10]. В работе [11] данная методика используется на стадии покрытия графа потока данных задачи шаблонами поддерживаемых

операций. В дальнейшем происходит оптимальное разделение графа задачи на отдельные кластеры для непосредственного отображения на КРВС путем решения задач целочисленного линейного программирования. С целью избавления от необходимости решения многочисленных ресурсоемких задач линейного программирования в [12] был представлен эвристический метод решения задачи кластеризации, SPKM (Split-Push Kernel Mapping). Данный метод базируется на основе подхода к построению графа произвольной сложности с использованием операций разделения дуг графа и образования новых вершин [13].

Во всех рассмотренных подходах к отображению задач на КРВС отсутствует потенциальная возможность применения метода к другому типу КРВС одновременно с независимостью метода от структурной организации КРВС. Кроме того особенность организации быстрых эвристических методов отображения задач на КРВС делает невозможным использование динамических параметров функционирования системы для дальнейшей оптимизации отображения.

Эвристический метод отображения программы на КРВС

Основная цель эвристического метода заключается в том, чтобы отобразить программу цифровой обработки сигналов или мультимедийных данных, написанных на языке высокого уровня, например, языке Си, на встроенную систему с процессором общего назначения и спроектированным КРВС, что значительно ускорит выполнение данной программы и уменьшит энергопотребление системы [14]. Разработанный эвристический метод отображения с использованием алгоритма «разделения и выталкивания» вершин графа (OPB) во многом основан на подходах из области составления оптимального расписания для многопроцессорных систем [15, 16] и эвристическом методе решения задачи кластеризации SPKM [12].

Предлагаемый метод состоит из четырех этапов.

1. Преобразование исходного кода в граф потока данных с тегированием управляющими командами с выделением вычислительно-интенсивных ядер и выполнении оптимизаций над полученным графом. На данной стадии используется платформа для исследования методов компиляции для высокопроизводительных машин SUIF [17]. После использования инструментария SUIF получается в общем случае неациклический граф потока данных и управления. В дальнейшем используются подходы к оптимизации на основе «гиперблоков», представленных в работе [18], а также выделение подходящих «гиперблоков» вычислительно интенсивных ядер с использованием набора заранее заданных критериев для последующего отображения их на КРВС. Участки исходного кода, которые оказались нецелесообразно исполнять на КРВС или которые не удалось отобразить в результате последующих этапов метода, будут исполняться на процессоре общего назначения.
2. Кластеризация полученного графа на полносвязанный граф неограниченного количества вычислительных элементов КРВС с использованием алгоритма покрытия графов [9]. Задача данного этапа состоит в том, чтобы отобразить полученные на предыдущем этапе графы на граф, состоящий из неограниченного количества вычислительных элементов КРВС, при этом все соседние вычислительные элементы в этом графе связаны друг с другом. Во время выполнения операции кластеризации осуществляется попытка уменьшить количество используемых вычислительных элементов КРВС, их различных конфигураций и критический путь исходного графа. В результате выполнения второй стадии получается граф отображения задач на вычислительные элементы КРВС.
3. Планирование отображения с сжатием полученного графа потока данных позволяет учесть количество вычислительных элементов КРВС и связи между ними. Кроме того, на данном этапе могут быть использованы динамические параметры функционирования системы для получения оптимального сжатия графа конкретной задачи. Предлагаемый метод планирования состоит из двух стадий:
 1. использование метода разделения графа с выталкиванием вершин [13] и нахождение минимального сечения графа на основе применения метода оптимального планирования задач на многопроцессорных системах CASS-II [19];
 2. применение разработанных эвристических подходов к оптимизации и упаковке графа для конкретной конфигурации КРВС.

Предлагаемые эвристические подходы к оптимизации и упаковке графа:

- осуществляется попытка изменения конфигурации транзитных вершин с использованием маршрутизации на основе дерева «Штейнера» [20];
- осуществляется избавление от транзитных вершин, которые могут быть убраны в определенных случаях;
- задействуются возможности контроллера вычислительного элемента второго уровня для последовательного выполнения в одной вершине нескольких операций, что позволяет уменьшить критический путь в отображенном графе задаче и уменьшить количество транзитных вершин;
- производится процедура «сжатия» графа по строкам и столбцам за счет коммутации входного и выходного банка памяти и использования контроллера вычислительного элемента второго уровня.

4. Финальное отображение на ресурсы КРВС исходных алгоритмов с генерацией конфигураций для каждого вычислительного элемента КРВС. Задачи этого этапа.

1. Уместить в имеющиеся ресурсы КРВС и в ресурсы ее отдельных вычислительных элементов подготовленные графы задач. Для достижения заданных целей используется разработанный алгоритм отображения распланированного графа на ресурсы КРВС, псевдокод которого приведен ниже:

/* На вход подается распланированный и сжатый на предыдущем шаге граф G

На выходе формируется конфигурация для каждого вычислительного элемента КРВС для каждого такта */

```

0 function AllocateCGRAResources(G) {
1   foreach row in G {
2     Allocate PEs of the current clock cycle;
3   do {
4     try to move each output to memory or internal
5     PE's register or stack memory
6   } if failed GOTO label_recover
7   do {
8     try to move each input to proper PE's outputs
9     at the clock cycle, which is N steps before, where
10    N in (1,2,3,4,...,starting_step)
11  } if failed GOTO label_recover
12  GOTO label_next_row
13  label_recover: insert one or more clock cycles
14  before the current one
15  label_next_row:
16  }
17  }

```

2. Формирование конфигурационных карт для каждого цикла исполнения отдельных вычислительных элементов КРВС. Множество таких конфигурационных карт представляет собой конфигурационную программу исполнения исходной задачи на КРВС.

Экспериментальные результаты

Для проведения экспериментальных измерений были выбраны различные методы эвристического отображения задач на КРВС: DRESC [4], AHN [11], SPKM [12], QEA [8], а также разработанный метод OPB. Все эксперименты проводились на упрощенной модели встроенной системы, реализованной на языке проектирования системного уровня SystemC, с вариантом КРВС MATRIX [21], которая поддерживает конфигурируемое распределение вычислительных ресурсов. КРВС MATRIX состоит из массива одинаковых 8-битных базовых вычислительных блоков 4×4, объединенных каналами передачи данных и конфигурационной сетью. Каждый базовый вычислительный блок содержит блок памяти на 256 Б, 8-битовое арифметико-логическое устройство и блок снижения логики управления. На рис. 1 показана доля успешных отображений для выборки из случайных задач в зависимости от количества вершин в отображаемом графе потока данных для различных методов отображения. Из графика видно, что разработанный метод OPB заметно опережает в эффективности методы AHN, DRESC и SPKM, находясь наравне с методом QEA.

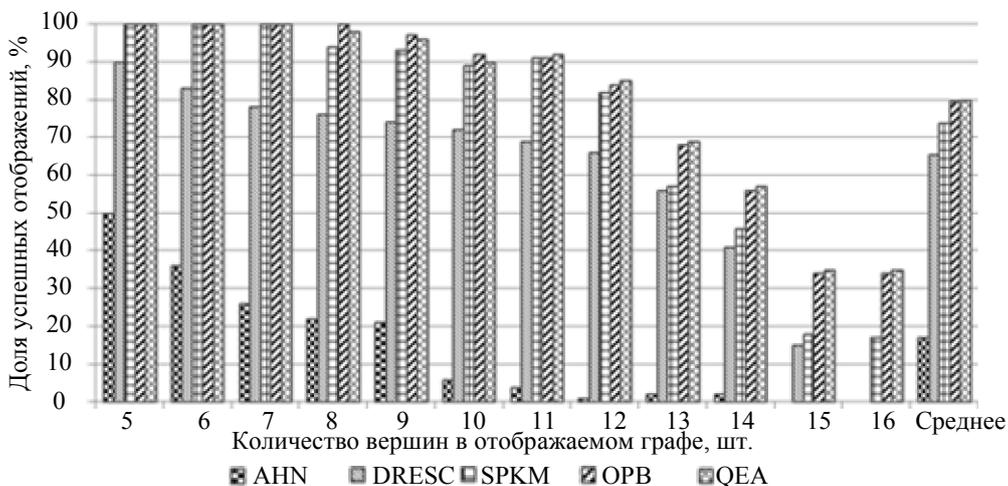


Рис. 1. Зависимость количества успешных отображений от количества вершин в графе

На рис. 2 показана зависимость времени выполнения метода от количества вершин в отображаемом графе потока данных. Разработанный метод ОРВ в среднем оказывается в 1,5 раза быстрее остальных эвристических методов. По времени выполнению отображенной задачи на КРВС из рис. 3 видно, что разработанный метод ОРВ показывает себя не хуже, а в 6 из 10 проведенных тестов даже лучше остальных методов.

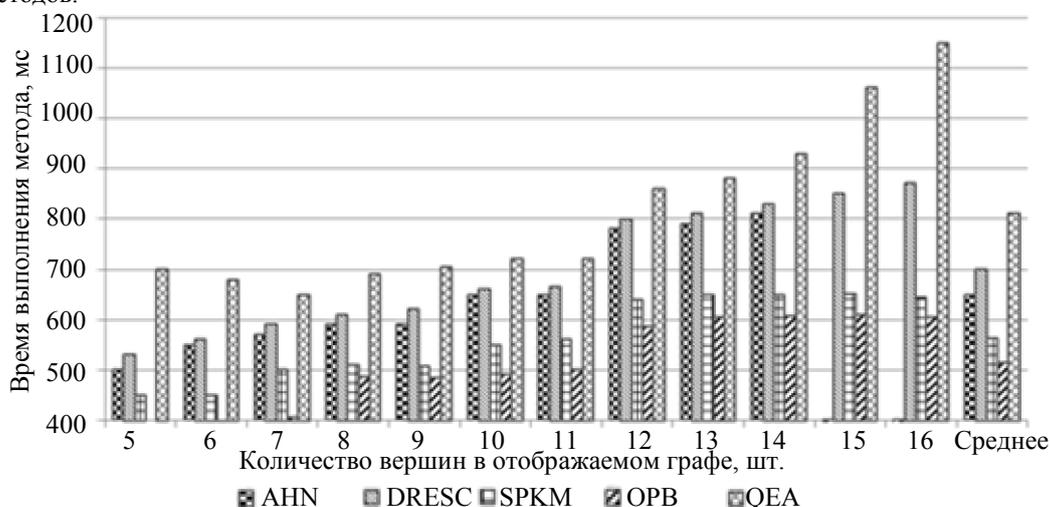


Рис. 2. Зависимость времени выполнения метода от количества вершин в графе

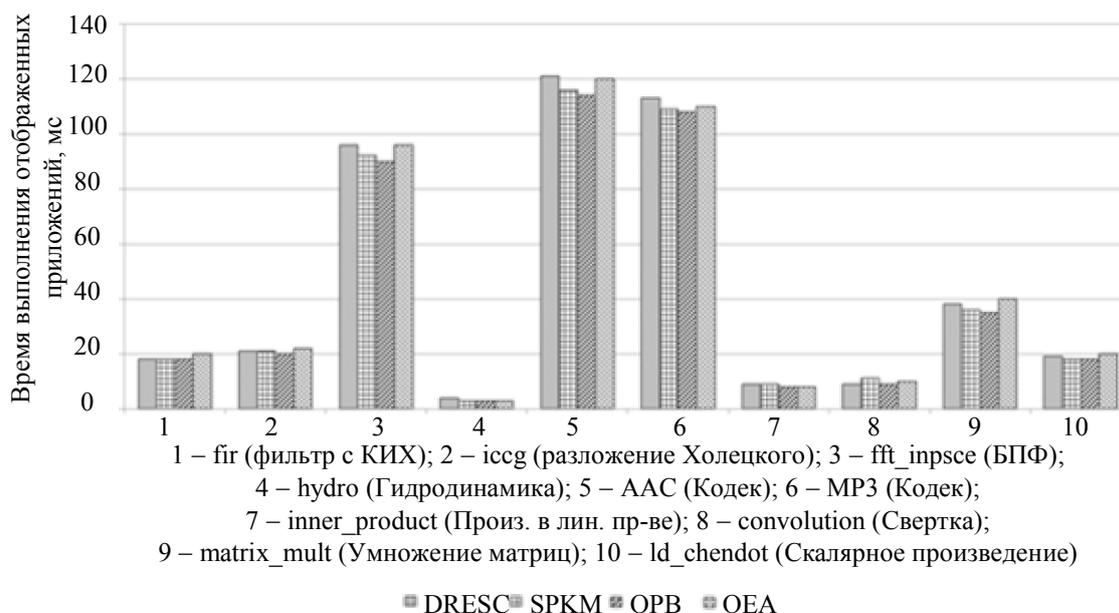


Рис. 3. Относительное время выполнения задач [22], отображенных на систему различными методами

Заключение

В рамках работы предложен новый эвристический метод отображения задач на крупногранулярные вычислительные системы ОРВ, который выполняется на 20–40% быстрее существующих эвристических методов отображения и обеспечивает оптимальное отображение задач более чем в 15% случаев. Кроме того, по эффективности отображенной задачи представленный метод оказывается не хуже существующих методов и позволяет использовать динамические параметры функционирования системы для дальнейшего улучшения получаемого отображения задачи.

Литература

1. Румянцев А.С. Организация и инструментальные средства реконфигурируемых вычислительных систем // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 4 (80). С. 80–86.
2. Hannig F., Dutta H., Teich J. Mapping of Regular Nested Loop Programs to Coarse-Grained Reconfigurable Arrays - Constraints and Methodology // Proc. 18th International Parallel and Distributed Processing Symposium (IPDPS'04). IEEE Computer Society, 2004. Santa Fe, USA (2004). Workshop 3. 8 p.

3. El-Rewini H., Lewis T., Ali H. Task Scheduling in Parallel and Distributed Systems. Englewood Cliffs, NJ: Prentice Hall, 1994. 290 p.
4. Mei B., Vernalde S., Verkest D., De Man H., Lauwereins R. DRESC: a retargetable compiler for coarse-grained reconfigurable architectures // Proc. of the IEEE International Conference on Field-Programmable Technology. Hong Kong, 2002. P. 166–173.
5. Rau B. R. Iterative Modulo Scheduling // Proc. of the 27th Annual International Symposium on Microarchitecture. San Jose, California, 1994. P. 63–74.
6. Park H., Fan K., Mahlke S.A., Oh T., Kim H. Edge-Centric Modulo Scheduling for Coarse-Grained Reconfigurable Architectures // Proc. of the 17th International Conference on Parallel Architecture and Compilation Techniques (PACT-2008). Toronto, Ontario, Canada, 2008. P. 166–176.
7. Lee G., Lee S., Choi K., Dutt N. Routing-Aware Application Mapping Considering Steiner Point for Coarse-Grained Reconfigurable Architectures // Proc. of the 6th International Conference on Reconfigurable Computing: Architectures, Tools and Applications (ARC-2010). Berlin-Heidelberg: Springer-Verlag, 2010. P. 231–243.
8. Han K.-H., Kim J.-H. Quantum-Inspired Evolutionary Algorithm for a Class of Combinatorial Optimization // IEEE Trans. Evolutionary Computation. 2002. V. 6. N 6. P. 580–593.
9. Guo Y., Smit G.J.M., Broersma H., Heysters P.M. A Graph Covering Algorithm for a Coarse Grain Reconfigurable System // ACM Sigplan Conference on Languages, Compilers, and Tools for Embedded Systems (LCTES'03). California, USA, 2003. P. 199–208.
10. Guo Y., Hoede C., Smit G.J.M. A Pattern Selection Algorithm for Multi-pattern Scheduling // Proc. of the 20th International Conference on Parallel and distributed processing (IPDPS' 06). IEEE Computer Society, 2006. P. 198–205.
11. Ahn M., Yoon J.W., Paek Y., Kim Y., Kiemb M., Choi K. A Spatial Mapping Algorithm for Heterogeneous Coarse-grained Reconfigurable Architectures // Proc. of the Conference on Design, Automation and Test in Europe (DATE'06). Leuven, Belgium, 2006. P. 363–368.
12. Yoon J.W., Shrivastava A., Park S., Ahn M., Paek Y. A Graph Drawing Based Spatial Mapping Algorithm for Coarse-Grained Reconfigurable Architectures // IEEE Transactions on Very Large Scale Integration Systems. 2009. V. 17. N 11. P. 1565–1578.
13. Battista G.D., Patrignani M., Vargiu F. A Split-and-Push Approach to 3D Orthogonal Drawing // 6th International Symposium on Graph Drawing. Lecture Notes in Computer Science. 1998. V.1547. Springer-Verlag, 1998. P. 87–101.
14. Петров С.В., Юрков К.В., Овсянников Е.П. Сравнительный анализ сложности реализации быстрых цифровых преобразований на RISC-процессорах // Изв. вузов. Приборостроение. 2011. Т. 54. № 9.С. 34–38.
15. Sarkar V. Partitioning and Scheduling Parallel Programs for Execution on Multiprocessors. Cambridge, MA: MIT Press, 1989. 213 p.
16. Liou J.-C., Palis M.A., Wei D.S.L. Performance analysis of task clustering heuristics for scheduling static DAGs on multiprocessor system // Parallel Algorithms and Applications. 1997. V. 12. N 1–3. P. 185–203.
17. Callahan T.J., Hauser J.R., Wawrzynek J. The Garp Architecture and C compiler // IEEE Computer. 2000. V. 33. N 4. P. 62–69.
18. Liou J.-C., Palis M.A. A New Heuristic for Scheduling Parallel Programs on Multiprocessor // Proc. of the International Conference on Parallel Architectures and Compilation Techniques (PACT'98). Paris, 1998. P. 358–365.
19. Patrignani M., Pizzonia M. The Complexity of the Matching-Cut Problem // Proc. of the 27th International Workshop on Graph-Theoretic Concepts in Computer Science (WG'01). Lecture Notes in Computer Science. 2001. V. 2204. P. 284–295.
20. Cohoon J.P., Richards D.S., Salowe J.S. A Linear-Time Steiner Tree Routing Algorithm for Terminals on the Boundary of a Rectangle // IEEE International Conference on Computer-Aided Design (ICCAD-88). Digest of Technical Papers. 1988. P. 402–405.
21. Mirsky E., DeHon A. MATRIX: A Reconfigurable Computing Architecture with Configurable Instruction Distribution and Deployable Resources // IEEE Symposium on FPGAs for Custom Computing Machines (FCCM'96). Napa Valley, California, 1996. P. 157–166.
22. DSP Compiler and Processor Evaluation – DSPstone [Электронный ресурс]. Режим доступа: <http://www.ice.rwth-aachen.de/research/tools-projects/entry/detail/dspstone/>, свободный. Яз. англ. (дата обращения 10.05.2013).

Румянцев Александр Сергеевич – аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, darkstreamray@gmail.com

Alexander Rumyantsev – postgraduate, Saint Petersburg State University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, darkstreamray@gmail.com

УДК 550.388. 2

ОЦЕНКА ПРИМЕНИМОСТИ МОДЕЛИ IRI-2012 ДЛЯ АВТОМАТИЗИРОВАННОЙ ОБРАБОТКИ ИОНОГРАММ ВЕРТИКАЛЬНОГО ЗОНДИРОВАНИЯ

В.Д. Николаева^{a, b}, М.В. Рыбаков^c, А.Л. Котиков^{b, c}, В.К. Кошелевский^c

^a Арктический и антарктический научно-исследовательский институт, Санкт-Петербург, Россия, vera_nik@list.ru

^b Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, vera_nik@list.ru

^c Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова РАН, Санкт-Петербург, Россия, mvr_spb@rambler.ru

Исследуется возможность применения глобальной эмпирической модели IRI-2012 для полуавтоматической обработки данных вертикального зондирования ионосферы. Проведено сопоставление данных вертикального зондирования на станции СПБФ ИЗМИРАН в Воейково за февраль 2013 г. с результатами расчетов по модели IRI-2012 для основных ионосферных характеристик. В обработку вошло 2688 модельных значений параметров f_{0F_2} , f_{0E} , h_{mF_2} , h_{mE} и 1866 реальных значений. По ионограммам определялись критические частоты (f_{0E} , f_{0F_2}) и высоты максимумов (h_{mF_2} , h_{mE}) слоев E и F_2 . По измеренным частотам и высотам с использованием модели IRI-2012 восстанавливались вертикальные профили электронной концентрации. Расчет проводился также по модели без включения реальных данных вертикального зондирования. По данным вертикального зондирования и модельным значениям были рассчитаны среднемесячные значения и стандартные отклонения (σ) для параметров f_{0F_2} , f_{0E} , h_{mF_2} , h_{mE} для каждого часа суток. Определены условия применимости модели для автоматизированной обработки субавроуральной ионосферы. Модель IRI-2012 на этапе первоначальной обработки субавроуральной ионограммы может быть применена в дневные часы при невозмущенных условиях и отсутствии спорадической ионизации. При этом модельные отклонения могут быть скорректированы по ближайшим по времени реальным данным. Модельные данные для f_{0E} (в дневные часы) и h_{mF_2} могут быть применены для сокращения вычислительных затрат в системах автоматического поиска параметров по ионограммам и для предварительного определения диапазонов поиска основных определяемых параметров на основе расчета по модели IRI-2012. В отсутствие отдельных реальных значений модель IRI-2012 может быть применена для более точной аппроксимации рядов реальных данных. В условиях спорадической ионизации необходимо применять модели высокоширотной ионосферы, включающие блок корпускулярного ионообразования.

Ключевые слова: вертикальное радиозондирование, ионограмма, эмпирическая модель, критическая частота, высота максимумов.

IRI-2012 MODEL ADAPTABILITY ESTIMATION FOR AUTOMATED PROCESSING OF VERTICAL SOUNDING IONOGRAMS

V. Nikolaeva^{d, e}, M. Rybakov^f, A. Kotikov^{e, f}, V. Koshelevsky^f

^d Arctic and Antarctic Research Institute, Saint Petersburg, Russia, vera_nik@list.ru

^e Saint Petersburg State University, Saint Petersburg, Russia, vera_nik@list.ru

^f Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences Saint Petersburg Branch (SPbF IZMIRAN), Saint Petersburg, Russia, mvr_spb@rambler.ru

The paper deals with possibility of IRI-2012 global empirical model applying to the vertical sounding of the ionosphere semi-automatic data processing. Main ionosphere characteristics from vertical sounding data at IZMIRAN Voeikovo station in February 2013 were compared with IRI-2012 model calculation results. 2688 model values and 1866 real values of f_{0F_2} , f_{0E} , h_{mF_2} , h_{mE} were processed. E and F_2 layers critical frequency (f_{0E} , f_{0F_2}) and the maximum altitudes (h_{mF_2} , h_{mE}) were determined from the ionograms. Vertical profiles of the electron concentration were restored with IRI-2012 model by measured frequency and height. The model calculation was also made without the inclusion of the real vertical sounding data. Monthly averages and standard deviations (σ) for the parameters f_{0F_2} , f_{0E} , h_{mF_2} , h_{mE} for each hour of the day were calculated according to the vertical sounding and model values. Model applicability conditions for automated processing of ionograms for subauroral ionosphere were determined. Initial IRI-2012 model can be applied in the sub-auroral ionograms processing at daytime with undisturbed conditions in the absence of sporadic ionization. In this case model calculations can be adjusted by the near-time vertical sounding data. IRI-2012 model values for f_{0E} (in daytime) and h_{mF_2} can be applied to reduce computational costs in the systems of automatic parameters search and preliminary determination of the searching area range for the main parameters. IRI-2012 model can be used for a more accurate approximation of the real data series in the absence of the real values. In view of sporadic ionization, ionosphere models of the high latitudes must be applied with corpuscular ions formation unit.

Keywords: vertical sounding of the ionosphere, ionogram, empiric model, critical frequency, maximum height.

Введение

Вертикальное радиозондирование – один из основных методов, позволяющий эффективно проводить мониторинг состояния ионосферы. Полученные данные используются для решения широкого круга практических и теоретических задач. В настоящее время ведутся работы в направлении автоматизации обработки ионограмм и передачи данных ионосферных наблюдений в международную сеть [1]. Помимо результатов вертикального зондирования, для анализа состояния ионосферы широко используются как теоретические [2–6], так и эмпирические модели [7–9]. Эмпирические модели основываются на совокупности данных экспериментальных наблюдений за ионосферой. Такие модели из-за усреднений не

могут воспроизводить распределения параметров, соответствующие нестандартным условиям, например, конкретному геомагнитному возмущению. Таким образом, для проведения эффективного мониторинга целесообразно использовать адаптацию ионосферных моделей данными непосредственных измерений, полученных, в частности, при помощи вертикального зондирования.

Наиболее широко используемой эмпирической моделью ионосферы является глобальная модель IRI (International Reference Ionosphere) [8], рекомендуемая Committee on Space Research (COSPAR) и International Union of Radio Science (URSI). IRI-2012 – последняя версия модели – основана на данных измерений мировой сети ионозондов, радаров некогерентного рассеяния, спутников и ракет. Модель рассчитывает электронную концентрацию, температуру электронов и ионов; ионный состав и дрейфовую скорость в диапазоне высот 50–1500 км (слой D – 50–90 км; слой E – 90–130 км; слой F1 – 130–180 км; слой F2 – от 180 км). Помимо модельных профилей, рассчитанных только по дате, времени и координатам, модель IRI-2012 дает возможность построения вертикальных профилей ионосферных параметров с использованием реально измеренных величин критических частот и высот слоев E и F2.

Сопоставление расчетов по модели IRI-2007 с данными вертикального зондирования, проведенное для высоких широт [10], показало, что модель IRI-2007 недостаточно точно описывает специфику физических процессов в авроральной ионосфере (корпускулярные высыпания). В средних широтах такие процессы оказывают меньшее влияние на формирование ионосферных слоев, что делает целесообразным проведение оценки применимости результатов расчетов по эмпирической модели IRI-2012 в качестве дополнительного инструмента на этапе первоначальной обработки ионограмм и при анализе данных для субавроральной станции Санкт-Петербургского филиала Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова Российской академии наук (СПбФ ИЗМИРАН) в Воейково.

Условия эксперимента

Было проведено сравнение расчетов по модели IRI-2012 с реальными данными ионозонда Canadian Advanced Digital Ionosonde (CADI) производства Scientific Instrumentation Limited (SIL, Канада) [11, 12] установленного в магнито-ионосферной обсерватории в Воейково СПбФ ИЗМИРАН (59°57' с.ш., 30°42' в.д.). В таблице приведены основные технические характеристики ионозонда CADI.

Характеристика	Значение
Мощность передатчика, Вт	600
Диапазон высот, км	90–1020
Диапазон частот, МГц	1–20
Разрешение по высоте, км	6
Максимальная длительность импульсов, мкс	40

Таблица. Технические характеристики ионозонда CADI

С января 2013 г. данные ионозонда – ионограммы вертикального зондирования с временным разрешением 15 мин – доступны в сети Internet [13]. На станции Воейково производится стандартная ручная обработка ионограмм.

Анализ данных вертикального зондирования и модельные расчеты

В рассмотрение вошли часовые данные вертикального зондирования за февраль 2013 г. Этот период характеризовался умеренной солнечной активностью: индекс F10,7 в среднем был равен 101,8 (максимум 111,0 [10^{-22} Вт/(Гц·м²)).

По ионограммам определялись критические частоты (f_{0F_2} , f_{0E}) и высоты максимумов (h_{mF_2} , h_{mE}) слоев E, F₂. По измеренным частотам и высотам с использованием модели IRI-2012 восстанавливались вертикальные профили электронной концентрации. Также проводился расчет по модели без включения реальных данных вертикального зондирования.

В обработку вошло 2688 модельных значений параметров f_{0F_2} , f_{0E} , h_{mF_2} , h_{mE} и 1866 реальных значений, при этом 822 отсутствующих реальных значения в расчете параметров не участвовали и для построения реального профиля с помощью IRI интерпретировались как 0.

По данным вертикального зондирования и модельным значениям были рассчитаны среднемесячные значения (рис. 1 – зеленая линия) и стандартные отклонения (σ) (рис. 1 – вертикальные отрезки) для параметров f_{0F_2} , f_{0E} , h_{mF_2} , h_{mE} для каждого часа суток. На рис. 1, а–г, представлено сопоставление критических частот (f_{0F_2} , f_{0E}) и высот максимумов (h_{mF_2} , h_{mE}) слоев E, F₂ по данным вертикального зондирования с модельными величинами (рис. 1 – синяя линия) этих параметров. Количество реальных данных, вошедших в обработку (n), варьируется в зависимости от времени суток и оцениваемых параметров.

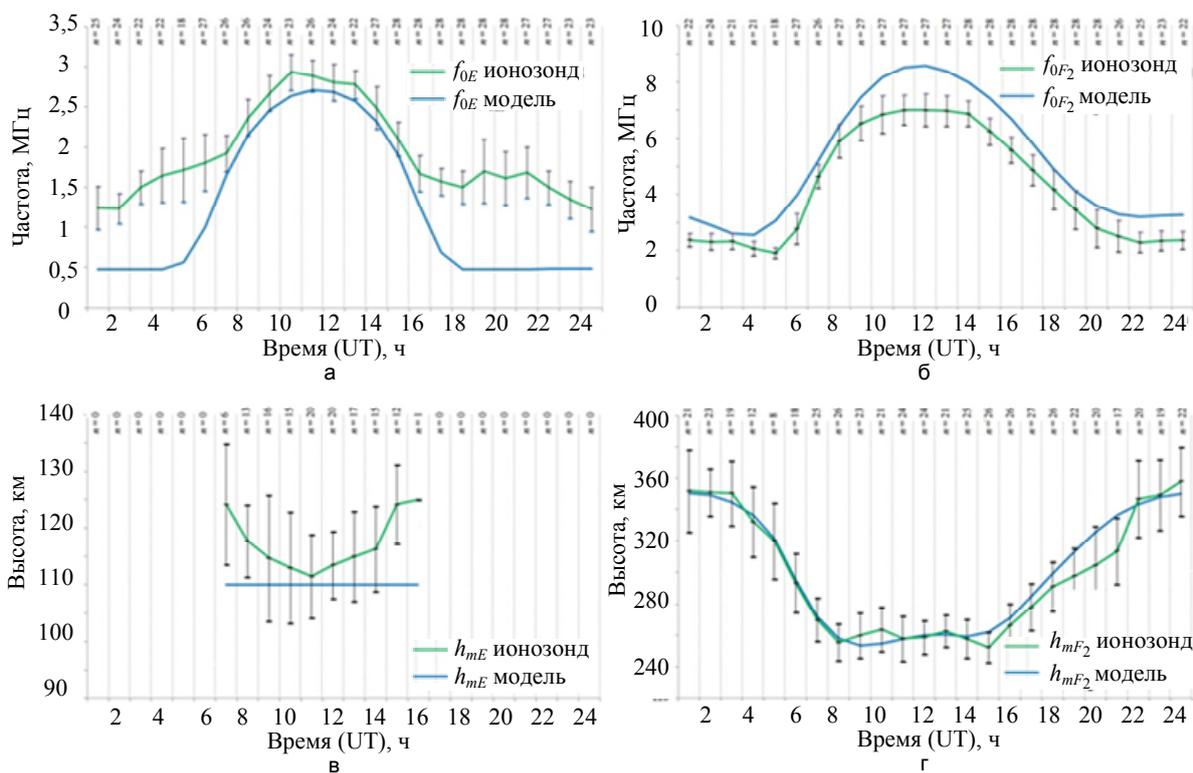


Рис. 1. Среднемесячные значения и стандартные отклонения (σ) для параметров: f_{0E} – (а); f_{0F_2} – (б); h_{mE} – (в); h_{mF_2} – (г) для каждого часа суток, полученных по данным ионозонда станции Воейково (зеленая линия) и рассчитанных по модели IRI (синяя линия)

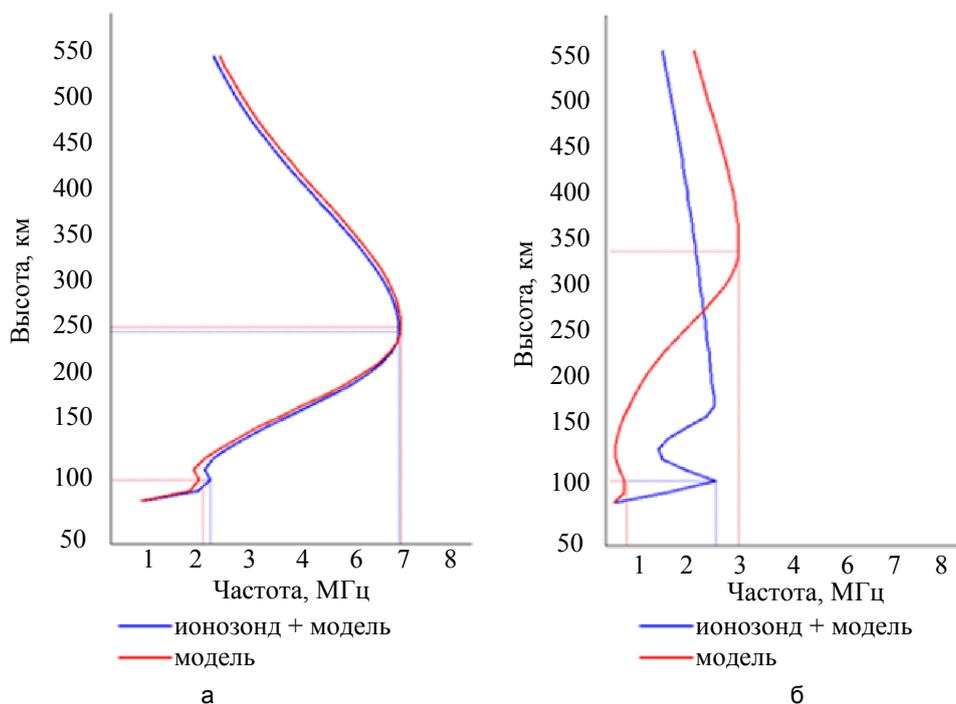


Рис. 2. Результаты сопоставления вертикального распределения частот, рассчитанного по модели IRI-2012 с коррекцией по данным вертикального зондирования (синяя кривая) и без коррекции (красная кривая); а – 8:00 UT 02.02.2013 г.; б – 19:00 UT 02.02.2013 г.

В дневные часы показано хорошее соответствие модельных расчетов и данных вертикального зондирования для критических частот слоя E. Однако от 17 до 7 ч Universal Time (UT) модель дает заниженные значения этого параметра (рис. 1, а). Подобный суточный ход электронной концентрации характерен для средних широт – выраженный полуденный максимум, связанный с суточным ходом ультрафиолето-

вого (УФ) потока солнечного излучения, и отсутствие корпускулярной ионизации в неосвященной ионосфере. Именно эти условия наилучшим образом обрабатываются моделью IRI.

Модельные значения h_{mE} лежат в пределах $\pm 2,1 \sigma$ реальных значений (рис. 1, в). Различие модельных и реальных значений составляет до 14,2 км (2,019 σ). Модельный алгоритм при отсутствии реального значения h_{mE} определяет $h_{mE} = 110$ км, что необходимо учитывать при анализе профиля, рассчитываемого моделью.

Наблюдается систематическое завышение модельных значений критической частоты f_{0F_2} (рис. 1, б) относительно реальных данных, что может быть связано с недостаточной точностью учета широтного распределения УФ излучения, задаваемого индексом F10,7. Диапазон разброса модельных значений f_{0F_2} от данных ионозонда – 0,4–1,8 МГц (тенденция завышения модельных значений). Наилучшее совпадение с экспериментом наблюдается при расчете максимальной высоты слоя F2 (рис. 1, г). Для h_{mF_2} модельные значения лежат в пределах $\pm 1,1 \sigma$ реальных значений. Наибольшие расхождения наблюдаются в 9 ч UT, 14 ч UT и 20 ч UT исследуемого периода. Максимальное различие модельных и реальных значений составляет 23 км.

На рис. 2 представлены результаты сопоставления вертикального распределения частот, рассчитанного по модели IRI-2012 с коррекцией по данным вертикального зондирования (синяя кривая) и без коррекции (красная кривая).

Рис. 2, а, демонстрирует практически полное совпадение двух профилей для 8:00 UT 02.02.2013 г. Во втором случае (рис. 2, б) при появлении спорадического слоя Es модель показывает полное несоответствие реальным данным. При этом скорректированный реальными данными вертикальный профиль частот выше слоя E моделью не воспроизводится.

Заключение

По результатам сопоставления данных вертикального зондирования и расчетов по модели IRI-2012 можно сделать следующие выводы:

- модель IRI-2012 на этапе первоначальной обработки субавроральной ионограммы может быть применена в дневные часы при невозмущенных условиях и отсутствии спорадической ионизации. При этом модельные отклонения могут быть скорректированы по ближайшим по времени реальным данным;
- модельные данные для f_{0E} (в дневные часы) и h_{mF_2} могут быть применены для сокращения вычислительных затрат в системах автоматического поиска параметров по ионограммам для предварительного определения диапазонов поиска основных определяемых параметров на основе расчета по модели IRI-2012;
- при отсутствии отдельных реальных значений модель IRI-2012 может быть применена для более точной аппроксимации рядов реальных данных;
- восстановление профиля электронной концентрации и получения параметров ионосферы при коррекции модели IRI-2012 данными вертикального зондирования ограничено появлением слоя Es;
- в условиях спорадической ионизации необходимо применять модели высокоширотной ионосферы, включающие блок корпускулярного ионообразования.

Полученные результаты используются при обработке ионограмм в СПбФ ИЗМИРАН и могут быть применены на магнитно-ионосферных станциях при обработке данных вертикального зондирования ионосферы.

Литература

1. Гришенцев А.Ю., Коробейников А.Г. Проектирование и технологическая подготовка сетей станций вертикального зондирования ионосферы // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 3 (85). С. 61–66.
2. Schunk R.W., Soika J.J. Ionospheric models // Modern Ionospheric Science / Eds H. Kohl, R. Rüster, K. Schlegel. Katlenburg-Lindau, Germany: European Geophysical Society, 1996. P. 181–215.
3. Namgaladze A.A., Korenkov Yu.N., Klimenko V.V., Karpov I.V., Bessarab F.S., Surotkin V.A., Glushchenko T.A., Naumova N.M. Global model of the thermosphere-ionosphere-protonosphere system // Pure and Applied Geophysics. 1988. V. 127. N 2–3. P. 219–254.
4. Namgaladze A.A., Namgaladze A.N., Yurik R.Yu. Global modeling of the quiet and disturbed upper atmosphere // Physics and Chemistry of the Earth. 2000. V. 25. N 5–6. P. 533–536.
5. Namgaladze A.N., Evstafiev O.V., Khudukon B.Z., Namgaladze A.A. Model interpretation of the ionospheric F-region electron density structures observed by ground-based satellite tomography at sub-auroral and auroral latitudes in Russia in January-May 1999 // Annales Geophysicae. 2003. V. 21. N 4. P. 1005–1016.

6. Mingalev O.V., Mingaleva G.I., Melnik M.N., Mingalev V.S. Numerical Simulation of the Time Evolution of Small-Scale Irregularities in the F-Layer Ionospheric Plasma // International Journal of Geophysics. 2011. V. 2011. P. 353640-1–353640-8.
7. Sergienko T.I., Ivanov V.E. A new approach to calculate the excitation of atmospheric gases by auroral electron impact // Annales Geophysicae. 1993. V. 11. N 8. P. 717–727.
8. Bilitza D. International Reference Ionosphere 2000 // Radio Science. 2001. V. 36. N 2. P. 261–275.
9. Turunen E., Matveinen H., Tolvanen J., Ranta H. D-region ion chemistry model // STEP Handbook of Ionospheric Models / Ed. R.W. Schunk. Logan: Utah State University, 1996. P. 1–25.
10. Николаева В.Д., Макарова Л.Н., Широчков А.В., Котиков А.Л. Сопоставление результатов расчетов модели IRI-2007 с данными вертикального зондирования и радара некогерентного рассеяния для авроральной ионосферы // Proc. of the XXXIV Annual Seminar «Physics of Auroral Phenomena». Apatity, 2011. P. 174–177.
11. Canadian Advanced Digital Ionosonde (CADI). Scientific Instrumentation Ltd. [Электронный ресурс]. Режим доступа: <http://www.sil.sk.ca/>, свободный. Яз. англ. (дата обращения 17.11.2013).
12. Canadian advanced digital ionosonde. System manuals. Saskatoon, Canada: Scientific instrumentation limited, 2009. 22 p.
13. Архив ионограмм. Институт земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова РАН. Ионосфера и распространение радиоволн [Электронный ресурс]. Режим доступа: <http://sphere.izmiran.ru/>, свободный. Яз. рус. (дата обращения 17.11.2013).

- | | |
|---|--|
| <i>Николаева Вера Дмитриевна</i> | – инженер, Арктический и антарктический научно-исследовательский институт; аспирант, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, vera_nik@list.ru |
| <i>Рыбаков Михаил Васильевич</i> | – ведущий инженер, Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова РАН, Санкт-Петербург, Россия, mvr_spb@rambler.ru |
| <i>Котиков Андрей Львович</i> | – кандидат физ.-мат. наук, старший научный сотрудник, Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова РАН; доцент, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, andkot54@mail.ru |
| <i>Коселевский Владимир Кириллович</i> | – кандидат физ.-мат. наук, зав. МИО Воейково, Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова РАН, Санкт-Петербург, Россия, koshel42@yandex.ru |
| <i>Vera Nikolaeva</i> | – leading engineer, Arctic and Antarctic Research Institute; postgraduate, Saint Petersburg State University, Saint Petersburg, Russia, vera_nik@list.ru |
| <i>Michael Rybakov</i> | – leading engineer, Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences Saint Petersburg Branch (SPbF IZMIRAN), Saint Petersburg, Russia, mvr_spb@rambler.ru |
| <i>Andrei Kotikov</i> | – PhD, senior scientific researcher, Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences Saint Petersburg Branch (SPbF IZMIRAN); associate professor, Saint Petersburg State University, Saint Petersburg, Russia, andkot54@mail.ru |
| <i>Vladimir Koshelevsky</i> | – PhD, Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences Saint Petersburg Branch (SPbF IZMIRAN), Saint Petersburg, Russia, koshel42@yandex.ru |

УДК 629.73.02; 629.73.05/06

ОЦЕНКА ТОЧНОСТИ ВИЗУАЛИЗАЦИИ МЕСТОПОЛОЖЕНИЯ ОБЪЕКТА В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ И СИСТЕМАХ ИНДИКАЦИИ НАВИГАЦИОННЫХ КОМПЛЕКСОВ ПИЛОТИРУЕМЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

М.О. Костишин^{a, b}, И.О. Жаринов^{a, b}, О.О. Жаринов^c, В.А. Нечаев^{a, b}, В.Д. Суслов^{a, b}

^a Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, job.max@me.com

^b ФГУП «Санкт-Петербургское ОКБ «Электроавтоматика» имени П.А. Ефимова», Санкт-Петербург, Россия, postmaster@elvat.spb.ru

^c Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, Россия, zharinov73@hotmail.ru

Авторами рассматривается проблема оценки точности отображения местоположения объекта в геоинформационных системах и системах индикации навигационных комплексов пилотируемых летательных аппаратов. Рассматриваются особенности применения различных типов жидкокристаллических экранов с разным числом пикселей по вертикали и горизонтали при отображении геоинформационных данных в различных масштабах. Отображение оценок значений навигационных параметров на борту летательного аппарата осуществляется двумя способами: непосредственно на экране многофункционального цветного индикатора отображается числовое значение параметра; на экране формируется силуэт объекта на фоне подложки, в качестве которой используется графическое изображение карты местности в зоне полета. Авторами рассмотрены различные масштабы отображения цифровой карты местности, используемые в настоящее время в авиационной промышленности. Приведены результаты расчета цены деления одного пикселя в зависимости от технических характеристик жидкокристаллического экрана и масштаба отображения карты местности на многофункциональном цифровом индикаторе. Приводятся результаты экспериментальных расчетов оценки точности отображения на местности положения летательного аппарата по данным от спутниковой навигационной системы и инерциальной навигационной системы, полученным в ходе выполнения программы полетов реального объекта. На основе произведенных расчетов построено семейство графиков ошибки точности отображения положения опорной точки объекта при использовании бортовых индикаторов с жидкокристаллическим экраном и различным разрешением экрана (в дюймах) (6"×8", 7,2"×9,6", 9"×12") для двух масштабов (1:0,25 км, 1:2 км) отображения карты. Приведенные семейства графиков зависимостей могут быть использованы как для оценки величины ошибки отображения положения объекта на местности в существующих навигационных системах, так и для расчета величины ошибки при модернизации объектов.

Ключевые слова: геоинформационные данные, отображение информации, цифровая карта местности.

ACCURACY EVALUATION OF THE OBJECT LOCATION VISUALIZATION FOR GEO-INFORMATION AND DISPLAY SYSTEMS OF MANNED AIRCRAFTS NAVI- GATION COMPLEXES

M. Kostishin^{d, e}, I. Zharinov^{d, e}, O. Zharinov^f, V. Nechaev^{d, e}, V. Suslov^{d, e}

^d Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, job.max@me.com

^e Saint Petersburg Scientific Design Bureau "Electroavtomatika" n.a. P. A. Efimov, Saint Petersburg, Russia, postmaster@elvat.spb.ru

^f Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, Russia, zharinov73@hotmail.ru

The paper deals with the issue of accuracy estimating for the object location display in the geographic information systems and display systems of manned aircrafts navigation complexes. Application features of liquid crystal screens with a different number of vertical and horizontal pixels are considered at displaying of geographic information data on different scales. Estimation display of navigation parameters values on board the aircraft is done in two ways: a numeric value is directly displayed on the screen of multi-color indicator, and a silhouette of the object is formed on the screen on a substrate background, which is a graphical representation of area map in the flight zone. Various scales of area digital map display currently used in the aviation industry have been considered. Calculation results of one pixel scale interval, depending on the specifications of liquid crystal screen and zoom of the map display area on the multifunction digital display, are given. The paper contains experimental results of the accuracy evaluation for area position display of the aircraft based on the data from the satellite navigation system and inertial navigation system, obtained during the flight program run of the real object. On the basis of these calculations a family of graphs was created for precision error display of the object reference point position using the onboard indicators with liquid crystal screen with different screen resolutions (6 "×8", 7.2 "×9.6", 9"×12") for two map display scales (1:0, 25 km, 1-2 km). These dependency graphs can be used both to assess the error value of object area position display in existing navigation systems and to calculate the error value in upgrading facilities.

Keywords: geographic information data, information display, digital area map.

Введение

В процессе полета летательного аппарата (ЛА) в бортовом навигационном комплексе производится оценивание значений пилотажно-навигационных параметров – широты и долготы позиционирования объекта в заданной системе координат, скорости самолета, относительного и барометрического давлений и др. Для информирования экипажа о местонахождении ЛА значения оценок ряда навигационных пара-

метров отображаются на бортовых средствах индикации. Основным элементом индикации современных навигационных комплексов являются бортовые индикаторы класса МФЦИ (многофункциональные цветные индикаторы), выполненные на базе плоских жидкокристаллических (ЖК) панелей [1, 2].

Отображение оценок значений навигационных параметров на борту ЛА может осуществляться двумя способами:

- непосредственно – на экране МФЦИ индицируется числовое значение параметра;
- мнемонически – на экране МФЦИ отображается силуэт объекта на фоне подложки, в качестве которой используется графическое изображение карты местности в зоне полетов (в проекциях «вид с земли на самолет» или «вид с самолета на землю»).

Непосредственный способ индикации значений оценок навигационных параметров прост в использовании экипажем, однако требует от штурмана выполнения дополнительных операций позиционирования ЛА на местности. Мнемонический способ отображения более нагляден для экипажа, так как в совмещенном режиме одновременно отображаются и карта местности, и местоположение объекта, однако он менее точен и используется сегодня для геоинформационной поддержки и повышения ситуационной осведомленности экипажа о географическом рельефе местности в зоне полетов.

Целью настоящей работы является представление широкому кругу читателей результатов научных исследований авторского коллектива в области получения оценок точности отображения местоположения объекта в геоинформационных системах и системах индикации современных навигационных комплексов пилотируемых ЛА для мнемонического способа.

Индикация геоинформационных данных на борту ЛА. Погрешность отображения навигационных параметров

Как было показано авторами ранее [3–9], синтез и отображение геоинформационных данных на экране МФЦИ, имеющем дискретную (пиксельную) структуру экрана, осуществляется в совмещенном режиме. Совмещенный режим (рис. 1, а) предполагает одновременное отображение на одном средстве индикации навигационной информации и картографической информации, представляющей собой цифровой массив карты местности в зоне полетов ЛА.

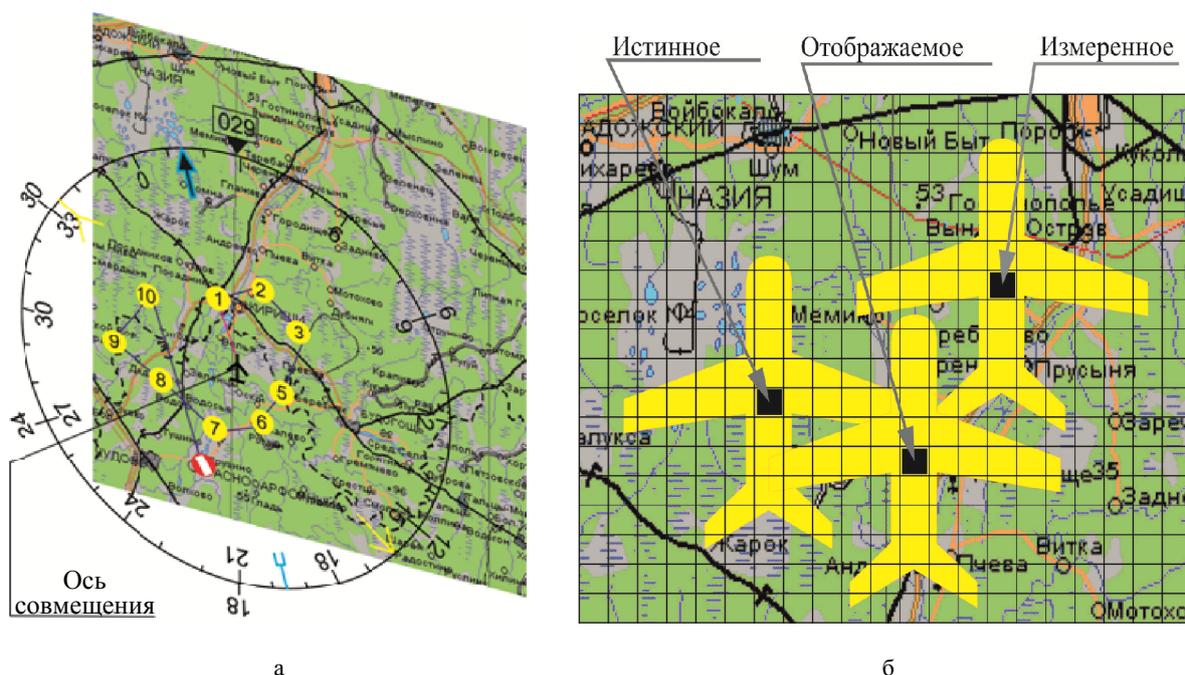


Рис. 1. Режим совмещения навигационной и картографической информации: режим отображения совмещенной навигационной информации и картографической информации (а); различия в оценке истинной, измеренной и отображаемой точек привязки положения летательного аппарата на местности в зоне полетов (б)

Для совмещения двух видов разнородных изображений используется опорная точка, в качестве которой могут выступать точка начала системы координат ЖК-панели, а также географические координаты широты и долготы объекта, измеряемые (оцениваемые) в процессе полета. При этом принято различать (рис. 1, б) истинное положение ЛА в географических координатах, оцениваемое положение ЛА в географических координатах и отображаемое положение ЛА в географических координатах. Истинное положение ЛА определяется проекцией центра масс объекта на физическую карту местности. Оцениваемое по-

ложение ЛА определяется проекцией оцененных значений широты и долготы положения ЛА на карту местности с учетом существующих погрешностей измерений и программно-алгоритмической обработки. Отображаемое положение ЛА определяется проекцией опорной точки мнемонического силуэта объекта на индицируемую карту местности. При отсутствии систематических погрешностей отображаемое и оцениваемое положения ЛА на экране МФЦИ совпадают.

В зависимости от масштаба отображения цифровой карты местности и типа применяемого в системе индикации навигационного комплекса ЖК-экрана, размера сторон ЖК-матрицы и числа пикселей экрана по вертикали и по горизонтали матрицы цена деления одного пикселя будет различной.

В настоящее время в авиационной промышленности получили широкое распространение следующие масштабы отображения цифровой карты местности [10]: 1:0,25 км; 1:2 км; 1:5 км; 1:10 км; 1:15 км. Применяемые сейчас на борту ЖК-экраны имеют размеры сторон (в дюймах): 5"×5"; 6"×6"; 6"×8"; 7,2"×9,6"; 9"×12" со следующим числом пикселей по сторонам матрицы: 480×480; 640×480; 768×576; 800×600; 768×768; 1024×768; 1400×1050; 1600×1200.

Масштаб, ед.:км	Число пикселей по вертикали и горизонтали	Размер экрана (длина, ширина), дюйм				
		5×5	6×6	8×6	9,6×7,2	12×9
		Цена деления одного пикселя ЖК-панели, м				
1:0,25	480×480	6,6	7,9	–	–	–
	640×480	–	–	7,9	9,5	11,9
	768×576	–	–	6,6	7,9	9,9
	800×600	–	–	6,4	7,6	9,5
	768×768	–	4,9	–	–	–
	1024×768	–	–	4,9	6,0	7,4
	1400×1050	–	–	3,6	4,4	5,4
	1600×1200	–	–	3,2	3,8	4,8
1:2	480×480	20,8	63,5	–	–	–
	640×480	–	–	63,5	76,2	95,3
	768×576	–	–	52,9	63,5	79,4
	800×600	–	–	50,8	61,0	76,2
	768×768	–	39,4	–	–	–
	1024×768	–	–	39,4	47,6	59,5
	1400×1050	–	–	29,0	34,8	43,5
	1600×1200	–	–	25,4	30,5	38,1
1:5	480×480	132,3	158,6	–	–	–
	640×480	–	–	158,6	190,5	238,1
	768×576	–	–	132,3	158,8	198,4
	800×600	–	–	127,0	152,4	190,5
	768×768	–	98,6	–	–	–
	1024×768	–	–	98,6	119,1	148,8
	1400×1050	–	–	72,6	87,1	108,9
	1600×1200	–	–	63,5	76,2	95,3

Таблица. Цена деления одного пикселя ЖК-экрана индикатора МФЦИ при отображении на борту ЛА цифровой карты местности

В таблице представлены результаты расчета цены деления одного пикселя (в метрах) в зависимости от технических характеристик ЖК-экрана и масштаба отображения на МФЦИ цифровой карты местности. Нетрудно видеть, что минимальное значение цены деления пикселя составит 3,2 м и достигается в комбинации: экран 6"×8" с числом пикселей по сторонам 1600×1200 и масштаб отображения 1:0,25 км. Аналогично, максимальная цена деления пикселя составит 238,1 м и достигается в комбинации: экран 9"×12" с числом пикселей по сторонам 640×480 и масштаб отображения 1:5 км. Расчет цены деления одного пикселя ЖК-панели производился по формуле

$$m = \frac{M \cdot L}{l \cdot c},$$

где m – цена деления одного пикселя ЖК-панели, m ; M – масштаб отображаемой карты местности, m ; L – сторона ЖК-матрицы, m ; l – число пикселей в стороне L ЖК-матрицы; $c = 0,01$ – размерный коэффициент, m . Прочерки в таблице соответствуют неиспользуемой комбинации параметров.

Оценка погрешности отображения положения ЛА на местности

Для оценки погрешности отображения положения ЛА на местности использовались результаты измерений оценочных значений навигационных параметров, зарегистрированные при выполнении программы полетов пилотируемого ЛА [11]. Результаты измерений для широты и долготы положения ЛА, полученные в процессе 3 ч 20 мин полета от спутниковой навигационной системы (СНС) и от инерциальной навигационной системы (ИНС), представлены на рис. 2, а, и рис. 2, б, соответственно.

Расчет значений широты и долготы полета в таблице произведен по формулам

$$\lambda^* = 111,12\lambda \cos \varphi, \quad \varphi^* = 111,12\varphi,$$

где φ – географическая широта в градусах; λ – географическая долгота в градусах; φ^* – географическая широта в километрах; λ^* – географическая долгота в километрах. В предположении, что «истинное» значение местоположения ЛА может с определенной погрешностью оцениваться по результатам измерений СНС, на рис. 2 штриховкой залита область разности результатов оценки значений географической широты и долготы для результатов измерений СНС и ИНС.

Анализ рис. 2 показывает, что в течение 3 ч 20 мин полета максимальное значение разности измерений между СНС и ИНС составило 612 м для широты и 22 км для долготы. При этом существенный рост разницы в результатах измерений СНС и ИНС начинается примерно с 45 мин полета (момент t_0 на графиках рис. 2), что объясняется собственной погрешностью ИНС в процессе инерциального счисления координат местоположения ЛА.

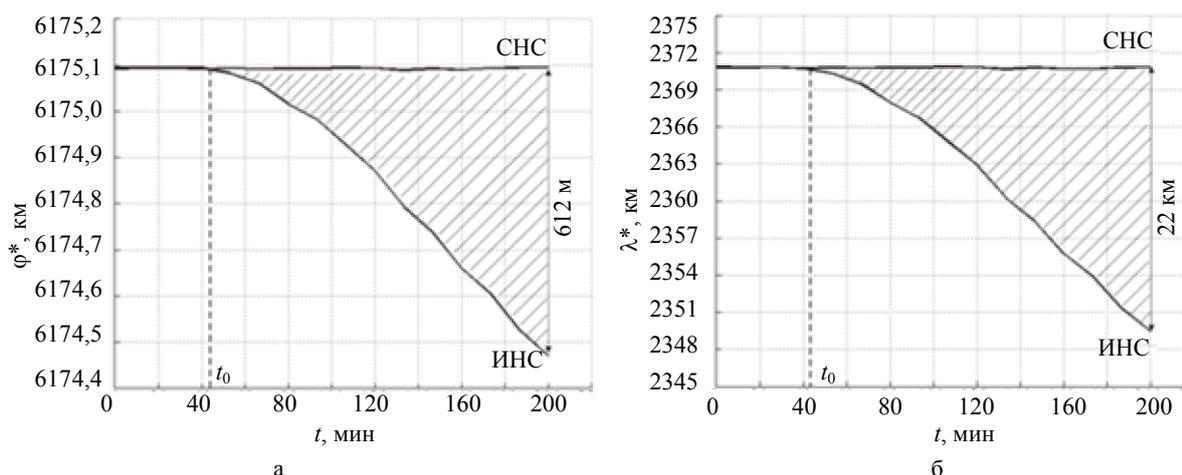


Рис. 2. Результаты измерения (оценки) навигационных параметров для широты (а) и долготы местоположения ЛА в процессе выполнения программы полетов (б), г. Жуковский, 2013 г.

На рис. 3 приведены результаты экспериментальных расчетов оценки точности Δ (в пикселях) отображения «измеренного» значения навигационных параметров по СНС, $\varphi^*_{\text{снс}}$, по отношению к значениям, оцененным навигационным комплексом по ИНС, $\varphi^*_{\text{инс}}$ [12–15]: $\Delta^*\varphi = \varphi^*_{\text{снс}} - \varphi^*_{\text{инс}}$. Следует заметить, что использование режима СНС на борту ЛА осложняется условиями эксплуатации объекта и не всегда возможно, в связи с чем семейство зависимостей на рис. 3 следует рассматривать как характеристики ошибки визуализации местоположения ЛА на карте местности, возникающие в режиме чистого счисления пути, начиная с момента времени t_0 [16, 17].

Заключение

В результате проведенного исследования и натурных экспериментов была совершена серия полетов пилотируемого летательного аппарата с использованием спутниковой навигационной системы и инерциальной навигационной системы. В качестве инерциальной навигационной системы использовалась лазерная инерциальная навигационная система ИНС-2000. Как следует из рис. 2, примерно через 45–50 мин полета ошибка измерения широты и долготы полета летательного аппарата начинает монотонно возрастать из-за влияния собственной погрешности инерциальной системы, вследствие чего опорная точка совмещения навигационных и геоинформационных данных смещается. Семейство зависимостей, приведенное на рис. 3, иллюстрирует разницу (в пикселях) между отображаемой и измеренной опорными

точками объекта на цифровой карте местности для многофункциональных цветных индикаторов с различными техническими характеристиками. Величина ошибки позиционирования объекта на карте местности, накопленная за 2,5 ч полета, варьируется от 16 до 190 пикселей жидкокристаллического экрана. Точные значения величины ошибки могут быть оценены по данным, представленным в таблице.

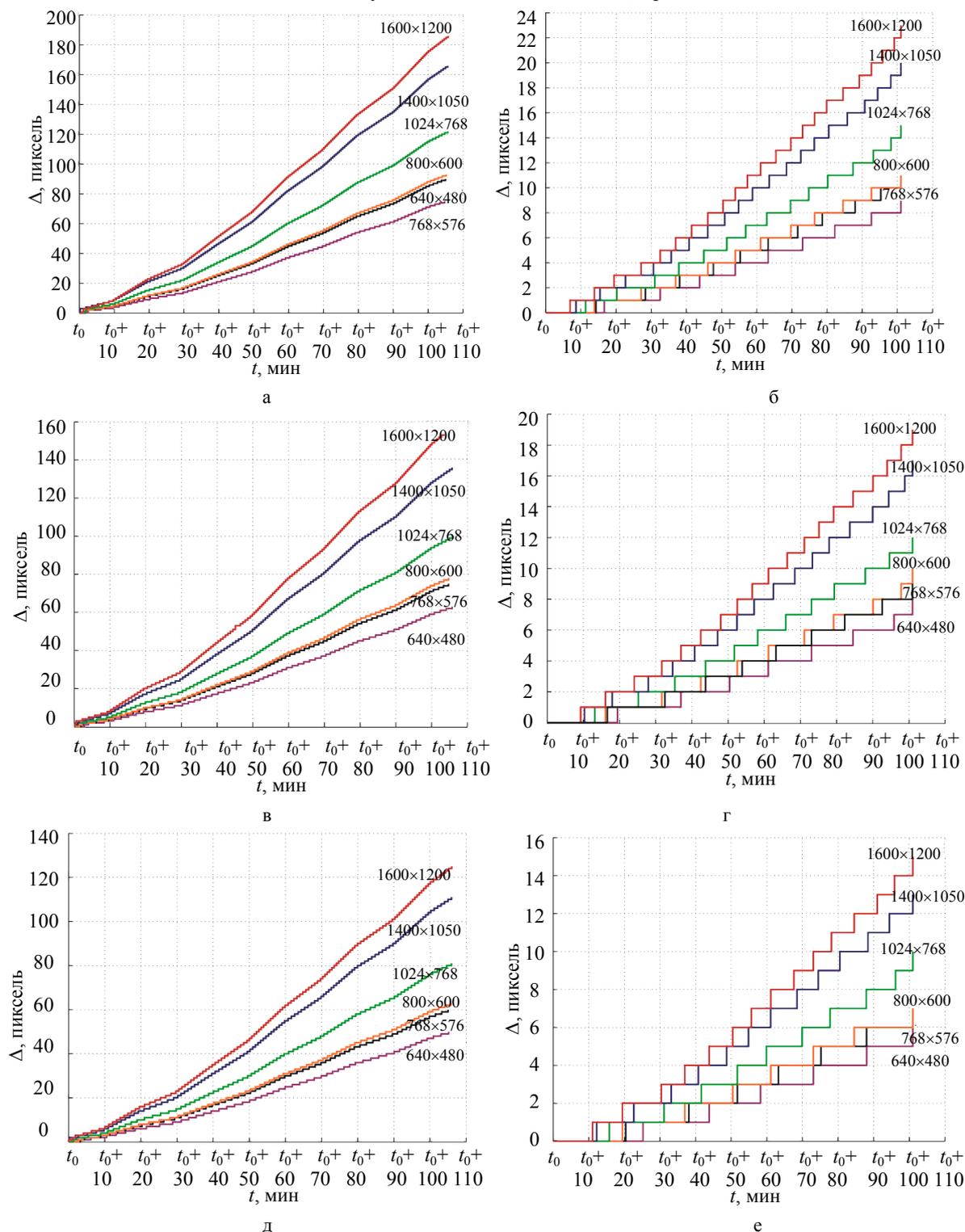


Рис. 3. Семейство зависимостей ошибки визуализации местоположения объекта в геоинформационных системах и системах индикации навигационных комплексов при оценке широты полета для бортовых индикаторов МФЦИ со следующими характеристиками: размер экрана 6"×8", масштаб 1:0,25 км (а); размер экрана 6"×8", масштаб 1:2 км (б); размер экрана 7,2"×9,6", масштаб 1:0,25 км (в); размер экрана 7,2"×9,6", масштаб 1:2 км (г); размер экрана 9"×12", масштаб 1:0,25 км (д); размер экрана 9"×12", масштаб 1:2 км (е)

Приведенные на рис. 3 семейства зависимостей могут быть использованы как для оценки величины ошибки отображения местоположения объекта на местности в существующих навигационных системах, так и для учета величины будущей ошибки при модернизации объектов с использованием инерциальных навигационных систем других типов:

- при замене многофункциональных индикаторов на индикаторы с более совершенными характеристиками жидкокристаллических матриц;
- при загрузке в бортовую систему картографической информации геоинформационных данных в других масштабах;
- при разработке системы индикации и навигационных подсистем на различных проектных предприятиях и при использовании более совершенных инерциальных навигационных систем [18].

Литература

1. Жаринов И.О., Жаринов О.О. Бортовые средства отображения информации на плоских жидкокристаллических панелях: Учеб. пособие. СПб: ГУАП, 2005. 144 с.
2. Копорский Н.С., Видин Б.В., Жаринов И.О. Бортовые средства отображения информации современных пилотируемых летательных аппаратов // Современные технологии: Сборник статей / Под ред. С.А. Козлова, В.Л. Ткалич. СПб: СПбГУ ИТМО, 2004. С. 154–165.
3. Парамонов П.П., Ильченко Ю.А., Жаринов И.О. Теория и практика статистического анализа картографических изображений в системах навигации пилотируемых летательных аппаратов // Датчики и системы. 2001. № 8. С. 15–19.
4. Парамонов П.П., Ильченко Ю.А., Жаринов И.О., Тарасов П.Ю. Структурный анализ и синтез графических изображений на экранах современных средств бортовой индикации на плоских жидкокристаллических панелях // Авиакосмическое приборостроение. 2004. № 5. С. 50–57.
5. Копорский Н.С., Видин Б.В., Жаринов И.О. Система бортовой картографической информации пилотируемых летательных аппаратов. Основные принципы построения // Сборник трудов 10-й Международной конференции «Теория и технология программирования и защиты информации». СПб: СПбГУ ИТМО, 2006. С. 18–23.
6. Парамонов П.П., Коновалов П.В., Жаринов И.О., Кирсанова Ю.А., Уткин С.Б. Реализация структуры данных, используемых при формировании индикационного кадра в бортовых системах картографической информации // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 2 (84). С. 165–167.
7. Парамонов П.П., Костишин М.О., Жаринов И.О., Нечаев В.А., Сударчиков С.А. Принцип формирования и отображения массива геоинформационных данных на экран средств бортовой индикации // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 6 (88). С. 136–142.
8. Жаринов И.О., Коновалов П.В. Классификация структуры данных, используемых при отображении геоинформационных ресурсов в бортовых системах картографической информации // 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. Ч. 1. С. 118–121.
9. Жаринов И.О., Емец Р.Б. Индикационное оборудование в авиации XXI века // Научно-технический вестник СПбГУ ИТМО. 2003. № 5 (11). С. 193–195.
10. Парамонов П.П., Копорский Н.С., Видин Б.В., Жаринов И.О. Многофункциональные индикаторы на плоских жидкокристаллических панелях: наукоемкие аппаратно-программные решения // Научно-технический вестник СПбГУ ИТМО. 2004. № 3 (14). С. 238–245.
11. Суслов В.Д., Шек-Иовсепенц Р.А., Видин Б.В., Жаринов И.О., Немолочнов О.Ф. К вопросу об унификации бортовых алгоритмов комплексной обработки информации // Изв. вузов. Приборостроение. 2006. Т. 49. № 6. С. 39–40.
12. Жаринов И.О. Системный подход при проектировании комплексов бортового радиоэлектронного оборудования // Научная сессия ГУАП: Сборник докладов. СПб: ГУАП, 2006. Ч. 2. С. 68–74.
13. Парамонов П.П., Видин Б.В., Есин Ю.Ф., Жаринов И.О., Колесников Ю.Л., Кофман М.М., Сабо Ю.И., Шек-Иовсепенц Р.А. Теория и практика системного проектирования авионики. Тула: Гриф и К°, 2010. 265 с.
14. Жаринов И.О., Жаринов О.О. Бортовые системы картографической информации. Принципы построения геоинформационных ресурсов: Учеб. пособие. СПб: СПбГУ ИТМО, 2008. 48 с.
15. Kostishin M.O., Zharinov I.O. Precision characteristics of the positioning of objects in aircraft geoinformation systems // Automation & Control: Proc. of the International Conference of Young Scientists «ISCAC-2013». St. Petersburg: National Research University St. Petersburg State Polytechnical University, 2013. P. 92–96.

16. Swail C., Jennings S. Enhanced and synthetic vision system concept for application to search and rescue missions // Symposium on «Sensor Data Fusion and Integration of the Human Element». Ottawa, 1998. P. 15-1–15-6.
17. Young S., Kakarlapudi S., Uijt de Haag M. A Shadow detection and extraction algorithm using digital elevation models and X-Band weather radar measurements // International Journal of Remote Sensing. 2005. V. 26. N 8. P. 1531–1549.
18. Honeywell's primus epic integrated avionics system provides advanced flight deck functionality. SmartView synthetic vision system [Электронный ресурс]. Режим доступа: http://www.honeywellbusinessaviation.com/primus_epic/performance/smartview, свободный. Яз. англ. (дата обращения 28.11.2013).

- Костишин Максим Олегович** – аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; инженер, ФГУП «Санкт-Петербургское ОКБ «Электроавтоматика» имени П.А. Ефимова», Санкт-Петербург, Россия, job.max@me.com
- Жаринов Игорь Олегович** – доктор технических наук, доцент, зав. кафедрой, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; руководитель учебно-научного центра, ФГУП «Санкт-Петербургское ОКБ «Электроавтоматика» имени П.А. Ефимова», Санкт-Петербург, Россия, igor_rabota@pisem.net
- Жаринов Олег Олегович** – кандидат технических наук, доцент, Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, Россия, zharinov73@hotmail.ru
- Нечаев Владимир Анатольевич** – доцент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; руководитель научно-исследовательского центра, ФГУП «Санкт-Петербургское ОКБ «Электроавтоматика» имени П.А. Ефимова», Санкт-Петербург, Россия, nil-12@mail.ru
- Суслов Владимир Дмитриевич** – доцент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; руководитель экспертного совета, ФГУП «Санкт-Петербургское ОКБ «Электроавтоматика» имени П.А. Ефимова», Санкт-Петербург, Россия, postmaster@elavt.spb.ru
- Maxim Kostishin** – postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; engineer, Saint Petersburg Scientific Design Bureau “Electroavtomatica” n.a. P. A. Efimov, Saint Petersburg, Russia, job.max@me.com
- Igor Zharinov** – D.Sc., Associate professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; Head of learning scientific center, Saint Petersburg Scientific Design Bureau “Electroavtomatica” n.a. P. A. Efimov, Saint Petersburg, Russia, igor_rabota@pisem.net
- Oleg Zharinov** – PhD, Associate professor, Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, Russia, zharinov73@hotmail.ru
- Vladimir Nechaev** – Associate professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; Head of scientific research center, Saint Petersburg Scientific Design Bureau “Electroavtomatica” n.a. P. A. Efimov Saint Petersburg, Russia, nil-12@mail.ru
- Vladimir Suslov** – Associate professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; Head of expert council, Saint Petersburg Scientific Design Bureau “Electroavtomatica” n.a. P. A. Efimov, Saint Petersburg, Russia, postmaster@elavt.spb.ru

УДК 007.51

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ БАКТЕРИЙ В ПРОДУКТАХ ПИТАНИЯ

А.П. Саенко^а, В.М. Мусалимов^а, Ш. Лерм^б, Г. Линц^б

^аСанкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, musvm@yandex.ru

^бТехнический университет Ильменау, Ильменау, Германия, steffen.lerm@tu-ilmenau.de

Рассматривается один из способов решения актуальной проблемы обеспечения контроля качества продуктов питания с использованием методов машинного обучения. Существующие в настоящее время методы анализа требуют специального лабораторного оборудования, значительного времени и сильно зависят от квалификации и некоторых физиологических особенностей эксперта, в то время как предлагаемый метод позволяет существенно снизить затраты за счет автоматизации процесса. Рассмотрено устройство, реализующее данный метод, принцип действия которого основан на флуоресцентной микроскопии. Для набора обучающих данных решена задача классификации объектов на

классы «бактерия» и «сторонний артефакт», проведен сравнительный анализ различных алгоритмов классификации (метод опорных векторов, случайный лес, деревья решений C4.5, метод k ближайших соседей, метод Байеса), который показал, что наилучшей эффективностью обладают метод опорных векторов и случайный лес.

Данная работа выполнена на кафедре мехатроники Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики и кафедре контроля качества и промышленной обработки изображений Технического университета г. Ильменау (Германия) в рамках совместной программы «Михаил Ломоносов» Министерства образования и науки Российской Федерации и Германской службы академических обменов.

Ключевые слова: машинное обучение, обнаружение бактерий.

DETECTION OF BACTERIA IN FOODSTUFF BY MACHINE LEARNING METHODS

A. Saenko^c, V. Musalimov^c, S. Lerm^d, G. Linss^d

^c Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, alexey.saenko@gmail.com

^d Ilmenau University of Technology, Ilmenau, Germany, steffen.lerm@tu-ilmenau.de

The paper deals with an actual problem of ensuring the control of foodstuff quality by means of machine learning methods. Existing analysis methods require special laboratory environment, significant time and depend on the qualification and some physiological characteristics of an expert while the suggested method gives the possibility to decrease significantly the costs due to automatization. The mobile analysis platform performing this method is based on the fluorescence microscopy. The problem of the object classification as either “bacterium” or “third-party artifact” was solved for the test data with some classification algorithms as support vector machine, random forest, decision tree C4.5, k -nearest neighbors, Bayes method. The analysis showed that the most effective algorithms are support vector machine and random forest. This research is performed on the Mechatronics Department of Saint Petersburg National Research University of Information Technologies, Mechanics and Optics and the Quality Assurance and Industrial Image Processing Department of Ilmenau University of Technology with the support of the program “Mikhail Lomonosov” of the Ministry of Education and Science of Russia and the German Academic Exchange Service.

Keywords: machine learning, bacteria detection.

Введение

В последние годы методы машинного обучения все чаще применяются для решения большого количества задач в различных отраслях науки и техники [1–7]. С другой стороны, одной из важнейших проблем современного мира является обеспечение контроля качества продуктов питания, что требует решения комплексной задачи, включающей в себя, в том числе, мероприятия по обнаружению бактерий в жидких образцах мясного сока. Оценка риска заражения мяса заключается в определении вида и количества бактерий в пробе. Существующие в настоящее время методы анализа требуют специального лабораторного оборудования, значительного времени и сильно зависят от квалификации и некоторых физиологических особенностей эксперта, обеспечивающего визуальный контроль микроскопических изображений исследуемых проб. В частности, международный стандарт [8] определяет процедуру контроля качества продуктов, состоящую из этапов, показанных на рис. 1. Выполнение всех этапов осуществляется в течение 3–5 дней.

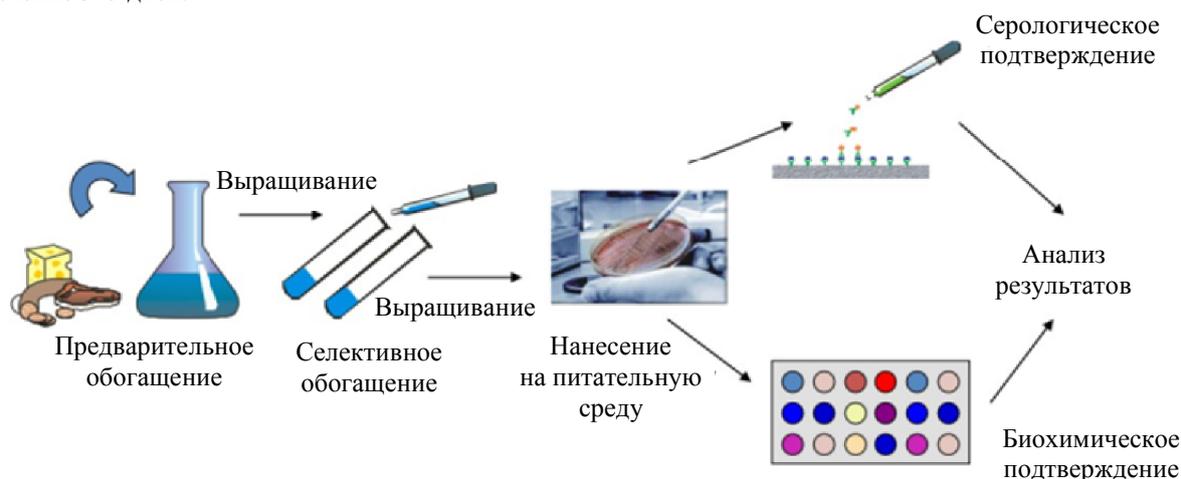


Рис. 1. Процедура контроля качества мясных продуктов питания, определенная стандартом [8] (источник: презентация о проекте LabChip IO на ежегодном собрании CoOPTICS)

Указанные недостатки являются основной предпосылкой для создания нового более совершенного способа контроля качества продуктов питания, способного обеспечивать своевременное предоставление результатов до момента доставки продуктов потребителям, а также до момента порчи продуктов. В каче-

стве варианта решения данной проблемы предлагается использование мобильной платформы для анализа продуктов питания (мясного сока) на наличие бактерий, описанной в работе [9] и изображенной на рис. 2.

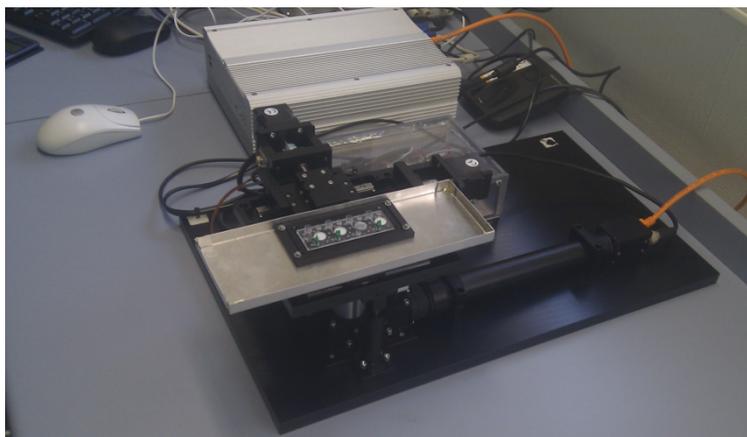
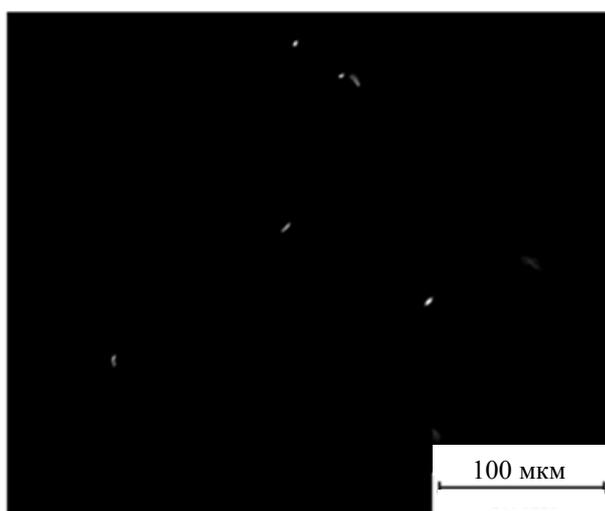


Рис. 2. Мобильная аналитическая платформа



а

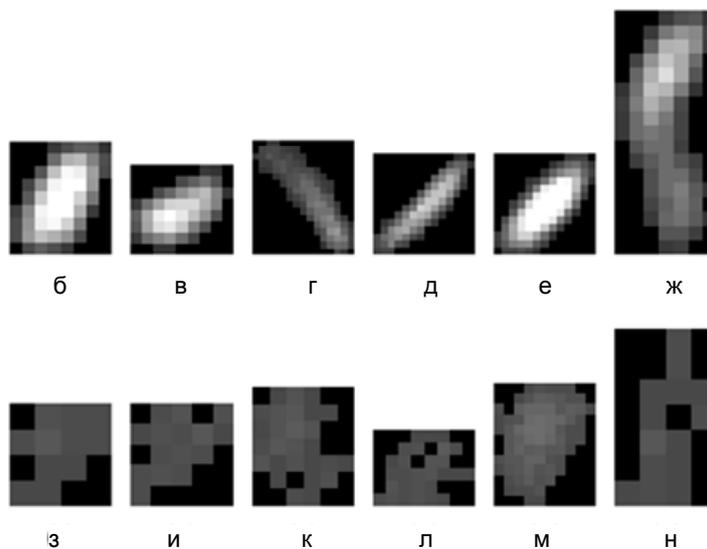


Рис. 3. Пример исходного изображения (а) и выделенных на нем объектов: бактерий (б–ж) и сторонних артефактов (з–н)

Все изложенное дает основание утверждать, что применение методов обработки изображений и машинного обучения с целью выявления бактерий в продуктах питания является актуальной задачей, решение которой позволит значительно ускорить и упростить процесс контроля. В соответствии с этой целью настоящей работы является решение поставленной задачи путем классификации объектов, а также сравнительный анализ эффективности различных алгоритмов.

Принцип действия мобильной аналитической платформы

Принцип действия рассматриваемой оптической измерительной системы основан на флуоресцентной микроскопии. Измеряемым объектом является мембрана с биологическими образцами, представляющая собой дискообразную полость диаметром около 10 мм и толщиной около 2 мм. Сама мембрана при этом располагается на подвижной платформе, которая осуществляет прецизионные перемещения по горизонтали и вертикали. Источник света, облучая мембрану светом определенной длины волны, возбуждает флуоресценцию, изображения которой записываются камерой. Испытанные образцы содержали бактерии вида *Legionella pneumophila* – палочковидное возбудителя легионеллеза длиной около 2 мкм и шириной 0,3–0,9 мкм. Каждое полученное изображение покрывает площадь $0,84 \times 7,07 \text{ мм}^2$, что существенно меньше общей области образца, поэтому для каждой мембраны осуществляется создание набора изображений для каждого положения по горизонтали и вертикали. Возможные способы оптимизации маршрута обхода образца оптической измерительной системой описаны в работах [9, 10].

Эмпирически установлено, что разница между информативностью цветных и монохромных изображений в рамках данной задачи незначительна, поэтому предпочтение отдано монохромным как менее объемным и более простым в обработке.

Обнаружение бактерий

Задача обнаружения бактерий сводится к задаче классификации на два непересекающихся класса («бактерия» и «сторонний артефакт») и, по сути, является частным случаем задачи машинного обучения, которая в общем виде заключается в необходимости при конечном множестве классов $Y = 1, 2, \dots, l$ построить алгоритм, который по объекту x определяет точное или достаточно точное значение $y(x)$ [3]. В качестве исходных данных принимается пространство допустимых объектов X , пространство меток Y , а также целевая функция $y(x)$, заданная в конечном множестве точек обучающей выборки $y(x_1), y(x_2), \dots, y(x_m)$. Обучающая выборка представляет собой матрицу с описанием объектов \mathbf{X} и вектор меток \mathbf{Y} :

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix},$$

где m – количество объектов, а n – количество признаков этих объектов. Таким образом, каждый ряд матрицы \mathbf{X} соответствует одному объекту x_i , представленному в виде вектора признаков, а каждый элемент $y_i \in \{0, 1\}$ определяет класс i -го объекта.

Функция потерь $L(A(x), y(x))$ показывает, насколько ответ $A(x)$ соответствует верному ответу $y(x)$ и определяется как

$$L(A(x), y(x)) = \begin{cases} 1, & A(x) \neq y(x) \\ 0, & A(x) = y(x) \end{cases}.$$

Соответственно, алгоритмы машинного обучения должны решать задачу оптимизации в виде

$$\frac{1}{m} \sum_{i=1}^m L(A(x), y(x)) \rightarrow \min.$$

Во время подготовки обучающей выборки для классификации объектов после получения снимков они были обработаны с целью сегментации объектов, которые затем классифицировались экспертами – сотрудниками микробиологической лаборатории с помощью программной системы бинаризации и сегментации изображений [11]. В результате была получена база, состоящая из более 200 изображений (60 бактерий и более 140 сторонних артефактов), которая впоследствии сократилась до 120 изображений (по 60 представителей каждого класса, рис. 3). Затем для каждого изображения извлекались 30 признаков, таких как длина и диаметр области, моменты области, большая и малая полуоси эллипса, описанного вокруг области, округлость и др. [12–15]. Гистограммы всех признаков (рис. 4) позволили выбрать релевантные, на основе которых осуществлялась классификация.

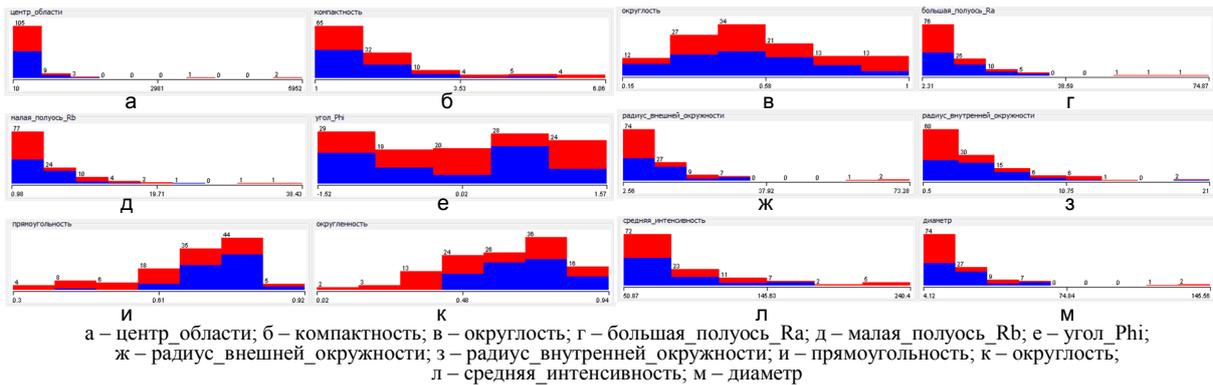


Рис. 4. Гистограммы признаков (красным цветом обозначены значения для бактерий, синим – для сторонних артефактов)

Оценка эффективности классификаторов

Обычно оценка эффективности алгоритмов классификации производится экспериментально, из-за существенной неформальности большинства задач распознавания, и выражается в способности выбранных алгоритмов принимать верные решения, что характеризуется уровнем ошибок первого («ложный пропуск» – событие ложно не обнаруживается) и второго рода («ложное обнаружение» – событие ошибочно считается произошедшим).

Если количество объектов для каждого класса в тестовом наборе $N = Np + Nn$, где N – общее количество объектов, Np – количество бактерий и Nn – количество сторонних артефактов, а количество ложных пропусков и ложных обнаружений равно FN и FP соответственно, то количество верных пропусков и верных обнаружений определяется как

$$TP = Np - FN,$$

$$TN = Nn - FP.$$

При этом уровни ошибок выражаются следующим образом:

$$nFN = \frac{FN}{Np} \cdot 100\%,$$

$$nFP = \frac{FP}{Nn} \cdot 100\%,$$

$$nTN = \frac{TN}{Nn} \cdot 100\%,$$

$$nTP = \frac{TP}{Np} \cdot 100\%.$$

Одним из способов оценки эффективности алгоритмов классификации в заданных условиях является мера расстояния до точки $(0, 1)$ на ROC-диаграмме (Receiver Operating Characteristic – операционная характеристика приемника), которая вычисляется как $E = \sqrt{FP_{rate}^2 + (1 - TP_{rate})^2}$ [1, 5, 6]. При этом минимальное возможное значение 0 соответствует наилучшей эффективности ($FP_{rate} = 0$, а $TP_{rate} = 1$, т.е. все бактерии верно определены как экземпляры класса «бактерия» и ни один сторонний артефакт не определен как экземпляр класса «бактерия»). Максимальное значение $\sqrt{2}$ отображает наихудшую эффективность при $FP_{rate} = 1$ и $TP_{rate} = 0$. Таким образом, для процентного выражения эффективности классификатора, принимая за 100% максимальную эффективность, получаем $E_{\%} = \left(1 - \frac{E}{\sqrt{2}}\right) \cdot 100\%$.

Проведенный анализ эффективности в работе [16] показал, что наилучшей эффективностью, около 82%, обладают метод опорных векторов и случайный лес. Для сравнения: деревья решений C4.5 и метод k ближайших соседей показали эффективность около 72%, а метод Байеса – 38%.

Заключение

В работе продемонстрирована эффективность решения проблемы обнаружения бактерий для контроля качества продуктов питания методами обработки изображений и машинного обучения, описан принцип действия мобильной аналитической платформы для мгновенного анализа образцов, предложены

пути решения проблемы путем классификации объектов. Кроме этого, приведен сравнительный анализ эффективности различных классификаторов.

Литература

1. Bigus J.P. Data Mining with Neural Networks. McGraw-Hill, 1996. P. 220.
2. The handbook of data mining / Ed. N.Ye. Lawrence Erlbaum Associates, 2003. 689 p.
3. Дьяконов А.Г. Анализ данных, обучение по прецедентам, логические игры, системы WEKA, RapidMiner и MatLab (Практикум на ЭВМ кафедры математических методов прогнозирования): Учеб.пособие. М.: Издательский отдел факультета ВМК МГУ им. М.В. Ломоносова, 2010. 278 с.
4. Anding K. Automatisierte Qualitätssicherung von Getreide mit überwachten Lernverfahren in der Bildverarbeitung: Dissertation zur Erlangung der akademischen Grades Doktoringenieur (Dr.-Ing.). Ilmenau, Germany: Technische Universität Ilmenau, 2010. 235 p.
5. Witten I.H., Frank E., Hall M.A. Data Mining: Practical Machine Learning Tools and Techniques. 3rd ed. Morgan Kaufmann, 2011. 629 p.
6. Bramer M. Principles of data mining. 2nd ed. Springer, 2013. 440 p.
7. Латышев С.М., Воронин А.А., Андинг К., Линц Э., Курицын П.А. Оптико-электронные методы и средства идентификации веществ и материалов // Изв. вузов. Приборостроение. 2013. Т. 56. № 10. С. 81–87.
8. ISO 6579:2002. Microbiology of food and animal feeding stuffs – Horizontal method for the detection of Salmonella spp. 08.08.2002. 32 p.
9. Lerm S., Holder S., Gopfert A., Futterer R., Linss G. Concepts of a scanning hardware platform for high-resolution image processing with Lab-on-a-chip analysis // Proc. of the 15th International Symposium «MECHATRONIKA». Prague, 2012. P. 1–4.
10. Lerm S. Objektsegmentierung von kompakten Schüttgut für die technische Erkennung: Dissertation zur Erlangung der akademischen Grades Doktoringenieur (Dr.-Ing.). Ilmenau, Germany: Technische Universität Ilmenau, 2012. 235 p.
11. Саенко А.П. Программная система бинаризации и сегментации изображений: Свидетельство о государственной регистрации программы для ЭВМ 2013613134. Заявл. 2013.
12. Shapiro L., Stockman G. Computer Vision. Prentice Hall PTR, 2001. 580 p.
13. Gonzalez R.C., Woods R.E. Digital Image Processing. 2nd ed. Prentice Hall, 2002. 793 p.
14. Erhardt A. Einführung in die Digitale Bildverarbeitung. Vieweg+Teubner Verlag, 2008. 248 p.
15. HALCON Version 11.0.1 – HALCON / HDevelop Reference Manual. MVTec Software GmbH, 2012. 2352 p.
16. Саенко А.П. Оценка эффективности обнаружения бактерий методами обработки цифровых изображений и интеллектуального анализа данных // Сборник научных трудов Одиннадцатой сессии международной научной школы «Фундаментальные и прикладные проблемы надежности и диагностики машин и механизмов». СПб: Институт проблем машиноведения РАН, 2013. С. 318–321.

<i>Саенко Алексей Петрович</i>	– аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, alexey.saenko@gmail.com
<i>Мусалимов Виктор Михайлович</i>	– доктор технических наук, профессор, зав. кафедрой, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, musvm@yandex.ru
<i>Лерм Штеффен</i>	– аспирант, Технический университет Ильменау, Ильменау, Германия, steffen.lerm@tu-ilmenau.de
<i>Линц Герхард</i>	– доктор технических наук, профессор, зав. кафедрой, Технический университет Ильменау, Ильменау, Германия, gerhard.linss@tu-ilmenau.de
<i>Alexey Saenko</i>	– postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, alexey.saenko@gmail.com
<i>Victor Musalimov</i>	– D.Sc., Professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, musvm@yandex.ru
<i>Steffen Lerm</i>	– postgraduate, Ilmenau University of Technology, Ilmenau, Germany, steffen.lerm@tu-ilmenau.de
<i>Gerhard Linss</i>	– D.Sc., Professor, Department head, Ilmenau University of Technology, Ilmenau, Germany, gerhard.linss@tu-ilmenau.de

УДК 004.912

ПРЕДСТАВЛЕНИЕ ДОКУМЕНТОВ В ЗАДАЧЕ КЛАСТЕРИЗАЦИИ АННОТАЦИЙ НАУЧНЫХ ТЕКСТОВ

С.В. Попова^{a, b}, В.В. Данилова^c

^a Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, svp@list.ru

^b Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, svp@list.ru

^c Автономный университет Барселоны, Барселона, Испания, maolve@gmail.com

Рассматривается проблема кластеризации узкотематических текстов короткой длины, таких как аннотации к научным публикациям. Цель решения данной задачи – группировка результатов запросов в поисковых системах по научным публикациям. Используются наблюдения, полученные при решении задачи извлечения ключевых фраз из документов. Был применен расширенный список стоп-слов, построенный автоматически для решения задачи извлечения ключевых фраз и позволивший значительно улучшить качество информации, получаемой из научных публикаций. Приводится описание процедуры построения данного списка стоп-слов. Основной задачей является исследование возможности повысить качество и (или) скорость кластеризации аннотаций с помощью вышеупомянутого списка стоп-слов, а также информации о частях речи лексем. В последнем случае для представления документов применяется словарь, содержащий не все слова коллекции, а только существительные и прилагательные, или словарь, состоящий из последовательностей существительных и прилагательных.

Использованы два базовых алгоритма кластеризации: *k*-means и иерархическая кластеризация (метод межгруппового среднего). Показано, что использование расширенного списка стоп-слов и представление документов на основе существительных и прилагательных из словаря коллекции позволяют улучшить качество и скорость работы алгоритма *k*-means. Для метода межгруппового среднего в аналогичном случае может наблюдаться ухудшение качества кластеризации. Показано, что использование для представления документов последовательностей из существительных и прилагательных снижает качество кластеризации для обоих алгоритмов и оправдано только в тех случаях, когда требуется значительное снижение размерности пространства признаков.

Ключевые слова: кластеризация документов; представление документов; использование ключевых фраз, существительных и прилагательных; построение расширенного списка стоп-слов, представления результатов поиска.

DOCUMENT REPRESENTATION FOR CLUSTERING OF SCIENTIFIC ABSTRACTS

S. Popova^{d, e}, V. Danilova^f

^d Saint Petersburg State University, Saint Petersburg, Russia, svp@list.ru

^e Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, svp@list.ru

^f Autonomous University of Barcelona, Barcelona, Spain, maolve@gmail.com

The key issue of the present paper is clustering of narrow-domain short texts, such as scientific abstracts. The work is based on the observations made when improving the performance of key phrase extraction algorithm. An extended stop-words list was used that was built automatically for the purposes of key phrase extraction and gave the possibility for a considerable quality enhancement of the phrases extracted from scientific publications. A description of the stop- words list creation procedure is given. The main objective is to investigate the possibilities to increase the performance and/or speed of clustering by the above-mentioned list of stop-words as well as information about lexeme parts of speech. In the latter case a vocabulary is applied for the document representation, which contains not all the words that occurred in the collection, but only nouns and adjectives or their sequences encountered in the documents. Two base clustering algorithms are applied: *k*-means and hierarchical clustering (average agglomerative method). The results show that the use of an extended stop-words list and adjective-noun document representation makes it possible to improve the performance and speed of *k*-means clustering. In a similar case for average agglomerative method a decline in performance quality may be observed. It is shown that the use of adjective-noun sequences for document representation lowers the clustering quality for both algorithms and can be justified only when a considerable reduction of feature space dimensionality is necessary.

Keywords: document clustering, document representation, key phrases application, use of nouns and adjectives, extended list of stop-words creation, results retrieval representation.

Введение и обзор состояния дел в области

В работе рассматривается проблема кластеризации текстовых данных в режиме реального времени. Исследование проводится в рамках решения задачи представления результатов поиска пользователю в виде кластеров [1–3], в частности, в системах, осуществляющих поиск по базам научных публикаций [4, 5]. Целью решения является представление в структурированном виде коллекции документов с автоматическим разделением ее на группы тематически близких текстов.

Исходные данные – аннотации к научным публикациям, доступные в научных электронных библиотеках. Большинство таких библиотек предоставляет свободный доступ к аннотациям, получение полных текстов статей, как правило, требует специальной подписки или оплаты доступа. Тем не менее, аннотации представляют собой лаконичную выдержку основного содержания работы и могут быть достаточными для кластеризации научных текстов и получения адекватного результата [6].

Работа с аннотациями связана с рядом сложностей: во-первых, это короткие тексты, обработка которых обычно выделяется в отдельную проблему *short-text clustering/classification* (кластеризация/классификация текстов короткой длины) [7–10], во-вторых, обрабатываемые тексты могут иметь одну общую тему, в связи с чем требуется решение проблемы кластеризации узкотематических текстов короткой длины [11–14] (*narrow-domain short text clustering*). В более ранних исследованиях, посвященных решению указанных проблем, ученые сосредоточены на разработке алгоритмов безотносительно их вычислительной сложности, основной акцент делается на повышение качества кластеризации.

Одной из основных проблем при кластеризации текстов короткой длины является сильная разреженность данных [11]. Так как мы имеем дело с короткими документами, а суммарный объем текста в коллекции может быть небольшим, затруднен сбор статистических характеристик, необходимых для обработки. Если документы взяты из одного источника (например, выбраны как релевантные одному запросу), велика вероятность того, что все они будут тематически близкими друг к другу. В этом случае, помимо небольшого размера, тексты могут иметь значительные перекрытия по общим словам, что еще больше усложняет задачу кластеризации [11]. Для тестирования и анализа качества работы алгоритмов кластеризации узкотематических текстов короткой длины был сформирован ряд коллекций, которые активно используются в области (CICling, EasyAbstracts, SEPLN-CICling) [6–8, 11–14]. Эксперименты в данной работе также базируются на данных коллекциях.

Выбор алгоритмов и мотивация

Поставлена задача исследовать возможность повысить качество и (или) скорость кластеризации коротких текстов узкой тематики, используя дополнительную информацию о частях речи слов и расширенный список стоп-слов. При постановке задачи для экспериментов мы опирались на исследования в области кластеризации/классификации текстов и на исследования в области извлечения ключевых понятий (ключевых фраз). Наблюдения показывают, что в обеих областях используются очень похожие подходы и инструменты [1–3, 15–17], что позволяет сделать вывод о возможном использовании достижений в одной области в рамках работ другого направления.

В области аннотирования документов ключевыми словами/фразами было показано, что существительные и прилагательные являются основными частями речи, необходимыми для извлечения ключевых понятий [18–21]. Данное наблюдение позволило сформировать гипотезу о том, что использование только существительных и прилагательных может положительно сказаться на качестве кластеризации и позволит понизить размерность пространства признаков. Поставленный эксперимент также связан с нашими исследованиями в области извлечения ключевых фраз [5, 22–24]. Показано, что последовательности, состоящие из существительных и прилагательных, часто содержат в себе основные понятия из документов и позволяют добиться неплохого качества аннотирования.

Второе наблюдение связано с исследованием, в рамках которого нам удалось показать эффективный способ автоматического построения списка стоп-слов для извлечения ключевых понятий из аннотаций к научным публикациям [24]. В настоящей работе мы исследуем возможность повысить качество кластеризации при использовании списка стоп-слов, полученного в работе [24].

Для кластеризации были выбраны алгоритмы *k-means* (*k-средних*) и *hierarchical agglomerative clustering* (иерархическая кластеризация), метод межгрупповых связей. Была использована известная в области библиотека Weka, предлагающая реализацию указанных выше алгоритмов. Алгоритмы являются классическими в задаче информационного поиска и хорошо описаны в литературе [25]. Авторы предполагают, что читатель знаком с данными методами или может самостоятельно ознакомиться с ними, например, в [25]. Отметим, что для иерархической кластеризации применяется алгоритм, оценивающий расстояние между кластерами как среднее расстояние между каждой парой объектов в разных кластерах (в [25] этот метод обозначен как «centroid», в Weka как «AVERAGE»). Данный метод не учитывает расстояния между объектами внутри объединяемых кластеров (в отличие от *Group-average agglomerative clustering* в [25]). В дальнейшем для выбранного алгоритма иерархической кластеризации мы будем использовать обозначение НАС.

При выборе алгоритмов кластеризации мы исходили из двух положений: вычислительная простота (*k-средних*) и потенциально возможный высокий результат кластеризации (НАС). В области автоматической группировки коротких текстов узкой тематики представлены результаты для таких алгоритмов иерархической кластеризации, как методы ближнего или дальнего соседа [11]. Нашей задачей было расширить данные по результатам тестирования алгоритмов иерархической кластеризации для указанной задачи.

Мотивацией для проведения исследования является задача кластеризации результатов запроса в поисковых системах [1–3], акцент был сделан на изучение данной проблемы для систем академического поиска (т.е. поиска по научным публикациям) [4, 5]. Введено предположение о неважности затрат на обработку данных, если данная обработка может быть проведена оффлайн, например, извлечение информации о части речи слов или извлечение из документов последовательностей из существительных и прила-

гательных. Основными требованиями к алгоритму кластеризации были высокая скорость работы (в пределах времени, допустимого для ожидания пользователя) и как можно более высокое качество работы.

Описание тестовых коллекций

Выбраны три коллекции (sites.google.com/site/merrecalde/resources), часто используемые для тестирования алгоритмов кластеризации документов короткой длины или узкотематических коротких текстов [7, 8, 11–14]. Все три коллекции представляют собой наборы аннотаций к научным публикациям, распределенные по четырем кластерам экспертом. Для каждой коллекции известен «золотой стандарт» – наилучший вариант группировки. Размер каждой из указанных коллекций – 48 документов. Коллекция CICling 2002 признана исследователями как одна из наиболее сложных коллекций для кластеризации [7, 8, 11–14].

Оценка качества кластеризации

В работе была использована классическая оценка качества кластеризации, основанная на комбинировании информации о точности (Precision) и полноте (Recall) полученных кластеров [25–27]:

$$FM = \sum_i \frac{G_i}{|D|} \max_j F_{ij}, \text{ где } F_{ij} = \frac{2 \cdot P_{ij} \cdot R_{ij}}{P_{ij} + R_{ij}}, \quad (1)$$

$$P_{ij} = \frac{|G_i \cap C_j|}{G_i}, \quad R_{ij} = \frac{|G_i \cap C_j|}{C_j},$$

$G = \{G_i\}_{i=1, \dots, m}$ – кластеры, полученные в результате кластеризации, $C = \{C_j\}_{j=1, \dots, n}$ – классы, построенные экспертом вручную, D – множество документов в коллекции.

Описание эксперимента

Исследование состояло из двух основных блоков: работа с алгоритмом k -средних и с алгоритмом НАС. Для каждой тестовой коллекции требовалась кластеризация по четырем кластерам (аналогично работам [7, 8, 11–14], для алгоритмов, требующих данные о числе кластеров, указывалось число классов в золотом стандарте). Для представления документов использовалась векторная модель: каждый документ был представлен вектором в пространстве признаков (словаря коллекции). Вес каждого признака (слова) w для документа d определялся с помощью $tf-idf$ [25] (кроме случая представления документов с помощью последовательностей слов):

$$tf-idf(w, d) = tf^d(w) \cdot \log \frac{|D|}{df(w)},$$

где D – множество документов в коллекции; $tf^d(w)$ – частота слова w в документе d ; $df(w)$ – число документов, в которых встретилось слово w .

В обоих алгоритмах для определения расстояния между объектами использовался косинус угла между векторами, представляющими документ.

Для каждого из алгоритмов проверялись две гипотезы:

- **гипотеза 1:** качество кластеризации можно повысить, используя для представления документов только существительные и прилагательные, а также расширенный список стоп-слов;
- **гипотеза 2:** повысить качество кластеризации можно, используя для представления документов последовательности из существительных и прилагательных.

Во всех экспериментах для представления документов и пространства признаков были использованы стемы слов, полученные при помощи стеммера Портера (The Porter Stemming Algorithm¹). Во всех экспериментах применялся стандартный список стоп-слов, кроме отдельно отмеченных экспериментов с расширенным списком.

Проверка гипотезы 1. В эксперименте для построения пространства признаков были использованы все существительные и прилагательные из словаря коллекции, встретившиеся хотя бы раз, по крайней мере, в одном тексте. Каждый документ был представлен вектором в полученном пространстве признаков. Также был рассмотрен случай, когда из пространства признаков были удалены все слова, попавшие в расширенный список стоп-слов. Способ построения данного списка приводится далее.

Построение расширенного списка стоп-слов. Нами проводились исследования в области извлечения ключевых фраз для аннотаций к научным публикациям [5, 22–24]. Для построения фраз были использованы только существительные и прилагательные, представленные в тексте. В работе [24] был предложен алгоритм автоматического построения словаря стоп-слов, включающего общепотребитель-

¹<http://tartarus.org/martin/PorterStemmer>

ные слова, которые не отражают тематической направленности публикации и, как следствие, не входят в состав ключевых фраз (например: предыдущий, новый, предшествующий, статья и т.д.). Была использована одна из основных в области извлечения ключевых фраз коллекция INSPEC [18, 19, 21–24, 28, 29]. Подробное описание коллекции приводится в [18]. INSPEC содержит в себе несколько подколлекций, в частности, «trial», которая используется для обучения алгоритмов, и «test» – для оценки качества работы алгоритмов аннотирования и сравнения алгоритмов. Каждая из коллекций имеет «золотой стандарт» – результат аннотирования ключевыми фразами каждого документа. Для построения списка стоп-слов использовалась подколлекция «trial»: последовательно отбиралось каждое существительное или прилагательное из словаря коллекции и добавлялось в стандартный список стоп-слов. Запускался алгоритм извлечения ключевых фраз (см. полное описание в [24]), и, если добавление этого слова приводило к повышению качества полученной информации более чем на заданную величину (0,0002, см. в [24]), то добавленное в список слово помечалось как «плохое». Примеры таких «плохих» слов приводятся в табл. 1.

№	Слово (оригинал извлеченного слова)	Русскоязычный перевод слова
1	Actual	Существующий, текущий
2	Possible	Вероятный, возможный
3	Excellent	Отличный, превосходный
4	Results	Результаты
5	Number	Число, количество, номер
6	Pure	Строгий, чистый, простой
...

Таблица 1. Примеры стоп-слов, попавших в расширенный список стоп-слов

После того как описанная выше процедура была выполнена для всех существительных и прилагательных из словаря коллекции, все слова, получившие отметку «плохое», были собраны вместе и добавлены к стандартному списку стоп-слов. Полученный таким образом расширенный список был использован для извлечения ключевых фраз из документов коллекции «test» и позволил значительно повысить качество полученной информации по сравнению со случаем использования стандартного списка. В настоящей работе мы применяем вышеописанный расширенный список стоп-слов.

Проверка гипотезы 2. В эксперименте каждый документ был описан с помощью последовательностей существительных и прилагательных, следующих в документе друг за другом (не более пяти слов подряд), разделенных словами других частей речи, пунктуацией, стоп-словами. Также в описание документов были добавлены пересечения полученных последовательностей. Например, если в тексте есть последовательности «русский язык» и «английский язык», то последовательность «язык» также будет добавлена в описание документа.

Для кластеризации было использовано пространство признаков, состоящее из всех последовательностей, встретившихся хотя бы раз, по крайней мере, в одном описании документа. Каждый документ был представлен бинарным вектором в полученном пространстве признаков, где вес признака оценивался как 1, если последовательность, соответствующая признаку, присутствует в описании документа, и 0 в противном случае.

Результаты экспериментов и обсуждение

Результаты экспериментов представлены в табл. 2, 3. В табл. 2 приводятся результаты для алгоритма k -средних, в табл. 3 – итоги вычислений для алгоритма НАС. В таблицах приняты следующие обозначения. В строке «Базовый эксперимент» представлены результаты оценки качества алгоритма для каждой из трех коллекций (CICling, SEPLN-CICling, EasyAbstracts). Для построения пространства признаков были использованы все слова из словаря коллекции. Так как в алгоритме k -means центроиды выбираются произвольным образом, то качество работы алгоритма меняется от эксперимента к эксперименту. Для k -means приводятся результаты, полученные на основе 1000 запусков алгоритма: средний результат (avg), худший результат (min), лучший результат (max). В строке «Эксперимент: только существительные и прилагательные» приводятся результаты экспериментов для случая, когда построение пространства признаков основано только на существительных и прилагательных из словаря коллекции. В строке «Эксперимент: только существительные и прилагательные, использован расширенный список стоп-слов» представлены результаты эксперимента, аналогичного последнему, но пространство признаков сужено за счет расширенного списка стоп-слов. В строке «Эксперимент с использованием словосочетаний» приведены результаты экспериментов, в рамках которых для построения пространства признаков и представления документов использовались только последовательности из существительных и прилагательных; в строке «Эксперимент с использованием словосочетаний и расширенного списка стоп-слов» представле-

ны итоги экспериментов с использованием расширенного списка стоп-слов. Оценка качества кластеров, полученных в результате работы каждого из алгоритмов, проводилась с помощью (1).

Название эксперимента	Название коллекции								
	CICling			SEPLN-CICling			EasyAbstracts		
	avg	min	max	avg	min	max	avg	min	max
Базовый эксперимент	0,48	0,34	0,68	0,56	0,34	0,80	0,57	0,35	0,83
Эксперимент: только существительные и прилагательные	0,48	0,33	0,68	0,57	0,36	0,83	0,57	0,37	0,83
Эксперимент: только существительные и прилагательные, использован расширенный список стоп-слов	0,49	0,34	0,73	0,58	0,35	0,83	0,58	0,37	0,86
Эксперимент с использованием словосочетаний	0,45	0,33	0,66	0,50	0,34	0,71	0,49	0,34	0,77
Эксперимент с использованием словосочетаний и расширенного списка стоп-слов	0,46	0,32	0,67	0,53	0,36	0,77	0,50	0,33	0,66

Таблица 2. Результаты оценки качества работы алгоритма *k*-means

Название эксперимента	Название коллекции		
	CICling	SEPLN-CICling	EasyAbstracts
Базовый эксперимент	0,56	0,84	0,92
Эксперимент: только существительные и прилагательные	0,58	0,84	0,92
Эксперимент: только существительные и прилагательные, использован расширенный список стоп-слов	0,54	0,82	0,92
Эксперимент с использованием словосочетаний	0,44	0,43	0,51
Эксперимент с использованием словосочетаний и расширенного списка стоп-слов	0,52	0,63	0,54

Таблица 3. Результаты оценки качества работы алгоритма НАС

Результаты, указанные в таблицах, показывают, что использование для представления документов последовательностей из существительных и прилагательных приводит к снижению качества кластеризации по сравнению со случаем, когда для представления документов и построения пространства признаков использованы все слова. Основная причина – краткость обрабатываемых документов и, как следствие, небольшое число последовательностей, представляющих текст, а также очень низкое пересечение по общим последовательностям между документами. Использование расширенного списка стоп-слов при представлении документов с помощью последовательностей из существительных и прилагательных в большинстве экспериментов позволяет повысить качество кластеризации по сравнению со случаем использования указанных последовательностей и стандартного списка стоп-слов. Наблюдение позволяет сделать предположение о целесообразности использования расширенного списка стоп-слов для описания документов с помощью последовательностей из существительных и прилагательных как в задаче извлечения ключевых фраз, так и в задаче кластеризации.

Результаты работы алгоритма НАС улучшают результаты *k*-means. НАС обеспечивает высокие результаты по сравнению с другими алгоритмами, используемыми в области [7, 8, 11–14]. Для сравнения приведем результаты работы других алгоритмов, опубликованные в [14]. Нужно учитывать, что НАС дает один стабильный результат кластеризации, который не изменяется от запуска к запуску и является одновременно средним, лучшим и худшим значением. При сравнении рекомендуем читателю ознакомиться с входными параметрами, которые были использованы для алгоритмов в [14]. Так, существует некоторое различие в результатах, представленных для *k*-means в настоящем исследовании и в [14]. В первую очередь влияет на это различное число запусков алгоритма *k*-means: 1000 в данной работе и 50 в [14], а также случайный выбор центроидов во время каждого из запусков. В табл. 4 представлены результаты из [14],

полученные по результатам 50 запусков каждого из алгоритмов: CLUDIPSO [7, 30], Ant-Tree [31], Major-Clust [32], DBSCAN, AntSA-CLU [14].

Algorithms	EasyAbstracts			SEPLN-CICling			CICling 2002		
	avg	min	max	avg	min	max	avg	min	max
Major-Clust	0,69	0,44	0,98	0,59	0,4	0,77	0,43	0,37	0,58
DBSCAN	0,66	0,62	0,72	0,63	0,4	0,77	0,47	0,42	0,56
Ant-Tree	0,6	0,5	0,67	0,49	0,41	0,64	0,41	0,38	0,48
CLUDIPSO	0,92	0,85	0,98	0,72	0,58	0,85	0,6	0,47	0,73
AntSA-CLU	0,96	0,92	0,98	0,75	0,63	0,85	0,61	0,47	0,75

Таблица 4. Результаты работы алгоритмов, представленные в [14]

Несмотря на хорошие результаты кластеризации HAC (табл. 3), нужно отметить бóльшую вычислительную сложность HAC по сравнению с k -means, что допускает использование первого только на небольших объемах данных в случае ввода ограничения на время работы алгоритма. Ту же особенность имеют алгоритмы CLUDIPSO и AntSA-CLU, качество работы которых значительно выше, чем у других алгоритмов, рассмотренных в табл. 4.

Сложность алгоритма k -means оценивается как $\Theta(I*K*N*M)$ [25], где M – временные затраты на расчет расстояния между парой объектов (документом и центроидом), $K*N$ – количество производимых расчетов расстояния между документом и центроидом (K – число центроидов, N – число документов), I – число итераций алгоритма. Отмечено [25], что число итераций, как правило, не является большой величиной, более того, может быть фиксировано в алгоритме (например, ограничено 10-ю итерациями). Если число итераций фиксировать, то оценка сложности станет $\Theta(K*N*M)$. Если ввести предположение, что число кластеров ограничено некоторым небольшим значением, и фиксировать максимально возможное число кластеров (например, в пределах 1000 штук), то сложность алгоритма можно оценить как $\Theta(N*M)$, т.е. сложность будет линейно зависеть от числа документов и скорости вычисления расстояния между документом и центроидом. Сложность иерархической кластеризации (HAC) оценивается как:

- время, необходимое для построения таблицы расстояний между каждой парой объектов $\Theta(N^2*M)$, где N – число документов, M – время расчета расстояния между парой документов;
- время, необходимое для работы алгоритма, когда таблица расстояний между объектами уже построена $\Theta(N^2 * \log N)$.

Как отмечено в [25], можно считать сложность HAC как $\Theta(N^2*M)$ (в [25] авторы пишут о $\Theta(N^2)$, указывая, что это сложность относительно временных затрат на расчет расстояний между каждой парой объектов). Обоснование: временные затраты, требуемые на расчет расстояний между парами объектов и построение таблицы расстояний, как правило, являются более значительными, чем временные затраты на последующую работу алгоритма. Таким образом, в отличие от k -means, в HAC наблюдается не линейная, а квадратичная зависимость от размера входных данных, что приводит к значительно более быстрому росту временных затрат на работу алгоритма с увеличением объема данных и делает невозможным использование HAC в режиме реального времени, начиная с некоторого момента. По этой причине при увеличении объема обрабатываемых данных потребуются перейти к использованию k -means. Аналогичное произойдет и в случае использования CLUDIPSO или AntSA-CLU.

Алгоритм CLUDIPSO, также лежащий внутри AntSA-CLU, является реализацией подхода к решению задачи кластеризации как задачи оптимизации. Задача алгоритма – разбить документы на кластеры так, чтобы доставить оптимум выбранной внешней оценки качества. Для решения задачи оптимизации в CLUDIPSO используется метод роя частиц (PSO) [7]. В экспериментах [14] в качестве оптимизируемой функции был выбран Global Silhouette coefficient (GS), расчет которого требует вычисления для каждого документа:

- средней непохожести этого документа с другими документами его кластера;
- средней непохожести документа с другими документами ближайшего кластера [7].

В худшем случае такой расчет даст $O(N^2*M)$, где M – время, затрачиваемое на расчет непохожести двух документов, а N – число документов, в отличие от k -means: $O(N*M)$, где M – время на расчет расстояния между парой объектов.

Фактором, влияющим на время работы алгоритма k -means, являются затраты на расчет расстояний между объектами и центроидами. Снижение пространства признаков уменьшает размерность векторов, представляющих документы, и, как следствие, затраты на расчет расстояний между объектами. Использование только существительных и прилагательных для построения пространства признаков позволяет снизить размерность для коллекции CICling на 27%, для коллекции SEPLN-CICling на 25%, для коллек-

ции EasyAbstracts на 23%. Можно ожидать снижения временных затрат на кластеризацию с помощью k -means при таком изменении размерности. Как показывает табл. 2, в этом случае не происходит потери в качестве кластеризации, напротив, наблюдается повышение качества для всех трех тестовых коллекций, особенно в случае применения расширенного списка стоп-слов. Таким образом, обосновывается использование только существительных и прилагательных, а также расширенного списка стоп-слов для представления документов в случае алгоритма k -means. Для получения экспериментального подтверждения мы провели оценку времени работы алгоритма:

- для случая, когда для задания пространства признаков были использованы все слова из словаря коллекции (далее – «базовый эксперимент»);
- для случая, когда были использованы только существительные и прилагательные, не попавшие в расширенный список стоп-слов (далее – «улучшенный эксперимент»).

Для каждого из двух случаев было поставлено по 20 экспериментов на каждой из трех тестовых коллекций. Результатом каждого из 20 экспериментов явилось среднее время работы алгоритма кластеризации по результатам 1000 запусков. В каждом блоке результатов из 20 экспериментов выделялось минимальное, максимальное и среднее значения времени. Эксперимент показал, что для каждой из трех тестовых коллекций во время «улучшенного эксперимента» наблюдается сокращение временных затрат не менее чем на 28% по сравнению с «базовым экспериментом». При этом сравнивались лучшие, худшие и средние значения, полученные по результатам 20 экспериментов, для обоих случаев. Теоретические выводы и экспериментальные результаты показывают, что использование для представления пространства признаков только существительных и прилагательных из словаря коллекции, не попавших в расширенный список стоп-слов, приводит к повышению качества работы алгоритма k -means и одновременно позволяет сократить время работы алгоритма.

Заключение

Для алгоритма k -means было выявлено преимущество описания пространства признаков при помощи существительных и прилагательных (за исключением слов, входящих в расширенный список стоп-слов) вместо всего словаря коллекции. Показано, что в этом случае одновременно достигается улучшение качества кластеризации и снижение временных затрат. Преимуществом k -means по сравнению с другими алгоритмами является вычислительная простота, благодаря которой возможно его применение на достаточно больших массивах данных в режиме реального времени, в то время как использование других алгоритмов оказывается затруднительным.

Показано, что алгоритм иерархической кластеризации, определяющий расстояние между кластерами как среднее расстояние между каждой парой объектов в разных кластерах, дает высокие результаты по сравнению с k -means, поэтому на небольших массивах данных использование метода межгруппового среднего является более выгодным. Применение для описания пространства признаков только существительных и прилагательных вместо всего словаря коллекции в случае иерархической кластеризации приводит к ухудшению качества работы, вследствие чего не является оправданным. Полученные результаты экспериментов дополняют имеющиеся данные о качестве работы алгоритмов иерархической кластеризации коротких текстов узкой тематики [11]. Показано, что метод «межгрупповое среднее» иерархической кластеризации дает высокие показатели по сравнению с рядом других алгоритмов, используемых в области [7, 8, 11–14].

В лучшем случае k -means показывает достаточно хорошие результаты для выбранных коллекций. Напомним, что качество работы данного алгоритма в каждом из экспериментов определялось выбором исходных центроидов. Можно ввести предположение о приемлемости использования алгоритма k -means для решения задачи кластеризации узкотематических текстов короткой длины при условии выработки стратегии поиска исходных центроидов. Такая стратегия должна позволить значительно повысить результат работы алгоритма в среднем и приблизить его к лучшим результатам. Тем не менее, выработка такой стратегии (особенно за линейное время) является отдельной нетривиальной задачей и требует специального исследования, на котором мы планируем сосредоточиться в следующей работе.

Литература

1. Bernardini A., Carpineto C., D'Amico M. Full-Subtopic Retrieval with Keyphrase-Based Search Results Clustering // Proc. of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology. IEEE Computer Society, 2009. V. 1. P. 206–213.
2. Zhang D., Dong Y. Semantic, Hierarchical, Online Clustering of Web Search Results // Proc. of the 6th Asia-Pacific Web Conference (APWeb 2004). Lecture Notes in Computer Science. 2004. V. 3007. P. 69–78.
3. Zeng H.-J., He Q.-C., Chen Z., Ma W.-Y., Ma J. Learning to cluster web search results // Proc. of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '04). NY: ACM Press, 2004. P. 210–217.

4. Gutwin C., Paynter G., Witten I., Nevill-Manning C., Frank E. Improving browsing in digital libraries with keyphrase indexes // *J. Decision Support Systems*. 1999. V. 27. N 1-2. P. 81–104.
5. Popova S., Khodyrev I., Egorov A., Logvin S., Gulyaev S., Karpova M., Muromtsev D. Sci-Search: Academic Search and Analysis System Based on Keyphrases // *Proc. of the 4th Conference on Knowledge Engineering and Semantic Web (KESW 2013)*. Communications in Computer and Information Science series. 2013. V. 394. P. 281–288.
6. Alexandrov M., Gelbukh A., Rosso P. An Approach to Clustering Abstracts // *Proc. of the 10th International Conference NLDB-05. Lecture Notes in Computer Science*. 2005. V. 3513. P. 8–13.
7. Cagnina L., Errecalde M., Ingaramo D., Rosso P. A discrete particle swarm optimizer for clustering short text corpora // *Proc. of the 3rd International Conference on Bioinspired Optimization Methods and their Applications (BIOMA08)*. Ljubljana, Slovenia, 2008. P. 93–103.
8. Errecalde M., Ingaramo D., Rosso P. ITSA: An Effective Iterative Method for Short-Text Clustering Tasks // *Proc. 23rd International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE 2010)*. Lecture Notes in Artificial Intelligence. 2010. V. 6096. P. 550–559.
9. Ramírez-de-la-Rosa G., Montes-y-Gómez M., Solorio T., Villaseñor-Pineda L. A document is known by the company it keeps: neighborhood consensus for short text categorization // *Lang Resources and Evaluation*. 2012. V. 47. P. 127–149.
10. Romero F.P., Julián-Iranzo P., Soto A., Ferreira-Satler M., Gallardo-Casero J. Classifying unlabeled short texts using a fuzzy declarative approach // *Lang Resources and Evaluation*. 2013. V. 47. P. 151–178.
11. Pinto D. Analysis of narrow-domain short texts clustering // Research report for «Diploma de Estudios Avanzados (DEA)». Department of Information Systems and Computation. UPV. 2007 [Электронный ресурс]. Режим доступа: <http://users.dsic.upv.es/~proso/resources/PintoDEA.pdf>, свободный. Яз. англ. (дата обращения 23.12.2013).
12. Pinto D., Rosso P., Jiménez H. A Self-Enriching Methodology for Clustering Narrow Domain Short Texts // *Computer Journal*. 2011. V. 54. N 7. P. 1148–1165.
13. Pinto D., Jimenez-Salazar H., Rosso P. Clustering abstracts of scientific texts using the transition point technique // *Proc. of the 7th International Conference CICLING 2006. Lecture Notes in Computer Science*. 2006. V. 3878. P. 536–546.
14. Errecalde M., Ingaramo D., Rosso P. A new AntTree-based algorithm for clustering short-text corpora // *J. Computer Sci. Technol.* V. 10. N 1. P. 1–7.
15. Stein B., Meyer zu Eissen S., Potthast M. Syntax versus Semantics: Analysis of Enriched Vector Space Models // *Third International Workshop on Text-Based Information Retrieval (TIR 06)*/ Eds B. Stein, O. Kao. Trento, Italy: University of Trento, 2006. P. 47–52.
16. Meyer zu Eissen S., Stein B., Potthast M. The Suffix Tree Document Model Revisited // *Proc. of the 5th International Conference on Knowledge Management (I-KNOW 05)*. Graz, Austria, 2005. P. 596–603.
17. You W., Fontaine D., Barhes J.-P. An automatic keyphrase extraction system for scientific documents // *Knowledge and Information Systems*. 2013. V. 34. N 3. P. 691–724.
18. Hulth A. Improved automatic keyword extraction given more linguistic knowledge // *Proc. of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP'03)*. Stroudsburg, 2003. P. 216–223.
19. Mihalcea R., Tarau P. TextRank: Bringing order into texts // *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP '04)*. Stroudsburg, 2004. P. 404–411.
20. Wan X., Xiao J. Exploiting Neighborhood Knowledge for Single Document Summarization and Keyphrase Extraction // *ACM Transactions on Information Systems*. 2010. V. 28. N 2. Article 8.
21. Zesch T., Gurevych I. Approximate Matching for Evaluating Keyphrase Extraction // *Proc. of the International Conference on Recent Advances in Natural Language Processing (RANLP 2009)*. 2009. P. 484–489.
22. Popova S., Khodyrev I. Ranking in keyphrase extraction problem: is it useful to use statistics of words occurrences? // *RuSSUR 2013*. Режим доступа: http://romip.ru/russiras/doc/2013_for_participant/russirysc2013_submission_18_1.pdf, свободный. Яз. англ. (дата обращения 27.12.2013).
23. Попова С.В., Ходырев И.А. Извлечение и ранжирование ключевых фраз в задаче аннотирования // *Научно-технический вестник информационных технологий, механики и оптики*. 2013. № 1 (83). С. 81–85.
24. Popova S., Kovriguina L., Muromtsev D., Khodyrev I. Stop-words in Keyphrase Extraction Problem // *Proc. of 14th Conference of Open Innovations Association FRUCT*. Helsinki, Finland, 2013. P. 113–121.
25. Manning C., Raghavan P., Schütze H. Introduction to Information Retrieval. Cambridge University Press, 2009. 544 p.
26. Meyer zu Eissen S., Stein B. Analysis of Clustering Algorithms for Web-based Search // *Proc. of the 4th International Conference on Practical Aspects of Knowledge Management (PAKM 2002)*. Lecture Notes in Artificial Intelligence. 2002. V. 2569. P. 168–178.

27. Stein B., Meyer zu Eissen S., Wißbrock F. On Cluster Validity and the Information Need of Users // Proc. of the 3rd IASTED International Conference on Artificial Intelligence and Applications (AIA 03). Benalmádena, Spain, 2003. P. 216–221.
28. Tsatsaronis G., Varlamis I., Norvag K. SemanticRank: Ranking Keywords and Sentences Using Semantic Graphs // Proc. of the 23rd International Conference on Computational Linguistics (Coling'10). Beijing, China, 2010. P. 1074–1082.
29. Hasan K.S., Ng V. Conundrums in Unsupervised Keyphrase Extraction: Making Sense of the State-of-the-Art // Proc. of the 23rd International Conference on Computational Linguistics: Posters (Coling'10). Beijing, China, 2010. P. 365–373.
30. Ingaramo D., Errecalde M., Cagnina L., Rosso P. Particle Swarm Optimization for lustering short-text corpora // Computational Intelligence and Bioengineering/ Eds F. Masulli, A. Micheli, A.Sperduti. IOS Press, 2009. P. 3–19.
31. Azzag H., Monmarche N., Slimane M., Venturini G. AntTree: A new model for clustering with artificial ants // Proc. of the 2003 Congress on Evolutionary Computation (CEC '03). IEEE Press, 2003. V. 4. P. 2642–2647.
32. Stein B., Meyer zu Eissen S. Document Categorization with MAJORCLUST // Proc. of the 12th Workshop on Information Technology and Systems (WITS 02) / Eds A. Basu, S. Dutta. Barcelona, Spain: Technical University of Barcelona, 2002. P. 91–96.

- Попова Светлана Владимировна** – инженер, ст. преподаватель, Санкт-Петербургский государственный университет; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, svp@list.ru
- Данилова Вера Владимировна** – аспирант, Автономный университет Барселоны, Барселона, Испания, maolve@gmail.com
- Svetlana Popova** – engineer, senior lecturer, Saint Petersburg State University; Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, svp@list.ru
- Vera Danilova** – postgraduate, Autonomous University of Barcelona, Barcelona, Spain, maolve@gmail.com

УДК 004.925.4

ИСПОЛЬЗОВАНИЕ КОНТЕЙНЕРА BC7 ДЛЯ ХРАНЕНИЯ ТЕКСТУР С ГЛУБИНОЙ ЦВЕТА 10 БИТ

И.В. Перминов^a, Т.Т. Палташев^{a, b, c}

^a Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, i.am.perminov@gmail.com

^b Северо-Западный политехнический университет, Фримонт, Калифорния, США, timpal@mail.npu.edu

^c Advanced Micro Devices (AMD), Калифорния, США, timpal@mail.npu.edu

Изображения с высокой глубиной цвета обладают гораздо лучшими возможностями воспроизведения цветов и плавных градиентов, особенно при использовании устройств с широким цветовым охватом. В работе рассматривается проблема сжатия текстур с глубиной цвета 10 бит, применяемых в трехмерной компьютерной графике. Предложен метод хранения подобных текстур с использованием стандартного формата сжатия текстур BC7, рассмотрен тип блока BC7, с помощью которого можно закодировать цвета с точностью, превышающей 8 бит, и показаны необходимые изменения в аппаратуре декодера. Предложенный подход обладает обратной совместимостью с существующими декодерами. В ходе исследования была разработана программная реализация кодека на основе компрессора bc7_gru. Сравнение исходного и предлагаемого кодека показало уменьшение ошибок сжатия для изображений с высокой глубиной цвета.

Ключевые слова: сжатие текстур, BC7, block compression, глубина цвета, Direct3D.

USAGE OF BC7 CONTAINER FOR STORING TEXTURES WITH 10-BIT COLOR DEPTH

I. Perminov^d, T. Paltashev^{d, e, f}

^d Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, i.am.perminov@gmail.com

^e Northwestern Polytechnic University, Fremont, California, USA, timpal@mail.npu.edu

^f Advanced Micro Devices (AMD), California, USA, timpal@mail.npu.edu

High color depth images can more accurately reproduce colors and smooth color transitions without banding artifacts, especially for wide color gamut devices. The paper deals with texture compression with 10-bit color depth applied in 3D graphics. A method for storing of such textures in standard BC7 blocks is proposed. The example of BC7 block applicable for storing of smooth texture information with an accuracy exceeding 8-bit is also given. The proposed approach has a backward compatibility with current hardware. Additional hardware cost to support 10-bit decoding is expected to be low. An overview of

developed software for compressor based on bc7_gpu is given. Comparison of PSNR and RMSE shows that the proposed compressor provides better quality for 10-bit textures; compressing errors become less for images with high color depth.

Keywords: texture compression, BC7, block compression, color depth, Direct3D.

Введение

Стандартным методом представления изображений в компьютерных системах является TrueColor, подразумевающий использование цветовой модели RGB с глубиной цвета 8 бит, поэтому интенсивности каждого цветового канала (красного, зеленого и синего) могут принимать значения от 0 до 255. Хотя приложения могут использовать для хранения и обработки изображений прочие форматы с повышенной точностью, вывод на экран, как правило, осуществляется в формате TrueColor. При этом стандартным цветовым пространством является sRGB [1]. Цветовое пространство определяет, какой реальный цвет соответствует каждому конкретному числовому значению цвета. И хотя пространство sRGB покрывает всего около 35% всех видимых человеком цветов [2], глубины цвета TrueColor не хватает для отображения всех цветов даже внутри sRGB [3]. Наибольшие проблемы возникают с отображением плавных градиентов, так как в некоторых случаях явно видны границы между соседними цветами. Данная проблема усугубляется при использовании более широких цветовых пространств, например AdobeRGB, где доступные в каждом цветовом канале 256 уровней еще сильнее отдаляются друг от друга.

Исходя из этого, в профессиональных сферах применения, таких как работа с медицинскими снимками, кинопроизводство, графический дизайн, обработка фотографий, зачастую используются мониторы с поддержкой глубины цвета 10 бит и более и соответствующие форматы представления цвета. Полноценная поддержка таких изображений при работе с трехмерной графикой также является важной задачей.

Текстуры в трехмерной компьютерной графике

Использование различных видов текстур является неотъемлемой частью современной компьютерной графики. Это позволяет существенно улучшить визуальную детализацию трехмерных объектов без усложнения геометрии. В простейшем случае текстуры представляют собой двухмерное изображение, накладываемое на трехмерный объект. Отдельные точки растра, составляющие текстуру, принято называть текселями (texel – texture element). Однако применение большого количества текстур крупных размеров предъявляет высокие требования к видеопамяти.

В связи с этим широкое распространение получили различные технологии сжатия текстур [4–8]. При этом текстура передается и хранится в памяти в сжатом виде, а распаковывается непосредственно на графическом процессоре, как правило, между кэшами L1 и L2. Такой подход позволяет не только уменьшить занимаемый текстурами объем памяти, но и сэкономить пропускную способность, которая в большинстве случаев является куда более дефицитным ресурсом. Таким образом, использование текстурного сжатия позволяет существенно повысить производительность системы визуализации. Кроме этого, использование сжатия положительно сказывается на энергопотреблении за счет уменьшения объема используемой памяти и снижения трафика на интерфейсе «графический процессор» ↔ «видеопамять».

Универсальные алгоритмы сжатия без потерь (RLE, LZW, Deflate) и стандартные методы сжатия изображений плохо подходят для текстур, так как при отрисовке трехмерных объектов крайне важен эффективный доступ к произвольному участку текстуры. По этой причине большинство существующих форматов сжатия – это блочные кодеки с фиксированным уровнем сжатия. При этом изображение разбивается на множество блоков одинакового размера, которые сжимаются независимо друг от друга. Размер отдельного блока, как правило, составляет 4×4 текселя.

В большинстве случаев текстуры сжимаются в режиме «offline», так как алгоритмы, позволяющие снизить ошибки сжатия, обладают высокими вычислительными затратами. Но существуют и алгоритмы быстрого сжатия [9], однако сжатие динамически генерируемых текстур ограничивается применяемыми программными моделями работы с текстурами [10].

Современные видеоадаптеры и программные интерфейсы OpenGL [11] и Direct3D [12] уже достаточно давно обладают поддержкой текстур с большой глубиной цвета. Однако возможность сжатия таких текстур отсутствует. Исключения составляют текстуры, хранящие изображение с широким динамическим диапазоном, или HDR (High Dynamic Range). Отметим, что можно использовать существующие форматы для хранения изображений с глубиной цвета 10 бит.

Формат сжатия текстур BC7

Форматы сжатия текстур, используемые в Direct3D 10 и Direct3D 11, называются BlockCompression. Всего доступно семь различных форматов, начиная с BC1 и заканчивая BC7, а все семейство форматов часто обозначают как BCn. Во всех форматах сжатия BCn используется блок размером 4×4 текселя, и применяются схожие базовые принципы компрессии, которые проще всего продемонстрировать на примере блока BC1 (рис. 1).

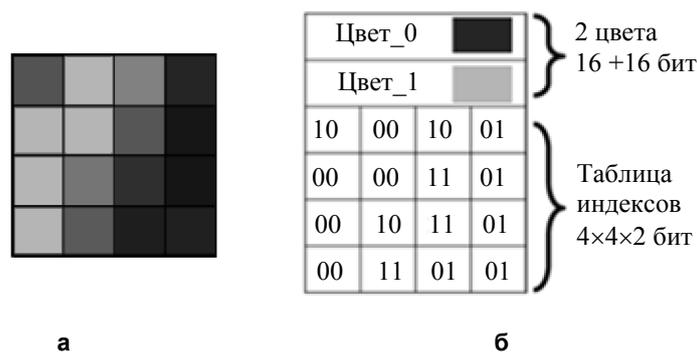


Рис. 1. Пример блока BC1: исходный блок 4×4 пикселя размером 512 бит (а); сжатый блок размером 64 бита (б)

В сжатом блоке (рис. 1, б) сохраняются 2 ключевых цвета в формате RGB565 (по 5 бит на красный и синий цветовые каналы, и 6 бит на зеленый канал) и таблица индексов, указывающая, в какой пропорции смешивать ключевые цвета при распаковке. Так как размер индекса составляет 2 бита, то внутри блока доступно только 4 различных цвета.

Формат BC7 [13, 14] появился в Direct3D 11. Блок BC7 занимает 128 бит и имеет гораздо более сложную структуру, чем BC1. Одним из основных отличий BC7 от BC1 является возможность сохранения до трех пар базовых цветов. Первые биты в блоке кодируют его тип. Всего существует восемь типов блоков, отличающихся набором и размером полей. В некоторых блоках, кроме цветовых каналов R, G и B, также сохраняется значение альфа-канала, что может, к примеру, использоваться для реализации эффекта полупрозрачности. Кроме этого, блоки отличаются размером таблицы индексов, количеством пар базовых цветов, точностью их представления и прочими полями. Размер индекса варьируется от 2 до 4 бит, а глубина цвета – от (4+1) бит до (7+1) бит. Подобная запись (7+1) используется потому, что дополнительный бит во всех цветовых каналах одной базовой точки обозначается как P и имеет одинаковое значение.

Распаковка сжатого блока осуществляется аппаратно. При этом, согласно требованиям Direct3D, результат работы аппаратного декодера BC7 должен побитово совпадать с результатом эталонного декодера. Упрощенно распаковка выглядит следующим образом: сначала осуществляется преобразование базовых цветов в формат с точностью представления 8 бит на канал (деквантование), после этого путем их смешивания с помощью линейной интерполяции формируется локальная палитра цветов, и индекс из таблицы используется для выборки из этой палитры.

Модификация этапов кодирования и декодирования

Идея предлагаемого подхода заключается в использовании 10-битного представления ключевых цветов и локальной палитры при распаковке, а также в модификации компрессора. При этом сам формат сжатого блока остается прежним.

Дело в том, что самые проблемные для 8-битной глубины места, а именно блоки с очень плавным изменением цвета, могут быть закодированы в BC7 с точностью, превышающей 8 бит. Для этого, к примеру, может использоваться тип блока BC7 Mode6 (рис. 2). В данном типе блока размер одного индекса составляет 4 бита (один бит в таблице индексов кодируется неявно). Фактически можно считать, что в данном блоке точность представления базовых цветов составляет 8 бит, а 4-битный индекс позволяет указать один из двух базовых цветов или любое из 14 промежуточных значений. Если значения базовых цветов достаточно близки, что как раз соответствует плавному градиенту, то промежуточные значения позволяют закодировать цвета, недоступные для глубины 8 бит. Если же декодер будет оперировать значениями с большей точностью, то подобные цвета смогут быть воспроизведены на выходе корректно. В связи с этим можно ввести новый режим работы декодера, в котором деквантование базовых цветов и последующие операции будут производиться с точностью 10 бит.

Аппаратная реализация дополнительного режима распаковки, выдающего на выходе значения с глубиной цвета 10 бит, не составит большого труда, так как логика работы декодера не изменится. В то же время сжатые текстуры остаются полностью совместимыми с обоими декодерами, так как формат сжатого блока остается прежним. Распаковка текстуры с глубиной 10 бит на «старом» декодере будет просто соответствовать понижению глубины цвета до 8 бит. И наоборот, текстуры с глубиной 8 бит при распаковке на предлагаемом декодере могут иметь лучшее качество по сравнению с обычной распаковкой и дальнейшим преобразованием до глубины цвета 10 бит за счет повышения точности представления промежуточных цветов в локальной палитре.

Естественно, для корректного сжатия текстур с глубиной цвета 10 бит необходима модификация компрессора, потому что существующие кодеры BC7 [15–17] принимают на вход изображения с точностью 8 бит.

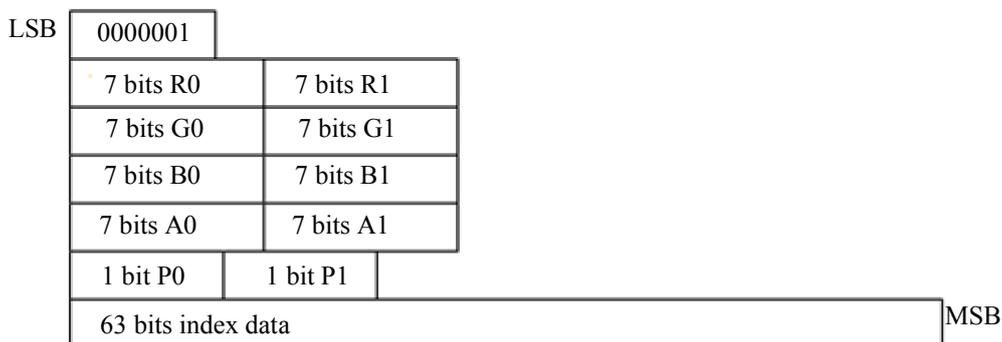


Рис. 2. Формат блока BC7 Mode6. LSB (Least Significant Bit) – младший значащий бит, MSB (Most Significant Bit) – старший значащий бит

Программная реализация

Для практической проверки состоятельности предложенного метода было произведено тестирование качества сжатия различных текстур с глубиной цвета 10 бит с использованием программного кодека. За основу был взят кодек bc7_gru [15], производящий сжатие текстуры в формат BC7 силами графического процессора с использованием технологий CUDA [18] и OpenCL [19].

Произведенные модификации заключались в изменении типов данных для возможности хранения цвета с повышенной точностью и модификации функций квантования/деквантования. Логика работы компрессора не изменялась.

Для тестирования качества сжатия были выбраны текстуры с глубиной цвета 10 бит и наличием плавных градиентов. При сжатии текстур с помощью оригинального компрессора производилась предварительная конвертация исходного изображения в формат с представлением цвета TrueColor. После распаковки с использованием оригинального декодера производилось преобразование в формат с точностью 10 бит. Для этого значения в каждом цветовом канале пересчитывались согласно формуле (1), где c – интенсивность цветового канала. Данная формула соответствует сдвигу на два разряда влево и копированию двух старших бит в высвободившиеся младшие разряды. Это стандартная целочисленная аппроксимация для деквантования, используемая, в том числе и в аппаратных блоках графических процессоров:

$$c = (c \ll 2) | (c \gg 6). \tag{1}$$

Значения метрик ошибок сжатия RMSE (Root Mean Square Error, среднеквадратическое отклонение) и PSNR (Peak Signal to Noise Ratio, соотношение сигнал/шум) для тестовых текстур приведены в таблице. Ошибки сжатия для текстуры № 1 также продемонстрированы на рис. 3.

	Оригинальный кодек		Предлагаемый кодек	
	RMSE	PSNR	RMSE	PSNR
Текстура № 1	175,2	51,4	144,7	53,1
Текстура № 2	178,3	51,3	150,4	52,8
Текстура № 3	170,1	51,7	99,3	56,4
Текстура № 4	249,9	48,3	222,3	49,4

Таблица. Качество сжатия текстур с глубиной цвета 10 бит

Результаты тестирования показывают, что предлагаемая модификация ощутимо увеличивает качество сжатия текстур, имеющих плавные градиенты с глубиной цвета 10 бит.

Заключение

Таким образом, предложенный метод позволяет использовать уже существующие форматы сжатия для хранения текстур с высокой глубиной цвета. При аппаратной модификации декодера это позволит получить при использовании таких текстур все преимущества сжатия, включая повышение производительности, снижение энергопотребления и требований к подсистеме памяти. При этом текстуры остаются обратно совместимыми с существующими декодерами, которые смогут корректно распаковать блок с использованием глубины цвета 8 бит. Также стоит отметить, что последующая модификация логики компрессора для работы с глубиной цвета 10 бит позволит еще сильнее снизить ошибки сжатия.

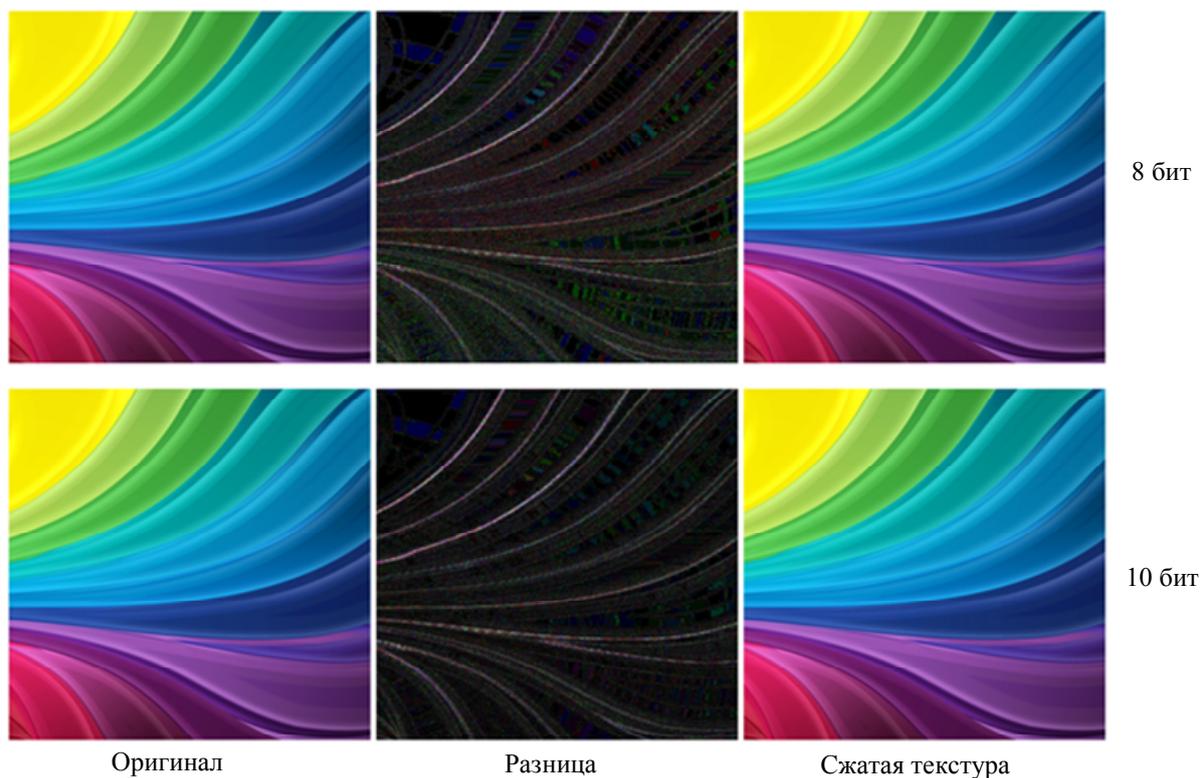


Рис. 3. Пример сжатия тестовой текстуры № 1, размер текстуры 512×512 пикселей

Литература

1. Stokes M., Anderson M., Chandrasekar S., Motta. R. A Standard Default Color Space for the Internet - sRGB [Электронный ресурс]. Режим доступа: <http://www.w3.org/Graphics/Color/sRGB.html>, свободный. Яз. англ. (дата обращения 23.11.2013).
2. sRGB VS. ADOBE RGB 1998 [Электронный ресурс]. Режим доступа: <http://www.cambridgeincolour.com/tutorials/sRGB-AdobeRGB1998.htm>, свободный. Яз. англ. (дата обращения 23.11.2013).
3. Is 8 bits enough? Of course not [Электронный ресурс]. Режим доступа: <http://19lights.com/wp/2011/09/30/is-8-bits-enough-of-course-not/>, свободный. Яз. англ. (дата обращения 23.11.2013).
4. Hong Z., Iourcha K.I., Nayak K.S. System and method for fixed-rate block-based image compression with inferred pixel values. Patent US 5956431 A; заявл. 02.10.97; опубл. 21.09.99.
5. Fenney S. Texture compression using low-frequency signal modulation // Proc. of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware. Eurographics Association, 2003. P. 84–91.
6. Strom J., Akenine-Moller T. iPACKMAN: High-Quality, Low-Complexity Texture Compression for Mobile Phones // Proc. of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware. Eurographics Association, 2005. P. 63–70.
7. Strom J., Pettersson M. ETC2: Texture Compression using Invalid Combinations // Proc. of the ACM SIGGRAPH/EUROGRAPHICS Symposium on Graphics Hardware. Eurographics Association, 2007. P. 49–54.
8. Nystad J., Lassen A., Pomianowski A., Ellis S., Olson T. Adaptive scalable texture compression // Proc. of the ACM SIGGRAPH/EUROGRAPHICS conference on High Performance Graphics. Eurographics Association, 2012. P. 105–114.
9. van Waveren J.M.P. Real-time DXT compression. Id Software, Inc., 2006. 42 p.
10. Перминов И.В. Повышение эффективности обработки динамически сжимаемых текстур // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 6 (88). С. 164–165.
11. OpenGL 4.4 Core Specification [Электронный ресурс]. Режим доступа: <http://www.opengl.org/registry/doc/glspec44.core.pdf>, свободный. Яз. англ. (дата обращения 23.11.2013).
12. Direct3D 11 Graphics [Электронный ресурс]. Режим доступа: <http://msdn.microsoft.com/enus/library/windows/desktop/ff476080.aspx>, свободный. Яз. англ. (дата обращения 23.11.2013).

13. BC7 Format Mode Reference [Электронный ресурс]. Режим доступа: <http://msdn.microsoft.com/en-us/library/windows/desktop/hh308954.aspx>, свободный. Яз. англ. (дата обращения 23.11.2013).
14. OpenGL A.R.B., ARB texture compression bptc. 2010 [Электронный ресурс]. Режим доступа: http://www.opengl.org/registry/specs/ARB/texture_compression_bptc.txt, свободный. Яз. англ. (дата обращения 23.11.2013).
15. Free high-quality BC7 GPU texture compressor [Электронный ресурс]. Режим доступа: <http://gpuscience.com/software/free-high-quality-bc7-gpu-texture-compressor/>, свободный. Яз. англ. (дата обращения 23.11.2013).
16. DirectXTex texture processing library [Электронный ресурс]. Режим доступа: <http://directxtex.codeplex.com/>, свободный. Яз. англ. (дата обращения 23.11.2013).
17. Krajcevski P., Lake A., Manocha D. FasTC: Accelerated Fixed-Rate Texture Encoding // Proc. of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D'13). NY: ACM, 2013. P. 137–144.
18. CUDA Parallel Computing Platform [Электронный ресурс]. Режим доступа: http://www.nvidia.com/object/cuda_home_new.html, свободный. Яз. англ. (дата обращения 03.12.2013).
19. OpenCL – The open standart for parallel programming of heterogeneous systems [Электронный ресурс]. Режим доступа: <http://www.khronos.org/opencl/>, свободный. Яз. англ. (дата обращения 03.12.2013).

- Перминов Илья Валентинович* – аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, i.am.perminov@gmail.com
- Палташев Тимур Турсунович* – доктор технических наук, профессор, зав. кафедрой; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия; профессор, Северо-Западный политехнический университет, Фримонт, Калифорния, США; начальник отдела моделирования и исследования графических архитектур, Advanced Micro Devices (AMD), timpal@mail.npu.edu

- Илья Perminov* postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, i.am.perminov@gmail.com
- Timour Paltashev* D.Sc., Professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia; Professor, Northwestern Polytechnic University, Fremont, California, USA; Senior Manager, Graphics Architecture Research and Modeling, Advanced Micro Devices (AMD), timpal@mail.npu.edu

УДК 621.397.3

ПРЕДШЕСТВУЮЩАЯ И ПОСЛЕДУЮЩАЯ ФИЛЬТРАЦИЯ ШУМОВ В АЛГОРИТМАХ ВОССТАНОВЛЕНИЯ ИЗОБРАЖЕНИЙ¹

В.С. Сизиков^{a, b}, Р.А. Экземпляров^b

^a Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Россия, Санкт-Петербург, sizikov2000@mail.ru

^b Санкт-Петербургский государственный политехнический университет, Санкт-Петербург, Россия, gehe@yandex.ru

Рассмотрено зашумление смазанных или (и) дефокусированных изображений. Определяется последовательность фильтрации шумов на таких изображениях – до устранения смазывания/дефокусирования или после него. Введены понятия предшествующей и последующей фильтрации шумов. Устранение смазывания/дефокусирования ряда изображений выполнено методами параметрической фильтрации Винера и регуляризации Тихонова, а фильтрация шумов – методами медианной фильтрации Тююки и адаптивной фильтрации Винера. На репрезентативных выборках проведена экспериментальная проверка, получены количественные оценки погрешностей восстановления изображений при различных типах шумов и очередностей их фильтрации. Показано, что методы параметрической фильтрации Винера и регуляризации Тихонова достаточно эффективно устраняют смазывание/дефокусирование, но недостаточно фильтруют шумы. Эффективность фильтрации шумов повышается при добавлении таких методов, как медианный фильтр Тююки, адаптивный фильтр Винера. При этом для импульсного шума важен порядок (очередность) его фильтрации (до или после устранения смазывания/дефокусирования в зависимости от помехо-сигнальной ситуации), а для гауссова шума порядок несущественен. Показано, что импульсный шум лучше фильтруется медианной, ранговой, адаптивной медианной фильтрацией, а гауссовый шум – адаптивной винеровской фильтрацией, среднеарифметическим фильтром. Дается объяснение этим эффектам.

Ключевые слова: зашумленное смазанное или дефокусированное изображение, предшествующая и последующая фильтрация шума, устранение смазывания и дефокусирования.

¹ Работа выполнена при поддержке РФФИ (грант № 13-08-00442).

PRELIMINARY AND SUBSEQUENT FILTERING OF NOISE IN IMAGE RESTORATION ALGORITHMS¹

V. Sizikov^{c, d}, R. Ekzemplyarov^d

^c Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, sizikov2000@mail.ru

^d Saint Petersburg State Polytechnic University, Saint Petersburg, Russia, rexe@yandex.ru

The paper deals with noising of smeared or/and defocused images. The sequence of the noise filtering on such images is stated before elimination of the smearing/defocusing or after it. The concepts of preliminary and subsequent filtering of noise are brought in. The smearing/defocusing of a series of images is eliminated by the methods of parametric Wiener filtering and Tikhonov regularization, while the noise is filtered out by the methods of median Tuckey filtering and adaptive Wiener filtering. These questions are considered by carrying out experimental testing with obtaining numerical estimates of image restoration errors for different noise types and sequence of their filtering on representative samples. The methods of parametric Wiener filtering and Tikhonov regularization are shown to eliminate the smearing/defocusing effectively enough but noise filtering is not effective. Effectiveness of noise filtering is raised by adding such methods as the median Tuckey filter, the adaptive Wiener filter. Furthermore, an order (sequence) of filtering is important for impulse noise (before or after elimination of the smearing/defocusing depending on noise-signal situation), and for Gaussian noise the order is unessential. It is shown that an impulse noise is filtered better by the median, order, adaptive median filtering, and the Gaussian noise - by adaptive Wiener filtering, average filtering. An explanation for these effects is given.

Keywords: noisy smeared or defocused image, preliminary and subsequent filtering of noise, elimination of smearing and defocusing.

Введение

В работе рассматриваются вопросы, связанные с зашумлением смазанных и (или) дефокусированных изображений. Такие вопросы уже анализировались в работах [1–9] и др. В настоящей работе акцент ставится на главном вопросе: каков наиболее эффективный порядок (очередность) фильтрации шума (до или после устранения смазывания/дефокусирования) [10, 11]. Рассматриваются также сопутствующие вопросы:

- в какой мере метод регуляризации Тихонова (МРТ), метод фильтрации Винера (МФВ), метод параметрической фильтрации Винера (МПФВ) и др. [2, 3, 5–9] не только устраняют смазывание или дефокусирование, но и фильтруют шумы;
- насколько улучшают фильтрацию шумов медианный фильтр Тьюки (МФТ), адаптивный фильтр Винера (АФВ) и др. [1, 3–9];
- какова эффективность перечисленных методов в зависимости от типа шума (импульсный, гауссовый и др.) [1, 3–5, 7, 11].

Введем понятия *пре-* и *пост-фильтрации* – предшествующей и последующей фильтрации шума, выполняемой до устранения смазывания/дефокусирования и после него соответственно (будем использовать также сокращения *pre* и *post*). Использование *пре-* и *пост-фильтрации* (при условии правильного выбора параметров фильтрации) должно уменьшить уровень случайной погрешности восстановления, что позволит уменьшить значение параметра регуляризации в МРТ и МПФВ, а это, в свою очередь, уменьшит систематическую (детерминированную) погрешность восстановления и в целом уменьшит среднеквадратическое отклонение (СКО) восстановленного изображения от точного.

Эти вопросы рассматриваются путем проведения экспериментальной проверки с получением количественных оценок погрешности восстановления изображений при различных типах шумов и очередностей их фильтрации на репрезентативных выборках.

Качественный анализ

Сначала обсудим вопрос о фильтрации шумов на качественном уровне. Если изображение смазано за счет (равномерного и прямолинейного) движения объекта (самолета, автомобиля [9, 12] и т.д.), а в среде между объектом и неподвижной камерой был шум (туман, пыль, капли, снежинки и т.п.), то на изображении объект будет смазанным, а шум – неподвижным (см. дальше рис. 1, б). В этом случае представляется «естественным и очевидным» выполнить сначала *предшествующую фильтрацию шума* (посредством МФТ, АФВ и др.), а затем устранить смазывание (с помощью МПФВ, МРТ и др.).

Если изображение объекта дефокусировано вместе с шумом (например, наблюдение в микроскоп зашумленных бактерий, см. рис. 3, б), то кажется «естественным» сначала устранить дефокусирование, а затем выполнить *последующую фильтрацию шума*.

Если смазывание изображения объекта произошло из-за сдвига камеры (например, рассогласования вращения телескопа и небесной сферы), а между объектом и камерой имелся импульсный шум (пыль, капли влаги и т.п.), то будут смазанными и объект, и шум. В этом случае представляется «естественным» сначала устранить дефокусирование, а затем выполнить *последующую фильтрацию шума*.

¹ The work was done with supporting from the Russian Foundation for Basic Research (grant № 13-08-00442)

венным» сначала устранить смазывание объекта и шума, а затем выполнить последующую фильтрацию шума.

Если изображение объекта смазалось за счет сдвига камеры (или дефокусировалось), а импульсный (аппаратурный) шум возник в результате сбоя в работе сенсоров ПЗС-матрицы, то изображение объекта будет смазанным (или дефокусированным), а импульсный шум представлен в виде так называемых «битых пикселей» [9, С. 130, 173]. В этом случае «естественно» сначала выполнить *предшествующую фильтрацию шума*, а затем устранить смазывание (или дефокусирование) изображения объекта.

Могут иметь место и иные ситуации с зашумлением, смазыванием или (и) дефокусированием [5, 9], причем шум может быть различных типов (импульсный, гауссовый, равномерный и др.), изображение (а также шум) – полутоновым (серым) или цветным, а искажение – более сложным, чем смазывание или дефокусирование [5, 13, 14].

Заметим, что порядок (очередность) фильтрации шума в вышеописанных ситуациях может казаться «естественным и очевидным», однако нижеприведенные результаты говорят о том, что отмеченный порядок не всегда является оптимальным (ср. [11]).

Основные математические соотношения

Прямая задача моделирования искажений изображений и *обратная задача* восстановления изображений описываются соотношениями [3, 5–7, 9–14]:

$$\int_{-\infty}^{\infty} h(x-\xi)w_y(\xi)d\xi = g_y(x), \quad (1)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x-\xi, y-\eta)w(\xi, \eta)d\xi d\eta = g(x, y), \quad (2)$$

где h – функция рассеяния точки (ФРТ, PSF, аппаратная функция – АФ), обычно (но не всегда) пространственно-инвариантная (разностная), w и g – распределение интенсивности по неискаженному и искаженному изображениям соответственно. В (1) оси x и ξ направлены вдоль смаза, а y – перпендикулярно ему (y играет роль параметра).

Полагаем, что вместо точных g и h известны \tilde{g} и \tilde{h} такие, что $\|\tilde{g} - g\| \leq \delta$, $\|\tilde{h} - h\| \leq \zeta$, где δ и ζ – оценки погрешностей (шумов) g и h .

Если рассматривается обратная задача, то соотношения (1) и (2) являются интегральными уравнениями (ИУ) Фредгольма I рода. Их решение есть некорректная задача [15–17], поэтому мы будем использовать для их решения устойчивые методы – МРТ и МПФВ [5–7, 9, 12, 15–17], а решения другими методами (Люси–Ричардсона, «слепой» деконволюции, итераций и т.д.), полученные в работах [1, 5, 13] и др., также учтем.

Решение ИУ типа свертки (1) *методом одномерного преобразования Фурье (ПФ) с регуляризацией Тихонова* имеет вид (при каждом y) [6, 7, 9, 12, 16, 17]

$$w_\alpha(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_\alpha(\omega) e^{-i\omega\xi} d\omega, \quad (3)$$

где

$$W_\alpha(\omega) = \frac{H^*(\omega)G(\omega)}{|H(\omega)|^2 + \alpha Q(\omega)}, \quad H(\omega) = \int_{-\infty}^{\infty} h(x)e^{i\omega x} dx, \quad G(\omega) = \int_{-\infty}^{\infty} g(x)e^{i\omega x} dx \quad (4)$$

– спектры (преобразования Фурье), $\alpha > 0$ – параметр регуляризации, $Q(\omega) = \omega^{2p}$, $p \geq 0$ – порядок регуляризации.

В работах [2, 6, 7, 18] предложен и развит другой метод решения ИУ (1), объединяющий МРТ и МФВ, который мы запишем в виде

$$\hat{w}_\alpha(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{W}_\alpha(\omega) e^{-i\omega\xi} d\omega, \quad \hat{W}_\alpha(\omega) = \frac{H^*(\omega)G(\omega)}{|H(\omega)|^2 + \alpha[1 + \theta|\hat{W}_\alpha(\omega)|^2] \cdot Q(\omega)}, \quad (5)$$

где $\theta = \zeta^2/\delta^2$. Такая модификация учитывает характеристики шумов δ и ζ уже на этапе построения регуляризованного решения (а не только при выборе α). Спектр $\hat{W}(\omega)$ можно найти, решая нелинейное уравнение в (5) итеративно.

В данной работе использован также МПФВ, в котором [5, 9, 11]

$$w_k(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_k(\omega) e^{-i\omega\xi} d\omega, \quad W_k(\omega) = \frac{H^*(\omega)G(\omega)}{|H(\omega)|^2 + K}, \quad (6)$$

где $K \geq 0$ – параметр, дающий оценку NSR – отношению шум/сигнал по мощности.

ИУ (1) может быть решено также *методом квадратур с регуляризацией Тихонова* (МКРТ) [15–17, 19], согласно которому ИУ (1) при каждом значении y приводится к системе линейных алгебраических уравнений (СЛАУ): $Aw_y = g_y$, где A – матрица, w_y и g_y – вектора, а решение СЛАУ с помощью МКРТ имеет вид (при каждом y)

$$w_{\alpha y} = (\alpha I + A^T A)^{-1} A^T g_y, \quad (7)$$

где I – единичная матрица (или LL^T , где L – дискретный лапласиан [16, 17]).

Что касается задачи восстановления дефокусированных изображений, то она описывается двумерным ИУ (2) и ее решение $w_\alpha(\xi, \eta)$ *методом 2-мерного ПФ с регуляризацией Тихонова* есть обобщение формул (3)–(5) на 2-мерный случай [5–7, 9, 16, 17], а решение $w_K(\xi, \eta)$ с помощью МПФВ есть обобщение формулы (6) [5, 9].

В данной работе рассмотрены два типа шумов – биполярный импульсный (типа 'salt & pepper', согласно обозначениям MatLab [4]) и гауссовый, а также специальные фильтры для фильтрации шумов – медианный фильтр Тьюки (МФТ) и адаптивный фильтр Винера (АФВ) [1, 3–7]. Что касается других типов шумов (равномерный, мультипликативный) и методов их фильтрации (адаптивный медианный фильтр Гонсалеса, ранговый фильтр и др.), то результаты их использования в различных публикациях [4, 5] также учтены в работе.

Остановимся на вопросе оценки обработки изображений. Качественно оценивать искаженные и восстановленные изображения будем визуально (такой способ при всей его субъективности часто бывает весьма эффективен). А количественно степень искажения и результат восстановления изображения будем оценивать с помощью относительного СКО изображения \tilde{w} от точного изображения \bar{w} [9, 11, 12]:

$$\sigma_{\text{rel}} = \frac{\|\tilde{w} - \bar{w}\|_{L_2}}{\|\bar{w}\|_{L_2}} = \frac{\sqrt{\sum_{j=1}^m \sum_{i=1}^n (\tilde{w}_{ji} - \bar{w}_{ji})^2}}{\sqrt{\sum_{j=1}^m \sum_{i=1}^n \bar{w}_{ji}^2}}. \quad (8)$$

Под \tilde{w} будем подразумевать искаженное (смазанное, дефокусированное, зашумленное) изображение g , а также восстановленные изображения w_K , w_α и др.

Вместо (8) можно использовать выражения [1, 9, 11]: $\sigma_{\text{rel}} = \|\tilde{w} - \bar{w}\|_{L_1} / \|\bar{w}\|_{L_1}$, $\sigma = \|\tilde{w} - \bar{w}\|_{L_2}$, $\sigma = \|\tilde{w} - \bar{w}\|_{L_1}$, $\text{PSNR} = 10 \lg(\bar{w}_{\text{max}} \cdot m \cdot n / \|\tilde{w} - \bar{w}\|_{L_2}^2)$, а также коэффициент эффективности [6, 18] $E_\alpha = \|w_{\alpha_{\text{opt}}} - \bar{w}\| / \|w_\alpha - \bar{w}\|$ и др. Однако в данной работе мы будем пользоваться удобным выражением (8) – типа отношения шум/сигнал.

Заметим, что приведенные выражения для σ_{rel} , σ и E_α можно использовать лишь тогда, когда изображение \bar{w} известно, т.е. при обработке модельных изображений. При обработке же реальных изображений (когда истинное изображение \bar{w} неизвестно) следует использовать способы выбора параметра регуляризации α : обобщенный принцип невязки [15, 17, 19], критерий L-кривой [16, 20], метод перекрестной значимости [15], алгоритм, основанный на критерии оптимальности [6], и др. Однако в этих способах делается выбор α , но не всегда делается оценка погрешности решения при конечных δ и ζ (в некоторых способах делается асимптотическая оценка погрешности решения). В способе вычислительных экспериментов (другие названия – способ эталонных, или модельных, примеров, способ псевдообратного оператора) [17, С. 239, 246; 21] выполняется выбор α и оценка погрешности решения (без знания \bar{w}), однако требуется составление и обработка модельного изображения, «близкого» к искомому.

В данной работе мы намеренно не используем какой-либо способ выбора α , стремясь показать потенциальные возможности фильтрации шумов на смазанных/дефокусированных изображениях. В последующих публикациях будут использованы некоторые способы выбора α (а также K).

Результаты численных экспериментов

Ниже приведены результаты обработки изображений некоторых объектов.

Смазанное изображение и несмазанный шум. На рис. 1, а, приведено истинное (неискаженное) полутоновое изображение подвижного объекта (самолета) plane.bmp 510×640. На рис. 1, б, – изображение этого объекта, смазанное (величина смаза $\Delta = 20$ пкс, угол смаза $\theta = 0$) и зашумленное несмазанным импульсным шумом (доля зашумленных пикселей $d = 0,02$), при этом добавлено размытие краев (для уменьшения эффекта Гиббса [9]). Смазывание устранялось с помощью МПФВ при различных K (без размытия краев) согласно (6) и МКРТ (с размытием) при различных α согласно (7). Фильтрация шума

выполнялась посредством МФТ и АФВ с различными масками, причем фильтрация шума выполнялась как предшествующая (pre), так и последующая (post).

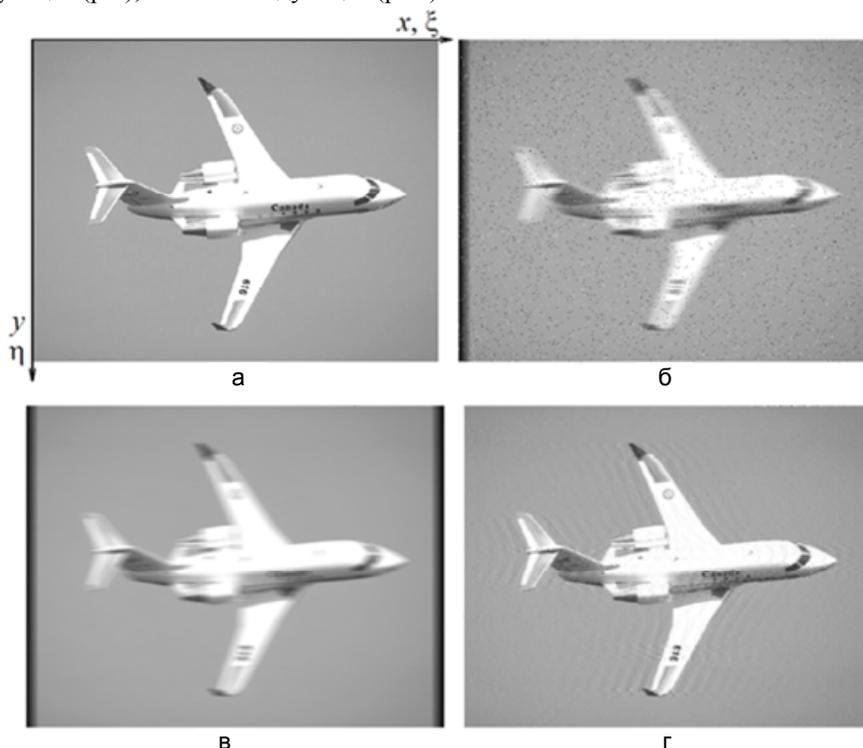


Рис. 1. Неискаженное (истинное) изображение самолета (а); изображение, смазанное с размытием краев и зашумленное несмазанным импульсным шумом (б); изображение после предшествующей МФТ с маской 3×3 ($\sigma = 0,1045$ без размытия краев и $\sigma = 0,0573$ с размытием) (в); изображение после устранения смазывания с помощью МКРТ ($\alpha = 10^{-2,5}$, $\sigma = 0,0352$) и МПФВ ($K = 10^{-2,5}$, $\sigma = 0,0353$) (г)

Следуя методике [6] (а также [18]), определялись характеристики случайной величины σ_{rel} согласно (8) путем формирования выборочной совокупности из значений $\sigma_{rel}^{(s)}(K)$ и $\sigma_{rel}^{(s)}(\alpha)$, где $s = 1, 2, \dots, S$ – номер реализации, а S – объем выборки ($S = 30$). Значения $\sigma_{rel}^{(s)}(K)$ и $\sigma_{rel}^{(s)}(\alpha)$ вычислялись по результатам восстановления изображения при $K = K_{opt}$ и $\alpha = \alpha_{opt}$. Изображение смазывалось при $\Delta = 20$. Затем на смазанное изображение накладывался несмазанный импульсный шум с долей зашумления $d \in [0, 1]$. Выполнялось S реализаций шума, причем все реализации имели одинаковые статистические характеристики – нулевое среднее и долю зашумления d (играющую роль СКО). Что касается зашумленности АФ h , то при решении прямой задачи искажения изображения величина смаза полагалась равной Δ , а при решении обратной задачи восстановления изображения смаз полагался равным $\tilde{\Delta}$, т.е. в АФ вносилась детерминированная (неслучайная) погрешность $\zeta = |\tilde{\Delta} - \Delta|$.

Вычислялись $\min \sigma_{rel}$, $\text{med } \sigma_{rel}$ и $\max \sigma_{rel}$ по выборочным совокупностям $\{\sigma_{rel}^{(s)}(K)\}$ и $\{\sigma_{rel}^{(s)}(\alpha)\}$ при разных d и $\tilde{\Delta}$. Результаты с маской 3×3 представлены в табл. 1.

Выполнялись также вычисления с масками 1×1 , 5×5 , 7×7 и 9×9 (табл. 2, 3).

Табл. 1 показывает:

1. при увеличении d (доли зашумленности изображения) погрешность σ растет;
2. при увеличении $\tilde{\Delta}$ (а значит, погрешности оператора ζ) погрешность σ монотонно растет при использовании МФТ, а при использовании АФВ погрешность σ ведет себя несистематически;
3. погрешность σ при использовании МФТ существенно меньше (в 2–3 раза), чем при АФВ.

На рис. 2 представлены полученные зависимости относительных погрешностей восстановления изображения $\sigma_{rel}(K) = \text{med } \sigma_{rel}(K)$ (под \tilde{w} в (8) подразумевается w_K согласно (6)) и $\sigma_{rel}(\alpha) = \text{med } \sigma_{rel}(\alpha)$ (под \tilde{w} подразумевается w_α согласно (7)) для случая, когда фильтрация импульсного шума выполнялась МФТ (с помощью m-функции medfilt2.m [4]). По минимумам кривых $\sigma_{rel}(K)$ и $\sigma_{rel}(\alpha)$ определялись K_{opt} и α_{opt} , а также $\sigma_{rel}(K_{opt})$ и $\sigma_{rel}(\alpha_{opt})$ – значения, при которых изображение восстанавливается с

наименьшей погрешностью согласно количественному критерию (хотя по визуальному критерию «наилучшему» восстановлению могут соответствовать несколько другие значения K_{opt} и α_{opt}).

Смазанное изображение plane.bmp, несмазанный импульсный шум				МПФВ		МКРТ	
				$\tilde{\Lambda} = \Delta = 20$	$\tilde{\Lambda} = 22$	$\tilde{\Lambda} = \Delta = 20$	$\tilde{\Lambda} = 22$
Медианный фильтр Тьюки	Предшествующая фильтрация шума	$d = 0,02$	min σ	0,0344	0,0421	0,0352	0,0502
			med σ	0,0352	0,0425	0,0359	0,0504
			max σ	0,0360	0,0430	0,0367	0,0506
	Последующая фильтрация шума	$d = 0,1$	min σ	0,0473	0,0506	0,0478	0,0573
			med σ	0,0490	0,0522	0,0506	0,0580
			max σ	0,0506	0,0538	0,0557	0,0593
Адаптивный фильтр Винера	Предшествующая фильтрация шума	$d = 0,02$	min σ	0,0726	0,0755	0,0793	0,0828
			med σ	0,0750	0,0768	0,0807	0,0834
			max σ	0,0769	0,0782	0,0819	0,0842
	Последующая фильтрация шума	$d = 0,1$	min σ	0,1430	0,1447	0,1478	0,1490
			med σ	0,1455	0,1456	0,1500	0,1504
			max σ	0,1480	0,1468	0,1540	0,1525
Адаптивный фильтр Винера	Предшествующая фильтрация шума	$d = 0,02$	min σ	0,1131	0,1220	0,1200	0,1250
			med σ	0,1145	0,1233	0,1221	0,1257
			max σ	0,1165	0,1246	0,1230	0,1266
	Последующая фильтрация шума	$d = 0,1$	min σ	0,1646	0,1627	0,1686	0,1677
			med σ	0,1668	0,1640	0,1701	0,1686
			max σ	0,1680	0,1659	0,1732	0,1699
Последующая фильтрация шума	$d = 0,02$	min σ	0,0997	0,0990	0,1083	0,1058	
		med σ	0,1016	0,1002	0,1103	0,1067	
		max σ	0,1027	0,1015	0,1122	0,1076	
Последующая фильтрация шума	$d = 0,1$	min σ	0,1556	0,1542	0,1572	0,1570	
		med σ	0,1570	0,1560	0,1590	0,1580	
		max σ	0,1586	0,1573	0,1625	0,1594	

Таблица 1. Значения $\min \sigma \equiv \min \sigma_{rel}$, $\text{med } \sigma \equiv \text{med } \sigma_{rel}$ и $\max \sigma \equiv \max \sigma_{rel}$

Как отмечено в [7], при уменьшении α (левые ветви графиков на рис. 2, а также на рис. 4, см. дальше) увеличивается случайная ошибка регуляризованного изображения w_α , но уменьшается систематическая ошибка (а также невязка $\|Aw_\alpha - g\|$), и, наоборот, при увеличении α (правые ветви графиков на рис. 2) уменьшается случайная ошибка (повышается гладкость решения w_α), но увеличивается систематическая ошибка (и невязка). Необходим такой алгоритм восстановления и такое (умеренное) α , чтобы как систематическая, так и случайная погрешности в восстановленном изображении w_α были близки к минимальным. В работе [7] изложен такой алгоритм (см. (5)) и способ выбора α , основанный на критерии оптимальности, а также комбинированный фильтр шума (который может содержать импульсный и низкоамплитудный шум). Это – последовательное соединение медианного фильтра (удаление импульсного шума) и модификации интервального фильтра скользящего среднего (удаление низкоамплитудного шума с сохранением контрастных деталей изображения).

Однако в настоящей работе выбран другой путь, а именно, при восстановлении изображения для выбора α (а также K) использован критерий минимума погрешности (8), а при фильтрации шумов использованы МФТ и АФВ, причем выполнены пре- и пост-фильтрации. В последующих публикациях будут использованы некоторые способы выбора α и комбинированный фильтр.

Рис. 2 показывает следующее:

- величина погрешности $\sigma_{rel}(K_{opt})$ и $\sigma_{rel}(\alpha_{opt})$ заметно зависит от очередности выполнения фильтрации шума (пре- или пост-фильтрация), а также от размера маски медианного фильтра;

- в данном примере пре-фильтрация импульсного шума при умеренной маске 3×3 ведет к погрешности σ , примерно в 2,1–2,3 раза меньшей, чем пост-фильтрация, что согласуется с вышеприведенным качественным анализом;
- МКРТ дает примерно такую же погрешность σ_{rel} , как и МПФВ.

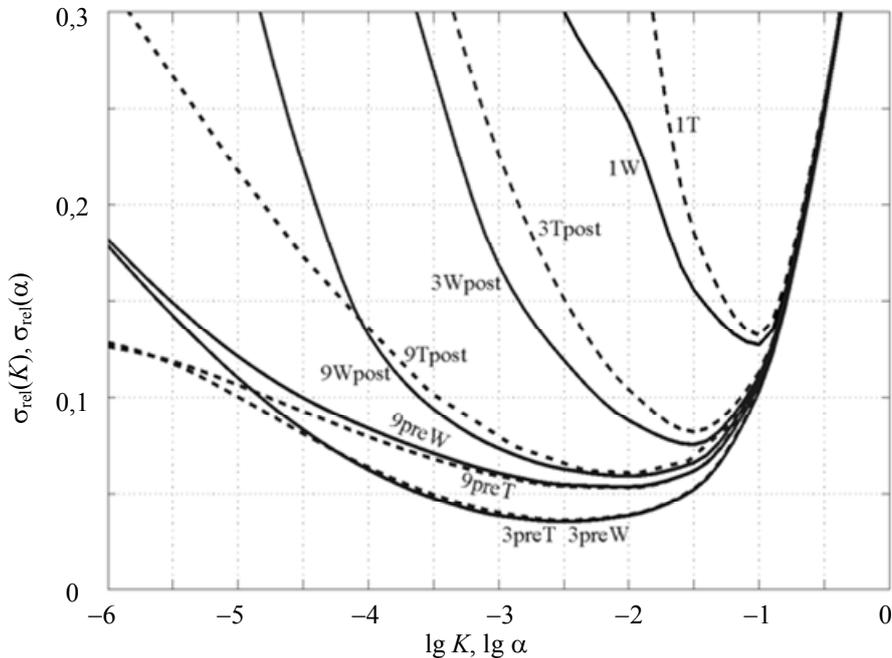


Рис. 2. Погрешность восстановления изображения самолета: $\sigma_{rel}(K)$ – с помощью МПФВ (W) и $\sigma_{rel}(\alpha)$ – с помощью МКРТ (Т) с использованием предшествующей (pre) и последующей (post) фильтрации импульсного шума посредством МФТ с масками 1×1, 3×3 и 9×9

При маске 1×1 медианная фильтрация, по существу, не выполняется, поэтому кривая 1W на рис. 2 аналогична кривым 1preW и 1Wpost, а 1T аналогична 1preT и 1Tpost.

На рис. 1, в, приведен результат пре-фильтрации шума посредством МФТ, а на рис. 1, г, – результат устранения смазывания с помощью МКРТ (МПФВ дает близкий результат). Рис. 1, г, показывает, что удалось не просто отфильтровать шум и устранить смазывание, но и восстановить опознавательные знаки на самолете. Данный пример был промоделирован также с использованием АФВ (с помощью m-функции wiener2.m [4]) применительно к фильтрации как импульсного, так и гауссова шума. В табл. 2 приведены результаты для импульсного шума, а в табл. 3 – для гауссова шума.

Смазанное изображение plane.bmp ($\Delta = 20, \theta = 0$), несмазанный импульсный шум ($d = 0,02$)		Маска 1×1		Маска 3×3		Маска 9×9	
		МПФВ	МКРТ	МПФВ	МКРТ	МПФВ	МКРТ
Медианный фильтр Тьюки	Пре-фильтрация шума	$K = 10^{-1}$ $\sigma = 0,1221$	$\alpha = 10^{-1}$ $\sigma = 0,1324$	$K = 10^{-2,5}$ $\sigma = 0,0353$	$\alpha = 10^{-2,5}$ $\sigma = 0,0352$	$K = 10^{-2,2}$ $\sigma = 0,0534$	$\alpha = 10^{-2,2}$ $\sigma = 0,0529$
	Пост-фильтрация шума	$K = 10^{-1}$ $\sigma = 0,1273$	$\alpha = 10^{-1}$ $\sigma = 0,1322$	$K = 10^{-1,5}$ $\sigma = 0,0752$	$\alpha = 10^{-1,5}$ $\sigma = 0,0814$	$K = 10^{-2}$ $\sigma = 0,0590$	$\alpha = 10^{-2}$ $\sigma = 0,0604$
Адаптивный фильтр Винера	Пре-фильтрация шума	$K = 10^{-1}$ $\sigma = 0,1273$	$\alpha = 10^{-1}$ $\sigma = 0,1322$	$K = 10^{-1,1}$ $\sigma = 0,1166$	$\alpha = 10^{-1,1}$ $\sigma = 0,1221$	$K = 10^{-1,5}$ $\sigma = 0,0843$	$\alpha = 10^{-1,5}$ $\sigma = 0,0932$
	Пост-фильтрация шума	$K = 10^{-1}$ $\sigma = 0,1273$	$\alpha = 10^{-1}$ $\sigma = 0,1322$	$K = 10^{-1,3}$ $\sigma = 0,1014$	$\alpha = 10^{-1}$ $\sigma = 0,1204$	$K = 10^{-1,5}$ $\sigma = 0,0783$	$\alpha = 10^{-1,4}$ $\sigma = 0,0841$

Таблица 2. Значения $K = K_{opt}$, $\sigma = \sigma_{rel}(K_{opt})$, $\alpha = \alpha_{opt}$ и $\sigma = \sigma_{rel}(\alpha_{opt})$

Табл. 2 показывает, что МКРТ сам не только устраняет смазывание изображения, но и подавляет шум ($\alpha = 10^{-1}$, $\sigma_{rel} = 0,1322$ при маске 1×1), однако делает это лишь частично (МПФВ дает близкий результат). Лучше результат получается, когда сначала выполняется предшествующий МФТ, причем с большим размером маски, а затем устраняется смазывание ($\alpha = K = 10^{-2,5}$, $\sigma_{rel} = 0,0352$ при маске 3×3). Табл. 2 показывает также, что если шум импульсный, то МФТ фильтрует его заметно лучше, чем АФВ (а

также среднеарифметический фильтр): погрешность σ примерно в 3,4 раза меньше в случае пре-фильтрации и маски 3×3 и в 1,5–1,7 раза меньше для маски 9×9.

В табл. 3 приведены аналогичные результаты для случая гауссова 1%-го шума. Табл. 3 показывает, что погрешность σ фильтрации гауссова шума посредством АФВ чуть меньше (примерно в 1,05–1,4 раза в данном примере), чем МФТ.

Смазанное изображение plane.bmp ($\Delta = 20, \theta = 0$), несмазанный 1%-ый гауссовый шум		Маска 1×1		Маска 3×3		Маска 9×9	
		МПФВ	МКРТ	МПФВ	МКРТ	МПФВ	МКРТ
Медианный фильтр Тююки	Пре-фильтрация шума	$K = 10^{-2}$ $\sigma = 0,0445$	$\alpha = 10^{-1,9}$ $\sigma = 0,0464$	$K = 10^{-2,2}$ $\sigma = 0,0405$	$\alpha = 10^{-2,1}$ $\sigma = 0,0410$	$K = 10^{-2,3}$ $\sigma = 0,0542$	$\alpha = 10^{-2,3}$ $\sigma = 0,0530$
	Пост-фильтрация шума	$K = 10^{-2}$ $\sigma = 0,0445$	$\alpha = 10^{-1,9}$ $\sigma = 0,0464$	$K = 10^{-2,3}$ $\sigma = 0,0382$	$\alpha = 10^{-2,2}$ $\sigma = 0,0387$	$K = 10^{-2,6}$ $\sigma = 0,0509$	$\alpha = 10^{-2,5}$ $\sigma = 0,0516$
Адаптивный фильтр Винера	Пре-фильтрация шума	$K = 10^{-2}$ $\sigma = 0,0445$	$\alpha = 10^{-1,9}$ $\sigma = 0,0464$	$K = 10^{-2,4}$ $\sigma = 0,0387$	$\alpha = 10^{-2,3}$ $\sigma = 0,0380$	$K = 10^{-2,4}$ $\sigma = 0,0462$	$\alpha = 10^{-2,2}$ $\sigma = 0,0473$
	Пост-фильтрация шума	$K = 10^{-2}$ $\sigma = 0,0445$	$\alpha = 10^{-1,9}$ $\sigma = 0,0464$	$K = 10^{-2,3}$ $\sigma = 0,0355$	$\alpha = 10^{-2,3}$ $\sigma = 0,0356$	$K = 10^{-2,5}$ $\sigma = 0,0358$	$\alpha = 10^{-2,5}$ $\sigma = 0,0369$

Таблица 3. Значения $K = K_{opt}$, $\sigma = \sigma_{rel}(K_{opt})$, $\alpha = \alpha_{opt}$ и $\sigma = \sigma_{rel}(\alpha_{opt})$

Однако табл. 2, 3 демонстрируют и непредсказуемый результат: если шум – несмазанный импульсный и используется АФВ (см. табл. 2), то пост-фильтрация шума дает практически такую же погрешность σ (лишь в 1,1 раза меньшую для маски 3×3 в данном примере), как и пре-фильтрация. Такой же эффект имеет место и в случае несмазанного гауссова шума и использования как МФТ, так и АФВ (погрешность пост-фильтрации лишь в 1,05–1,1 раза меньше, чем погрешность пре-фильтрации в случае маски 3×3, см. табл. 3). Это можно объяснить различием свойств импульсного и гауссова шумов, а также тем, что АФВ – линейный фильтр (для него порядок фильтрации неважен), а МФТ – нелинейный.

Дефокусированное изображение и дефокусированный шум. На рис. 3, а, представлено искаженное изображение бактерий, наблюдаемых в микроскоп.

На рис. 3, б, показано дефокусированное изображение (ФРТ – в виде однородного диска радиуса $r = 5$ пкс), зашумленное импульсным шумом при $d = 0,02$ (шум также дефокусирован и поэтому на рис. 3, б, невидим). Было добавлено размытие краев (темная рамка). Далее устранено дефокусирование с помощью метода двумерного ПФ с регуляризацией Тихонова (обобщение формул (3), (4)), а также МПФВ (обобщение формулы (6)) для ряда значений параметров α и K и затем при каждом α и K выполнена пост-фильтрация шума посредством МФТ с масками 1×1, 3×3, 9×9.

На рис. 4 представлены полученные зависимости $\sigma_{rel}(K)$ и $\sigma_{rel}(\alpha)$ – относительные погрешности восстановления изображения бактерий.

По минимумам этих зависимостей определены оптимальные значения $K = K_{opt}$, $\sigma = \sigma_{rel}(K_{opt})$, $\alpha = \alpha_{opt}$ и $\sigma = \sigma_{rel}(\alpha_{opt})$: при маске 1×1 имеем $K = 10^{-2}$, $\sigma = 0,0567$, $\alpha = 10^{-1,3}$, $\sigma = 0,0771$; при маске 3×3 имеем $K = 10^{-3,8}$, $\sigma = 0,0458$, $\alpha = 10^{-8}$, $\sigma = 0,0117$; при маске 9×9 имеем $K = 10^{-3,6}$, $\sigma = 0,0763$, $\alpha = 10^{-8}$, $\sigma = 0,0696$.

На рис. 3, в, дано изображение, восстановленное методом двумерного ПФ с регуляризацией Тихонова при $\alpha = 10^{-8}$, $\sigma = 0,1479$, а на рис. 3, г, – изображение, полученное пост-фильтрацией импульсного шума посредством МФТ с маской 3×3 ($\sigma = 0,0117$). Восстановление изображения с помощью МПФВ хуже: погрешность $\sigma = 0,0458$, т.е. в 3,9 раза больше, чем методом Тихонова.

Смазанное изображение и смазанный шум. Было обработано также изображение астрономического объекта – галактики М83 (файл m83.jpg 378×400×3, цветное изображение, переведенное в полутоновое 378×400). Особенностью данного изображения является наличие мелких деталей – звезд, что предъявляет повышенные требования к обработке изображения. Изображение было смазано и зашумлено смазанным импульсным шумом (смаз $\Delta = 10$ пкс, угол смаза $\theta = 31^\circ$, доля зашумления $d = 0,02$). Такое смазывание может возникнуть в случае рассогласования вращения телескопа и небесной сферы при длительной экспозиции, а зашумление может возникнуть из-за наличия пыли и капель в атмосфере Земли. Звезды и импульсный шум превратились в штрихи, и по ним представилась возможность оценить параметры смазывания Δ и θ , а значит, ФРТ h .

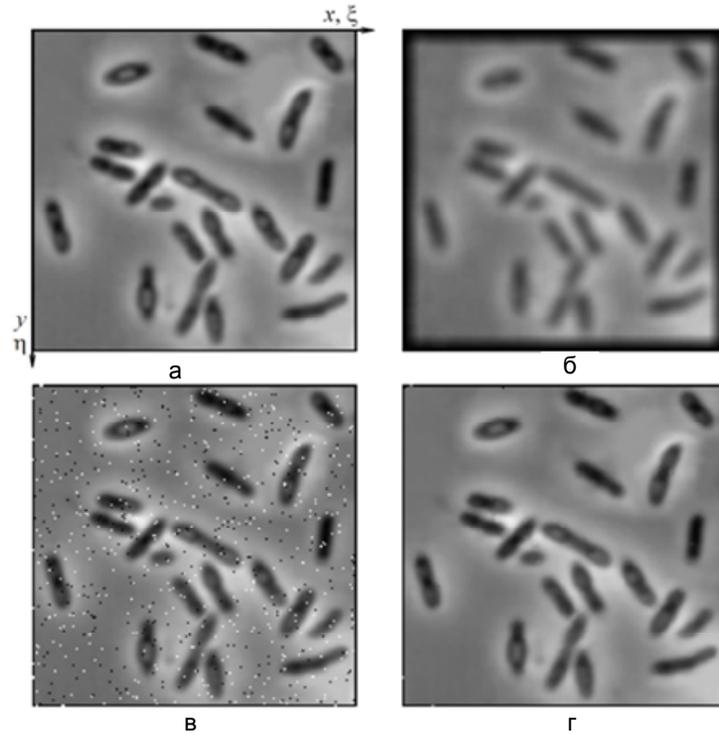


Рис. 3. Неискаженное изображение бактерий (а); изображение, зашумленное импульсным шумом и дефокусированное вместе с шумом (б); изображение (б), рефокусированное методом 2-мерного ПФ с регуляризацией Тихонова (в); изображение (в) после последующей МФТ с маской 3×3 ($\sigma = 0,0117$) (г)

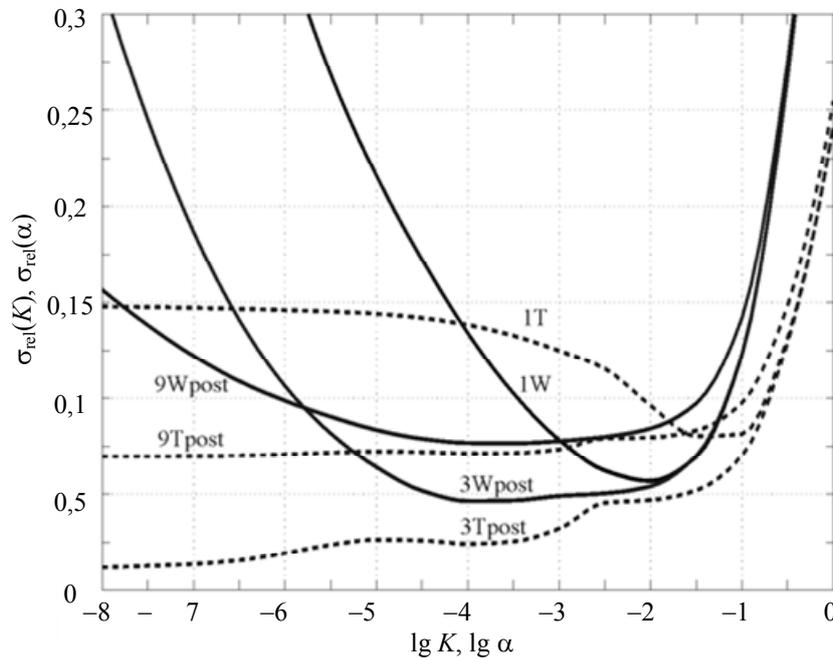


Рис. 4. Погрешность восстановления искаженного изображения бактерий: $\sigma_{rel}(K)$ – методом ПФВ (W) и $\sigma_{rel}(\alpha)$ – методом КРТ (Т) с использованием последующей (post) фильтрации импульсного шума посредством МФТ с масками 1×1, 3×3 и 9×9

Для устранения смазывания изображения был использован МПФВ, а также МКРТ, и при каждом значении параметров K и α выполнена пост-фильтрация посредством МФТ с маской 1×1, 3×3, 9×9 и др. По минимумам кривых $\sigma_{rel}(K)$ и $\sigma_{rel}(\alpha)$ получены значения K_{opt} , $\sigma_{rel}(K_{opt})$, α_{opt} и $\sigma_{rel}(\alpha_{opt})$. При K_{opt} и α_{opt} восстановились галактика, звезды и импульсы: с маской 3×3 $\sigma_{rel}(K_{opt}) = 0,106$, $\sigma_{rel}(\alpha_{opt}) = 0,108$.

Для репрезентативности исследования были обработаны также другие изображения: полутонные girl.jpg, cameraman.tif, pout.tif (портретные изображения), moon.tif (астрономический объект), coins.png, eight.tif (монеты), mrt-1-02.bmp (томограмма), textp.bmp, textn.bmp (тексты), а также цветные flower.bmp, pears.png, peppers.png, onion.png, gantrycrane.png, football.jpg, calculator.png (предметы), saturn.png (астрономический объект), kids.tif (портрет), auto.jpg (подвижный объект), board.tif (микросхема) – модельные изображения (фантомы), а также pair.jpg, rehoitus.jpg – реальные (натурные) изображения [9–12, 22].

Выводы

На основе полученных результатов, а также результатов различных публикаций ([1, 3–14, 18, 20, 22] и др.) можно сделать следующие выводы в отношении задачи фильтрации шумов на искаженных (смазанных или дефокусированных) изображениях.

1. Если устранять смазывание/дефокусирование, а также фильтровать шум на изображении лишь каким-то одним методом (параметрической фильтрации Винера, регуляризации Тихонова, максимального правдоподобия Люси–Ричардсона, «слепой» деконволюции и т.п. [1, 5, 6–9]), то этого может оказаться недостаточно для получения наилучшего результата (см. рис. 3, в, а также рис. 2 и 4 и табл. 2 и 3 с маской 1×1). Более точная обработка может получиться, если указанные методы использовать в основном для устранения смазывания/дефокусирования, а для фильтрации шума дополнительно использовать такие специальные методы, как медианный фильтр Тьюки, адаптивный фильтр Винера, адаптивный медианный фильтр Гонсалеса, комбинированный фильтр Воскобойникова, среднеарифметический фильтр, ранговая фильтрация и т.п. [1, 3–5, 7, 8] (см. рис. 1, г, и рис. 3, г).
2. Погрешность обработки σ зависит от очередности фильтрации шума на смазанном/дефокусированном изображении – до устранения смазывания/дефокусирования (пре-фильтрация) или после него (пост-фильтрация). Например, если на смазанное изображение наложен несмазанный импульсный шум (рис. 1, б), то нужно сначала выполнить пре-фильтрацию шума медианным фильтром (рис. 1, в), а потом устранить смазывание (рис. 1, г). Если же эти операции поменять местами (смазывание – пост-фильтрация), то это может привести к заметному увеличению погрешности σ (см. табл. 2, медианный фильтр, маска 3×3 , увеличение в 2,1–2,3 раза).
3. Если импульсный несмазанный шум фильтровать адаптивным фильтром Винера, то пост-фильтрация шума дает чуть меньшую погрешность σ , чем пре-фильтрация (табл. 2). Такой же результат получается для гауссова несмазанного шума (для любого фильтра, см. табл. 3). Это – несколько неожиданный результат, не согласующийся с вышеприведенным качественным анализом. Он связан, во-первых, с различием свойств импульсного и гауссова шумов: биполярный импульсный шум типа 'salt & pepper' имеет интенсивности лишь 0 и 255 (импульсы), присутствует лишь в d -ой доле ($d \in [0, 1]$) точек изображения и замещает интенсивности на изображении, а гауссовый шум имеет различные интенсивности, присутствует в каждой точке изображения и добавляется к изображению. Во-вторых, медианная и винеровская адаптивная фильтрации имеют также различия: в медианной фильтрации практически исключаются импульсы, а в винеровской адаптивной фильтрации импульсы с большими значениями сильно влияют на результат фильтрации. В-третьих, медианная фильтрация – это нелинейная фильтрация, и для нее важен порядок операций, а винеровская адаптивная фильтрация – линейная фильтрация, и порядок несущественен.
4. Импульсный шум лучше фильтруется медианным фильтром Тьюки, чем адаптивным фильтром Винера или среднеарифметическим фильтром, а гауссовый шум, наоборот, лучше фильтруется адаптивным фильтром Винера, чем медианным фильтром (табл. 1–3).

Эти выводы подтвердились обработкой иных изображений: girl.jpg, cameraman.tif, pout.tif, moon.tif, coins.png, eight.tif, mrt-1-02.bmp, textp.bmp, textn.bmp, flower.bmp, pears.png, peppers.png, onion.png, gantrycrane.png, football.jpg, calculator.png, saturn.png, kids.tif, auto.jpg, board.tif, pair.jpg, rehoitus.jpg и др.

Заключение

В работе рассмотрена фильтрация шумов на смазанных/дефокусированных изображениях. Показано следующее. Методы параметрической фильтрации Винера и регуляризации Тихонова достаточно эффективно устраняют смазывание/дефокусирование, но недостаточно фильтруют шумы. Эффективность фильтрации шумов повышается, если добавить такие методы, как медианная фильтрация, адаптивная фильтрация Винера и т.п. При этом для импульсного шума важен порядок (очередность) его фильтрации (до или после устранения смазывания/дефокусирования в зависимости от помехо-сигнальной ситуации), а для гауссова шума порядок несущественен. Кроме того, показано, что импульсный шум лучше фильтруется медианной, ранговой, адаптивной медианной фильтрацией и т.п., а гауссовый шум – адаптивной винеровской фильтрацией, среднеарифметическим фильтром и т.п. Дается объяснение этим эффектам.

Авторы благодарят Ю.Е. Воскобойникова за полезные обсуждения.

Литература

1. Lim J.S. Two-Dimensional Signal and Image Processing. Englewood Cliffs, NJ: Prentice Hall PTR, 1990. 694 p.
2. Mesarović V.Z., Galatsanos N.P., Katsaggelos A.K. Regularized constrained total least squares image restoration // IEEE Trans. Image Processing. 1995. V. 4. N 8. P. 1096–1108.
3. Грузман И.С., Киричук В.С., Косых В.П., Перетягин Г.И., Спектор А.А. Цифровая обработка изображений в информационных системах. Новосибирск: Изд-во НГТУ, 2002. 352 с.
4. Дьяконов В., Абраменкова И. MATLAB. Обработка сигналов и изображений: Специальный справочник. СПб: Питер, 2002. 608 с.
5. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М.: Техносфера, 2006. 1072 с.
6. Воскобойников Ю.Е., Литасов В.А. Устойчивый алгоритм восстановления изображения при неточно заданной аппаратной функции // Автометрия. 2006. Т. 42. № 6. С. 3–15.
7. Воскобойников Ю.Е. Комбинированный нелинейный алгоритм восстановления контрастных изображений при неточно заданной аппаратной функции // Автометрия. 2007. Т. 43. № 6. С. 3–16.
8. Яне Б. Цифровая обработка изображений. М.: Техносфера, 2007. 584 с.
9. Сизиков В.С. Обратные прикладные задачи и MatLab. СПб: Лань, 2011. 256 с.
10. Экземпляров Р.А., Сизиков В.С. Устранение смазывания и дефокусирования изображений с предварительной фильтрацией шумов в рамках системы MatLab // Материалы XXXIX Международной конференции «Неделя науки СПбГПУ». СПб: Изд-во Политехн. ун-та, 2010. Ч. XIII. С. 223–225.
11. Сизиков В.С., Экземпляров Р.А. Последовательность операций при фильтрации шумов на искаженных изображениях // Оптический журнал. 2013. Т. 80. № 1. С. 39–48.
12. Petrov Yu.P., Sizikov V.S. Well-Posed, Ill-Posed, and Intermediate Problems with Applications. Leiden–Boston: VSP, 2005. 234 p.
13. Горшков А.В. Улучшение разрешения изображений при обработке данных физического эксперимента и нахождение неизвестной аппаратной функции по программам пакета REIMAGE // Приборы и техника эксперимента. 1995. № 2. С. 68–78.
14. Donatelli M., Estatico C., Martinelli A., Serra-Capizzano S. Improved image deblurring with anti-reflective boundary conditions and re-blurring // Inverse Problems. 2006. V. 22. N 6. P. 2035–2053.
15. Воскобойников Ю.Е., Преображенский Н.Г., Седельников А.И. Математическая обработка эксперимента в молекулярной газодинамике. Новосибирск: Наука, 1984. 240 с.
16. Engl H.W., Hanke M., Neubauer A. Regularization of Inverse Problems. Dordrecht: Kluwer, 1996. 328 p.
17. Верлань А.Ф., Сизиков В.С. Интегральные уравнения: методы, алгоритмы, программы. Киев: Наук. думка, 1986. 544 с.
18. Воскобойников Ю.Е., Литасов В.А. Регуляризирующий алгоритм непараметрической идентификации при неточных исходных данных // Научный вестник Новосибирского государственного технического университета. 2005. № 2 (20). С. 33–45.
19. Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г. Численные методы решения некорректных задач. М.: Наука, 1990. 232 с.
20. Hansen P.C. Discrete Inverse Problems: Insight and Algorithms. Philadelphia: SIAM, 2010. 213 p.
21. Верлань А.Ф., Сизиков В.С., Мосенцова Л.В. Метод вычислительных экспериментов для решения интегральных уравнений в обратной задаче спектроскопии // Электронное моделирование. 2011. Т. 33. № 2. С. 3–12.
22. Захаров Д.Д., Сизиков В.С., Шемплинер В.В., Щекотин Д.С. Новые способы устранения артефактов на томографических и иных изображениях // Научно-технический вестник СПбГУ ИТМО. 2006. № 9 (32). С. 138–143.

- | | |
|-------------------------------------|--|
| Сизиков Валерий Сергеевич | – доктор технических наук, профессор, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия; профессор, Санкт-Петербургский государственный политехнический университет, Санкт-Петербург, Россия, sizikov2000@mail.ru |
| Экземпляров Роман Алексеевич | – аспирант, Санкт-Петербургский государственный политехнический университет, Санкт-Петербург, Россия, rexe@yandex.ru |
| Valery Sizikov | – D.Sc., Professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia; Professor, Saint Petersburg State Polytechnic University Institute of International Educational Programs (SPb SPU IIEP), Saint Petersburg, Russia, sizikov2000@mail |
| Roman Ekzemplyarov | – postgraduate, Saint Petersburg, Russia, Saint Petersburg State Polytechnic University, Saint Petersburg, Russia, rexe@yandex.ru |

ОЦЕНКА ВОЗМОЖНОСТИ ЭКРАННОЙ РЕПРОДУКЦИИ НАСЫЩЕННЫХ ПИГМЕНТОВ

Л.В. Черевань^a, В.Т. Тозик^{a,b}

^aСанкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Россия, Санкт-Петербург, 4erevanb@mail.ru

^b Межотраслевой институт повышения квалификации, Россия, Санкт-Петербург

Рассматривается проблема возможности воспроизведения на экранах видеомониторов полного диапазона цветовых тонов классической масляной живописи. Показаны модели стандартных цветовых охватов современных мониторов. Модели визуализированы в цветовой координатной системе CIE Lab. Доказано, что чистые минеральные пигменты имеют большую хроматическую насыщенность по сравнению с масляной краской, исходя из этого, возможность воспроизведения насыщенных пигментов обеспечивает воспроизведение всего диапазона масляных красок. Методом спектрофотометрии получены 75 хроматических координат насыщенных пигментов. Показано положение измеренных координат относительно цветовых охватов стандартных мониторов. Разработана количественная методика оценки достоверности воспроизведения на экранах современных мониторов полного диапазона тонов классической масляной живописи. Методика включает расчет цветовых отличий ряда выходящих за пределы цветового охвата монитора цветовых тонов (исходной координаты и координаты, преобразованной в диапазон охвата), результаты рассмотрены относительно стандартизованных пороговых цветовых отличий. Доказано, что видеомонитор стандарта sRGB не позволяет воспроизводить полный диапазон тонов масляной живописи, а видеомонитор стандарта AdobeRGB воспроизводит их удовлетворительно (с отличием ниже порога различимости). Разработанная количественная методика может применяться для оценки любого монитора.

Ключевые слова: видеомониторы, цвет, цветовые координаты, MCO, CIE XYZ, CIE Lab, экранная репродукция.

RELIABILITY ESTIMATION FOR SCREEN REPRODUCTION OF SATURATED PIGMENTS

L. Cherevan^c, V. Tozik^{c,d}

^cSaint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, 4erevanb@mail.ru

^dIntersectoral Institute for Advanced Training, Saint Petersburg, Russia

The paper deals with the ability of modern display monitors to reproduce full color range of classical oil paintings. The models of the standard color gamuts of modern monitors are represented by visualization in CIE Lab color coordinate system. Pure mineral pigments are proved to be more saturated than oil colors. Therefore, reproduction ability of pure saturated pigments ensures the full range reproduction of oil colors. 75 color coordinates of pure pigments are obtained by means of the spectrophotometry method. The measured coordinates displacement is shown as compared with the color gamuts of standard monitors. The quantitative method of reliability estimation for the reproducing of full color ranges of classical oil paintings is developed. The method consists of color differences calculating for out-of-gamut colors (measuring that of the reference coordinate against the closest one inside the gamut) and comparing the results with the standard threshold color difference. It is proved that sRGB monitor does not represent effectively the full color range of classical oil paintings, and AdobeRGB monitor reproduces satisfactorily the full range of colors (with the difference below the threshold of human perception). Developed quantitative method can be applied for any kind of monitor estimation.

Keywords: monitors, color, color coordinates, CIE, CIE XYZ, CIE Lab, screen reproduction.

Введение

В связи с тем, что многообразие цветовых тонов в природе и возможности человеческого цветового восприятия превосходят возможности современных цветовоспроизводящих устройств, остро стоит проблема цветовых охватов. Для достоверной экранной репродукции живописных произведений одним из важнейших требований является способность цветовоспроизводящего устройства обеспечить весь диапазон цветовых тонов, присутствующих в живописных полотнах, что позволит свести цветовые отличия между репродукцией и оригиналом до уровня ниже физиологического порога различимости.

Проблемой точности цветовоспроизведения в течение последних лет занимались крупнейшие лаборатории, в том числе лаборатории Лондонской Национальной галереи (К. Мартинез, Дж. Купитт, Д. Саундерс, Р. Пиллэй) [1], Стэнфордского университета, США (Б. Ванделл, Д. Сильверстейн) [2], университета Невады, США (М. Вебстер, Д. Леонард) [3], Рочестерского университета, США (М. Ферчайлд, Р. Бернс, И. Чен) [4], университетов Чикаго, Нагана, Япония (С. Томинага, Н. Танака) [5]. Подход к контролю качества цвета также описан в работах группы исследователей НИУ ИТМО (А.Д. Яськов, К.А. Акмаров, Н.П. Белов, Ю.Ю. Смирнов, А.С. Шерстобитова, Е.Ю. Щербаква) [6].

В основу любых измерений цвета положена принятая в 1931 г. МКО (Международная комиссия по освещению, Comite Europeen d'Imprimerie – CIE) координатная система CIE XYZ. Данная система является абстрактной системой описания цветового стимула и используется, в том числе, при стандартизации хроматических свойств красок [7]. Однако, помимо XYZ-значений самого стимула, для прогнозирования цветового восприятия необходим учет данных об условиях просмотра (обычно координаты источника освещения) [4, 8, 9]. Это повлекло за собой необходимость разработки моделей цветового восприятия, позволяющих строить прогнозы, касающиеся восприятия цветовых стимулов зрительной системой чело-

века, производить расчеты цветовых отличий, а также перцептивные сдвиги при смене статуса хроматической адаптации.

Координатной системой, стандартизованной МКО для расчета цветовых отличий, является система CIELab. Она же как система, не зависящая от свойств каких-либо устройств и моделирующая цветовое восприятие человека, широко используется для визуализации цветовых охватов мониторов, принтеров, стандартных колориметров.

В координатной системе CIELab оси a и b отображают оппонентное кодирование цветовых стимулов (оси a – красный/зеленый, b – желтый/синий) и формируют плоскость цветовых тонов; вертикальная ось L – относительная яркость цвета от 0 до 100. Цветовое пространство Lab формируется исходя из фотометрической яркости и спектрального распределения энергии (СРЭ) осветителя, принятого для данной модели за хроматически нейтральную белую точку максимальной яркости (100; 0; 0). Модель может быть построена для источника света с любым СРЭ. Утвержден ряд стандартных источников, одним из которых является $D50$ (рассеянный дневной свет), на рис. 1 представлены цветовые охваты, нормированные под этот источник освещения.

Для данной работы также необходимо было обратиться к системе, разработанной Альбертом Х. Манселлом [4, 10]. Эта система интересна тем, что дает примерное представление о цветовом охвате человеческого зрения, так называемое «цветовое тело человека». На рис. 1 представлено «цветовое тело человека», пересчитанное из системы Манселла в систему CIELab (сетчатая фигура). Сравнив охват человеческого зрения с охватом цветопроизводящих устройств, можно приблизительно оценить «проблемные области» цветопроизведения [11, 12].

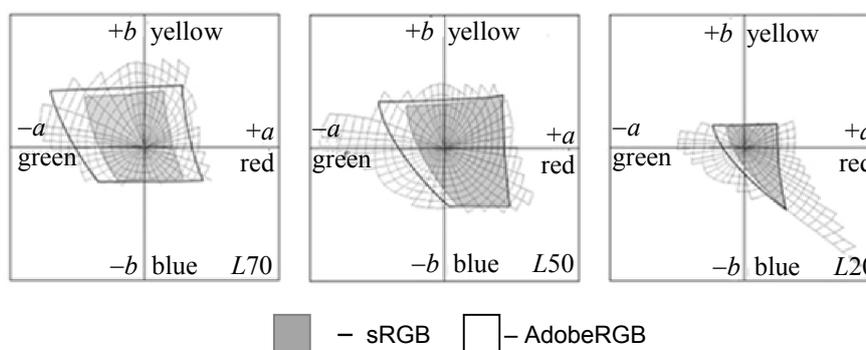


Рис. 1. Поперечные сечения фигур цветовых охватов sRGB, AdobeRGB по разным уровням L и примерного охвата человеческого зрения [12] в координатах CIELab, визуализированные GretagMacbeth ProfileEditor

В настоящее время один из наиболее важных критериев качества профессионального монитора с расширенным цветовым охватом является степень приближенности его тонового диапазона к цветовому охвату абстрактного колориметра AdobeRGB (аппроксимация до 98%). Большинство мониторов среднего уровня аппроксимируют цветовое пространство sRGB (телевизионный стандарт).

Охват AdobeRGB несколько шире, чем sRGB, однако не покрывает охват человеческого зрения. Особенно видны потери в области насыщенных сине-зеленых тонов средней яркости и в области насыщенных темно-синих и фиолетовых. В тонах высокой яркости за пределами охватов оказываются высоко насыщенные желтые и оранжевые.

Итак, при решении задачи достоверной экранной репродукции возникает проблема невозможности ряда насыщенных оттенков [13]. Однако художники, реставраторы, специалисты в области технологии производства масляных красок и репродукции масляной живописи отмечают невысокую насыщенность красок минерального происхождения – наиболее стойких и чаще всего используемых в красочной промышленности [1, 14, 15]. Кроме того, насыщенность красок с течением времени падает, поэтому следует, прежде всего, оценить воспроизводимость свежих колеров.

Таким образом, была поставлена задача определения положения в цветовой координатной системе CIELab хроматических координат насыщенных пигментов, используемых при производстве масляных красок, оценки их положения относительно хроматических охватов современных мониторов, а также расчет цветового отличия до и после репродукции их в тоновых диапазонах пространств sRGB и AdobeRGB как аппроксимирующих охваты большинства мониторов.

Определение цветовых координат

Со времен возникновения масляной живописи набор пигментов, использующихся для изготовления красок, изменился незначительно [14, 15]. Большинство пигментов – это минеральные соли, некогда добываемые в природе, а в настоящее время получаемые химическим путем. Краска, включающая один

пигмент, всегда имеет бóльшую насыщенность, чем смешанные краски. По этой причине производители сохраняют исторически сложившуюся технологию, по которой в каждый колер входит только по одному пигменту, что позволяет получать краски исключительной интенсивности и чистоты цвета. С течением времени некоторые пигменты разрушаются, а также темнеет связующее вещество – льняное масло. Соответственно наибольшую хроматическую насыщенность имеют свежие колеры, не содержащие лишних примесей.

Поскольку целью работы является анализ пигментов, связанных льняным маслом, которое не обладает идеальной хроматической нейтральностью (современное техническое масло – до 20 мг йода по йодометрической шкале, исторические связующие – еще более цветные [7, 14]). В этой связи из двух методов подготовки красочного образца, предложенных в [7], наиболее подходящим является метод накрасок.

Наименование образца	<i>L</i>	<i>a</i>	<i>b</i>	chroma
Кадмий лимонный	94,2	-13,3	90,4	82,5
Стронциановая желтая	86,1	-6,6	82,2	91,4
Кадмий желтый светлый	84,0	-3,5	86,9	87,0
Хром желтый	79,1	36,3	89,9	65,6
Кадмий желтый темный	75,2	11,3	77,1	77,9
Неаполитанская желтая	75,2	19,1	75,4	91,7
Кадмий оранжевый	75,1	34,4	85,0	97,0
Киноварь	62,3	61,2	57,6	96,4
Кобальт зеленый светлый	54,7	-53,4	-2,9	79,6
Изумрудная зеленая	54,6	-54,8	3,5	69,1
Кадмий красный светлый	51,7	63,5	48,0	74,6
Ализарин красный	50,5	80,2	53,4	58,4
Сурик свинцовый	50,1	65,4	45,6	62,7
Хром кобальт зелено-голубой	49,0	-35,8	5,3	58,6
Кобальт зеленый светлый холодный	44,8	-32,5	-21,8	67,5
Церулеум	41,6	-14,5	-38,4	66,2
Английская красная	40,8	52,6	51,7	54,7
Краплек розовый прочный	39,0	58,2	4,3	40,6
Кобальт синий средний спектральный	34,4	11,9	-60,6	55,0
Кадмий желтый средний	33,4	51,8	40,2	49,1
Краплек фиолетовый прочный	31,5	50,2	-30,2	79,7
Кобальт синий средний	30,0	20,6	-61,9	84,0
Краплек красный прочный	28,7	54,1	31,6	73,8
Хром кобальт сине-зеленый	28,7	-19,5	-3,2	39,1
Марганцевая фиолетовая светлая	28,2	26,6	-8,8	53,5
Ультрамарин розовый	27,6	38,7	12,2	33,7
Окись хрома	26,4	-11,1	31,0	32,9
Ультрамарин светлый	23,1	13,8	-38,9	41,3
Индийская красная	22,5	38,0	31,1	41,1
Кобальт фиолетовый темный	21,7	41,8	-51,3	65,2
Марганцевая фиолетовая темная	21,3	40,4	-36,8	61,3
Травяная зеленая	16,8	-7,9	5,2	41,3

Таблица. Цветовые координаты 32 наиболее насыщенных пигментов

Накраски на бумажной основе, описанные в стандарте [7], предназначены для методов визуального сравнения. Такие образцы непригодны для инструментального метода, поскольку луч спектрофотометра проникает сквозь красочный слой и основу, и в результате множественных подповерхностных переломлений хроматические характеристики могут исказиться. Изучение влияния подповерхностных

эффектов на цвет образца не входило в задачи работы, поэтому образцы были подготовлены таким образом, чтобы максимально исключить пропускание света.

В качестве основы для нанесения пигментов, вместо предложенной в стандарте чертежной бумаги, была выбрана алюминиевая пластина с нанесенным на нее малопрозрачным грунтом (толщина слоя – 5 мкм, прозрачность – 3,6, рассчитанная по [16]). Грунт обеспечил хроматически нейтральную основу со сдвигом не более 4,2 ΔE в область диапазона положительных значений по оси b в системе Lab и светлотой $L92 \pm 2$. Далее подготовка образцов производилась в соответствии с [7]. Колеры были нанесены в несколько слоев до уровня, после которого величина насыщенности по Манселлу перестала увеличиваться. Для большинства красок минимальная толщина красочного слоя, дающая полную насыщенность, составила около 50 мкм, что при сравнительно большом окрашенном поле (400 мм²) и непрозрачной подложке исключило влияние окружающих объектов на хроматическую характеристику образца [1, 16].

В качестве измерительного прибора был использован спектрофотометр GretagMachbeth Eye-One Pro, измеряющий СРЭ света, отраженного от образца с шагом 10 нм (заявленная производителем точность $\Delta E \leq 0,1$). Для обработки результатов использовалась программа GretagMachbeth Eye-One Share, выводящая результаты измерений как в виде графиков СРЭ, так и в виде координат в цветовых координатных системах, в том числе в системе CIELab.

В таблице представлены примеры хроматических координат наиболее насыщенных пигментов (показано 32 из 75 замеренных). В таблицу внесена также насыщенность по Манселлу – chroma. Обратившись к фигуре «цветового тела человека» на рис. 1, можно проследить место насыщенности данных колеров по сравнению с максимальной, с учетом того, что каждое последующее большее кольцо сетчатой фигуры обозначает увеличение насыщенности на 20.

Положение координат пигментов относительно AdobeRGB, sRGB

Оценка положения координат показала, что большинство пигментов вошли как в охват AdobeRGB, так и в охват sRGB. На рис. 2 представлено положение цветовых координат наиболее насыщенных пигментов. За пределами охвата sRGB оказались высоко насыщенные светлые колеры – кадмий желтый светлый, кадмий оранжевый. В области средней светлоты выпали колеры сине-зеленой области – церулеум, изумрудная зелень. За пределами охвата в глубоких тенях остались высоко насыщенный синий (кобальт синий спектральный), а также насыщенные темные тона красной области, которые способны дать лишь органические красители, в нашем эксперименте – краплек розовый прочный. Расширенный цветовой охват AdobeRGB включил в себя все полученные хроматические координаты, за исключением двух высоко насыщенных темных тонов – кобальт синий спектральный, краплек розовый прочный (рис. 2).

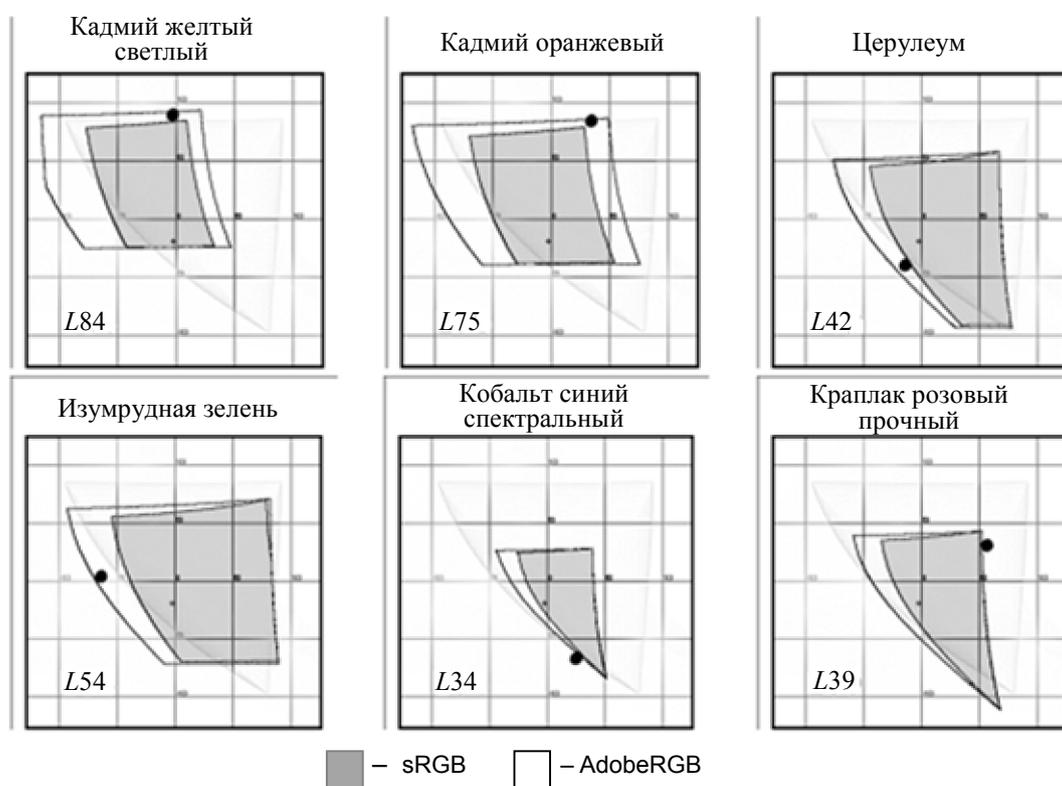


Рис. 2. Визуализация охватов sRGB, AdobeRGB по разным уровням L и координат некоторых пигментов средствами программы GretagMachbeth ProfileEditor

Расчет цветовых отличий

Следующим этапом работы было определение хроматических координат «внеохватных» («out-of-gamut») тонов после отсечения невозпроизводимых областей цветового пространства («gamut-clipping») для sRGB и AdobeRGB, определение их отличия от исходных хроматических характеристик пигментов и сравнение этих отличий со стандартизованными допустимыми цветовыми отличиями.

Расчет цветовых отличий производился по формуле, утвержденной CIE в 2000 г.:

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C S_C}\right) \left(\frac{\Delta H'}{k_H S_H}\right)},$$

где $\Delta L'$, $\Delta C'$, $\Delta H'$ – абсолютная разница по светлоте, насыщенности, цветовому тону; S_L , S_C , S_H – компенсирующие коэффициенты для светлоты, насыщенности, тона; R_T – смещение по синей области цветовых тонов [17].

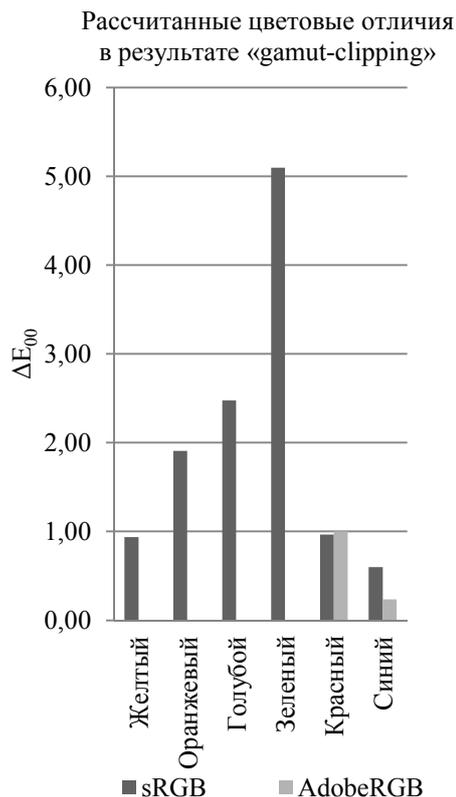


Рис. 3. ΔE_{00} для различных областей насыщенных тонов, вышедших за пределы охватов sRGB, AdobeRGB

Величины допустимых цветовых отличий утверждены в [18]. Для тонов высокой насыщенности среднее допустимое цветовое отличие цветной репродукции составляет $3\Delta E$, для тонов, близких к серому, эта величина снижена до $1,5\Delta E$ [18]. Стоит отметить, что физиологический порог цветового отличия значительно ниже утвержденного [18] и составляет около $2\Delta E$ для насыщенных оттенков и десятые доли – для близких к серому [13, 19, 20]. На диаграмме (рис. 3) представлены рассчитанные величины ΔE_{00} для различных областей насыщенных тонов, вышедших за пределы охватов sRGB, AdobeRGB.

После визуализации насыщенных цветовых тонов в пространстве sRGB изменения хроматических координат голубой и зеленой области оказались выше, чем допускает стандарт, отличия значительно превосходят порог различимости. Это говорит о том, что монитор, аппроксимирующий sRGB, непригоден для репродукции насыщенных красок из-за значительного отсечения голубых и зеленых областей цветового пространства. Отличия визуализированных в AdobeRGB «внеохватных» цветовых тонов не превышают $1\Delta E$, таким образом, на мониторе, аппроксимирующем AdobeRGB, возможна репродукция насыщенных красок с отличием ниже физиологического порога различимости.

Заключение

Измерения и расчет показали, что за пределами охвата абстрактного колориметра sRGB (охват большинства средних мониторов) оказались хроматические координаты ряда насыщенных пигментов, в результате отсечения которых происходят цветовые искажения, превышающие порог различимости. Это

говорит о том, что при таком охвате цветовоспроизводящее устройство не способно обеспечить удовлетворительную достоверность репродукции. Монитор с расширенным цветовым охватом, аппроксимирующий AdobeRGB, воспроизводит насыщенные оттенки гораздо более удовлетворительно, и минимально достижимые цветовые отличия по «внеохватным» цветовым тонам не превосходят физиологического порога различимости.

В заключение стоит отметить, что в настоящей работе цветовой охват красочных пигментов сравнивался с охватами идеальных устройств. В ходе подготовки конкретного монитора для экранной репродукции живописного произведения полученные хроматические координаты пигментов следует сравнить с охватом уже реального устройства. Для этого после калибровки и характеристики экрана можно использовать данную методику, заменив профайлы абстрактных колориметров sRGB, AdobeRGB на профайл реального монитора. Если цветовые отличия после отсечения невоспроизводимых областей цветового пространства по насыщенным тонам не превысят $2\Delta E$, можно перейти к минимизации цветовых отличий живописного произведения и его изображения на экране.

Дальнейшие исследования будут направлены на разработку методики расчета и минимизации цветовых отличий оригинала и его экранной репродукции.

Литература

1. Saunders D., Hamber A. From Pigments to Pixels: Measurement and Display of the Colour Gamut of Paintings // Proc. of SPIE. 1990. V. 1250. P. 90–102.
2. Wandell B., Silverstein D. Digital Color Reproduction // The Science of Color / Ed. S. Shevell. 2nd ed. Optical Society of America, 2003. P. 281–316.
3. McDermott K.C., Webster M.A. Uniform color spaces and natural image statistics // J. Opt. Soc. Am. A. 2012. V. 29. N 2. P. A182–A187.
4. Fairchild M.D. Color Appearance Models. 2nd ed. John Wiley & Sons, Ltd, 2005. 408 p.
5. Tominaga S., Tanaka N. Spectral image acquisition, analysis, and rendering for art paintings // J. Electronic Imaging. 2008. V. 17. N. 4. P. 6–13.
6. Акмаров К.А., Белов Н.П., Смирнов Ю.Ю., Шерстобитова А.С., Щербакова Е.Ю., Яськов А.Д. Лабораторный спектрофотометр для видимой области спектра // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 5 (87). С. 39–44.
7. ГОСТ 16873-92 (ИСО 787/1-82). Пигменты и наполнители неорганические. Методы определения цвета и белизны. Введ. 01.07.93. М.: Изд-во стандартов, 1992. 10 с.
8. Юстова Е.Н. Цветовые измерения (Колориметрия). СПб: Изд-во СПбГУ, 2000. 397 с.
9. Ohta N., Robertson A. Colorimetry: Fundamentals and Applications. John Wiley & Sons, Inc., 2005. 350 p.
10. Hunt R.W.G. The Reproduction of Colour. 6th ed. John Wiley & Sons, Ltd, 2004. 724 p.
11. Gijzenij A., Gevers T., Weijer J. Generalized Gamut Mapping Using Image Derivative Structures for Color Constancy // Int. J. Computer Vision. 2010. V. 86. N 2. P. 127–139.
12. Lindbloom B.J. Useful Color Information, Studies and Files. 2001–2013 [Электронный ресурс]. Режим доступа: <http://www.brucelindbloom.com/>, свободный. Яз. англ. (дата обращения 26.10.2013).
13. Robertson Ph., Schönhut J. Color in Computer Graphics // IEEE Computer Graphics and Applications. 1999. V. 19. N 4. P. 18–19.
14. Лентовский А.М. Технология живописных материалов. Л. –М.: Искусство, 1949. 220 с.
15. Филатов В.В. Реставрация настенной масляной живописи. М.: Изобразительное искусство, 1995. 248 с.
16. ГОСТ 7086-75. Краски печатные. Методы определения прозрачности. Введ. 01.01.77. М.: Изд-во стандартов, 2003. 7 с.
17. Sharma G., Wu W., Dalal E.N. The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations // Submitted to Color Research and Application. 2004. V. 16. N 3. P. 217–218.
18. ISO 12647-7:2007(E). Graphic technology – Process control for the production of half-tone colour separations, proof and production prints. Part 7: Proofing processes working directly from digital data. 20 p.
19. Hansen T., Olkkonen M., Walter S., Gegenfurtner K. Memory Modulates Color Appearance // Nature Neuroscience. 2006. V. 9. N 11. P. 1367–1368.
20. Lau C., Heidrich W., Mantiuk R. Cluster-based color space optimizations // IEEE International Conference on Computer Vision (ICCV). 2011. P. 1172–1179.

Черевань Любовь Владимировна	– аспирант, старший преподаватель, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, 4erevanb@mail.ru
Тозик Вячеслав Трофимович	– кандидат технических наук, доцент; директор, Межотраслевой институт повышения квалификации, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, tozikvt@mail.ru

Lyubov' Cherevan

– postgraduate, senior lecturer, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, 4erevanb@mail.ru

Vyacheslav Tozik

– PhD, Associate Professor, Head of Intersectoral Institute for Advanced Training, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, tozikvt@mail.ru

УДК 004.056.53

ЭКСПЛУАТАЦИОННЫЕ ХАРАКТЕРИСТИКИ РИСКА НАРУШЕНИЙ БЕЗОПАСНОСТИ ИНФОРМАЦИОННОЙ СИСТЕМЫ

К.А. Щеглов^{а, б}, А.Ю. Щеглов^{а, б}

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, info@npp-itb.spb.ru

^б ЗАО «НПП «Информационные технологии в бизнесе», Санкт-Петербург, Россия, info@npp-itb.spb.ru

Рассмотрены известные подходы к оценке эффективности и к проектированию средств защиты информационных систем, в результате чего выявлен их общий недостаток, заключающийся в том, что данные подходы не позволяют рассматривать угрозы информационной безопасности в качестве эксплуатационной характеристики информационной системы. Это не позволяет построить адекватную модель информационной системы, в том числе защищенной, как системы с отказами и восстановлениями, характеризующимися возникновением и устранением угроз в процессе эксплуатации системы, не позволяет выявить и соответствующие взаимосвязи, в том числе временные, между угрозами и их совокупностями, эксплуатируемыми атаками, ввести количественные меры актуальности угроз и эффективности атак как эксплуатационных характеристик системы и, как следствие, количественно оценить уровень эксплуатационной безопасности информационной системы и эффективности средств защиты. Предложены основы теории эксплуатационной информационной безопасности, введены основные эксплуатационные параметры и характеристики, предложен подход к оцениванию эксплуатационных характеристик информационной системы, введены эксплуатационные характеристики риска нарушений безопасности информационной системы, в том числе риска потерь.

Ключевые слова: информационная безопасность, защита информации, информационная система, угроза, атака, эксплуатационная безопасность, уровень безопасности, несанкционированный доступ, оценка эффективности, теория массового обслуживания, теория надежности, средство защиты, риск.

OPERATIONAL CHARACTERISTICS OF INFORMATION SYSTEM SECURITY THREATS RISK

К. Shcheglov^{с, d}, A. Shcheglov^{с, d}

^с Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, info@npp-itb.spb.ru

^d JSC “Information Technologies in Business”, Saint Petersburg, Russia, info@npp-itb.spb.ru

The paper deals with widely used methods for effectiveness evolution and information systems security tools development. Their general disadvantage that consists in not giving the possibility to consider security threats as information system operational characteristic is revealed. Therefore, an adequate information system model can't be created, including secure information system like system with failure and recovery parameters, characterized by threats appearing and elimination during system operation. Also it doesn't make it possible to identify appropriate relationship (including time ones), between threats and their exploited aggregates and doesn't give the possibility to introduce quantitative measures of the threats relevance and attacks effectiveness (like system operational characteristic). Consequently, it doesn't make it possible to estimate the level of information system operational security and security tools effectiveness. The principles of operational security theory are suggested, base operational parameters and characteristics are introduced, a method for operational characteristics of information system estimation is proposed, operational characteristics of risks for information system security threats and loss risks are stated.

Keywords: information security, information system, threat, attack, operational security, security level, unauthorized access, effectiveness evaluation, queueing theory, reliability theory, security tools, risk.

Введение

Под математической теорией риска формально понимается совокупность моделей и методов теории вероятностей, применяемых к анализу случайных величин и их распределений [1]. Это обусловлено тем, что величина возможного ущерба в стохастической ситуации до осуществления этой ситуации неизвестна и потому случайна. Риском же называют совокупность значения возможного ущерба в некоторой стохастической ситуации и его вероятности [1, 2].

Применению теории риска в области информационной безопасности, в частности, для оценки эффективности средств защиты информации и при проектировании защищенной информационной системы (далее ИС), ввиду актуальности данной задачи, в настоящее время посвящено достаточно много работ,

например [3–8]. Общий недостаток известных подходов состоит в том, что они не позволяют рассматривать угрозы информационной безопасности в качестве эксплуатационной характеристики ИС. Это не позволяет построить адекватную модель ИС, в том числе защищенной, как системы с отказами и восстановлениями, характеризуемыми возникновением и устранением угроз в процессе эксплуатации системы, не позволяет выявить и соответствующие взаимосвязи, в первую очередь, временные, между угрозами и их совокупностями, эксплуатируемыми атаками, ввести количественные меры актуальности угроз и эффективности атак как эксплуатационных характеристик системы и, как следствие, адекватно количественно оценить эксплуатационные характеристики риска нарушений безопасности ИС.

Эксплуатационные характеристики риска

Прежде всего, в рамках основ излагаемой теории эксплуатационной информационной безопасности введем основные понятия, определим ключевые эксплуатационные параметры и характеристики ИС.

Под *эксплуатационной информационной безопасностью* понимается свойство ИС функционировать в безопасном состоянии – без обнаруженных в ней либо внесенных в нее в процессе эксплуатации и не устраненных «каналов» несанкционированного доступа к информации.

Под *«каналом» несанкционированного доступа к информации* понимается появившийся в ИС по каким-либо причинам способ осуществления злоумышленником несанкционированного доступа с какой-либо целью к обрабатываемой в системе информации во время эксплуатации ИС.

Под *нарушением эксплуатационной информационной безопасности* понимается обнаружение в ИС либо внесение в нее в процессе эксплуатации «каналов» несанкционированного доступа к информации.

Под *угрозой безопасности ИС* понимается нарушение ее эксплуатационной информационной безопасности в результате обнаружения в ИС либо внесения в нее в процессе эксплуатации «каналов» несанкционированного доступа к информации.

Замечание. Нарушение эксплуатационной информационной безопасности (угроза безопасности ИС) – это не есть осуществление несанкционированного доступа к информации, это возникновение условий для успешной атаки на ИС с использованием существующего (выявленного) в системе неустраненного «канала» («каналов») несанкционированного доступа к информации.

Введем основные эксплуатационные параметры и характеристики ИС, которые будут далее нами использованы для определения характеристики риска как эксплуатационной характеристики ИС, после чего определим эксплуатационные характеристики риска, в том числе риска потерь.

Под *интенсивностью нарушения эксплуатационной информационной безопасности (возникновения угроз)* λ будем понимать количество обнаружений в ИС либо внесений в нее в процессе эксплуатации «каналов» несанкционированного доступа к информации в единицу времени.

Под *интенсивностью восстановления эксплуатационной информационной безопасности (нейтрализации угроз)* μ будем понимать количество устранений выявленных в ИС либо внесенных в нее «каналов» несанкционированного доступа к информации в единицу времени.

Под *уровнем эксплуатационной информационной безопасности ИС* будем понимать количественную оценку готовности ИС к безопасной эксплуатации – значение вероятности P_0 того, что информационная система готова к безопасной эксплуатации.

Данный подход к оцениванию уровня эксплуатационной безопасности ИС, $P_0 = f(\lambda, \mu)$, рассмотренный нами впервые в [9], позволяет применить математический аппарат теории массового обслуживания для получения требуемой количественной оценки [9, 10]. При этом важным является тот факт, что при расчете характеристики эксплуатационной информационной безопасности используются не некие гипотетические, заданные экспертным путем, значения параметров ИС, а реальные значения параметров, полученные на основании анализа соответствующей статистики угроз. Это позволяет говорить об адекватности получаемых подобным образом моделей безопасности информационных систем.

Под *атакой на информационную систему* понимается воздействие злоумышленником на ИС посредством использования (эксплуатации) обнаруженных в ИС либо внесенных в нее в процессе эксплуатации «каналов» несанкционированного доступа к информации (угроз безопасности ИС).

В общем случае атака многоступенчатая и предполагает последовательное использование атаккой некой совокупности угроз, присутствующих в системе. Формальное описание атаки (атаки, реализуемой с определенной целью) предлагается представлять взвешенным ориентированным графом атаки на ИС.

Под *графом атаки на информационную систему* понимается взвешенный ориентированный граф (или оргграф), вершинами которого являются угрозы, характеризуемые («взвешенные») вероятностью наличия обнаруженных или внесенных в систему и не устраненных «каналов» несанкционированного доступа к информации, создающих угрозу, которые должны эксплуатироваться (использоваться) злоумышленником при осуществлении атаки; дуги графа определяют последовательность использования злоумышленником угроз (направленного перехода между вершинами) при осуществлении атаки.

Пример графа атаки на ИС приведен на рис. 1. На рис. 1 P_{0r} , где $r = 1, \dots, R$, обозначает вероятность отсутствия в системе r -й угрозы (информационная система готова к безопасной эксплуатации в отноше-

нии r -й угрозы), одной из R угроз, последовательно эксплуатируемых (используемых) злоумышленником при осуществлении описываемой графом атаки. Эту характеристику, как отмечали ранее, мы можем считать с использованием аппарата теории массового обслуживания.

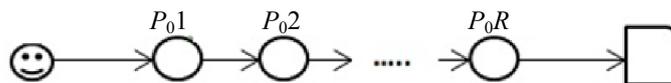


Рис. 1. Пример графа атаки на ИС

С учетом сказанного может быть сделан крайне важный вывод: каждая последующая угроза, используемая при осуществлении атаки, может рассматриваться, в терминологии теории надежности [11], в качестве параллельно включенного в систему резерва (для осуществления атаки необходимо использовать всю совокупность угроз, приведенных на графе, см. рис. 1). Как следствие, вероятность того, что защищенная информационная система готова к безопасной эксплуатации в отношении атаки P_{0a} , последовательно использующей R угроз, $r = 1, \dots, R$, определяется следующим образом:

$$P_{0a} = 1 - \prod_{r=1}^R (1 - P_{0r}).$$

Аналогичным образом (параллельно включенный резерв) может рассматриваться и средство защиты, направленное на нивелирование угрозы, используемой атакой, которое может быть охарактеризовано аналогичными параметрами и характеристиками.

Под *готовностью защищенной информационной системы к безопасной эксплуатации в отношении атаки* понимается вероятность того, что защищенная ИС готова к безопасной эксплуатации – в ней отсутствуют обнаруженные либо внесенные в нее и не устраненные «каналы», как собственно создающие угрозы, используемые атакой, так и в средстве защиты информации, призванном нивелировать угрозу (угрозы), эксплуатируемую атакой, позволяющие осуществить несанкционированный доступ к информации.

Соответственно под *уровнем эксплуатационной информационной безопасности защищенной информационной системы в отношении атаки* понимается количественная оценка готовности защищенной ИС к безопасной эксплуатации – значение вероятности того, что защищенная в отношении атаки информационная система готова к безопасной эксплуатации.

Все сказанное можно отнести и к ИС в целом. Атаки в этом случае могут рассматриваться, в терминологии теории надежности [11], в качестве последовательно включенного в систему резерва (для осуществления атаки на систему в целом злоумышленнику достаточно реализовать одну из атак). Как следствие, вероятность того, что информационная система готова к безопасной эксплуатации в целом, $P_{0исц}$, при возможной совокупности из D атак, $d = 1, \dots, D$, каждая из которых соответственно характеризуется P_{0ad} , определяется следующим образом:

$$P_{0исц} = \prod_{d=1}^D P_{0ad}.$$

С учетом того, что каждая угроза, используемая при осуществлении атаки, может рассматриваться в терминологии теории надежности в качестве параллельно включенного в систему резерва, а каждая из совокупности возможных атак на ИС может рассматриваться в качестве последовательно включенного в систему резерва, в общем случае, применительно к ИС в целом, можем говорить о последовательно-параллельном резервировании. Таким образом, мы можем количественно оценить готовность ИС к безопасной эксплуатации (вероятность того, что информационная система готова к безопасной эксплуатации) как в отношении отдельной атаки, так и в целом.

Под *средним временем наработки на нарушение эксплуатационной информационной безопасности информационной системы* (среднее время наработки на отказ в теории надежности) понимается средний интервал времени, в течение которого ИС готова к безопасной эксплуатации (между моментами нарушения эксплуатационной информационной безопасности ИС). Данная характеристика определяется следующим образом:

$$T_0 = \frac{P_0}{\mu (1 - P_0)}.$$

Это – параметры и характеристики непосредственно ИС, определяющие риск осуществления успешной атаки на ИС.

Под *риском нарушения безопасности информационной системы* будем понимать вероятность наступления условия нарушения эксплуатационной информационной безопасности ИС, позволяющего злоумышленнику осуществить успешную атаку на ИС с целью получения несанкционированного доступа к обрабатываемой в ней информации.

Риск нарушения безопасности ИС определяется значением вероятности P_0 (соответственно применительно к атаке P_{0a} , к ИС в целом $P_{0исц}$). Это эксплуатационные характеристики непосредственно ИС, определяющие риск осуществления успешной атаки на ИС.

Рассмотрим еще один важный эксплуатационный параметр.

Под *интенсивностью осуществления злоумышленником атак на защищенную информационную систему* будем понимать количество совершаемых им атак на ИС в единицу времени λ_a . Это уже эксплуатационный параметр нарушителя, характеризующий его заинтересованность в получении несанкционированного доступа к информации, обрабатываемой в системе (это характеристика ценности обрабатываемой информации для нарушителя).

Отметим, что значение параметра λ_a должно задаваться при проектировании системы защиты ИС с целью формирования требований к уровню эксплуатационной информационной безопасности проектируемой защищенной ИС (проектируемой системы защиты ИС). Требуемый для этого параметр λ_a может быть определен либо на основании опыта эксплуатации подобных (обрабатывающих аналогичную информацию) защищенных информационных систем, либо экспертным путем. Важным является уточнение данного параметра в процессе всего времени эксплуатации ИС с применением средств аудита – протоколирования событий, в данном случае – осуществления атак (и их отражения) на защищенную ИС, что, в свою очередь, определяет соответствующие требования к средствам аудита событий, реализуемым в ИС. Определение значения параметра λ_a как эксплуатационной характеристики (в процессе эксплуатации ИС) позволяет оценить корректность сделанных исходных предположений о ценности для злоумышленников обрабатываемой в ИС информации, а также учесть изменение данного параметра в процессе эксплуатации системы, в результате чего, при необходимости, доработать исходные требования к уровню эксплуатационной информационной безопасности ИС.

С точки зрения эксплуатационной информационной безопасности успешная атака будет реализована злоумышленником в том случае, если она будет проведена в тот интервал времени, когда информационная система не готова к безопасной эксплуатации. С учетом сказанного определим важнейшую характеристику эксплуатационной информационной безопасности – вероятность реализации успешной атаки на ИС.

Под *вероятностью реализации успешной атаки на информационную систему* будем понимать вероятность того, что атака будет осуществлена злоумышленником в промежуток времени, когда ИС не готова к безопасной эксплуатации.

Вероятность реализации успешной атаки на защищенную ИС P_a , определяемая вероятностью того, что атака (по крайней мере, одна из атак) придется на интервал времени $1/\mu$, может быть получена из закона Пуассона [10]:

$$P_a = 1 - e^{-\lambda_a/\mu}.$$

С учетом сказанного определим эксплуатационную характеристику риска.

Под *риском осуществления успешной атаки на информационную систему* будем понимать вероятность реализации успешной атаки на ИС с целью получения несанкционированного доступа к обрабатываемой в ней информации.

Риск осуществления успешной атаки на ИС определяется значением вероятности P_a .

Важным моментом является то, что мы рассматриваем восстанавливаемую систему, для которой можно вести понятие периода эксплуатации T , задаваемого следующим образом:

$$T = T_0 + 1/\mu.$$

В этой формуле T_0 – среднее время наработки на нарушение эксплуатационной информационной безопасности ИС, по истечении которого в течение времени $1/\mu$ безопасное состояние ИС восстанавливается – информационная система в течение этого промежутка времени уязвима для атак. Таким образом, если вероятность успешной атаки на ИС, P_a , определяется вероятностью того, что атака (по крайней мере, одна из атак) придется на интервал времени $1/\mu$, то с вероятностью, равной $1 - P_a$, в данном интервале времени успешная атака на ИС осуществлена не будет. Следующая подобная возможность злоумышленнику «будет предоставлена» в следующем периоде эксплуатации системы T – через промежуток времени T_0 . При этом значение характеристики P_a определяется эксплуатационными параметрами ИС и нарушителя, как следствие, P_a одинаковы для различных периодов эксплуатации системы $T_i, i=1, \dots, I$.

Распределение характеристики P_a во времени (качественная оценка), определяемом периодами эксплуатации системы $T_i, i=1, \dots, I$, представлено на рис. 2.

Таким образом, если вероятность успешной атаки на ИС в периоде эксплуатации T_i в течение интервала времени $1/\mu$ на этом периоде, см. рис. 2 (в различных периодах это события независимые), равна P_a , то вероятность успешной атаки на ИС в i -ом периоде эксплуатации $T_i, i = 1, \dots, I$, равна

$$P_{ai} = \begin{cases} 0, & \text{если } i = 1; \\ P_a(1 - P_a)^{i-2}, & \text{если } i \geq 2. \end{cases}$$

Замечание. Для удобства формального описания за первый период эксплуатации в этой формуле и далее (что необходимо учитывать при использовании приведенных ниже расчетных формул) принимаем $T_{i-1} = T_0$ за последующие периоды эксплуатации:

$$T_{i \geq 2} = T_0 + 1/\mu.$$

Данная эксплуатационная характеристика определяет крайне важное эксплуатационное свойство ИС, которое должно быть нами далее учтено. Поскольку информационная система – это система восстанавливаемая, т.е. восстанавливаются свойства ее информационной безопасности в процессе эксплуатации (с интенсивностью μ), несмотря на значение характеристики P_0 , информационная система может с соответствующими вероятностями P_{ai} , $i = 1, \dots, I$, безопасно эксплуатироваться в различных периодах (на различных временных интервалах T_i). С учетом этого может быть введена следующая эксплуатационная характеристика ИС – вероятность того, что на ИС будет осуществлена успешная атака за время ее эксплуатации, характеризуемое I периодами эксплуатации системы T_i с номерами $i = 1, \dots, I$, P_{aI} . Эта характеристика рассчитывается по следующей формуле:

$$P_{aI} = \sum_{i=1}^I P_{ai}.$$

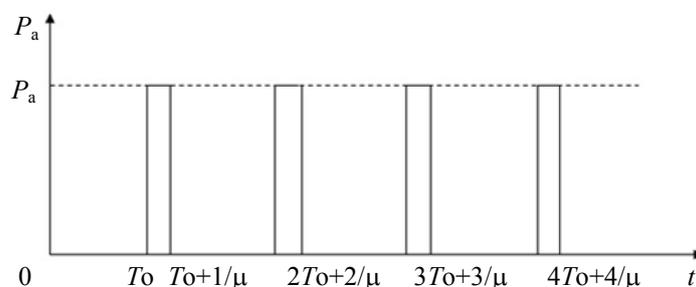


Рис. 2. Распределение характеристики P_a во времени

Для любого произвольного времени эксплуатации ИС t характеристика P_{aI} (в данном случае $P_a(t)$) может быть определена следующим образом:

$$P_a(t) = \begin{cases} 0, & \text{если } 0 < t \leq T_0; \\ \sum_{i=2}^I P_{ai}, & \text{если } (I-1)T_0 + \frac{I-2}{\mu} < t \leq IT_0 + \frac{I-1}{\mu}. \end{cases}$$

С учетом изменения во времени характеристики P_a и соответственно P_{ai} , (рис. 2), качественная зависимость изменения характеристики $P_a(t)$ во времени представлена на рис. 3. Таким образом, зная эксплуатационные параметры ИС T_0 и μ , для любого произвольного времени эксплуатации ИС t можно определить соответствующее ему число периодов эксплуатации системы I продолжительностью T_i , для которого уже определить искомое значение $P_a(t)$.

Под *риском осуществления успешной атаки на информационную систему в течение времени ее эксплуатации t* будем понимать вероятность реализации успешной атаки на ИС с целью получения несанкционированного доступа к обрабатываемой в ней информации за время ее эксплуатации t .

Риск осуществления успешной атаки на ИС в течение времени ее эксплуатации t определяется значением вероятности $P_a(t)$ на соответствующем времени t числе периодов эксплуатации системы I продолжительностью T_i .

Акцентируем внимание читателя на особенностях введенной эксплуатационной характеристики риска. Как правило, на практике риск оценивается вероятностью взлома (либо вероятностью успешной атаки, которая, кстати, также не рассматривается в качестве эксплуатационной характеристики) ИС. При этом не рассматривается распределение характеристики риска во времени эксплуатации системы. Естественно, возникает вопрос, насколько адекватна подобная модель риска и что на самом деле ею описывается – для какого момента времени эксплуатации ИС определяется риск, с учетом того, что его значение (значение соответствующей вероятности) изменяется во времени? Насколько и как применима подобная характеристика, если непонятно, что ею определяется?

Если, по аналогии с теорией надежности [11], ранее нами рассматривалась характеристика T_0 – среднее время наработки на нарушение эксплуатационной информационной безопасности ИС, то с ис-

пользованием параметра λ_a можем ввести важнейшую временную эксплуатационную характеристику ИС – среднюю продолжительность безопасной эксплуатации ИС или среднее время наработки ИС до успешной атаки.

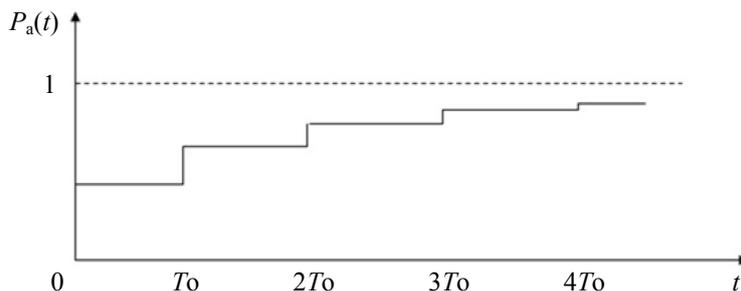


Рис. 3. Качественная зависимость изменения характеристики $P_a(t)$ во времени

Под *средней продолжительностью безопасной эксплуатации информационной системы* (или под *средним временем наработки информационной системы до успешной атаки*) будем понимать среднее время, в течение которого ИС будет находиться в безопасном состоянии с начала ее эксплуатации.

Данная характеристика $T_{бэ}$ может быть рассчитана по следующей формуле:

$$T_{бэ} = T_0 + \sum_{i=2}^{\infty} (i-1)TP_{ai}.$$

Замечание. $T_i = T, i \geq 2$.

Итак, нами были введены важнейшие вероятностные и временные характеристики риска, позволяющие оценить безопасность эксплуатации ИС с учетом интенсивности осуществления злоумышленником атак на ИС и уровня эксплуатационной информационной безопасности ИС. В ряде случаев подобных оценок риска будет достаточно для ответа на вопрос, следует ли реализовывать защиту ИС с целью повышения уровня ее эксплуатационной информационной безопасности.

Однако при высокой заинтересованности злоумышленника в получении несанкционированного доступа к обрабатываемой в ИС информации (характеризуемой большим значением параметра λ_a) и, как следствие, при принятии решения о необходимости реализации защиты ИС уже целесообразно рассматривать такие характеристики ИС (в том числе, защищенной ИС), как потери (экономическую характеристику эксплуатационной информационной безопасности). Ведь потери в защищенной ИС связаны как с потенциальной возможностью несанкционированного доступа злоумышленника к обрабатываемой информации (потенциальные потери, которые будут снижаться при реализации защиты), так и с реализацией и эксплуатацией системы защиты в ИС.

Прежде чем говорить о потерях, введем еще одну важнейшую эксплуатационную характеристику ИС – характеристику сложности реализации атаки на ИС для злоумышленника.

Как уже отмечалось, в рамках рассматриваемой теории эксплуатационной информационной безопасности каждая последующая угроза, используемая при осуществлении атаки, может рассматриваться в терминологии теории надежности в качестве параллельно включенного в систему резерва. При этом успешная реализация атаки возможна лишь при условии актуальности (выявленных и неустраненных «каналов» несанкционированного доступа к информации применительно к каждой угрозе) всех угроз, используемых атакой. Таким образом, для осуществления успешной атаки злоумышленник должен обладать определенным количеством информации – информацией о выявленных и неустраненных (существующих) «каналах» несанкционированного доступа к информации применительно к каждой угрозе, используемой атакой. Очевидно, что именно количество информации, которым должен обладать злоумышленник в отношении актуальности угроз, используемых атакой, и определяет сложность реализации успешной атаки на ИС.

Под *характеристикой сложности реализации атаки на информационную систему для злоумышленника* будем понимать количество информации, которым должен обладать злоумышленник в отношении актуальности угроз, используемых атакой.

Обратимся к основам теории информации. Следуя К. Шеннону [12], вероятностная мера количества информации определяется следующим образом. Пусть можно получить n сообщений по результатам некоторого опыта (т.е. у опыта есть n исходов), причем известны вероятности получения каждого сообщения (исхода) – p_i . Тогда в соответствии с идеей Шеннона, количество информации I в сообщении i определяется по формуле $I = -\log_2 p_i$, где p_i – вероятность i -го сообщения (исхода).

В нашем случае в отношении атаки возможны два исхода, определяющие то количество информации, которым должен обладать злоумышленник для реализации успешной атаки – готова или нет информационная система к безопасной эксплуатации.

С учетом же того, что вероятность готовности защищенной информационной системы к безопасной эксплуатации в отношении атаки, последовательно использующей R угроз, $r = 1, \dots, R$, определяется следующим образом:

$$P_{0a} = 1 - \prod_{r=1}^R (1 - P_{0r}),$$

количество информации (как вероятностная мера этого исхода), необходимое злоумышленнику для осуществления успешной атаки – характеристика сложности реализации атаки на ИС для злоумышленника, обозначим ее через Sa – может быть определена по следующей формуле:

$$Sa = -\log_2 \prod_{r=1}^R (1 - P_{0r}).$$

Это важнейшая эксплуатационная характеристика ИС, позволяющая количественно оценить и достаточно информативно представить сложность реализации злоумышленником успешной атаки. Например, рассмотрим, как изменится сложность реализации атаки, в предположении, что исходное для ИС значение P_{0a} составляло 0,7, а в результате использования средства защиты стало равно 0,99. Видим, что в первом случае $Sa = 1,74$, во втором случае $Sa = 6,64$, т.е. использование данного средства защиты в ИС усложнило злоумышленнику реализацию успешной атаки на ИС в 3,82 раза.

Замечание. За единицу сложности реализации атаки $Sa = 1$ при этом принимается условие $P_{0a}=0,5$ (система с равной вероятностью защищена или нет).

Введенную характеристику сложности реализации атаки на ИС для злоумышленника можно использовать и в процессе эксплуатации ИС для количественной оценки того, насколько упростится атака на ИС в случае актуализации (обнаружения «канала» несанкционированного доступа) одной или нескольких угроз (для актуальной угрозы r справедливо соотношение $P_{0r}=0$).

Данная важнейшая характеристика применима не только в отношении конкретной атаки, но и в отношении ИС в целом, обозначим ее $S_{ис}$. С учетом вероятности того, что информационная система готова к безопасной эксплуатации в целом при возможной совокупности из D атак, $d = 1, \dots, D$, каждая из которых соответственно характеризуется P_{0ad} , определяемой как

$$P_{0исц} = \prod_{d=1}^D P_{0ad},$$

характеристика $S_{ис}$ может быть рассчитана по следующей формуле:

$$S_{ис} = -\log_2 \left(1 - \prod_{d=1}^D P_{0ad} \right).$$

Теперь о потерях. Потери для незащищенной ИС связаны исключительно с успешной атакой на ИС, в результате которой злоумышленник получает несанкционированный доступ к обрабатываемой в ней информации (в данном случае речь идет о потенциальных потерях).

Пусть потери от несанкционированного доступа к информации (в результате ее хищения, удаления или модификации) составляют $C_{инф}$.

Под *риском потерь от осуществления успешной атаки на информационную систему в течение времени ее эксплуатации t* будем понимать экономическую эксплуатационную характеристику ИС, определяемую как потенциальные потери от успешной атаки на ИС в результате реализации злоумышленником несанкционированного доступа к обрабатываемой в ней информации за время ее эксплуатации t .

С учетом сказанного ранее (с учетом полученной расчетной формулы для $P_a(t)$), для любого произвольного времени эксплуатации ИС t риск потерь от осуществления успешной атаки на ИС, $R_{C_{инф}}(t)$, может быть рассчитан следующим образом:

$$R_{C_{инф}}(t) = \begin{cases} 0, & \text{если } 0 < t \leq T_0; \\ C_{инф} \sum_{i=2}^I P_{ai}, & \text{если } (I-1)T_0 + \frac{(I-2)}{\mu} < t \leq IT_0 + \frac{(I-1)}{\mu}. \end{cases}$$

Очевидно, что функция распределения риска потерь от осуществления успешной атаки на ИС во времени имеет тот же вид, что и соответствующая функция распределения вероятности успешной атаки на ИС (см. рис. 2), причем в пределе риск потерь стремится к значению $C_{инф}$.

Рассмотрим возможные способы учета ценности информации. Для этого нами были введены два параметра – λ_a и $C_{инф}$. Отметим, что данные параметры характеризуют альтернативные подходы к определению ценности обрабатываемой информации: λ_a – с точки зрения злоумышленника, $C_{инф}$ – с точки зрения владельца информации (или оператора, обрабатывающего информацию). В общем случае – это альтернативные подходы к определению ценности информации, которые должны использоваться совместно, поскольку, как бы ни была ценна обрабатываемая информация для ее владельца, на деле она может не представлять никакой ценности для потенциальных злоумышленников, и наоборот.

**Эксплуатационные характеристики риска для защищенной ИС.
Экономическая оценка эффективности системы защиты**

Как говорилось ранее, реализация в ИС системы защиты может, в терминологии теории надежности, рассматриваться в качестве параллельно включенного в систему резерва. Это обуславливается тем, что для осуществления успешной атаки на ИС злоумышленнику потребуется использовать как возникшие и неустраненные угрозы, эксплуатируемые атакой, собственно в ИС (например, выявленные уязвимости операционной системы и приложений), так и соответствующие угрозы, выявляемые непосредственно в системе защиты.

Несложно проиллюстрировать и то, что высокий уровень эксплуатационной информационной безопасности собственно системы защиты (значение вероятности готовности системы защиты к безопасной эксплуатации 0,95 и выше) вполне достижим на практике. Для этого оценим потенциальные возможности средств защиты в части повышения уровня эксплуатационной информационной безопасности информационных систем. С этой целью определим граничную оценку (при худших условиях уровень эксплуатационной информационной безопасности ИС полностью определяется уровнем эксплуатационной безопасности средства защиты). Расчетные значения приведены в таблице.

Интенсивность устранения «каналов» несанкционированного доступа к информации в средстве защиты	Готовность к безопасной эксплуатации при различных интенсивностях обнаружения «каналов» несанкционированного доступа к информации в средстве защиты			
	1/ 3 месяца	1/ 6 месяцев	1/ 12 месяцев	1/ 18 месяцев
1/3 дня	0,97	0,98	0,99	0,99
1/7 дней	0,93	0,96	0,98	0,99
1/14 дней	0,87	0,93	0,96	0,98

Таблица. Оценка уровня эксплуатационной информационной безопасности защищенной ИС

Поскольку приведенные в таблице эксплуатационные характеристики средств защиты достижимы на практике, за счет их использования в качестве параллельно включенного резерва могут быть существенно улучшены эксплуатационные характеристики ИС – в данном случае уже защищенной.

Некоторые технические решения, иллюстрирующие возможности средств защиты, применение которых направлено на повышение уровня эксплуатационной информационной безопасности информационных систем, описаны, например, в [13–15] (заметим, что данные технические решения на сегодняшний день авторами патентуются).

Для обозначения соответствующих эксплуатационных параметров и характеристик защищенной ИС далее будем использовать индекс «з». При этом исходные значения эксплуатационных параметров непосредственно ИС λ и μ не изменятся, но добавятся соответствующие эксплуатационные параметры системы защиты информации, $\lambda_{сзи}$, $\mu_{сзи}$, соответственно имеем $P_{0з} = f(\lambda, \mu, \lambda_{сзи}, \mu_{сзи})$. Отметим, что при расчете характеристик защищенной ИС, $T_{0з}$, T_3 , $P_{а3}$ и соответственно $P_{аиз}$, в качестве параметра μ при этом уже используется $\mu_{сзи}$ (при этом будем определять граничные, худшие значения соответствующих характеристик).

Не изменится и исходное (при реализации защиты) значение параметра λ_a , поскольку им характеризуется ценность для злоумышленника обрабатываемой в ИС информации, вне зависимости от того, защищена информационная система или нет.

С учетом сказанного приведенные ранее формулы для расчета основных введенных характеристик эксплуатационной информационной безопасности, применительно к защищенной ИС, примут следующий вид:

$$P_{аиз} = \begin{cases} 0, & \text{если } i = 1; \\ P_{а3}(1 - P_{а3})^{i-2}, & \text{если } i \geq 2; \end{cases}$$

$$P_{а3}(t) = \begin{cases} 0, & \text{если } 0 < t \leq T_{0з}; \\ \sum_{i=2}^I P_{аиз}, & \text{если } (I - 1)T_{0з} + \frac{(I - 2)}{\mu_{сзи}} < t \leq IT_{0з} + \frac{(I - 1)}{\mu_{сзи}}; \end{cases}$$

$$T_{0зз} = T_{0з} + \sum_{i=2}^{\infty} (i - 1)T_3 P_{аиз}.$$

Распределения характеристик P_a и P_{a3} во времени (качественная оценка), определяемых периодами эксплуатации системы $T_i, i = 1, \dots, I$, иллюстрирующие причины положительного эффекта от внедрения системы защиты, представлены на рис. 4.

Обсудим теперь дополнительные потери. Риск потерь для незащищенной ИС связан исключительно с успешной атакой на ИС, в результате которой злоумышленник получает несанкционированный доступ к обрабатываемой в ней информации. Внедрение в ИС системы защиты связано с дополнительными затратами (потерями), определяемыми собственно стоимостью внедряемой системы защиты $C_{сзи}$ и удельной стоимостью (стоимостью в единицу времени) ее эксплуатации $C_{уэсзи}(t)$ (техническая поддержка от разработчика, увеличение штата сотрудников, эксплуатирующих средство защиты и т.п. (предполагается, что она характеризуется линейной зависимостью), также являющейся эксплуатационной характеристикой уже средства защиты).

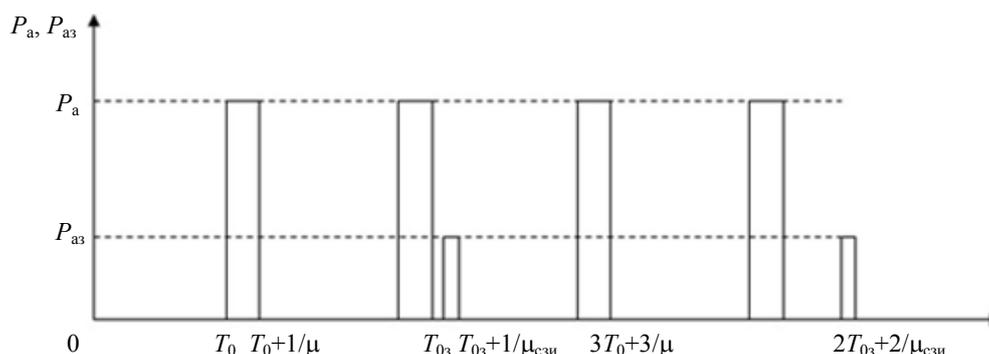


Рис. 4. Распределения характеристики P_a и P_{a3} во времени

Под риском потерь от осуществления успешной атаки на защищенную информационную систему будем понимать экономическую эксплуатационную характеристику защищенной ИС, определяемую совокупностью потерь – затрат, определяемых как стоимостью системы защиты и ее эксплуатации в системе, так и потенциальными потерями от успешной атаки на защищенную ИС.

С учетом сказанного ранее (с учетом полученной расчетной формулы для $P_{a3}(t)$), для любого произвольного времени эксплуатации ИС t риск потерь от осуществления успешной атаки на защищенную ИС, $R_{с_{инф3}}(t)$, может быть рассчитан следующим образом:

$$R_{с_{инф3}}(t) = \begin{cases} C_{сзи} + tC_{уэсзи}(t), & \text{если } 0 < t \leq T_{03}, \\ C_{сзи} + tC_{уэсзи}(t) + C_{инф} \sum_{i=2}^I P_{ai3}, & \text{если } (I-1)T_{03} + \frac{(I-2)}{\mu_{сзи}} < t \leq IT_{03} + \frac{(I-1)}{\mu_{сзи}}. \end{cases}$$

Используя полученные результаты, соответствующим образом определим оценки эффективности системы защиты как количественные меры снижения риска осуществления успешной атаки на ИС в течение времени ее эксплуатации t , $\Delta P_{a3}(t)$, и снижения риска потерь от осуществления успешной атаки на ИС в течение времени ее эксплуатации t , $\Delta R_{с_{инф3}}(t)$, которые могут быть рассчитаны по следующим формулам:

$$\Delta P_{a3}(t) = P_a(t) - P_{a3}(t),$$

$$\Delta R_{с_{инф3}}(t) = R_{с_{инф}}(t) - R_{с_{инф3}}(t).$$

Отметим, что, если значение $R_{с_{инф}}(t)$ в пределе стремится к $C_{инф}$, то значение $R_{с_{инф3}}(t)$ в пределе стремится к ∞ . Это позволяет утверждать, что величина выигрыша в потерях от реализации в ИС защиты информации будет сильно изменяться в процессе эксплуатации системы, причем по мере увеличения продолжительности эксплуатации системы в общем случае будет иметь две фазы распределения – фазу увеличения и фазу уменьшения (начальную фазу эксплуатации, которая, естественно, проигрышна, так как связана с дополнительными затратами на реализацию в ИС системы защиты информации, здесь не рассматриваем).

Качественная оценка изменения (распределения) характеристики $\Delta R_{с_{инф3}}(t)$ во времени проиллюстрирована на рис. 5.

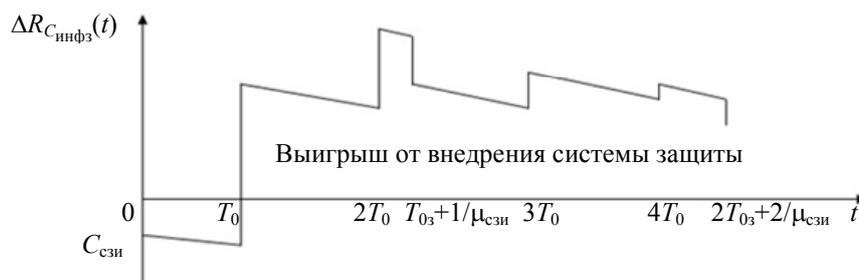


Рис. 5. Качественная оценка изменения характеристики $\Delta R_{C_{инфз}}(t)$ во времени

Как видим из рис. 5, характеристика $\Delta R_{C_{инфз}}(t)$ имеет пиковые значения в процессе эксплуатации системы. Это обусловливается различием в общем случае характеристик T и T_3 (соответственно для незащищенной и защищенной информационных систем).

Таким образом, используя введенные эксплуатационные параметры и характеристики информационной (в том числе защищенной) системы, эксплуатационные характеристики риска, применяя представленные расчетные формулы, можно количественно оценить выигрыш (проигрыш) от внедрения системы защиты для различных временных интервалов эксплуатации защищенной ИС, что позволяет получить экономическое обоснование целесообразности и эффективности реализации в ИС системы защиты и соответствующим образом, исходя из подобной оценки, сформировать требования к эксплуатационным параметрам и характеристикам системы защиты для конкретной ИС.

Заключение

Предложенный в работе подход к оцениванию риска, в том числе риска потерь, реализуемый в рамках излагаемой авторами теории эксплуатационной информационной безопасности, предоставляет возможность оценить распределение потерь в процессе эксплуатации информационной (в том числе защищенной) системы. Использование введенных эксплуатационных характеристик риска позволяет построить адекватную модель ИС с учетом нарушений и восстанавливаемости свойств ее информационной безопасности в процессе эксплуатации и соответственно получить адекватные требования к параметрам и характеристикам (эксплуатационным) системы защиты, предполагаемой к использованию в конкретной ИС.

Литература

1. Королев В.Ю., Бенинг В.Е., Шоргин С.Я. Математические основы теории риска. М.: ФИЗМАТЛИТ, 2011. 591 с.
2. Шапкин А.С., Шапкин В.А. Теория риска и моделирование рискованных ситуаций. М.: Дашков и К, 2005. 880 с.
3. Мельников В.П. Защита информации в компьютерных системах. М.: Финансы и статистика, 1997. 368 с.
4. Павлухин Д.В. Теория информационной безопасности и методология защиты информации: Учебно-методическое пособие. Тамбов: Изд-во ТГУ им. Г.Р. Державина, 2005. 104 с.
5. Малюк А.А. Информационная безопасность: концептуальные и методологические основы защиты информации: Учеб. пособие для вузов. М.: Горячая линия – Телеком, 2004. 280 с.
6. Петренко С.А., Симонов С.В. Управление информационными рисками. Экономически оправданная безопасность. М.: Компания АйТи; ДМК Пресс, 2004. 384 с.
7. Корченко А.Г. Построение систем защиты информации на нечетких множествах. Теория и практические решения. Киев: МК-Пресс, 2006. 320 с.
8. Мизнов А.С., Шевяков М.Ю. Некоторые подходы к оценке информационных рисков с использованием нечетких множеств // Системный анализ в науке и образовании. 2010. № 1. С. 54–60.
9. Щеглов К.А., Щеглов А.Ю. Защита от атак на уязвимости приложений. Модели контроля доступа // Вопросы защиты информации. 2013. № 2 (101). С. 36–43.
10. Саати Т. Элементы теории массового обслуживания и ее приложения. М.: Сов. радио, 1965. 511 с.
11. Половко А.М., Гуров С.В. Основы теории надежности. СПб: БХВ-Петербург, 2006. 704 с.
12. Шеннон К.Е. Математическая теория связи // Работы по теории информации и кибернетике: Пер. с англ. М.: ИИЛ, 1963. С. 243–332.
13. Щеглов К.А., Щеглов А.Ю. Практическая реализация дискреционного метода контроля доступа к создаваемым файловым объектам // Вестник компьютерных и информационных технологий. 2013. № 4. С. 43–49.

14. Щеглов К.А., Щеглов А.Ю. Система защиты от запуска вредоносных программ // Вестник компьютерных и информационных технологий. 2013. № 5. С. 38–43.
15. Щеглов К.А., Щеглов А.Ю. Реализация метода мандатного доступа к создаваемым файловым объектам системы // Вопросы защиты информации. 2013. № 4 (103). С. 15–20.

<i>Щеглов Константин Андреевич</i>	–	студент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; менеджер по развитию, ЗАО «НПП «Информационные технологии в бизнесе», Санкт-Петербург, Россия, schegl_70@mail.ru
<i>Щеглов Андрей Юрьевич</i>	–	доктор технических наук, профессор, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; генеральный директор, ЗАО «НПП «Информационные технологии в бизнесе», Санкт-Петербург, Россия, info@npp-itb.spb.ru
<i>Konstantin Shcheglov</i>		Student, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; development manager, JSC “Information Technologies in Business”, Saint Petersburg, Russia, Scheglov.konstantin@gmail.ru
<i>Andrei Shcheglov</i>		D.Sc., Professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; General director, JSC “Information Technologies in Business”, Saint Petersburg, Russia, info@npp-itb.spb.ru

УДК 004.89

ИДЕНТИФИКАЦИЯ АНОНИМНЫХ ПОЛЬЗОВАТЕЛЕЙ ИНТЕРНЕТ-ПОРТАЛОВ НА ОСНОВАНИИ ТЕХНИЧЕСКИХ И ЛИНГВИСТИЧЕСКИХ ХАРАКТЕРИСТИК ПОЛЬЗОВАТЕЛЯ¹

А.А. Воробьева^а, А.В. Гвоздев^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Alice_w@mail.ru

Задача идентификации анонимных пользователей Интернет-порталов становится все более актуальной научной задачей, это обусловлено ростом числа интернет-пользователей, в том числе анонимных, ростом числа случаев совершения противоправных действий (например, анонимных угроз и экстремистских высказываний) и несовершенством существующих подходов и алгоритмов идентификации анонимных пользователей.

В контексте работы под идентификацией пользователя понимается распознавание анонимного пользователя в Интернете [1–5]. Распознавание производится путем соотнесения набора характеристик анонимного пользователя с характеристиками, собранными ранее и уже имеющимися в базе данных. К характеристикам пользователя относятся ряд технических (IP-адрес, версия операционной системы и пр.) и лингвистических (стиль письменной речи автора сообщения) характеристик. В работе рассматривается возможность идентификации пользователей по различным наборам таких характеристик (техническим, лингвистическим и комбинированным). Анализируется возможность применения различных методов классификации (метод опорных векторов, нейросети, логическая регрессия) для решения задачи по идентификации анонимных пользователей.

Проведенные эксперименты показали, что использование лингвистических характеристик совместно с техническими позволяет повысить точность идентификации анонимного пользователя Интернет-портала.

Ключевые слова: идентификация анонимных пользователей, атрибуция текстов, авторство сообщений, компьютерная лингвистика, информационная безопасность.

ANONYMOUS WEBSITE USER IDENTIFICATION BASED ON COMBINED FEATURE SET (WRITING STYLE AND TECHNICAL FEATURES)²

A. Vorob'yeva^b, A. Gvozdev^b

^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Alice_w@mail.ru

The task of anonymous web users identification becomes more and more important research task. The number of users is increased dramatically and usage of the Internet for criminal purposes (such as anonymous threats and extremist statements) becomes more frequent. Existing approaches and algorithms for identifying anonymous users are not enough efficient. In the context of this work, user identification means recognizing of an anonymous user on the Internet. Identification is performed by correlating the set of anonymous user features with stored in the database features collected previously. Feature set of the user consists of technical features (IP- address, OS version, etc.) and writing-style features of the user (for short texts in the Russian language). We compared the discriminating power of three feature sets (technical, writing-style and combined) and

¹ Работа выполнена в рамках НИР «Идентификация пользователей порталов сети Интернет».

² Done in the framework of S&R work «Identification of Internet portals users»

of three classification methods (Support Vector Machines, neural networks, logistic regression). Results of the experiment showed that the usage of combined feature set (writing-style and technical features) improves the identification accuracy of an anonymous user of the Internet.

Keywords: anonymous users' identification, text attribution, authorship of text messages, author attribution, computational linguistics, information security.

Введение

Вопрос анонимности в Интернете вызывает большое число дискуссий и споров, нельзя сказать однозначно, является ли она положительной или отрицательной его чертой. Необходимость идентификации анонимных пользователей возникает в случае, когда через Интернет были совершены какие-либо противоправные действия, например, в случае, когда производятся анонимные «вбросы» [6] (сознательная дезинформация), либо преступник с помощью Интернет-переписки готовит почву для совершения преступления. Новые методы решения задачи идентификации пользователей могут быть применимы в таких сферах, как компьютерная криминалистика (например, определение автора анонимных злонамеренных сообщений), противодействие терроризму (отнесение текстов с экстремистским содержанием к какому-то известному террористу) и др. [1–13]. Важность задачи идентификации обоснована в работах многих авторов.

В контексте настоящей работы под пользователем портала сети Интернет понимается конкретное физическое лицо, которое своими действиями с ресурсами портала обнаруживает некоторые признаки (характеристики пользователя), определяющиеся особенностями стиля его письменной речи или техническими средствами, которые он использует для доступа в Интернет.

Целью работы является оценка возможности применения различных наборов характеристик и алгоритмов классификации для идентификации пользователей Интернет-порталов.

Задача идентификации пользователя Интернет-порталов

Задача идентификации пользователя в общем случае сводится к решению задачи многоклассовой классификации. Имеется множество объектов – наборов характеристик пользователей $F = \{f_1, \dots, f_n\}$, а также множество пользователей $U = \{u_1, \dots, u_k\}$, $u_j = f_k$. Для некоторого подмножества характеристик $F' = \{f_1, \dots, f_m\} \subseteq F$ известна их принадлежность определенному пользователю, т.е. существует множество пар «характеристики–пользователь». Необходимо установить, кому из множества U действительно принадлежат остальные характеристики $F' = \{f_{m+1}, \dots, f_n\} \subseteq F$, т.е. требуется построить алгоритм $a: F \rightarrow U$, способный классифицировать произвольный набор характеристик из исходного множества – $f_i \in F$. Множество U содержит множество «характеристики–известные пользователи», F_{tr} – обучающая выборка, F' – классифицируемые характеристики.

Характеристики пользователей

При идентификации пользователя одной из главных задач является определение набора характеристик пользователя, выступающих в качестве сопоставляемых ему атрибутов (признаков), и позволяющие получить наиболее высокие показатели качества классификации. Существует достаточно большое количество возможных характеристик, начиная с самых простых технических (IP-адрес и версия операционной системы (ОС) [4]) и лингвистических (частоты слов определенной длины). Также используются более сложные лингвистические характеристики, требующие некоторого синтаксического или семантического анализа [2]. Вторая задача – это разработка методов идентификации пользователя, или классификации набора характеристик анонимного пользователя к известным пользователям. Авторами был отобран ряд моделей для проведения идентификации, в предыдущих исследованиях показавших наилучшую точность классификации [1–13].

Группируем характеристики пользователя $U = \{F_t, F_l\}$ по их типу – технические (F_t), лингвистические (F_l). В табл. 1 приводятся характеристики, используемые в качестве атрибутов (признаков) пользователя, а также описание особенностей их использования при идентификации.

Полагая, что пользователю свойственно использовать несколько устройств для выхода в Интернет, а также сопоставляя ему как конкретному физическому лицу определенные характеристики текста на естественном языке, определим пользователя u_i как кортеж, координатам которого сопоставляется множество характеристик технических средств f_{t_i} и характеристик, обусловленных стилистическими особенностями его письменной речи, f_{l_i} . Пользователь представляется как

$$u_i = f_i = \{f_{t_i}, f_{l_i}\}, u_i \in U, f_i \in F, f_{t_i} \in F_t, f_{l_i} \in F_l,$$

где f_i – совокупные характеристики пользователя; F – множество возможных совокупностей характеристик пользователей; f_{t_i} – характеристики технических средств пользователя; F_t – множество возможных значений характеристик технических средств; f_{l_i} – лингвистические характеристик текста; F_l – множество возможных значений характеристик текста.

Существуют определенные ограничения при использовании характеристик технических средств пользователя. Для каждого идентифицируемого пользователя они не являются уникальными и характери-

зуют лишь устройство (аппаратное и программное окружение), но никак не самого пользователя. Выбор оптимальной совокупности таких признаков приведен многими авторами, например, в [4, 11].

Отдельно стоит отметить, что при проведении постинцидентного анализа или расследования технические характеристики пользователя получить невозможно, исключения составляют только те случаи, когда сбор этих характеристик ведется на сервере заблаговременно. Лингвистические характеристики этим недостатком не обладают, их использование возможно во всех случаях, пока сообщения пользователя существуют на Интернет-портале и не удалены из базы данных.

Характеристики пользователя	Особенности использования при идентификации
Технические характеристики (F_T)	
Версия ОС Часовой пояс Версия обозревателя Интернет Язык браузера по умолчанию Разрешение экрана	В зависимости от способа получения достоверность определения этих значений может меняться. Легко подделать, поскольку, как правило, их легче всего получить из заголовков HTTP-запроса. Имеет смысл применение активных компонентов на странице для получения более достоверных сведений.
IP-адрес	Извлечение значения этого признака осуществляется на уровне веб-сервера. Несколько физических лиц могут использовать один и тот же IP-адрес 4-й версии. Применение активных компонентов не рассматривается (небезопасно для ОС конечного пользователя).
Лингвистические характеристики (F_L)	
Общее число символов (S) Частота буквенных символов Частота заглавных букв Частота чисел Частота пробелов Частоты управляющих символов Общее число слов (W) Частоты длин слов Частота коротких слов Средняя длина слова Средняя длина предложения в буквах Средняя длина предложения в словах Частота коротких предложений Частота средних предложений Частота длинных предложений Частоты знаков пунктуации Частота использования ссылок Частота использования изображений и другие.	<p>Выбор характеристик текста в качестве классификационных признаков. Самыми важными отличительными характеристиками являются слова с самыми высокими частотными характеристиками, т.е. наиболее часто употребляемые автором.</p> <p>Недостаточное количество и краткость текстов обучающей выборки. В реальности решение задачи идентификации пользователя усложняется тем, что, как правило, доступно всего лишь несколько достаточно коротких текстов, бесспорно принадлежащих пользователям-кандидатам.</p> <p>Дисбаланс классов. Стандартной является ситуация, когда существует неравномерное распределение длин текстов обучающей выборки различных пользователей, что может вызвать некоторый дисбаланс и привести к некорректным результатам. Недостаточная длина текстов обучающей выборки одного пользователя по сравнению с другими не должна снижать вероятность того, что данный пользователь будет выбран как истинный автор текста.</p>

Таблица 1. Характеристики пользователя Интернет-порталов и особенности их использования в задачах идентификации

Эксперимент

Корпус текстов для проведения данного исследования был составлен из текстов на русском языке. Этот корпус был специально составлен для проведения исследований по авторской идентификации пользователя, он представляет собой ограниченную коллекцию текстов различной тематики, для которых автор доподлинно известен, тексты не проходили никакой иной дополнительной обработки. Все тексты относились к различной тематике. Отбор по тематике не производился для того, чтобы эксперимент был максимально приближен к реальной жизненной ситуации, хотя это позволило бы минимизировать влияние тематики на классификацию.

Для составления набора текстов обучающей выборки выбиралось 25 последних сообщений пользователей, все сообщения были произвольной длины. Наиболее частыми были сообщения длиной до 2700 символов (73%), сообщения длиной 2700–5500 и 5500–8200 составляли 10% и 6,5% случаев соответственно. Корпус являлся несбалансированным. Обучающая и тестовая выборки не пересекались.

При выборе технических характеристик для идентификации в комбинации с лингвистическими характеристиками были отобраны только те, которые можно получить этичными методами, безопасными для пользователя, что накладывает определенные ограничения их применения для идентификации исключи-

тельно в неагрессивной среде, поскольку все их значения легко подделать при просмотре веб-страниц. Сопоставление признаков Etag и Cookie с идентификатором пользователя сопряжено с применением специальных методов выработки уникальных значений этих признаков для пользователя, применением специальных средств аутентификации в агрессивной среде, когда пользователь намеренно пытается обойти систему безопасности портала. Такой случай в настоящей работе не рассматривается, как это, например, сделано другими авторами в [4]. Извлечение же MAC-адреса (адреса устройства Media access control) вообще неэтично и сопряжено с техническими трудностями применения активных компонентов на странице; к тому же MAC однозначно определяет только устройство, но не пользователя. При этом несколько лиц могут пользоваться одним и тем же устройством. Для классификации были выбраны следующие признаки, которые легко получить из заголовков HTTP и журнала веб-сервера: версия ОС, версия интернет-браузера, языка интернет-браузера по умолчанию, часовой пояс, разрешение экрана, IP-адрес версии 4.

Для решения задачи идентификации анонимных пользователей было протестировано несколько алгоритмов $a: F \rightarrow U$, способных классифицировать произвольный набор характеристик из исходного множества характеристик тестовой выборки: метод опорных векторов (Support Vector Machine, SVM), многослойный перцептрон (NN), наивный байесовский классификатор (NB) и методы логистической регрессии (LR) [12].

В качестве классификационных признаков использовались характеристики, приведенные в табл. 1. Основой проверки является тестовая выборка, в которой отражено соответствие между документами и их классами. Для проверки качества работы алгоритма классификации необходимо применить его на тестовой выборке и соотнести его решение с заведомо известным правильным решением.

Результаты

При решении задачи идентификации пользователя на качество классификации могут оказывать влияние различные факторы: общее количество пользователей, типы используемых характеристик и сам метод классификации. Именно поэтому далее приводится сравнение точности классификации при использовании различных наборов характеристик в качестве признаков пользователя.

Проверка качества работы алгоритмов классификации производилась на тестовой выборке, в которой проставлено соответствие между характеристиками и пользователями, которым они принадлежат. Численная оценка качества работы алгоритма производилась путем оценки его точности (ассигуры). За точность идентификации (A) принимается отношение количества правильно классифицированных наборов характеристик ($f_i \in F'$) – C_c , к общему числу классифицируемых наборов характеристик C_t (размер тестовой выборки) [14].

$$A = \frac{C_c}{C_t} 100\%.$$

Для каждого набора характеристик проводилось тестирование каждого выбранного ранее алгоритма классификации и производилось сравнение с результатами, полученными при использовании других алгоритмов. Результаты для каждого набора характеристик (лингвистические, технические, комбинированные) приведены в табл. 2. Напомним, что во всех случаях используется один тестовый корпус, так что результаты являются сопоставимыми. Использование технических характеристик для идентификации пользователя является традиционным подходом, однако существуют значительные ограничения для использования этих характеристик при решении задачи идентификации пользователя. Идентификация пользователей на основе лингвистических характеристик для русского языка является достаточно новой задачей, требующей дополнительных углубленных исследований, но именно это направление признается наиболее перспективным. Из рисунка и табл. 2 видно, что использование предложенного комбинированного набора характеристик позволяет достичь гораздо большего качества классификации по сравнению с другими характеристиками, точность составляет 90,4%.

	Технические характеристики (Ft)	Лингвистические характеристики (Fl)	Комбинированные характеристики ($Ft+Fl$)
SVM	72,0	60,0	18,4
NN	72,1	61,6	90,4
NB	66,0	68,8	77,6
LR	68,9	79,2	89,6

Таблица 2. Точность идентификации (%) при использовании различных наборов характеристик пользователя

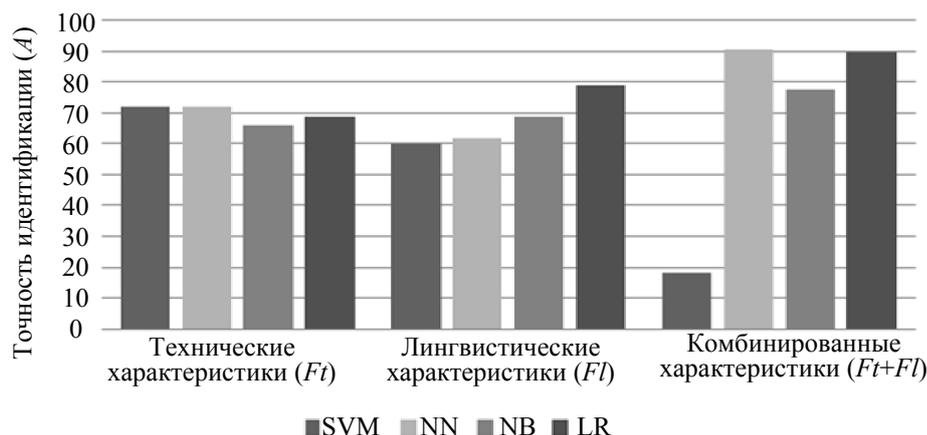


Рисунок. Точность идентификации пользователя при использовании разных наборов характеристик

Заключение

В работе предложен новый подход к решению задачи идентификации автора на основании комбинированного набора характеристик – лингвистических признаков, характеризующих его письменную речь, и признаков, характеризующих терминал его технических средств, которые он использует для доступа к веб-сайту. Предложенный подход может быть использован для классификации в условиях малого объема обучающей выборки. Идентификация пользователей – это один из наиболее характерных примеров, отражающих суть данной проблемы, так как обычно имеется крайне мало сведений о предыдущих характеристиках пользователя, особенно когда дело касается криминалистики. Совместное применение этих признаков позволяет значительно повысить точность идентификации с использованием предложенных алгоритмов. Неожиданно низким является результат применения метода опорных векторов на комбинированных характеристиках. Проведенные эксперименты доказали эффективность идентификации с использованием комбинированного набора характеристик при работе с ограниченной тестовой выборкой.

Сравнение предложенных методов классификации и наборов характеристик показало, что выбор наиболее информативных характеристик является определяющим фактором для получения высокой точности идентификации. С другой стороны, открытым остается вопрос о том, какой из предложенных методов является лучшим. Здесь требуются дальнейшие исследования.

Литература

1. de Vel A., Anderson O., Corney M., Mohay G. Mining e-mail content for author identification forensics // SIGMOD Record. 2001. V. 30. N 4. P. 55–64.
2. Zheng R., Li J., Chen H., Huang Z. A Framework for Authorship Identification of Online Messages: Writing Style Features and Classification Techniques // Journal of the American Society of Information Science and Technology. 2006. V. 57. N 3. P. 378–393.
3. Iváncsy R., Juhász S. Analysis of Web User Identification Methods // International Journal of Computer Science. 2007. V. 2. N 3. P. 172–177.
4. Бессонова Е.Е., Зикратов И.А., Росков В.Ю. Анализ способов идентификации пользователя в сети Интернет // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 6 (82). С. 128–129.
5. Романов А.С., Шелупанов А.А., Бондарчук С.С. Обобщенная методика идентификации автора неизвестного текста // Доклады Томского государственного университета систем управления и радиоэлектроники. 2010. № 1 (21). Ч. 1. С. 108–112.
6. Гвоздев А.В., Лебедев И.С., Зикратов И.А. Вероятностная модель оценки информационного воздействия // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 2 (78). С. 99–103.
7. Abbasi A., Chen H. Applying Authorship Analysis to Extremist-group Web Forum Messages // IEEE Intelligent Systems. 2005. V. 20. N 5. P. 67–75.
8. Park T., Li J., Zhao H., Chau M. Analyzing writing styles of bloggers with different opinions // Proc. of the 19th Annual Workshop on Information Technologies and Systems (WITS 2009). Phoenix, Arizona, USA, 2009. P. 151–156.
9. Layton R., Watters P., Dazeley R. Authorship attribution for twitter in 140 characters or less // Second Cyber-crime and Trustworthy Computing Workshop (CTC-2010). Ballart, VIC, Australia, 2010. P. 1–8.

10. Zheng R., Li J., Chen H., Huang Z. Authorship analysis in cybercrime investigation // Proc. of the 1st NSF/NIJ conference on Intelligence and security informatics (ISI'03). Berlin-Heidelberg: Springer-Verlag, 2003. P. 59–73.
11. Eckersley P. How Unique is Your Web Browser? // Lecture Notes in Computer Science. 2010. V. 6205. P. 1–18 [Электронный ресурс]. Режим доступа: <https://panopticlick.eff.org/browser-uniqueness.pdf>, свободный. Яз. англ. (дата обращения 26.11.2013).
12. Stamatatos E. A Survey of Modern Authorship Attribution Methods // Journal of the American Society for Information Science and Technology. 2009. V. 60. N 3. P. 538–556.
13. Nawrot M. Automatic Author Attribution for Short Text Documents // Lecture Notes in Computer Science. 2011. V. 6562. P. 468–477.
14. Маннинг К.Д., Рагхаван П., Шютце Х. Введение в информационный поиск. М.: Вильямс, 2011. 528 с.

<i>Воробьева Алиса Андреевна</i>	– ассистент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Alice_w@mail.ru
<i>Гвоздев Алексей Вячеславович</i>	– ассистент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, a.gvozdev@rcwg.net
<i>Alice Vorob'yeva</i>	assistant, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Alice_w@mail.ru
<i>Aleksei Gvozdev</i>	assistant, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, a.gvozdev@rcwg.net

УДК 520.224.2. 224.4

**МЕТОДИКА РАСЧЕТА КОЭФФИЦИЕНТОВ ОБЛУЧЕННОСТИ
ЦИЛИНДРИЧЕСКОГО КОСМИЧЕСКОГО ОБЪЕКТА ПОДСВЕТКОЙ ЗЕМЛИ**А.М. Дзитолев^а, С.И. Ханков^а^а Военно-космическая академия им. А.Ф. Можайского, Санкт-Петербург, Россия, Dzitoi8@gmail.com

Разработана приближенная методика расчета эффективных коэффициентов облученности космического объекта цилиндрической формы, которая типична для наружных корпусов космических телескопов, а также фрагментов космических аппаратов. В настоящее время отсутствуют аналитические методики расчета интегральных и эффективных коэффициентов облученности для такой конфигурации космических объектов. Интегральные коэффициенты облученности определяют средний тепловой баланс на цилиндрической поверхности, а эффективные – на всей поверхности цилиндра с учетом его торцов. Расчеты эффективных коэффициентов облученности космических объектов подсветкой Земли необходимы для определения важнейшей компоненты энергетического баланса – удельной мощности собственного теплового излучения Земли. Такие расчеты используются для определения средних температур космических объектов. Методика основана на обобщении и аппроксимации результатов численных расчетов в диапазоне изменения высоты орбиты от 200 км до 40000 км, а также во всем возможном диапазоне изменения угла наклона оси цилиндра. Показаны неопределенности в математической модели при небольших отклонениях положения оси цилиндра от линии зенит–надир. Исследованы погрешности приближенного решения при определении интегрального коэффициента облученности боковой цилиндрической поверхности. Допустимая итоговая погрешность полученного приближенного решения оценивалась по отклонению температуры объекта в тени Земли от ее точного значения. При больших отношениях высоты цилиндра к диаметру погрешности расчета интегрального коэффициента облученности максимальны и могут приводить к погрешности определения температур в тени Земли до 1,5 К. Погрешности расчета температуры цилиндра уменьшаются с уменьшением отношения его высоты к диаметру. Они минимальны для модели тонкого диска. Погрешности расчета температур значительно снижаются для случая нахождения цилиндрического объекта на подсвеченном Солнцем участке траектории, а еще более – при расчетах нестационарного теплового режима. Предложенная методика расчета достаточно точна, проста и удобна для расчетов.

Ключевые слова: космический объект цилиндрической формы, коэффициент облученности космического объекта, удельный тепловой поток излучения Земли, тепловой баланс космического объекта.

**CALCULATION METHODS FOR IRRADIANCE COEFFICIENTS
OF CYLINDRICAL SPACE OBJECT BY THE EARTH RADIATION**A. Dzitoev^b, S. Khankov^b^b Military Space Academy n.a. A.F. Mozhaisky, Saint Petersburg, Russia, Dzitoi8@gmail.com

The paper deals with approximate calculation methods for effective irradiance coefficients of cylindrical form space object typical for external cases of space telescopes and fragments of spacecrafts as well. There are no analytical calculation methods nowadays for integrated and effective irradiance coefficients for space objects of such configuration. Integrated irradiance coefficients define the average thermal balance on a cylindrical surface and effective ones – all over the cylinder surface taking into account its butt ends. Calculations of effective coefficients for space objects irradiance by the Earth radiation are necessary for definition of the major component of power balance – the specific power of the Earth inherent thermal radiation. Such calculations are used for determination of average temperatures of space objects. The technique is based on generalization and approximation of results for numerical calculations in the range of orbit height change from 200 km to 40000 km and also in all possible range of a tilt angle change of a cylinder axis. Uncertainties in mathematical model at small location deviations for cylinder axis from zenith-nadir line are shown. Errors of the approximate solution are analyzed during determination of integrated irradiance coefficient for a lateral cylindrical surface. The admissible total error of the received approximate solution was estimated at a temperature deviation of an object in the Earth shadow from its exact value. At the large ratios of the cylinder height to its diameter calculation, errors of integrated irradiance coefficient are maximum and can cause temperature error definitions in the Earth shadow up to 1,5 K. Errors of cylinder temperature calculation decrease with reduction of its height-to-diameter ratio. They have minimum values for a thin disk model. Temperatures calculation errors are considerably decreased for a case of cylindrical object location on the trajectory section illuminated by the Sun, and even more – at calculations of a non-stationary thermal mode. The offered calculation technique is rather accurate, simple and convenient for calculations.

Keywords: cylindrical form space object, irradiance coefficient of space object, specific thermal radiation flux of the Earth, space object thermal balance.

Введение

При расчетах нестационарных тепловых режимов космических объектов (КО) необходимо определение величин падающих и поглощаемых поверхностью КО тепловых потоков от прямого и отраженного Землей солнечного излучения, а также собственного теплового излучения Земли [1–4]. Первые две ком-

поненты теплового баланса КО вычисляются следующим образом. Прямая солнечная засветка поверхности КО определяется через солнечную постоянную с учетом площади его миделя и коэффициента поглощения солнечного излучения [1]. Эта компонента теплового баланса вычисляется без особых затруднений. Вторая компонента вычисляется через так называемый комбинированный коэффициент облученности, расчет которого связан с большими затруднениями [1] и, как правило, осуществляется численными методами. Третья компонента теплового баланса определяется через величину удельного теплового потока в полном спектре собственного теплового излучения Земли, а также через коэффициент облученности, который представляет собой отношение падающего на поверхность КО удельного теплового потока к тепловому потоку, излучаемому Землей с верхней границы атмосферы. В этом потоке учитывается тепловое излучение атмосферы и излучение подстилающей земной поверхности, проходящее через окна прозрачности атмосферы в спектре инфракрасного теплового излучения [1].

Известно, что спектр собственного теплового излучения Земли с учетом атмосферы имеет ярко выраженный селективный характер. Однако на практике принято пользоваться для расчетов теплового баланса интегральными удельными тепловыми потоками [1]. По этой причине значительные средства затрачиваются на экспериментальные определения глобальной (усредненной по всей поверхности Земли) величины излучаемой Землей в космическое пространство удельной мощности [5, 6]. Такие экспериментальные исследования, проводимые с космических аппаратов (КА), позволили уточнить эту величину с 235 Вт/м^2 [5] до 239 Вт/м^2 [6].

Для расчетов тепловых потоков, поглощаемых КО, необходимо определение интегрального коэффициента облученности (ИКО) [7–12] КО земной подсветкой в зависимости от его конфигурации. ИКО определяет удельный тепловой поток, который усредняется по всей поверхности КО и вычисляется интегрированием локальных коэффициентов облученности для каждого элемента поверхности. Для сферического объекта коэффициент облученности описывается просто [1, 12], что облегчает расчет нестационарных температур сферического КО на любой орбите [12–15]. Большой практический интерес представляет решение той же задачи для КО цилиндрической формы, которую имеют многие реальные объекты, в том числе корпуса телескопов, базируемых на борту КА [16]. Сами КА также содержат фрагменты цилиндрической формы. Это определяет актуальность данной задачи.

Для объектов с конфигурацией, составленной из разного рода поверхностей (параллелепипед, цилиндр, конус) используется эффективный коэффициент облученности (ЭКО), который описывает удельный тепловой поток, поглощенный всей поверхностью КО. Частью расчета ЭКО является определение ИКО цилиндрической боковой поверхности. Для цилиндрического КО (ЦКО) в эффективный коэффициент облученности входят также коэффициенты облученности Землей торцов цилиндра [14, 15].

Целью настоящей работы являлось получение приближенного аналитического описания зависимости эффективного коэффициента облученности ЦКО от высоты орбиты, угла наклона его оси относительно плоскости местного горизонта и отношения высоты цилиндра к его диаметру.

Разрабатываемая методика должна основываться на аналитических формулах, составляющих алгоритм расчетов, исключаящий необходимость проведения интегрирования громоздких выражений для локального коэффициента облученности.

Физическая и математическая модели процесса подсветки цилиндра собственным тепловым излучением Земли

Стационарная температура T КО произвольной выпуклой формы в тени Земли может быть описана формулой [10–14]

$$T = T_e \sqrt[4]{\frac{\varphi_e}{(1 - \varphi_e)}}; \quad T_e = \sqrt[4]{\frac{Q}{\sigma}}, \quad (1)$$

где $T_e = 255 \text{ К}$ – эффективная температура Земли [15, 16]; φ_e – ЭКО объекта; $Q = 239 \text{ Вт/м}^2$ – удельная мощность собственного теплового излучения Земли [15, 16]; $\sigma = 5,67 \cdot 10^{-8} \text{ Вт/(м}^2 \cdot \text{К}^4)$ – постоянная Стефана–Больцмана.

Для ЦКО ЭКО определяется на основании данных [12, 13] в виде

$$\varphi_c = \frac{\varphi_b + 2n\varphi_{int}}{1 + 2n}; \quad n = \frac{L}{d}, \quad (2)$$

где φ_b – коэффициент облученности торцевой поверхности цилиндра (в расчетных формулах учитывается наличие двух торцов); φ_{int} – интегральный коэффициент облученности боковой поверхности цилиндра; L – высота цилиндра; d – диаметр его основания; φ_c – эффективный коэффициент облученности.

На рис. 1 представлены варианты ориентации оси цилиндра относительно линии зенит–надир; α – угол наклона оси ЦКО относительно этой вертикальной линии.

При вертикальной ориентации цилиндра, т.е. при $\alpha = 0$ (рис. 1, а), верхний торец цилиндра не подсвечен Землей, коэффициент облученности нижнего торца вычисляется как для площадки, лежащей в плоскости местного горизонта [1, 10]

$$\varphi_0 = (1+a)^{-2}; \quad a = \frac{h}{R}, \quad (3)$$

где h – высота площадки над поверхностью Земли; $R = 6371$ км – радиус Земли.

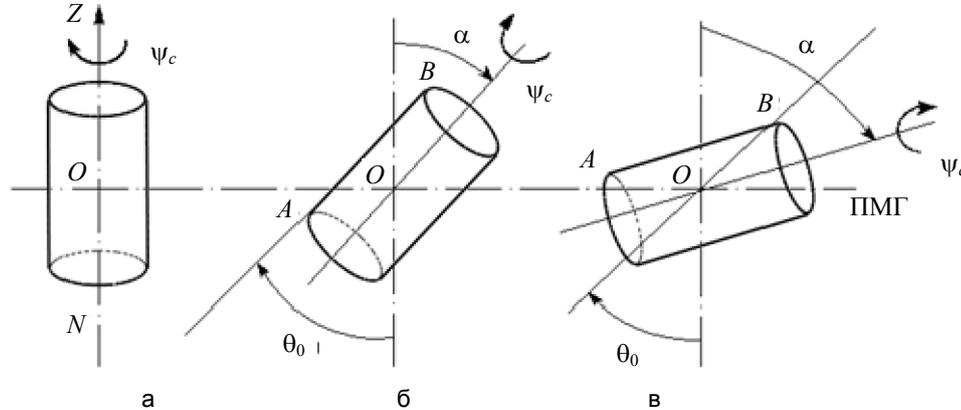


Рис. 1. Варианты ориентаций оси цилиндра относительно линии зенит–надир (ZN) и плоскости местного горизонта (ПМГ): а – вертикальная ориентация оси цилиндра; б – ориентация оси цилиндра под углом α , равным углу θ_0 ; в – ориентация оси цилиндра под углом α , большим угла θ_0 и меньшим 90° . Углы соответствуют: ψ_c – угловая координата по цилиндрической поверхности; α – угол наклона оси цилиндра, отсчитываемой от оси ZN; θ_0 – половинный угол, под которым из любой точки цилиндра видна Земля.

Линия АВ на образующей цилиндрической поверхности – наиболее удаленная от Земли

Для боковой поверхности цилиндра в этом случае расчет проводится как для вертикально ориентированной площадки по формуле [1, 10]

$$\varphi_{bc} = \frac{1}{\pi} \left(\arcsin \sqrt{\varphi_0} - \sqrt{\varphi_0} \cdot \sqrt{1 - \varphi_0} \right), \quad (4)$$

где φ_{bc} – коэффициент облученности вертикальной площадки, одновременно локальный и интегральный; индексы b и c относятся к боковой поверхности цилиндра.

Для горизонтально ориентированного цилиндра также имеются расчетные формулы для определения ЭКО [15, 16]. Для промежуточных ориентаций оси $0 < \alpha < \pi/2$ (рис. 1, б, в) расчеты существенно усложняются. При наклоне оси цилиндра относительно оси ZN на угол α (рис. 1) угол наклона ψ элемента поверхности цилиндра, задаваемого угловой координатой на цилиндрической поверхности ψ_c , отсчитываемой от образующей, наиболее близкой к Земле, может быть определен с использованием данных [1, 2] с помощью соотношения

$$\psi = \arccos x; \quad x = \sin \alpha \cdot \cos \psi_c. \quad (5)$$

Важным параметром, определяющим характеристики облученности элемента поверхности КО любой формы, является половинный угол θ_0 , под которым с этого элемента поверхности видна Земля, описываемый формулой [1, 12]

$$\theta_0 = \arcsin(\sqrt{\varphi_0}).$$

Зависимость локального коэффициента облученности плоского элемента поверхности собственным тепловым излучением Земли от угла наклона касательной к этому элементу, а также от высоты этого элемента над поверхностью Земли описывается известными соотношениями [1, 12], которые в наиболее компактной форме могут быть выражены через x в формуле (5). При этом в зависимости от величины угла наклона ψ коэффициент облученности описывается одной из трех формул:

$$\varphi_1 = \varphi_0 \cdot x \text{ при } 0 \leq \psi \leq \frac{\pi}{2} - \theta_0; \quad \varphi_3 = 0 \text{ при } \frac{\pi}{2} + \theta_0 \leq \psi \leq \pi; \quad (6)$$

$$\varphi_2 = \frac{\varphi_1}{2} \left[1 + \frac{2}{\pi} \arcsin \left(\sqrt{\frac{1}{\varphi_0} - 1} \cdot \frac{x}{\sqrt{1-x^2}} \right) \right] + \frac{1}{\pi} \arcsin \sqrt{\frac{\varphi_0 - x^2}{1-x^2}} - \frac{1}{\pi} \sqrt{(1-\varphi_0) \cdot (\varphi_0 - x^2)}. \quad (7)$$

Формула (7) используется при $\frac{\pi}{2} - \theta_0 \leq \psi \leq \frac{\pi}{2} + \theta_0$.

ИКО цилиндра определяется путем интегрирования (6) и (7) в указанных при этих формулах пределах. Затем полученные интегралы суммируются и делятся на π .

Для расчетов ЭКО ЦКО по формуле (2) необходимо определить φ_b и φ_{int} .

При определении коэффициентов облученности торцов ЦКО введены обозначения: верхний торец, наиболее удаленный от поверхности Земли, характеризуется коэффициентом облученности φ_{bu} , нижний торец, ближний к поверхности Земли, – коэффициентом облученности φ_{bl} . Тогда в (2) вместо φ_b подставляется

$$\varphi_b = 0,5(\varphi_{bu} + \varphi_{bl}). \quad (8)$$

Для расчета коэффициентов облученности торцов ЦКО в формулы (6) и (7) подставляется значение $x = \cos \psi$. Зависимость угловой координаты ψ от угла наклона оси ЦКО относительно плоскости местного горизонта для нижнего и верхнего торцов определяются соотношениями $\psi_l = \alpha$; $\psi_u = \pi - \alpha$. Использование формулы (8) обеспечивает предельные переходы и непрерывность ЭКО во всем диапазоне изменения угла наклона оси цилиндра α .

Методика расчета интегрального коэффициента облученности боковой поверхности цилиндра

При расчете ИКО ЦКО в пределах изменения угла наклона оси цилиндра $0 < \alpha < \theta_0$ возникает принципиальная неопределенность в математической модели. При $\alpha = 0$ (рис. 1, а) расчет φ_{int} в формуле (2) заменяется расчетом коэффициента облученности по формуле (4) – вариант модели 1 (ВМ 1). Эта модель предполагает использование локальных коэффициентов облученности φ_1 и φ_2 из (7) при малых величинах x . Однако границы допустимых значений α и соответственно x неопределенны. При значительных величинах угла наклона оси, вплоть до $\alpha = \theta_0$, ВМ 1 неприменима. В то же время до $\alpha \leq \theta_0$ неправомерно использование уравнений (6) и (7) для определения ИКО на основе интегрирования φ_1 и φ_2 . Для $\alpha > \theta_0$ (рис. 1, в) ИКО определяется интегрированием по участкам координат φ_1 и φ_2 – вариант модели 2 (ВМ 2).

Проведенные расчетные исследования подтвердили указанную неопределенность в выборе математической модели на участке $0 < \alpha < \theta_0$. При использовании ВМ 1 и ВМ 2 вычисленные значения коэффициентов облученности боковой поверхности ЦКО на одних и тех же диапазонах изменения α недостаточно точно стыкуются друг с другом. Это обосновывает необходимость получения хотя бы приближенного единого описания ИКО во всем диапазоне изменения угла наклона $0 \leq \alpha \leq \pi$ на основе получения аппроксимационных функций с определением погрешности получаемого конечного решения.

Зависимость коэффициента облученности φ_a боковой поверхности цилиндра от высоты и угла наклона оси аппроксимировалась формулами

$$\varphi_a = \varphi_b [1 - f(\alpha)] + \varphi_m \cdot f(\alpha); \quad \varphi_b = \varphi(\alpha = 0); \quad \varphi_m = \varphi(\alpha = \pi/2), \quad (9)$$

где φ_b и φ_m – функции только высоты; $f(\alpha)$ – функция только угла наклона оси ЦКО.

Зависимость $f(\alpha)$ была аппроксимирована функцией вида

$$f(\alpha) = a_1 \cdot \sin(b_1 \cdot \alpha + c_1) + a_2 \cdot \sin(b_2 \cdot \alpha - c_2), \quad (10)$$

которая изменяется от нуля при $\alpha = 0$ до единицы при $\alpha = \pi/2$.

Зависимости $\varphi_m(h)$ аппроксимирована формулой

$$\varphi_m = a_3 \exp(-b_3 H) + a_4 \exp(-b_4 H); \quad H = \frac{h}{1000}. \quad (11)$$

Величина H соответствует высоте траектории, исчисляемой в тысячах километров.

Для обеспечения минимальных погрешностей аппроксимационных зависимостей (9)–(11) относительно исходных расчетных зависимостей ИКО, определяемых по формулам (4), (6) и (7), коэффициенты в аппроксимационных формулах определялись для пяти интервалов высот H . Значения коэффициентов аппроксимации представлены в таблице.

Погрешность аппроксимации в максимуме зависимости, при $\alpha = \pi/2$, не превышает 1%. В то же время максимальная погрешность, реализуемая в области $20^\circ < \alpha < 60^\circ$, достигает 2,3%, 4%, 4,5%, –5,7%, –4,7% при H равном 0,2–2, 2–5, 5–10, 10–20, 20–40 соответственно. Для вытянутого цилиндрического стержня ($n > 10$) из (2) следует: $\varphi_c \approx \varphi_{int}$, поэтому для расчетов температур по формуле (1) достаточно использования ИКО. Погрешности определения стационарных температур ЦКО в тени Земли составляют для указанных диапазонов 1,2 К, 1,5 К, 1,4 К, –1,2 К и –0,8 К, что вполне удовлетворительно для практических расчетов. На подсвеченном Солнцем участке траектории погрешности расчетов температур будут еще меньше, а при расчетах нестационарных температур при движении ЦКО по траектории с попеременной сменой солнечных и теневых участков погрешности расчетов температур будут дополнительно сглаживаться за счет термической инерции ЦКО при его достаточно большой полной теплоемкости.

На рис. 2 представлены типовые зависимости φ_{int} и φ_c от угла наклона оси цилиндра α для разных высот ЦКО над поверхностью Земли в наиболее важном диапазоне изменения h . На рис. 2 представлены

вычисленные по разработанным формулам зависимости от угла наклона оси цилиндра для высот от 200 км до 2000 км. Сплошными линиями указаны зависимости ЭКО для отношения высоты цилиндра к диаметру его основания $n = 3$, штриховыми – ИКО боковой цилиндрической поверхности. Там же приведены для сравнения значения ИКО сферического КО.

№ п/п	Коэффициенты аппроксимации	Диапазон высот H в тысячах километров				
		0,2–2	2–5	5–10	10–20	20–40
1	a_1	0,7753	0,4995	0,4944	0,4063	0,3164
2	b_1	0,01652	0,0001099	0,0003042	0,0001966	0,0009266
3	c_1	0,08263	1,591	1,863	1,491	1,879
4	a_2	0,2248	0,5114	0,5458	0,6072	0,7152
5	b_2	0,04872	0,03407	0,03065	0,0269	0,02303
6	c_2	2,822	1,502	1,201	0,8528	0,5199
7	a_3	0,09587	0,1882	0,2149	0,1875	0,1162
8	b_3	0,002238	0,0006294	0,1171	0,0002392	0,0001389
9	a_4	0,3555	0,2039	0,1171	0,05989	0,02373
10	b_4	0,0002796	0,0001494	0,0001	0,0001	0

Таблица. Значения коэффициентов аппроксимации в формулах (10) и (11)

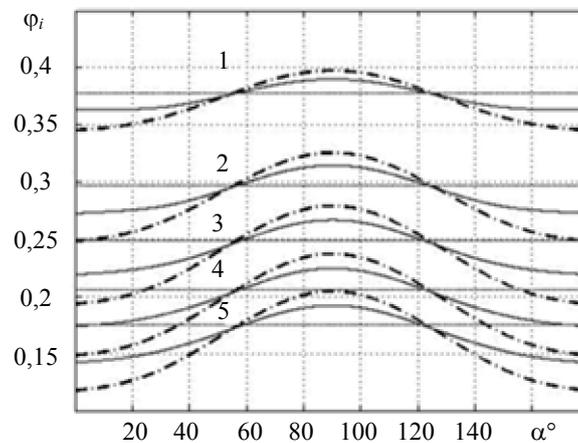


Рис. 2. Зависимость интегрального (штрихпунктирные линии) и эффективного (сплошные линии) коэффициентов облученности от угла наклона оси цилиндра при высотах h , км, равных: 200 (1), 600 (2), 1000 (3), 1500 (4), 2000 (5). Горизонтальными сплошными линиями указаны интегральные коэффициенты облученности сферического объекта

Как и ожидалось, погрешности расчета ЭКО меньше, чем вычисленных значений ИКО. Для вышеуказанных диапазонов изменения высоты ЦКО максимальные погрешности составляют 2%, -2,7%, 3,8%, -3,9% и 3,8%. Соответственно для цилиндра с умеренным отношением n погрешности определения среднеповерхностной температуры цилиндра будут меньше, чем в случае вытянутого цилиндра с большим n .

Обсуждение результатов

Предложенная методика расчета основывается на следующем алгоритме определения ЭКО для ЦКО. На первом этапе определяются значения функций $f(\alpha)$ и $\phi_m(H)$ для заданных угла наклона оси цилиндра и высоты ЦКО над поверхностью Земли по формулам (10) и (11). При этом используются значения коэффициентов в формулах из таблицы. Затем определяется значение ИКО боковой поверхности цилиндра ϕ_a по формуле (9), в которую подставляются также формулы для $\phi_b = \phi_{bc}$ из (4) и $\phi_m = \phi_0$ из (3). Затем полученное значение ИКО ϕ_a подставляется в (2) вместо значения ϕ_{int} . В ту же формулу (2) подставляется значение ϕ_b из (8), а также конкретная величина n . В результате получается значение ЭКО, которое для определения температуры ЦКО в тени Земли подставляется в формулу (1). При любой высоте ЦКО над поверхностью Земли погрешность расчета ИКО по формулам (9)–(11) при горизонтальном положении оси цилиндра, соответствующем максимальному значению ИКО, не превышает 1%. Максимальные погрешности реализуются в диапазоне изменения углов $20^\circ < \alpha < 60^\circ$, в котором осуществлялось «сшивание» зависимостей, вычисленных по двум моделям – ВМ 1 и ВМ 2. В указанной области изменения углов погрешности определения стационарных температур по формуле (1) не превышают 1,5 К для случая длин-

ного стержня (при $n > 10$). С уменьшением n погрешности определения ЭКО по формуле (2) снижаются, а при $n < 0,1$ (модель тонкого диска) расчеты становятся точными. Погрешности значительно снижаются также на солнечной стороне траектории и при расчетах нестационарного теплового режима. При α , равном 60° и 120° , значения ИКО и ЭКО цилиндра при $n = 3$, а также ИКО сферы примерно равны, причем ИКО сферы занимает промежуточное положение между значениями ИКО цилиндра при $\alpha = 0$ и $\alpha = \pi/2$.

Заключение

Предложенная приближенная аналитическая методика расчета эффективного коэффициента облученности для космического объекта цилиндрической формы реализует единственный способ преодоления неопределенностей в физической и математической моделях в области изменения углов наклона оси цилиндра до $\alpha \leq \theta_0$. Методика обеспечивает высокую для практических расчетов точность. Она проста в использовании и существенно облегчает расчеты нестационарных температур цилиндрического космического объекта. Альтернативой может быть численный расчет, обеспечивающий определение распределений поглощаемых лучистых потоков, а также температур по поверхности цилиндрического космического объекта. Такой расчет оправдан для определения температурных полей в объеме цилиндрических оболочечных конструкций с учетом прямой солнечной подсветки и отраженного Землей солнечного излучения. Однако на начальной стадии теплового проектирования телескопов космического базирования аналитические методы расчета незаменимы, поскольку позволяют быстро и надежно определять требования к основным параметрам конструкции на основе расчетов интегрального теплового баланса цилиндрического космического объекта.

Литература

1. Моделирование тепловых режимов космического аппарата и окружающей его среды / Под ред. Г.П. Петрова. М.: Машиностроение, 1971. 382 с.
2. Дубошин Г.Н. Небесная механика. Основные задачи и методы. М.: Наука, 1968. 800 с.
3. Чеботарев Г.А. Аналитические и численные методы небесной механики. М.–Л.: Наука, 1965. 367 с.
4. Сمارт У.М. Небесная механика. М.: Мир, 1965. 502 с.
5. Keihl J.T., Trenberth K.E. Earth's Annual Global Mean Energy Budget // Bull. Am. Meteorol. Soc. 1997. V. 78. N 2. P. 197–208.
6. Trenberth K.E., Fasullo J.T., Keihl J. Earth's global energy budget // Bull. Am. Meteorol. Soc. 2009. V. 90. N 3. P. 311–323.
7. Дульнев Г.Н., Семяшкин Э.М. Теплообмен в радиоэлектронных аппаратах. Л.: Энергия, 1968. 360 с.
8. Исаченко В.П., Осипова В.А., Сукомел А.С. Теплопередача. М.: Энергия, 1975. 488 с.
9. Уонг Х. Основные формулы и данные по теплообмену для инженеров: Справочник: Пер. с англ. М.: Атомиздат, 1979. 216 с.
10. Зигель Р., Хауэлл Дж. Теплообмен излучением: Пер. с англ. М.: Мир, 1975. 934 с.
11. Абдусаматов Х.И., Лаповок Е.В., Ханков С.И. Методы обеспечения термостабильности космического телескопа – солнечного лимбографа. СПб: Изд-во Политехн. ун-та, 2008. 195 с.
12. Каменев А.А., Лаповок Е.В., Ханков С.И. Аналитические методы расчета тепловых режимов и характеристик собственного теплового излучения объектов в околоземном космическом пространстве. СПб: НТЦ им. Л.Т. Тучкова, 2006. 186 с.
13. Баёва Ю.В., Лаповок Е.В., Ханков С.И. Методика расчета нестационарных температур космического объекта, движущегося по эллиптической орбите // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 6 (88). С. 67–72.
14. Баёва Ю.В., Лаповок Е.В., Ханков С.И. Аналитическая методика расчета тепловых потоков в околоземном космическом пространстве, формирующих тепловой режим космических телескопов // Оптический журнал. 2013. Т. 80. № 5. С. 30–37.
15. Баёва Ю.В., Лаповок Е.В., Ханков С.И. Метод поддержания заданного температурного диапазона космического аппарата, движущегося по круговой орбите с заходом в тень Земли // Изв. вузов. Приборостроение. 2013. Т. 56. № 7. С. 56–61.
16. Баёва Ю.В., Ханков С.И. Обеспечение термостабильности телескопа дистанционного зондирования Земли за счет выбора параметров бленды // Изв. вузов. Приборостроение. 2013. Т. 56. № 5. С. 76–79.

Дзитоев Азамат Миронович

– адъюнкт, Военно-космическая академия им. А.Ф. Можайского, Россия, Санкт-Петербург, Dzitoi8@gmail.com

Ханков Сергей Иванович

– доктор технических наук, ведущий научный сотрудник, Военно-космическая академия им. А.Ф. Можайского, Санкт-Петербург, Россия, Leva0007@rambler.ru

Azamat Dzitoev

Postgraduate in a military academy, Military Space Academy n.a. A.F. Mozhaisky, Saint Petersburg, Russia, Dzitoi8@gmail.com

Sergei Khankov

D.Sc., leading scientific researcher, Military Space Academy n.a. A.F. Mozhaisky, Saint Petersburg, Russia, Leva0007@rambler.ru

УДК 531

АВТОНОМИЗАЦИЯ НЕЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМС.Е. Иванов^а, Г.И. Мельников^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, sivanov@mail.ifmo.ru

Математической моделью многих механических систем является система динамических уравнений полиномиальной структуры с периодическими или постоянными параметрами. Такие механические системы широко применяют в динамике виброзащиты приборов и устройств. Исследование нелинейных систем с конечным числом степеней свободы представляет сложную актуальную проблему по сравнению с линейными системами. Исследование нелинейных систем не сводится к определению конечного числа частных решений, поскольку нелинейные системы не обладают свойством суперпозиции решений. Рассматривается математическая модель нелинейной динамической системы с тремя степенями свободы, которая содержит многочлены до четвертой степени от фазовых переменных. Для исследования данной модели представлены алгоритмические формулы метода автономизации нелинейных динамических систем. Нелинейная математическая модель динамической системы преобразуется к автономной форме, и определяются главные параметры динамической системы. Представлен алгоритм метода автономизации для исследования виброзащитной нелинейной системы с тремя степенями свободы. При решении задач виброзащиты широко применяются линейные системы, хотя линейность функций недостаточно точно аппроксимирует характеристики системы, внося погрешности при анализе. В работе решена задача получения и исследования более точной нелинейной модели виброзащитной системы. Рассмотрена нелинейная виброзащитная система с тремя степенями свободы с нелинейными правыми частями в виде многочлена третьей степени от фазовых переменных с постоянными и периодическими параметрами. Система состоит из объекта виброзащиты, установленного на две платформы, находящиеся одна под другой, нижняя из которых поставлена на вибрирующее основание. Внешнее гармоническое возмущение воздействует на основание. Предполагается, что упругие элементы системы описываются многочленами третьей степени, демпфирующие элементы имеют нелинейную кубическую характеристику. В результате применения метода нелинейная система преобразуется к более простому автономному виду, и существенно сокращается количество параметров нелинейной динамической системы без ухудшения качества решения. Применение метода существенно упрощает исследование переходных и установившихся процессов нелинейных динамических систем.

Решаемая задача исследования виброзащитной системы в нелинейной постановке является новой, имеет теоретическое и практическое значение.

Ключевые слова: автономизация нелинейных систем, методы исследования, нелинейные системы с тремя степенями свободы.

OFF-LINE INTERACTION OF THE NONLINEAR DYNAMIC SYSTEMSS. Ivanov^b, G. Melnikov^b

^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, sivanov@mail.ifmo.ru

Mathematical model of many mechanical systems is the system of dynamic equations with polynomial structure and periodic or constant parameters. Such mechanical systems are widely applied in the dynamics of a vibration guard of devices. Research of nonlinear systems with a finite number of freedom degrees represents a complicated actual problem in comparison with linear systems. Research of nonlinear systems does not come to definition of a finite number of private solutions because nonlinear systems do not possess superposition property of solutions. Mathematical model of nonlinear dynamic system with three degrees of freedom which contains polynomials to the fourth degree from phase variables is considered. Algorithmic formulas of an off-line interaction method of nonlinear dynamic systems are presented for the given model research. The nonlinear mathematical model of dynamic system is transformed to the autonomous form, and principal parameters of dynamic system are defined. The algorithm of an off-line interaction for research of nonlinear system protected from vibrations with three degrees of freedom is presented. In case of vibration guard solution of problems linear systems are widely used though linearity of functions does not approximate precisely enough the system performance, causing analysis errors. The problem of deriving and research of more exact nonlinear system model protected from vibrations is solved. The nonlinear system protected from vibrations with three degrees of freedom and nonlinear right members in the form of a polynomial of the third degree from phase variables with constant and periodic parameters is considered. The system consists of a vibration guard plant established on two platforms, one under another, the lower one is put on the vibrating foundation. Exterior harmonious perturbation exerts influence upon the foundation. Elastic system elements are supposed to be described by polynomials of the third degree and damping elements have nonlinear cubic performance. As a result of method application the nonlinear system will be transformed to the more simple autonomous form and the amount of parameters of nonlinear dynamic system is essentially reduced. The autonomous system contains essentially less parameters, than initial system, without making worse solution quality. Method application simplifies essentially the research of transitive and settled processes of nonlinear dynamic systems. The solved research problem for the system protected from vibrations in nonlinear statement is new and has theoretical and practical value.

Keywords: off-line interaction of nonlinear systems, research methods, nonlinear systems with three degrees of freedom.

Введение

Математической моделью многих механических систем является система динамических уравнений полиномиальной структуры с периодическими или постоянными параметрами. Такие механические системы широко применяют в динамике виброзащиты приборов и устройств [1]. Исследование нелиней-

ных систем с конечным числом степеней свободы представляет сложную актуальную проблему по сравнению с линейными системами. Исследование нелинейных систем не сводится к определению конечного числа частных решений, поскольку нелинейные системы не обладают свойством суперпозиции решений.

Предметом рассмотрения является математическая модель нелинейной динамической системы с тремя степенями свободы [2]. Целью работы является приведение математической модели нелинейной динамической системы к автономному виду. Рассматривается математическая модель в виде системы дифференциальных уравнений шестого порядка с нелинейными многочленами четвертой степени относительно фазовых переменных [3]. В результате применения метода автономизации математическая модель динамической системы преобразуется к автономному виду, и определяются главные параметры динамической системы. Метод применен для исследования приборной динамической системы с малыми нелинейными частями [4]. Проведенная проверка полученных результатов и сравнение с решением, полученным численным методом Рунге–Кутты [5], показали точность метода автономизации. В результате применения метода нелинейная система преобразуется к более простому автономному виду, и существенно сокращается количество параметров нелинейной динамической системы без ухудшения качества решения. Применение метода существенно упрощает исследование переходных и установившихся процессов нелинейных динамических систем.

Математическая модель динамической системы

Объектом исследования является математическая модель динамической системы, в виде обыкновенных дифференциальных уравнений с нелинейными частями в форме многочленов с постоянными и периодическими коэффициентами. Рассматривается система трех дифференциальных уравнений второго порядка с правыми частями в виде многочленов четвертых степеней относительно фазовых переменных. Запишем рассматриваемую систему в матричном виде:

$$\mathbf{A}\ddot{\mathbf{q}} + \mathbf{B}\dot{\mathbf{q}} + \mathbf{C}\mathbf{q} = \sum_{|\nu|=1}^4 h_\nu \cos(\omega t)^{\nu_1} \sin(\omega t)^{\nu_2} q_1^{\nu_3} q_2^{\nu_4} q_3^{\nu_5} \dot{q}_1^{\nu_6} \dot{q}_2^{\nu_7} \dot{q}_3^{\nu_8}, \quad (1)$$

где $\mathbf{q} = [q_1, q_2, q_3]^T$ – вектор координат, $\mathbf{A}, \mathbf{B}, \mathbf{C}$ – матрицы третьего порядка, $\mathbf{v} = (v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8)$ – векторный индекс, $|\nu| = \nu_1 + \nu_2 + \dots + \nu_8$.

Предполагается, что система содержит малые нелинейные части с постоянными и периодическими параметрами $|\mathbf{h}^s_\nu| < \varepsilon$.

Метод автономизации

Приведем алгоритм метода автономизации, который применим для программирования на языке высокого уровня.

После записи математической модели системы в виде (1) выполним следующие преобразования. Введем комплексно-сопряженные переменные для записи периодических функций в исходной системе:

$$q_0 = \exp(i\omega t) \text{ и } \bar{q}_0 = \exp(-i\omega t), \quad \lambda_1 = i\omega. \quad (2)$$

Запишем в новых переменных (2) периодические функции:

$$\cos(\omega t) = \frac{1}{2}(q_0 + \bar{q}_0) \text{ и } \sin(\omega t) = \frac{1}{2i}(q_0 - \bar{q}_0).$$

Систему дифференциальных уравнений (1) можно представить в нормальной форме Коши.

$$\dot{\mathbf{X}} = \begin{bmatrix} \mathbf{W} & 0 & 0 \\ 0 & 0 & \mathbf{I} \\ \mathbf{A}^{-1}\mathbf{H} & -\mathbf{A}^{-1}\mathbf{C} & -\mathbf{A}^{-1}\mathbf{B} \end{bmatrix} \mathbf{X} + \begin{bmatrix} 0 \\ 0 \\ \mathbf{A}^{-1}\mathbf{G} \end{bmatrix}, \quad (3)$$

где фазовый вектор имеет вид $\mathbf{X} = [q_0, \bar{q}_0, q_1, q_2, q_3, \dot{q}_1, \dot{q}_2, \dot{q}_3]^T$. Вначале применим линейное преобразование к системе (3):

$$\mathbf{Y} = \mathbf{D}\mathbf{X}. \quad (4)$$

Посредством линейного преобразования (4) система приводится к диагональному виду:

$$\dot{\mathbf{Y}} = \mathbf{\Lambda}\mathbf{Y} + \mathbf{R} \Big|_{\mathbf{X} \rightarrow \mathbf{D}^{-1}\mathbf{Y}}, \quad (5)$$

где $\mathbf{\Lambda} = \text{diag}[\lambda_1, \bar{\lambda}_1, \dots, \lambda_4, \bar{\lambda}_4]$, \mathbf{R} – преобразованная правая часть. Выполним многочленное преобразование системы (5), которое учитывает многочлены до четвертых степеней:

$$y_s = z_s + \sum_{|\nu|=2}^4 a_\nu^s \mathbf{Z}^\nu, \quad (s = 3, \dots, 8), \quad \mathbf{Z}^\nu \equiv z_1^{\nu_1} z_2^{\nu_2} \dots z_8^{\nu_8}. \quad (6)$$

В результате преобразования (6) система (5) приводится к автономному виду:

$$\dot{z}_s = \lambda_s z_s + \sum_{|\nu|=2}^4 q_\nu^s \mathbf{Z}^\nu, \quad (s = 3, \dots, 8). \quad (7)$$

Для нахождения коэффициентов преобразования a_ν^s и преобразованной системы q_ν^s получены формулы (8), удобные для программирования [6]:

$$\sum_{|\nu|=2}^4 q_\nu^s \mathbf{Z}^\nu + \sum_{|\nu|=2}^4 (a_\nu^s \mathbf{Z}^\nu (\sum_{k=1}^8 \lambda_k v_k - \lambda_s)) + \sum_{|\nu|=2}^4 (a_\nu^s \mathbf{Z}^\nu \sum_{k=3}^8 v_k z_k^{-1} \sum_{|\mu|=2}^4 q_\mu^k \mathbf{Z}^\mu) = \mathbf{R}_s(\mathbf{Z}) \quad (s = 3, \dots, 8). \quad (8)$$

При неособых значениях индексов коэффициенты преобразованной системы q_ν^s приравнивают к нулю и при таких значениях вычисляют коэффициенты преобразования a_ν^s . При особых значениях индексов коэффициенты преобразования a_ν^s приравнивают к нулю и вычисляют коэффициенты преобразованной системы q_ν^s . Для нахождения особых значений векторного индекса при фиксированном s необходимо найти решение двух уравнений:

$$\sum_{k=1}^8 \lambda_k v_k - \lambda_s \approx 0, \quad \sum_{k=1}^8 v_k = 2, 3, 4.$$

Произведем следующую замену переменных:

$$z_s = \rho_s \exp(i(t \operatorname{Im} \lambda_s + \theta_s)); \bar{z}_{s+1} = \rho_s \exp(i(t \operatorname{Im} \lambda_{s+1} - \theta_s)); s = 3, 5, 7; \quad z_{1,2} \equiv \exp(\pm it\omega). \quad (9)$$

В результате замены переменных (9) получим аналитическое решение автономной системы (7):

$$\begin{aligned} \dot{\rho}_s &= \operatorname{Re}(\lambda_s) \rho_s + \operatorname{Re}(\psi_s); \quad \dot{\theta}_s = \rho_s^{-1} \operatorname{Im}(\psi_s); \quad s = 3, 5, 7; \\ \psi_s &= \sum_{|\nu|=2}^4 q_\nu^s \rho_3^{v_3+v_4} \rho_5^{v_5+v_6} \rho_7^{v_7+v_8} \exp(i(\theta_3(v_3 - v_4) + \theta_5(v_5 - v_6) + \theta_7(v_7 - v_8) - \theta_s)). \end{aligned} \quad (10)$$

Для нахождения стационарных решений приравниваем к нулю правые части системы (10). Для нахождения фазовых координат системы (1) выразим вектор \mathbf{Y} по формулам, обратным линейной замене (4):

$$\mathbf{X} = \mathbf{D}^{-1} \mathbf{Y}.$$

Исследование нелинейной виброзащитной системы

При решении задач виброзащиты широко применяются линейные системы, хотя линейность функций не достаточно точно аппроксимирует характеристики системы, внося погрешности при анализе [7]. В работе решена задача получения и исследования более точной нелинейной модели виброзащитной системы. Решаемая задача виброзащиты в нелинейной постановке является новой, имеет теоретическое и практическое значение.

Применим представленный метод к нелинейной динамической системе с тремя степенями свободы. Рассмотрим виброзащитную систему (рис. 1), состоящую из объекта виброзащиты массой m_1 , установленного на платформы массой m_2 и m_3 , нижняя из которых закреплена на вибрирующем основании [8]. Предполагается, что упругие элементы системы описываются полиномом третьей степени $kx + lx^2 + px^3$, демпфирующие элементы имеют нелинейную кубическую характеристику $c\dot{x} + d\dot{x}^3$.

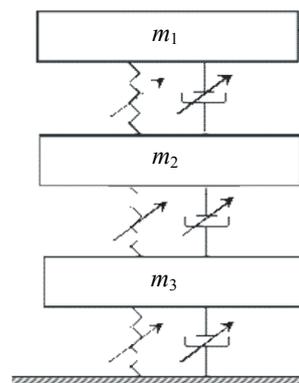


Рис. 1. Схема нелинейной виброзащитной системы

Математическая модель рассматриваемой системы представима в виде системы дифференциальных уравнений:

$$\begin{aligned}
 m_1 \ddot{x}_1 + c_1(\dot{x}_1 - \dot{x}_2) + d_1(\dot{x}_1 - \dot{x}_2)^3 + k_1(x_1 - x_2) + l_1(x_1 - x_2)^2 + p_1(x_1 - x_2)^3 &= 0, \\
 m_2 \ddot{x}_2 + c_1(\dot{x}_2 - \dot{x}_1) + d_1(\dot{x}_2 - \dot{x}_1)^3 + k_1(x_2 - x_1) - l_1(x_2 - x_1)^2 + p_1(x_2 - x_1)^3 + \\
 c_2(\dot{x}_2 - \dot{x}_3) + d_2(\dot{x}_2 - \dot{x}_3)^3 + k_2(x_2 - x_3) + l_2(x_2 - x_3)^2 + p_2(x_2 - x_3)^3 &= 0, \\
 m_2 \ddot{x}_3 + c_2(\dot{x}_3 - \dot{x}_2) + d_2(\dot{x}_3 - \dot{x}_2)^3 + k_2(x_3 - x_2) - l_2(x_3 - x_2)^2 + p_2(x_3 - x_2)^3 + \\
 c_3(\dot{x}_3 - \dot{f}) + d_3(\dot{x}_3 - \dot{f})^3 + k_3(x_3 - f) + l_3(x_3 - f)^2 + p_3(x_3 - f)^3 &= 0,
 \end{aligned} \tag{11}$$

где x_1, x_2, x_3 – абсолютные перемещения по отношению к положению равновесия системы.

На основании действуют вертикальные колебания

$$f(t) = a \sin(\omega t).$$

Введем относительное перемещение:

$$\tilde{x}_1 = x_1 - f, \quad \tilde{x}_2 = x_2 - f, \quad \tilde{x}_3 = x_3 - f. \tag{12}$$

Запишем уравнения движения (11) в новых переменных (12):

$$\begin{aligned}
 m_1 \ddot{\tilde{x}}_1 + c_1(\dot{\tilde{x}}_1 - \dot{\tilde{x}}_2) + d_1(\dot{\tilde{x}}_1 - \dot{\tilde{x}}_2)^3 + k_1(\tilde{x}_1 - \tilde{x}_2) + l_1(\tilde{x}_1 - \tilde{x}_2)^2 + p_1(\tilde{x}_1 - \tilde{x}_2)^3 &= -m_1 \ddot{f}, \\
 m_2 \ddot{\tilde{x}}_2 + c_1(\dot{\tilde{x}}_2 - \dot{\tilde{x}}_1) + d_1(\dot{\tilde{x}}_2 - \dot{\tilde{x}}_1)^3 + k_1(\tilde{x}_2 - \tilde{x}_1) - l_1(\tilde{x}_2 - \tilde{x}_1)^2 + p_1(\tilde{x}_2 - \tilde{x}_1)^3 + \\
 c_2(\dot{\tilde{x}}_2 - \dot{\tilde{x}}_3) + d_2(\dot{\tilde{x}}_2 - \dot{\tilde{x}}_3)^3 + k_2(\tilde{x}_2 - \tilde{x}_3) + l_2(\tilde{x}_2 - \tilde{x}_3)^2 + p_2(\tilde{x}_2 - \tilde{x}_3)^3 &= -m_2 \ddot{f}, \\
 m_2 \ddot{\tilde{x}}_3 + c_2(\dot{\tilde{x}}_3 - \dot{\tilde{x}}_2) + d_2(\dot{\tilde{x}}_3 - \dot{\tilde{x}}_2)^3 + k_2(\tilde{x}_3 - \tilde{x}_2) - l_2(\tilde{x}_3 - \tilde{x}_2)^2 + p_2(\tilde{x}_3 - \tilde{x}_2)^3 + \\
 c_3 \dot{\tilde{x}}_3 + d_3 \dot{\tilde{x}}_3^3 + k_3 \tilde{x}_3 + l_3 \tilde{x}_3^2 + p_3 \tilde{x}_3^3 &= -m_3 \ddot{f}.
 \end{aligned} \tag{13}$$

Здесь

$$\ddot{f}(t) = -a\omega^2 \sin(\omega t).$$

Выполним многочленное преобразование системы (13) согласно приведенному алгоритму [9]. С точностью до членов четвертого порядка получаем автономную дифференциальную систему

$$\begin{cases} \dot{z}_3 = (\lambda_3 + q_{11100000}^3)z_3, & z_4 = \bar{z}_3, \\ \dot{z}_5 = (\lambda_5 + q_{11001000}^5)z_5, & z_6 = \bar{z}_5, \\ \dot{z}_7 = (\lambda_7 + q_{11000010}^7)z_7, & z_8 = \bar{z}_7 \end{cases} \tag{14}$$

Решение системы (14) записывается в виде

$$\begin{aligned}
 z_{3,4} &\equiv \rho_{01} \exp(t \operatorname{Re}(\lambda_3 + q_{11100000}^3) \pm i(\theta_{01} + t \operatorname{Im}(\lambda_3 + q_{11100000}^3))), \\
 z_{5,6} &\equiv \rho_{02} \exp(t \operatorname{Re}(\lambda_5 + q_{11001000}^5) \pm i(\theta_{02} + t \operatorname{Im}(\lambda_5 + q_{11001000}^5))), \\
 z_{7,8} &\equiv \rho_{03} \exp(t \operatorname{Re}(\lambda_7 + q_{11000010}^7) \pm i(\theta_{03} + t \operatorname{Im}(\lambda_7 + q_{11000010}^7))).
 \end{aligned} \tag{15}$$

С помощью пакета программ рассчитан переходный и установившийся режимы движения виброзащитной системы при следующих параметрах:

$$\begin{aligned}
 m_1 &= 1,13, c_1 = 0,23, d_1 = 0,01, k_1 = 1,23, l_1 = 0,04, p_1 = 0,02, \\
 m_2 &= 3,17, c_2 = 0,61, d_2 = 0,03, k_2 = 3,13, l_2 = 0,13, p_2 = 0,06, \\
 m_3 &= 8,71, c_3 = 1,73, d_3 = 0,09, k_3 = 9,11, l_3 = 0,37, p_3 = 0,17.
 \end{aligned}$$

На систему действует внешнее возмущение: $\omega = 2, a = 0,5$.

Определены коэффициенты преобразованной автономной системы (14)

$$\begin{aligned}
 q_{11100000}^3 &= -0,086 + 0,011i, \\
 q_{11001000}^5 &= -0,154 + 0,012i, \\
 q_{11000010}^7 &= -0,044 + 0,001i.
 \end{aligned}$$

Решение системы (15) записывается в виде

$$\begin{aligned}
 z_{3,4} &\equiv \rho_{01} \exp(-0,297t \pm i(\theta_{01} + 1,480t)), \\
 z_{5,6} &\equiv \rho_{02} \exp(-0,273t \pm i(\theta_{02} + 1,130t)), \\
 z_{7,8} &\equiv \rho_{03} \exp(-0,083t \pm i(\theta_{03} + 0,635t)).
 \end{aligned}$$

Установившийся полигармонический режим движения системы имеет вид:

$$\begin{aligned}
 x_1 &= -0,011 - 0,035 \cos(\omega t) - 0,470 \sin(\omega t) + 0,00001 \cos(2\omega t) - 0,00001 \sin(2\omega t), \\
 x_2 &= -0,011 + 0,113 \cos(\omega t) - 0,526 \sin(\omega t) + 0,00016 \cos(2\omega t) + 0,00007 \sin(2\omega t), \\
 x_3 &= -0,008 - 0,261 \cos(\omega t) - 0,612 \sin(\omega t) - 0,00045 \cos(2\omega t) + 0,00056 \sin(2\omega t).
 \end{aligned}$$

Колебания системы происходят с частотами, кратными частоте внешней силы, воздействующей на основание. На рис. 2 показано абсолютное перемещение объекта виброзащиты при установлении колебаний нелинейной динамической системы. Представленный график показывает, что вынужденные колебания в начале установления являются квазипериодическими с частотами, кратными внешней силе.

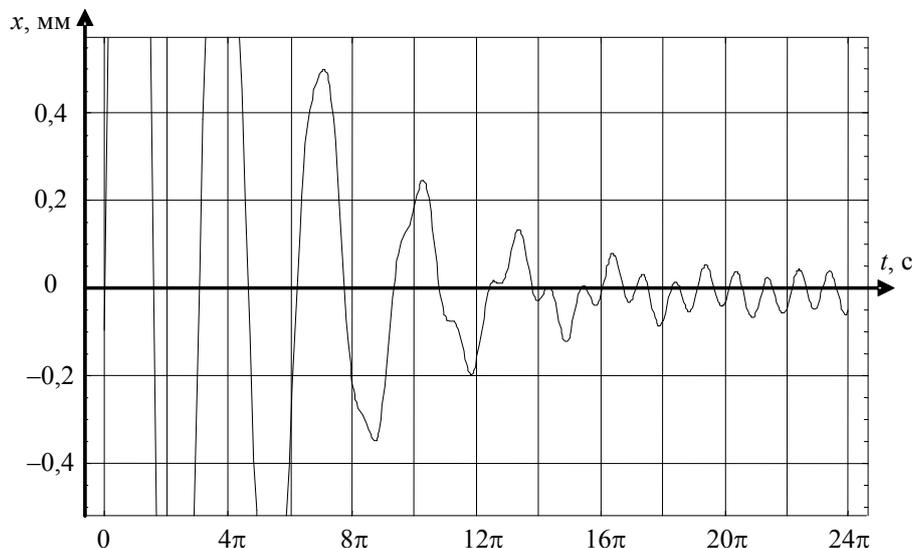


Рис. 2. Абсолютное перемещение объекта виброзащиты при установлении колебаний

В результате применения метода для анализа виброзащитной системы с тремя степенями свободы получены графики переходных и установившихся режимов виброзащитной системы. Уравнения движения системы приведены к автономному виду. Найдены существенные константы, характеризующие переходные процессы и установившиеся режимы колебаний. Приведена схема метода для систем с тремя степенями свободы, описывается алгоритм программной реализации метода.

Метод позволяет получить достаточно подробные качественные и количественные характеристики изучаемых движений [10], исследовать установившиеся режимы колебаний для систем, находящихся в условиях периодического внешнего воздействия, а также изучать переходные процессы.

Заключение

Рассмотрена нелинейная динамическая система с тремя степенями свободы и нелинейными правыми частями в виде многочлена третьей степени от фазовых переменных с постоянными и периодическими параметрами. Система состоит из объекта виброзащиты, установленного на две платформы, находящиеся одна под другой, нижняя из которых поставлена на вибрирующее основание [11]. Внешнее гармоническое возмущение воздействует на основание. Предполагается, что упругие элементы системы описываются многочленами третьей степени, демпфирующие элементы имеют нелинейную кубическую характеристику [12]. Система уравнений движения приведена к автономному виду. Определены существенные константы, характеризующие переходные процессы и установившиеся режимы колебаний.

В работе представлен алгоритм метода автономизации нелинейных динамических систем с тремя степенями свободы [13]. Разработан алгоритм метода для исследования нелинейных динамических задач. В результате применения метода нелинейная периодическая система преобразуется к автономному виду, и определяются параметры установившихся и переходных режимов движения. Для повышения точности метода в работе [14] предложено применение аппроксимации Чебышева для членов высоких степеней. Приведенные алгоритмические формулы метода позволяют приводить к автономному виду нелинейные динамические системы [15], математическая модель которых представима в виде системы дифференциальных уравнений шестого порядка с нелинейными многочленами до четвертых степеней относительно фазовых переменных.

Литература

1. Бабаков И.М. Теория колебаний. М.: Дрофа, 2004. 591 с.
2. Иванов С.Е. Исследование нелинейных динамических систем с тремя степенями свободы // Научно-технический вестник СПбГУ ИТМО. 2011. № 4 (74). С. 62–64.
3. Бибиков Ю.Н. Курс обыкновенных дифференциальных уравнений. СПб: Лань, 2011. 304 с.
4. Мельников В.Г. Многочленные преобразования нелинейных систем управления // Изв. вузов. Приборостроение. 2007. Т. 50. № 5. С. 20–25.

5. Демидович Б.П., Марон И.А., Шувалова Э.З. Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения. СПб: Лань, 2010. 400 с.
6. Иванов С.Е. Алгоритмическая реализация метода исследования нелинейных динамических систем // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 4 (80). С. 90–92.
7. Матросов В.М., Румянцев В.В., Карапетян А.В. Нелинейная механика. М.: ФИЗМАТЛИТ, 2001. 432 с.
8. Мельников В.Г., Мельников Г.И., Иванов С.Е. Компьютерные технологии в механике приборных систем: Учеб. пособие / Под ред. В.Г. Мельникова. СПб: СПбГУ ИТМО, 2006. 127 с.
9. Мельников Г.И. Динамика нелинейных механических и электромеханических систем. Л.: Машиностроение, 1975. 198 с.
10. Иванов С.Е. Определение установившихся режимов работы виброзащитной системы с двумя степенями свободы // Научно-технический вестник СПбГУ ИТМО. 2010. № 4 (68). С. 44–46.
11. Каток А.Б., Хасселблат Б. Введение в теорию динамических систем с обзором последних достижений. М.: МЦНМО, 2005. 464 с.
12. Hamming R.W. Numerical methods for scientists and engineers. NY: Dover, 1986. 721 p.
13. Мельников В.Г. Энергетический метод параметрической идентификации тензоров инерции тел // Научно-технический вестник СПбГУ ИТМО. 2010. № 1 (65). С. 59–63.
14. Мельников В.Г. Преобразование динамических многочленных систем с применением аппроксимации Чебышева // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 4 (80). С. 85–89.
15. Melnikov V.G. Chebyshev economization in Poincare-Dulac transformations of nonlinear systems // Nonlinear Analysis. 2005. V. 63. N 5–7. P. e1351–e1355.

Иванов Сергей Евгеньевич – кандидат физ.-мат. наук, доцент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, sivanov@mail.ifmo.ru

Мельников Геннадий Иванович – доктор физ.-мат. наук, профессор, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, melnikov@mail.ifmo.ru

Sergei Ivanov PhD, Associate professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, sivanov@mail.ifmo.ru

Gennady Melnikov D.Sc., Professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, melnikov@mail.ifmo.ru

УДК 629.78

СОВРЕМЕННОЕ СОСТОЯНИЕ И ПЕРСПЕКТИВЫ РАЗВИТИЯ ОСНОВНЫХ ПОНЯТИЙ В ОБЛАСТИ МЕХАТРОНИКИ

Е.В. Шалобаев^а, Р.-Т. Толочка^б

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, shalobaev47@mail.ru

^б Каунасский технологический университет, Каунас, Литва, tadas.tolocka@ktu.lt.

Проанализировано современное состояние базовых терминов в области мехатроники, которое зафиксировано в документах комиссии по стандартизации и терминологии Международной федерации по теории машин и механизмов. Указаны новые термины, которые были введены в научный оборот с начала XXI века. К таким терминам относятся «мехатронизированный объект», «мехатронный класс», объединяющий «мехатронные» и «мехатронизированные» объекты. Дано понятие об «уровневом подходе» к мехатронике, который позволяет увязать между собой собственно мехатронику, микросистемную технику и наноиндустрию. Изложены некоторые соображения о применении термина «мехатронный комплекс», которые стали особенно актуальным в связи с появлением орбитальных группировок малых космических аппаратов.

Внесены коррективы по таким критериям оценки мехатронных объектов, как ремонтпригодность и долговечность, а также по использованию термина «мехатронные модули». Поставлены вопросы о расширении понимания мехатроники от компьютерного управления движением до управления состоянием объекта, о терминах «авионика» и «автоника», о взаимосвязи мехатроники и логистики, которую стали относить к предметной области мехатроники.

Ключевые слова: мехатроника, мехатронные комплексы, модули и узлы, орбитальная группировка, малые космические аппараты, ремонтпригодность, долговечность, уровневый подход, мехатронизированные объекты.

MODERN STATE AND DEVELOPMENT PROSPECTS OF THE BASIC CONCEPTS IN THE FIELD OF MECHATRONICS

E. Shalobaev^c, R.-T. Tolochka^d

^c Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, vshalobaev47@mail.ru

^d Kaunas University of Technology, Kaunas, Lithuania, tadas.tolocka@ktu.lt.

The paper deals with modern state analysis of the basic terms in the field of mechatronics documented by the Commission on standardization and terminology of the International Federation for the Theory of Machines and Mechanisms. New terms introduced into scientific circulation since the beginning of the 21st century are indicated. They are: “mechatronized object”, “mechatronic class”, combined “mechatronic” and “mechatronized” objects. The notion of “tiered approach” to mechatronics is given, which makes it possible to connect the mechatronics proper, micro-system technique and nanotechnologies. Some considerations on the use of “mechatronic complex” term are stated especially topical in the implementation process for the groups of small spacecrafts. That is why the term “mechatronic complex” is filled with real content. Such criteria, characteristic of mechatronic objects, as maintainability and durability are amended. Corrections to the “mechatronic modules” term application are proposed. The question is raised on the necessity to expand the mechatronics meaning from the movement computer management to object state control and to define such terms as “avionics” and “autonics”. Moreover, interconnection between mechatronics and logistics, which is now referred to the subject field of mechatronics, must be defined.

Keywords: mechatronics, mechatronic complexes, group, small spacecrafts, maintainability, durability, mechatronic modules and units.

Введение

Мехатроника, как и всякая наука, развивается, и со временем появляется необходимость вводить некоторые коррективы, учитывая современное развитие техники. В этом году исполняется 25 лет со дня издания на русском языке монографии японских ученых «Мехатроника» [1] – первого издания по данной тематике. Вполне понятно, что за это время произошли серьезные изменения в определении взаимосвязи мехатроники с рядом смежных научно-технических направлений, в установлении базовой терминологии, в наполнении предложенных терминов реальным содержанием, во введении в научный оборот новых терминов и понятий.

В своей книге [2], выдержавшей два издания, проф. Ю.В. Подураев высказал мнение, что «до сих пор мехатроника является уделом практиков», но «потребность в теоретическом инструментарии, который имеет эвристическую ценность и позволяет вести целенаправленный поиск новых мехатронных решений, сегодня весьма велика». Именно поэтому в этой же работе появились параграфы «Понятия о мехатронике» и «Определения и терминология мехатроники».

Таким образом, пришло понимание необходимости и даже потребности теоретического инструментария, что реализовалось в ряде работ, касающихся разных аспектов теоретических основ мехатроники [3–17]. Вместе с тем в вопросах терминологии необходимо четко определить взаимосвязь между мехатроникой и пограничными или смежными научно-техническими направлениями, такими как кибернетика, робототехника, микро- и наносистемная техника. Существуют проблемы в определении взаимосвязи между мехатроникой и авионикой, автоникой, логистикой. С другой стороны, введены в научный оборот и успешно используются в публикациях термины и понятия, которые необходимо зафиксировать в нормативной литературе по мехатронике. В настоящей статье авторами, членами технического комитета по стандартизации и терминологии Международной федерации по теории машин и механизмов (IFToMM), делается попытка ответа на указанные вопросы.

Связь мехатроники с другими научно-техническими направлениями

Покажем, как связана мехатроника с другими научно-техническими направлениями.

Взаимосвязь мехатроники и робототехники. «Колыбелью» мехатроники принято считать станкостроение [1, 3, 18], хотя с этим утверждением спорят специалисты из таких областей, как ракетно-космическая, военная техника и робототехника.

Например, известный специалист в области робототехники проф. А.С. Ющенко (МГТУ им. Н.Э. Баумана) [19] еще в 1990-е гг. в беседе с одним из авторов данной публикации утверждал, что «мехатронику придумали робототехники, чтобы не остаться без работы». Действительно, связь между этими научно-техническими направлениями стала проявляться в следующих фактах. В 1996 г. вышел первый сборник трудов «Робототехника и мехатроника» [20], который, хотя и не стал периодическим, но заложил основу для выхода в начале 2000 г. всероссийского журнала «Мехатроника». В марте того же года междисциплинарная специальность «Мехатроника» вошла, наряду с уже ставшей классической специальностью «Робототехника», в одно направление подготовки специалистов «Мехатроника и робототехника». Позднее официальную позицию о связи мехатроники и робототехники сформулировали патриарх отечественной робототехники проф. Е.И. Юревич и сменивший его на посту директора ЦНИИ робототехники и технической кибернетики (РТК) член-корр. РАН В.А. Лопота. В работах [21, 22] они указали, что мехатроника входит в родственное робототехнике научно-техническое направление и наиболее близка ей. При

этом они согласились с точкой зрения автора работы [18], что робототехника определена по функциональному назначению, а мехатроника – по физическому составу. Несмотря на такое начальное различие, ряд робототехнических систем можно отнести к полноценным мехатронным. С другой стороны, мехатронные системы в наиболее развитом виде реализуют робототехническую «триаду» сенсоры–контроллеры–актуаторы [19]. Из изложенного следует, что базовые принципы построения и методы проектирования мехатроники и робототехники имеют много общего. В первую очередь, это относится к системному подходу при проектировании и блочно-модульному принципу построения.

Претензии на первородство представителей ракетно-космической отрасли подтверждаются тем, что еще в 1960-х гг. академики РАН А.Ю. Ишлинский и Д.Е. Охочимский создали первые отечественные мехатронные системы инерциального управления баллистической ракетой и управления входом космической ракеты в плотные слои атмосферы [23, 24]. В более традиционной военной технике также имеются публикации в области мехатроники, например, работы проф. Ю.М. Сазыкина [25] и его последователей.

Взаимосвязь мехатроники и кибернетики. В работе [5] сделан вывод о том, что мехатроника сложилась как парадигма технической кибернетики. По своему характеру работа относится к фундаментальным трудам, определяющим теоретическую базу мехатроники [8, 9, 11].

Аргументация авторов следующая. «Кибернетика – это типичная метанаука и, естественно, при развитии из нее извлекаются «чистые» науки, которые продолжают жить самостоятельно. А сама «мета» в конечном итоге перестает существовать как наука». Именно данная особенность развития метанаук явилась ключевой метафорой генезиса мехатроники: период зарождения мехатроники – закат кибернетики и рассвет информатики. Среди определений мехатроники есть и такое – «наука о компьютерном управлении в технических системах». В этом контексте термин «мехатроника» эквивалентен современной, «компьютерной» трактовке ушедшего из научно-технического лексикона термина «техническая кибернетика», означающего управление в технических системах.

Непрерывная эволюция предмета технической кибернетики, обусловленная постепенным усложнением кибернетических систем, а также применением современных средств автоматизации и вычислительной техники, привела в конечном итоге к радикальному изменению ее содержания. В начале 1980-х гг. техническая кибернетика претерпевает период компьютеризации, связанный с изобретением микропроцессорных технологий, и становится основой широкомасштабной автоматизации технических объектов и технологических процессов на базе встроенных мини- и микрокомпьютеров. Принципиальной фундаментальной инновацией здесь является переход от механических и электрических к микроэлектронным вычислительным средствам обработки информации и управления на основе микропроцессоров и интегральных схем. В связи с этим в технической кибернетике наметились два самостоятельных перспективных направления дальнейшего развития по пути разработки, создания и использования компьютерной техники: первое – для обработки информации (техническая информатика), второе – для управления (мехатроника).

Кибернетика стала материнской наукой для информатики и прародительницей мехатроники. Мехатроника придала «второе дыхание» оказавшейся почти бездыханной кибернетике, обеспечив ее «второе пришествие». В аннотации к первой монографии по мехатронике [1] прямо указано, что «она посвящена актуальной проблеме построения сложных автоматических систем», что по смыслу совпадает с классическим определением кибернетики, данным А.И. Бергом [26]: «Кибернетика – наука об управлении сложными динамическими системами». К сожалению, в рассматриваемой статье не приводятся работы, где высказываются мнения, что мехатронике «еще предстоит сформироваться как самостоятельной науке, предстоит обрести обязательные атрибуты новой науки».

В работе [5] ее авторы утверждают, что мехатроника уже 20 лет назад приобрела статус самостоятельной фундаментальной технической науки, и в подтверждение этого тезиса ссылаются на свои работы [27, 28]. Сюда с полным основанием можно отнести и работы [2–26].

Мехатроника естественным образом отражает целостное восприятие информационных, вычислительных и управляющих процессов, выраженное известным шведским специалистом в области компьютерного управления техническими системами К. Острёмом следующей афористичной формулой [29]:

$$\text{Communication} + \text{Computation} + \text{Control} = C3.$$

В подтверждение своей идеи авторы приводят такой факт: еще в 1954 г. Н. Винер дал удивительно точный прогноз именно мехатронной парадигмы развития технической кибернетики, подчеркивая, что ЭВМ будут активно использоваться для непосредственного управления исполнительными механизмами.

В работе [5] резюмируется, что информатика и мехатроника проходят этап своего тридцатилетнего становления и рассматриваются как перспективные «точки роста» науки и техники всего XXI века, что в полной мере отражает позицию такого авторитетного периодического издания в области мехатроники, как журнал «Мехатроника, автоматизация, управление».

Взаимосвязь мехатроники с микро- и наносистемной техникой. В настоящее время возникла насущная потребность остановиться на процессе признания уровневого характера мехатроники в целом, что определяется масштабом рассматриваемых объектов. Такой характер или подход заключается в том,

что существуют макро-, микро- и наноуровни [18, 30, 31]. И если границу между макро- и микромиром представить достаточно просто – она определяется возможностями человеческого зрения, то граница между микро- и наномиром не так однозначна. Существует много мнений, которые можно свести в две группы. Первая группа определяет эту границу численно – наиболее часто встречается цифра 100 нм. Авторы разделяют мнение о том, что граница микро- и наномира определяется качественно – при нарушении «сплошности» материала, т.е. при переходе к рассмотрению групп молекул, отдельных молекул и атомов. При этом надо помнить, что размер некоторых молекул (особенно органических) может достигать долей миллиметров и даже единиц миллиметров.

Один из авторов данной публикации уже в начале 2000-х г.г. в работах [18, 30] обратил внимание на то, что мехатроника к тому времени уже имела два уровня – макро- и микро-. К первому уровню относились собственно объекты мехатроники, такие как станочные приводы, бытовая и офисная техника; ко второму уровню – объекты микросистемной техники, в частности, микроэлектромеханические системы (МЭМС), микрооптоэлектромеханические системы (МОЭМС) и т.п. Таким образом, в работах [18, 30] были заложены основы уровневого подхода, который был расширен и узаконен в международном трансляторе по ТММ [31] (разрабатывался в 2002 г. [32, 33]), где закреплены понятия «мехатроника», «микро-мехатроника» (подобласть мехатроники, относящаяся к устройствам и системам с размерами в несколько миллиметров и меньше, т.е. синоним термина «микросистемная техника») и «наномехатроника» (подобласть мехатроники, относящаяся к устройствам и системам с размерами, соизмеримыми с молекулами, т.е. синоним термина «наносистемная техника»). Такой подход нашел положительный отклик у ряда отечественных ученых, которые стали использовать термины «микромехатроника» и «наномехатроника», например, в работах [7, 18, 27, 30, 31, 34].

Особо необходимо отметить, что существуют отечественные журналы «Мехатроника. Автоматизация. Управление» (первоначально «Мехатроника»), «Нано- и микросистемная техника» (первоначально «Микросистемная техника»), которые выпускаются издательством «Новые технологии», созданным на базе столичного издательства «Машиностроение». Это издательство также приняло подобный уровневый подход. И это видно не только по названию. В первом журнале печатаются статьи по микро- и наносистемной технике, а во втором – по мехатронике. Так что проблема отсутствия связи между уровнями мехатроники постепенно решается, хотя и не так быстро, как хотелось бы [5].

Кстати, уровневый подход вводится не только в мехатронике, но и в других научно-технических направлениях, например, в трибологии [35–40], где появились такие термины, как микро- и нанотрибология. В зависимости от масштаба, на котором изучается трение, в современной трибологии принято выделять три раздела: макротрибологию (или просто трибологию), микротрибологию и нанотрибологию. Макротрибология имеет дело с большими объектами и не рассматривает особенностей строения вещества. Основным соотношением макротрибологии является закон Амонтона–Кулона, который пропорционально связывает силу трения и нормальную реакцию. С другой стороны, нанотрибология в соответствии со своей предметной областью имеет дело с взаимодействием отдельных атомов и молекул или их групп.

Микротрибология является переходным разделом трибологии и занимается исследованием реальных, но элементарных контактов, например, в МЭМС [41]. Как правило, подразумевается, что небольшой выступ взаимодействует с некоторой поверхностью. В данном случае аналогом закона Амонтона–Кулона является формула Баудена–Табора.

Необходимо зафиксировать такой уровневый подход и к мехатронике, учитывая масштабный фактор. Иначе говоря, мехатроника относится к макромиру, микромехатроника – к микромиру, наномехатроника – к наномиру [7, 12, 16, 18, 27, 30, 31]. Требуется установить взаимосвязь между микро- и наносистемной техникой и мехатроникой, поскольку в отечественных журналах «Мехатроника, автоматизация, управление» и «Нано- и микросистемная техника» взаимосвязи между их предметными областями не оговариваются. В работах одного из авторов данной публикации [18, 30] такой уровневый подход к мехатронике был предложен, что нашло фактическое отражение в рассматриваемом международном трансляторе, но формально зафиксировано не было и требует такого закрепления. Также необходимо закрепление тождества терминов «микросистемная техника» и «микромехатроника» [30].

Введение в научный оборот новых терминов в мехатронике

Необходимо коснуться ряда основополагающих вопросов терминологии в мехатронике, которые в той или иной мере отражались в работах [1, 7–18, 27, 28, 30–35].

Наполнение термина «мехатронные комплексы» реальным содержанием. Коснемся следующих вопросов терминологии. В работах одного из авторов данной публикации [42, 43] и в ряде других работ [44–48] использовался термин «мехатронные комплексы». В первой работе речь шла чисто теоретически о наличии такого перспективного уровня мехатронных объектов (имелась в виду мехатронная распределенная система с отсутствием единого корпуса или оболочки), а в остальных работах название применено без четкого определения такого термина.

К настоящему времени развитие космической техники, а также миниатюризация объектов привели к реализации идеи орбитальных группировок космических аппаратов (КА). Появилась возможность нового подхода к проведению космических полетов и исследований. В частности, реализуется идея создания нового класса космических систем (КС), возникающих вследствие снижения массы КА как системообразующего элемента этих систем. Упомянутые выше КС представляют собой орбитальные группировки малых космических аппаратов (ОГ МКА), объединенных одной целевой задачей и выполняющих совместный полет на небольшом удалении друг от друга (от десятков метров до десятков километров), взаимное положение и движение которых контролируется и управляется [49, 50].

В ОГ МКА реализован принцип декомпозиции для получения качественного эффекта, который заключается в снижении общей массы. Необходимо отметить, что прямое, уменьшающее габариты масштабирование [51, 52] лишь увеличивает трудоемкость изготовления, например, малогабаритных приводов или реактивных двигателей, при очевидном снижении только затрат на материалы. Однако, применяя современные достижения в микроэлектронике, нанотехнологии, материаловедении и нетрадиционные подходы к конструированию (например, мехатронные), удастся создать качественно новые спутники, значительно отличающиеся в меньшую сторону по массе и размерам от традиционных аппаратов [18, 51, 52].

Отличительной особенностью ОГ из сверхмалых КА является большое число аппаратов в ее составе, структура ОГ может быть сложной. Например, главная ОГ может состоять из нескольких подчиненных ОГ, выполняющих частные подзадачи и содержащие КА разного типа и назначения. Иначе говоря, на каждом МКА, входящем в орбитальную группировку, может быть размещена полностью или частично сенсорная, управляющая и исполнительная, а также энергетическая подсистемы общей КС. Такая ОГ МКА или просто КС представляет собой высший уровень мехатронных объектов – мехатронный комплекс (т.е. распределенную мехатронную систему) [17]. И, естественно, при таком подходе к созданию КС кардинально меняются подходы к созданию систем управления, к созданию самих КА, к целевому применению КС в целом. Подробности такого подхода изложены в цикле работ [49–57].

Таким образом, снижение массы КА на порядки дало увеличение на порядки и численности ОГ, что, в свою очередь, привело к появлению нового класса КС. Иначе говоря, реализован закон перехода количества в качество – известный закон диалектики. Миниатюризация в этом случае входит в арсенал разработчиков новых КС. Такое направление использования малых спутников особенно важно для России, ибо космическая отрасль, будучи областью приложения высоких технологий из многих отраслей науки и техники, может и должна служить одним из локомотивов инновационного развития России, о чем сейчас так часто говорят с высоких трибун (хотя в начале 1990-х г.г. с тех же трибун раздавались предложения о ликвидации этой отрасли как таковой). К счастью для страны, развитие ее лишь затормозилось, хотя и значительно. А в последнее время число техногенных катастроф с ракетами-носителями МКА существенно возросло и наносит ощутимый вред как экономике, так и престижу страны [58].

Можно констатировать, что современная космическая техника выходит на первое место в плане наполнения реальным содержанием термина «мехатронные комплексы» [17]. Его можно найти и в других отраслях промышленности, например, в измерительной технике (системы климат-контроля [59]), в медицинской технике (сканирующие лазерные терапевтические установки [60]), в станкостроении (непоточные технологические линии [47]), в авиационной технике (беспилотные летательные аппараты с пультами управления [52]), в военной технике (управляемые снаряды [25]).

Введение в научный оборот термина «мехатронизированный объект». Термин «мехатронизированный» объект, появившийся в отечественной технической литературе в работах [10, 30, 33], за последние годы так прочно вошел в фундаментальные понятия мехатроники, что используется как общепризнанный в работах [2, 61–68]. Термин «мехатронизированный объект» позволяет применять мехатронные методы, принципы и подходы к объектам, не полностью отвечающим канонам мехатроники. В связи с появлением указанного термина появилось предложение о создании класса мехатронных машин [62–66], включающих в себя и мехатронизированные. Вопрос требует фиксации в документах типа MMS-Terms-2003 [31].

Коррекция использования термина «мехатронный модуль». Широкое использование термина «мехатронный модуль», который весьма популярен в научной литературе (см., например, [67, 68]), не совсем корректно. Дело в том, что модуль, согласно определению, – унифицированный узел [42], а мехатронный объект в настоящее время в основном предполагает эксклюзивный, креативный характер. В этой связи логичнее использовать термин «мехатронный узел». Из вышеизложенного логично вытекает, что для мехатронных объектов одним из характерных признаков является отсутствие ремонтпригодности, а долговечность носит ресурсный характер [69]. При этом для производства мехатронных объектов используют следующие типы производств: непоточное [47], индивидуальное или мелкосерийное [18].

Взаимосвязь мехатроники с авионикой, автотехникой, логистикой. В связи с широким внедрением мехатронных технологий в авиа- и автомобилестроении необходимо более четко установить взаимосвязь мехатроники с авионикой [18, 43, 52, 70] и автотехникой [6, 18, 71, 72].

Термин «авионика» появился на Западе в начале 1970-х гг., когда электронная техника достигла такого уровня развития, что стало возможно применять электронные устройства в бортовых авиационных системах и за счет этого существенно улучшать качественные показатели применения авиации. В том числе, в этот период времени появились первые бортовые электронные вычислители (компьютеры), а также принципиально новые автоматизированные и автоматические системы управления и контроля [43]. В энциклопедии [70] авионика трактуется как приборная часть летательного аппарата.

Широкое внедрение электроники привело к возникновению такой области техники, как автомобильная электроника. В иностранной литературе в качестве синонима автомобильной электроники используются также термины «мехатроника» и «электроника» или «автоника» (объединение слов «автомобильная электроника»). Понятие «автомобильная электроника» широко распространено в современной технической литературе [71, 72], но его общепринятое определение отсутствует. Авторы предлагают трактовать эти понятия как мехатронику в авиационной и автомобильной технике.

Логистика – управление потоками движения товаров, а также диспетчирование – управление группами объектов, имеют некоторую взаимосвязь с мехатроникой, которую необходимо вскрыть более подробно.

Отдельные замечания. К проблемам терминологии в области мехатроники можно отнести и следующий факт. В некоторых научных журналах имеются разделы под названиями «мехатронные системы», однако в них печатаются статьи, в которых термин «мехатроника» не упоминается, таким образом, авторы априори считают исследуемые системы мехатронными (например, [73]).

В работе [2] отмечено, что даже после того, как ИFToMM в лице своего комитета по стандартизации терминологии приняла ряд базовых терминов в области мехатроники [16, 31] некоторые ученые продолжают считать, что мехатроника – просто «хорошее проектирование» или даже «это то, что мы уже давно делаем, не давая нового имени».

Заключение

Из вышеизложенного следует, что:

- завершается процесс создания базовой терминологии в области мехатроники (о чем свидетельствуют работы [1–18, 21–23, 27, 28, 30–35, 42, 43, 62–65]);
- установлена взаимосвязь мехатроники с робототехникой и кибернетикой [5, 14, 16, 18–22];
- успешно идет процесс признания уровневого характера мехатроники (т.е. взаимосвязи мехатроники с микро- и наносистемной техникой) [7, 18, 27, 30, 31, 34];
- наполняется реальным содержанием термин «мехатронные комплексы» [17];
- получил признание введенный в научный оборот термин «мехатронизированный объект», позволяющий применять мехатронные методы, принципы и подходы проектирования к объектам, не полностью отвечающим канонам мехатроники [2, 16, 30, 42, 43, 61–66].

Перечислим тезисно ряд положений, которые нуждаются в уточнении для более успешного развития мехатроники:

- необходимо расширить понимание мехатроники от компьютерного управления движением до управления физическим состоянием объекта [5, 74, 75];
- следует добиться того, чтобы все определения отвечали основным критериям научно-технических терминов (однозначность, точность, четкость, краткость) [3, 5];
- следует сформировать отношение к введенному в научный оборот термину «мехатронный класс», включающему в себя мехатронные и мехатронизированные объекты [61–63];
- важно определить взаимосвязь мехатроники и логистики (т.е. управления потоками движения товаров, а также диспетчирования – управления группами объектов);
- необходимо определить такое понятие, как электромехатроника [76].

Изложенные в публикации некоторые предложения по корректировке понятий являются необходимым этапом развития мехатроники.

Авторы рассматривают данную публикацию в качестве апробации своих предложений, которые выносятся на рассмотрении 25-го рабочего заседания комиссии по стандартизации и терминологии ИFToMM, которое планируется провести на базе НИУ ИТМО в июне 2014 г.

Литература

1. Исии Т., Симояма И., Иноуэ Х., Накадзима Н. Мехатроника / Пер. с яп. С.Л. Масленникова; Под ред. В.В. Василькова. М.: Мир, 1988. 318 с.
2. Подураев Ю.В. Мехатроника: основы, методы, применение: Учеб. пособие для вузов. 2-е изд. М.: Машиностроение, 2007. 256 с.

3. Филимонов Н.Б. Эволюция мехатроники // Известия Тульского государственного университета. Сер. Вычислительная техника. Информационные технологии. Системы управления. 2006. Т. 2. В. 3. С. 277–281.
4. Подураев Ю.В. Актуальные проблемы мехатроники // Мехатроника, автоматизация, управление. 2007. № 4. С. 50–54.
5. Теряев Е.Д., Филимонов Н.Б., Петрин К.В. Мехатроника как компьютерная парадигма развития технической кибернетики // Мехатроника, автоматизация, управление. 2009. № 6. С. 2–10.
6. Алексеев О.П., Алексеев В.О., Логачов Е.П., Туренко А.И. Телематика, мехатроника и синергетика на автомобильном транспорте // Автомобильный транспорт (Харьковский национальный автомобильно-дорожный университет). 2009. № 25. С. 266–269. [укр.]
7. Теряев Е.Д., Филимонов Н.Б. Наномехатроника: состояние, проблемы, перспективы // Мехатроника, автоматизация, управление. 2010. № 1. С. 2–14.
8. Осипов Ю.М. К вопросу о развитии понятия «мехатроника» // Доклады Томского государственного университета систем управления и радиоэлектроники. 2010. № 1 (21). Ч. 2. С. 193–198.
9. Кориков А.М. О развитии понятия «мехатроника» // Доклады Томского государственного университета систем управления и радиоэлектроники. 2010. № 1 (21). Ч. 2. С. 199–202.
10. Shalobaev E.V. Mechatronics: Today Problems and Development trends of Terminology // Proc. 23rd Working Meeting of the IFToMM Permanent Commission for Standardization of Terminology on MMS. Minsk–Gomel, Belarus, 2010. P. 111–118.
11. Кориков А.М. Еще раз о мехатронике как науке // Мехатроника, автоматизация, управление. 2011. № 5. С. 2–8.
12. Толочка Р.-Т., Шалобаев Е.В. Терминология в области мехатроники // Материалы Десятой сессии Международной научной школы «Фундаментальные и прикладные проблемы надежности и диагностики машин и механизмов» (Фридендеровские чтения). ВУР-2011. СПб: Институт проблем машиноведения РАН, 2011. С. 38–44.
13. Сергеев С.Ф. Мехатроника как конвергентная научно-практическая дисциплина // Мехатроника, автоматизация, управление. 2012. № 1. С. 2–6.
14. Шалобаев Е.В., Толочка Р.-Т. К вопросу о терминологии в области мехатроники // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 5 (81). С. 148–151.
15. Осипов Ю.М., Щербинин С.В. О мехатронике как научной основе создания высокотехнологичной продукции // Мехатроника, автоматизация, управление. 2012. № 8. С. 2–6.
16. Шалобаев Е.В., Толочка Р.-Т. О рекомендациях IFToMM по терминологии в области мехатроники // Мехатроника, автоматизация, управление. 2013. № 2. С. 2–5.
17. Шалобаев Е.В. Вопросы терминологии и миниатюризации космических аппаратов // Мехатроника, автоматизация, управление. 2013. № 10. С. 60–66.
18. Шалобаев Е.В. Микросистемная техника и мехатроника: особенности соотношения микро- и макроуровней // Микросистемная техника. 2000. № 4. С. 5–9.
19. Зенкевич С.Л., Ющенко А.С. Основы управления манипуляционными роботами: учебник для вузов. 2-е изд., испр. И доп. М.: изд-во МГТУ им. Н.Э. Баумана, 2004. 480 с.
20. Робототехника и мехатроника: Сборник трудов / Под ред. А.М. Потапова. М.–СПб: Балтийский государственный технический университет, 1996. 235 с.
21. Юревич Е.И., Игнатова Е.И. Основные принципы мехатроники // Мехатроника, автоматизация, управление. 2006. № 3. С. 10–12.
22. Лопота В.А., Юревич Е.И. Экстремальная робототехника и мехатроника. Принципы и перспективы развития // Мехатроника, автоматизация, управление. 2007. № 4. С. 37–42.
23. Александров В.В. Несколько слов о мехатронике // Мехатроника. 2000. № 1. С. 4.
24. Энеев Т.М., Овчинников М.Ю., Голиков А.Р. Прикладная небесная механика и управление движением: Сборник статей, посвященный 90-летию со дня рождения Д.Е. Охочимского. М.: Институт прикладной математики им. М.В. Келдыша, 2010. 368 с.
25. Сазыкин Ю.М., Филиппов С.И. Мехатронные системы управления огнем и наведением огневых средств артиллерии // Мехатроника, автоматизация, управление. 2009. № 4. С. 52–57.
26. Кибернетика, мышление, жизнь / Под ред. А.И. Берга, Б.В. Бирюкова, И.Б. Новика, И. В. Кузнецова, А. Г. Спиркина. М.: Мысль, 1964. 511 с.
27. Теряев Е.Д., Филимонов Н.Б., Петрин К.В. Современный этап развития мехатроники и грядущая конвергенция с нанотехнологиями // Материалы 5-й Научно-технической конференции «Мехатроника, автоматизация, управление». СПб: ГНЦ ЦНИИ «Электроприбор», 2008. С. 9–20.
28. Теряев Е.Д., Петрин К.В., Филимонов Н.Б. От кибернетики к информатике и мехатронике: эволюция современных представлений // Труды X Международной конференции «Проблемы управления и моделирования в сложных системах». Самара: Самарский научный центр РАН, 2008. С. 22–25.
29. Острём К. Введение в стохастическую теорию управления. М.: Мир, 1973. 324 с.

30. Шалобаев Е.В. Теоретические и практические проблемы развития мехатроники // Современные технологии: Сборник / Под ред. С.А. Козлова. СПб: ГИТМО (ТУ), 2001. С. 46–66.
31. Delft University of Technology (Netherlands). Standardization of terminology for the mechanism and machine science (MMS-Terms-2003) [Электронный ресурс]. Режим доступа: <http://www.iftomm.3me.tudelft.nl/1049/frames.html>, свободный. Яз. англ. (дата обращения 25.11.2013).
32. Аршанский М.М., Шалобаев Е.В. Мехатроника: основы глоссария // Мехатроника. 2001. № 4. С.47–48.
33. Шалобаев Е.В. К вопросу о международном трансляторе по мехатронике // Мехатроника. 2002. № 2. С. 13–14.
34. Лопота В.А., Юревич Е.И. Мехатроника – основа интеллектуальной техники будущего // Микросистемная техника. 2003. № 1. С. 36–38.
35. Шалобаев Е.В. Проблемы и тенденции развития терминологии в современных условиях // Датчики и системы. 2004. № 4. С. 29–32.
36. Scherge M., Gorb S. Biological micro- and nanotribology: Nature's solutions. Berlin–Heidelberg: Springer-Verlag, 2001. 304 p.
37. Шилько С.В., Старжинский В.Е., Бабин А.П., Зернин М.В., Шалобаев Е.В. Особенности расчета сопряжений компонентов МЭМС // Микросистемная техника. 2003. № 6. С. 16–20.
38. Шалобаев Е.В. Микросистемная техника и тенденции развития современной трибологии // Микросистемная техника. 2003. № 9. С. 26–27.
39. Шалобаев Е.В., Старжинский В.Е., Шилько С.В. Тенденции развития современной трибологии на микро- и наноуровнях // Труды Международного симпозиума «Гидродинамическая теория смазки»: В 2-х т. М.: Машиностроение, 2005. Т. 2. С. 289–295.
40. Мышкин Н.К., Петроковец М.И. Трение, смазка, износ. М.: ФИЗМАТЛИТ, 2007. 368 с.
41. Шилько С.В., Старжинский В.Е., Бабин А.П., Зернин М.В., Шалобаев Е.В. Особенности расчета сопряжений компонентов МЭМС // Микросистемная техника. 2003. № 6. С. 16–20.
42. Шалобаев Е.В. К вопросу об определении мехатроники и иерархии мехатронных объектов // Датчики и системы. 2001. № 7. С. 62–64.
43. Шалобаев Е.В. Теоретические и практические проблемы развития авионики и мехатроники // Проблемы интеллектуального управления в авионике. СПб: СПбГУ ИТМО, 2005. С. 12–43.
44. Управление мехатронными вибрационными установками / Под ред. И.И. Блехмана, А.Л. Фрадкова. СПб: Наука, 2001. 278 с.
45. Паршин Д.Я. Анализ и синтез робототехнических и мехатронных комплексов для крупнопанельного и монолитного строительства. Дис. ... докт. техн. наук: 05.02.05. Новочеркасск, 2006. 328 с.
46. Шахнин В.А. Мехатронные комплексы магнитной локации технических объектов с элементами из нанокристаллических магнитомягких сплавов. Дис. ... докт. техн. наук: 05.02.05, 05.02.11. Владимир, 2009. 365 с.
47. Белоновская И.Д., Цветкова К.Е., Осадчий Ю.С. К вопросу выбора технологического оборудования непоточного производства // Известия Самарского научного центра РАН. 2011. Т. 13. № 1 (3). С. 521–524.
48. Бобцов А.А., Боргуль А.С., Зименко К.А., Маргун А.А. Применение мехатронных комплексов в обучении автоматизации и робототехнике // Дистанционное и виртуальное обучение. 2013. № 2. С. 22–30.
49. Фатеев В.Ф. Концепция создания орбитальных группировок сверхмалых космических аппаратов в интересах информационного обеспечения северных территорий России // Изв. вузов. Приборостроение. 2004. Т. 47. № 3. С. 5–9.
50. Smirnov G.V., Ovchinnikov M.Ya., Guerman A. Use of solar radiation pressure to maintain a spatial satellite formation // Acta Astronautica. 2007. V. 61. N 7–8. P.724–728.
51. Овчинников М.Ю. Малыши завоевывают мир [Электронный ресурс]. Режим доступа: http://ru.wikipedia.org/wiki/%CC%E0%EB%FB%E5_%F1%EF%F3%F2%ED%E8%EA%E8, свободный. Яз. рус. (дата обращения 25.11.2013).
52. Распопов В.Я. Микросистемная авионика. Тула: Гриф и К°, 2010. 248 с.
53. Овчинников М.Ю. Малые мира сего // Компьютерра. 2007. № 15. С. 37–43.
54. Трошин Е.В. О проблеме классификации космических аппаратов, порожденной миниатюризацией // Нано- и микросистемная техника. 2005. № 1. С. 45–48.
55. Малые спутники [Электронный ресурс] // Википедия [Официальный сайт]. Режим доступа: http://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D0%BB%D1%8B%D0%B5_%D1%81%D0%BF%D1%83%D1%82%D0%BD%D0%B8%D0%BA%D0%B8, свободный. Яз. рус. (дата обращения 25.11.2013).
56. Ардашов А.А. Основные тенденции в конструировании МКА // Изв. вузов. Приборостроение. 2004. Т. 47. № 3. С. 9–15.

57. Фатеев В.Ф., Шутов В.С., Кремез Г.В. Межуниверситетская программа развития технологий на основе малых и сверхмалых КА // Изв. вузов. Приборостроение. 2004. Т. 47. № 3. С. 15–17.
58. Крылов А. Орбитальная группировка космических аппаратов гражданского и двойного назначения России за первые 10 лет XXI века // Технологии и средства связи. 2011. № 4. С. 58–59.
59. Петров С.Ю., Шалобаев Е.В. Универсальные регистрирующие и показывающие приборы как элемент иерархии мехатронных объектов // Мехатроника. 2001. № 5. С. 29–34.
60. Шалобаев Е.В., Юркова Г.Н., Ефименко В.Т., Ефименко А.В., Леонтьева Н.В. Сканирующие лазерные установки в медицине // Научно-технический вестник СПб ГИТМО (ТУ). 2001. № 4. С. 147–150.
61. Подураев Ю.В. Анализ и проектирование мехатронных систем на основе критерия функционально-структурной интеграции. Ч. I. Функционально-структурный подход к проектированию мехатронных систем // Мехатроника. 2002. № 4. С. 28–34.
62. Горбатов П.А., Петрушкин Г.В., Лысенко Н.М., Павленко С.В., Косарев В.В. Горные машины для подземной добычи угля. Донецк: ДонНТУ, 2006. 669 с.
63. Горбатов П.А. Параллельное проектирование выемочных машин мехатронного класса как энергетических, динамических и автоматизированных систем // Наукові праці ДонНТУ. Сер. Гірничо-електромеханічна. 2010. В. 18 (172). С. 13–25. [укр.]
64. Горбатов П.А., Лысенко Н.М., Подобедов Н.И. Математическая и структурная модели для оптимального проектирования подсистем подвески и перемещения исполнительного органа проходческих комбайнов мехатронного класса // Горное оборудование и электромеханика. 2010. № 7. С. 9–15.
65. Горбатов П.А. Особенности параллельного проектирования горных выемочных машин новых поколений как энергетических систем мехатронного класса // Горное оборудование и электромеханика. 2010. № 10. С. 39–42.
66. Краснов С.В., Гурьянов Д.А., Краснов С.С. Интеллектуальное управление в мехатронных системах // Вестник Волжского университета им. В.В. Татищева. 2011. № 17. С. 100–104.
67. Казмиренко В.Ф. Электрогидравлические мехатронные модули движения: Основы теории и системное проектирование: Учеб. пособие. М.: Радио и связь, 2001. 432 с.
68. Егоров О.Д., Подураев Ю.В. Конструирование мехатронных модулей: Учебник. М.: МГТУ «СТАНКИН», 2004. 360 с.
69. Козырев Ю.Г. Промышленная роботизация, мехатроника и проблемы автоматизации сборочных операций // Сборка в машиностроении, приборостроении. 2006. № 2. С. 24–26.
70. Авионика России: Энциклопедический справочник / Под общ. ред. С.Д. Бодрунова. СПб: НАОП, 1999. 780 с.
71. Сига Х., Мидзутани С. Автомобильная электроника: Учеб. пособие. М.: Мир, 1989. 227 с.
72. Алексієв О.П., Алексієв В.О., Богаєвські О.Б., Бороденко Ю.М., Биков О.М., Неронов С.М., Калмиков В.І., Рожкова С.Е., Серіков С.А., Сергієнко О.Ю., Смирнов О.П., Тирса В.В., Чаплигін Е.О., Шапошнікова О.П. Електронний варіант термінологічного словника. Атомобільна мехатроніка. Харків: Харківський національний автомобільно-дорожній університет, 2011. 39 с.
73. Пупков К.А. Седьмой форум по мехатронике // Мехатроника. 2001. № 3. С. 46–47.
74. Мехатронные системы // Научно-технический вестник СПбГУ ИТМО. 2008. № 3 (48). С. 21–107.
75. Крайнов А.Ф. Механика машин: Фундаментальный словарь. М.: Машиностроение, 2000. 904 с.
76. Герман-Галкин С.Г., Ермилов А.С. Классификация и анализ электромеханотронных систем // Изв. вузов. Приборостроение. 1996. Т. 39. № 3. С. 13–21.

Шалобаев Евгений Васильевич

– кандидат технических наук, профессор, доцент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, shalobaev47@mail.ru

Толочка Римантас-Тадас Антонасович

– доктор технических наук, профессор, Каунасский технологический университет, Каунас, Литва, tadas.tolocka@ktu.lt.

Evgeniy Shalobaev

– PhD, Associate professor, Associate professor, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, shalobaev47@mail.ru

Rimantas-Tadas Tolockka

– D.Sc., Professor, Kaunas University of Technology, Kaunas, Lithuania, tadas.tolocka@ktu.lt.

УДК 681.7.023; 681.7.03; 681.7.06

ОПЫТ ПРИМЕНЕНИЯ И ПЕРСПЕКТИВЫ ТЕХНОЛОГИИ АЛМАЗНОГО МИКРОТОЧЕНИЯ

В.М. Медунецкий^а, С.В. Солк^б

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, vm57med@yandex.ru

^б ОАО «НИИ оптико-электронного приборостроения», Сосновый Бор, Россия, solk@sbor.net

На основе опыта применения технологии алмазного микроточения в Научно-исследовательском институте оптико-электронного приборостроения и анализа литературных источников рассмотрены области использования и направления дальнейшего развития технологии алмазного микроточения. Отмечается, что на сегодняшний день технология в основном применяется для формообразования рабочих поверхностей оптических элементов, изготовления киноформ, деталей точной механики, а также декоративной обработки. К новым областям применения алмазного микроточения относятся: изготовление рабочих поверхностей пресс-форм; технологии, позволяющие совместить технологические процессы формообразования и сборки оптико-механических систем; исследование технологической наследственности путем анализа дифракционной структуры поверхности, обработанной алмазным микроточением; получение рабочих поверхностей оптических элементов с астигматизмом заданной величины и направления, используемых в глубокоохлаждаемых оптико-механических системах. Приведены перспективные направления развития алмазного микроточения: создание станков на основе новых кинематических схем, химическая и электрохимическая дополировка рабочих поверхностей оптических элементов после алмазного микроточения с целью уменьшения шероховатости и создания нерегулярного рельефа для работы в видимом и ближнем инфракрасном диапазоне спектра, разработка технологии алмазного микроточения для новых материалов и совершенствование известных технологий на основе современных представлений о состоянии поверхностного слоя, создание упорядоченного нанорельефа для генерации поверхностных электромагнитных волн для управления свойствами поверхности.

Ключевые слова: алмазное микроточение, совмещенные технологии, исследование технологической наследственности, пресс-форма, поверхностные электромагнитные волны, химическая и электрохимическая полировка.

APPLICATION EXPERIENCE AND PROSPECTS OF DIAMOND MICRO-TURNING TECHNOLOGY

V. Medunetskiy^с, S. Solk^д

^с Saint Petersburg National University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, vm57med@yandex.ru

^д PLC «Scientific Research Institute for Optoelectronic Instrument Engineering», Sosnovy Bor, Russia, solk@sbor.net

The paper deals with application and further progress of diamond micro-turning technology (DT) based on the experience obtained in S&R Institute for Optoelectronic Instrument Engineering and compiled from literary sources. It is pointed out that the diamond turning technology is nowadays mainly used for shaping of the optical elements (OE) working surfaces, manufacturing of kinoforms, elements of fine mechanics and decorative processing. New DT applications are: shaping of the die mold working surfaces, combined technologies, making it possible to combine the shaping technologies and assembling of the optical mechanical systems (OMS), research into technological hereditary, consisting in studies of the surface diffraction structure machined by diamond turning, shaping of the OE working surfaces with astigmatism of the desired value and direction being used in the deeply cooled OMS. Perspective lines of DT development are given, such as production of machines based on the new kinematic schemes, chemical and electrochemical polishing of the OE working surfaces after DT with the aim to decrease roughness and to make irregular relief in visible and near-infrared spectral ranges, development of DT technologies for creation of new materials and improvement of the known technologies taking into account the concepts about the state of the surface layer, creation of regular nanorelief for generation of the surface electromagnetic waves to provide the surface properties control.

Keywords: diamond micro-turning, combined technologies, research into technological hereditary, die mold, surface electromagnetic waves, chemical and electrochemical polishing.

Введение

Одним из наиболее эффективных способов обеспечения требуемых качественных показателей высокоточной промышленной продукции является использование современных прогрессивных технологий. В оптической промышленности при производстве оптических элементов (ОЭ) и оптических систем (ОС) такой технологией является алмазное микроточение (АМТ). Под АМТ понимается лезвийная обработка различных материалов на сверхпрецизионных станках режущим инструментом из природных монокристаллических алмазов. Несмотря на высокую стоимость используемого оборудования, жесткие требования к помещениям АМТ (вибрационные, температурные, акустические, противопопылевые ограничения), специфические требования к режущему инструменту, данная технология получила существенное развитие [1].

Можно отметить, что с каждым годом растет число материалов, для которых разработана технология АМТ. Это медь, алюминий и их сплавы, серебро, золото, платина, свинец, гальванический никель, германий, кремний, селенид цинка, полиметилметакрилат (ПММК) и многие другие. Совершенствуются станки АМТ, режущий инструмент, контрольно-измерительное оборудование.

В настоящей работе рассматриваются области применения АМТ, как традиционные, широко используемые в промышленности, так и новые, на сегодняшний день недостаточно известные, разработанные сравнительно недавно. Рассмотрены перспективные направления развития АМТ.

Традиционные области применения АМТ

По сравнению с альтернативными технологиями в первую очередь с традиционным полированием, а также с ионной обработкой, вакуумными, термопластическими и другими технологиями, АМТ обладает значительными преимуществами, которые и определяют основные области его применения.

1. Изготовление ОЭ с рабочими поверхностями сложной формы. Посредством АМТ можно изготовить такие сложные поверхности, получение которых другими способами крайне сложно или практически невозможно: асферические поверхности высших порядков, *W*-аксиконы, аксиконы с асферическими, например, параболическими поверхностями.
2. Изготовление лазерных зеркал. При АМТ отсутствует шаржирование обрабатываемой поверхности зернами абразива, что существенно повышает стойкость отражающей поверхности к лазерному излучению высокой мощности.
3. Изготовление металлооптических элементов (МОЭ). АМТ осуществляется на сверхпрецизионных станках (радиальное и осевое биение шпинделя, точность направляющих – доли микрометра). Это позволяет получать на конструкционных материалах, в первую очередь металлах и их сплавах, не только высококачественные оптические поверхности сложной формы, но и базовые поверхности, элементы крепления, а также элементы, обеспечивающие контроль, сборку и юстировку этих элементов. Базовые поверхности позволяют однозначно ориентировать обрабатываемую деталь на станке, что дает возможность совместить геометрическую ось детали с осью вращения шпинделя станка. После обработки рабочей поверхности нет необходимости в операции центрирования. Такие МОЭ широко используются в «насыпных» объективах. Возможность изготовления базовых поверхностей с микрометрическими допусками позволяет получить неюстируемые системы высокого качества. Известны конструкции глубокоохлаждаемых ОЭ [2], устанавливаемые в корпус объектива с «натягом» 5–15 мкм (конструкции с упругими кинематическими элементами), не вызывающие деформации оптической поверхности. Эти конструкции используются в нерасстраиваемых системах, работающих в условиях высоких ударных и вибрационных нагрузок.
4. Изготовление киноформной оптики для инфракрасных (ИК) систем. Киноформы являются эффективным средством коррекции хроматических аберраций, что крайне важно для ИК линзовых систем [3, 4]. Это связано с ограниченностью количества материалов ИК диапазона. Технология АМТ позволяет совместить получение асферической поверхности и дифракционной структуры на ней в рамках единого технологического процесса.
5. Изготовление деталей, относящихся к области точной механики. Высокая точность станков АМТ позволяет изготавливать детали с минимальными значениями некруглости, нецилиндричности, неплоскостности обработанных поверхностей, с их точным взаимным расположением, а также эталоны для контрольно-измерительных приборов.
6. Декоративная обработка деталей из металлических и полимерных материалов.
7. Технология АМТ позволяет сразу получать зеркальную поверхность, что дает возможность применять интерференционные методы контроля на всех стадиях обработки, исключив трудоемкие и недостаточно точные контактные методы.

Отметим, что технология АМТ весьма высокопроизводительна. В ряде случаев алмазным точением оптический элемент может быть изготовлен за несколько часов, в то время как процесс полирования занимает не меньше месяца.

Новые области применения АМТ

В последние годы в Научно-исследовательском институте оптико-электронного приборостроения (ОАО «НИИ ОЭП») был разработан и апробирован ряд методик и технологий, использующих АМТ.

1. Изготовление ОЭ при разработке новых ОС для проверки теоретических расчетов. ОЭ изготавливают посредством АМТ, даже если в дальнейшем они будут изготавливаться по другой технологии, например, вакуумным прессованием [5]. Это обусловлено тем, что изготовление элемента АМТ из ряда материалов, например, ПММК, гораздо быстрее и дешевле изготовления пресс-форм (ПФ) с последующим прессованием того же ОЭ.
2. Изготовление ПФ [6]. Для получения десятков тысяч ОЭ методом прессования ПФ изготавливаются из специальных сортов сталей и имеют рабочие поверхности оптического качества. Стоимость таких ПФ может составлять десятки тысяч долларов. Конструкция ПФ не поддается точному расчету и обычно требует поэтапной доработки из-за усадки материала и других факторов. На стадии проектирования и изготовления макетного образца бывает важно изготовить и испытать хотя бы один образец устройства. Применение АМТ позволяет изготовить ПФ из алюминиевых или медных сплавов, при-

- годных для вакуумного прессования нескольких образцов ОЭ. Следует отметить, что использование стали для изготовления ПФ невозможно из-за быстрого разрушения алмазного резца.
3. Изготовление ОЭ с рабочими поверхностями, имеющими астигматизм заданной величины и направления [7]. Дисбаланс заготовки ОЭ при АМТ вызывает астигматическую ошибку волнового фронта излучения, отраженного от рабочей поверхности. В ряде случаев, вводя дисбаланс заданной величины и направления (на практике на заготовку в заданной точке крепится груз с расчетной массой), можно получить направленный «астигматизм». Например, в глубоко охлаждаемых объективах при рабочих температурах часто возникает астигматизм, связанный с анизотропией свойств материала, из которого изготавливаются оптические элементы. Вводя искусственный астигматизм при нормальной температуре, можно улучшить форму оптической поверхности при рабочей температуре.
 4. Изготовление оптико-механических систем (ОМС) по совмещенным технологиям [6]. В оптическом приборостроении технологический процесс изготовления ОМС, например, объектива, включает этапы изготовления ОЭ, механических деталей и сборки. На стадии сборки возможна расточка корпуса объектива и оправ линз, центрировка ОЭ. Формообразование рабочих поверхностей ОЭ к моменту сборки всегда завершено. При использовании АМТ с целью упрощения технологии и повышения точности взаимного расположения поверхностей в ряде случаев целесообразно совместить сборку с формообразованием ОЭ.
 5. Исследование технологической наследственности [6]. Поверхность, обработанная АМТ, представляет собой дифракционную решетку, параметры которой зависят от режимов резания (скорости резания, подачи и глубины резания) и используемого инструмента (радиуса заточки резца). Заготовки из одних и тех же материалов, обработанные на станке АМТ при одинаковых режимах одним и тем же резцом, имеют разные характеристики шероховатости, если образцы имеют разную технологическую наследственность, т.е. прошли разную механическую и термическую предварительную обработку. При оценке качества обработанной поверхности могут быть использованы методики и приборы, применяемые для контроля дифракционных структур.

Перспективные направления развития АМТ

На основе анализа литературных источников и опыта проектирования, конструирования и изготовления ОЭ и оптико-механических систем опытных образцов приборов в ОАО «НИИ ОЭП» можно обозначить следующие перспективные направления развития АМТ.

1. Создание мелкомасштабного упорядоченного нанорельефа на поверхности металла. АМТ может использоваться для создания структур, необходимых для генерации поверхностных электромагнитных волн [6, 8]. Такие структуры могут использоваться как для определения оптических свойств поверхности металла, так и для управления этими свойствами.
2. Разработка технологий АМТ для оптических стекол, карбида кремния и других материалов. Количество материалов, формообразование которых возможно посредством АМТ, постоянно растет. Так, в работе [9] описывается технология изготовления АМТ зеркал из магния. Но в этой и других подобных работах разработка технологии АМТ сводится к выбору режимов резания, вида смазывающе-охлаждающей жидкости и способу ее подачи для станка определенного типа. В то же время в литературе описываются и другие подходы к организации технологического процесса АМТ на новых принципах. Например, в [10] предлагается обрабатывать хрупкие материалы, например, оптическое стекло, АМТ при нагреве заготовки до температуры, близкой к точке размягчения. При этом должна образовываться сливная непрерывная стружка. В частности, в работе [11] предлагается осуществлять выбор режимов резания и параметров режущего инструмента на основе термодинамического анализа напряженно деформированного состояния поверхностного слоя рабочей поверхности ОЭ.
3. Создание следующих поколений станков АМТ на основе новых кинематических схем. На сегодняшний день и в России, и за рубежом выпускаются станки АМТ токарного и фрезерного типа, предназначенные для формообразования осесимметричных поверхностей. Информация о станках, позволяющих получать поверхности, задаваемые разными уравнениями в разных сечениях, либо отсутствует, либо носит противоречивый характер. Исключения составляют, пожалуй, станки для изготовления тороидальных поверхностей, но и то лишь в силу своей относительной (две сферические поверхности разного радиуса) простоты. В то же время применение неосесимметричных поверхностей крайне эффективно в брахитах и других ОС.
4. Дополнительная обработка металлических поверхностей с целью уменьшения их шероховатости. Поверхности, обработанные с применением АМТ, имеют невысокую шероховатость. Так, детали из алюминиевого сплава АМг-6 и бескислородной меди МОб, проточенные на станках МО 200 и МО 600 ПЛ [1], имеют среднеквадратичную шероховатость Rq 0,003–0,008 и 0,002–0,006 мкм соответственно. В то же время шероховатость имеет регулярный рельеф, что вызывает дифракционные эффекты для

излучения видимого и ближнего ИК диапазонов. На практике МОЭ с такой шероховатостью используются в ОС, работающих в спектральном диапазоне 3–14 мкм.

Для работы ОЭ в видимом и ближнем ИК диапазонах света необходимо уменьшение шероховатости и создание нерегулярного рельефа рабочих поверхностей. Проведенные эксперименты показали, что при дополировке МОЭ со сферическими, плоскими и асферическими рабочими поверхностями малоразмерным и полноразмерным инструментом происходит «матирование» обрабатываемых поверхностей, после чего требуется полноценная полировка. Проведенные эксперименты по дополировке с использованием технологий химического и электрохимического полирования позволяют говорить о возможности их использования. Перспективной представляется также ионная обработка поверхностей [12].

Опыт применения алмазного микроточения при обработке оптических деталей

Рассмотрим возможности АМТ на примере ряда ОЭ, изготовленных в ОАО «НИИ ОЭП». Традиционно формообразование подобных деталей осуществлялось либо полированием полноразмерным и малоразмерным инструментом, либо ионной обработкой. Сложность процесса полирования заключается в отсутствии зон для выхода инструмента. Ионная обработка не предназначена для формообразования большой площади обрабатываемой поверхности. Также следует учитывать плохую обрабатываемость для этих технологий использованных материалов. В ОАО «НИИ ОЭП» были разработаны методики и соответственно технологии, которые позволили решить вышеуказанные проблемы [13].

На рис. 1, а, показана линза сложной конструкции из ПММК, имеющая три рабочих поверхности: *A*, *B* и *C*. Поверхность *A* – сферическая, поверхности *B* и *C* – асферические высших порядков. Поверхности *B* и *C* обрабатываются за один установ, и поэтому при диаметре линзы 200 мм несовпадение фокусов ($F'_{B,C}$) этих поверхностей составляет доли микрометра.

На рис. 1, б, показан *W*-аксикон диаметром 350 мм из алюминиевого сплава 1201 с рабочими поверхностями *D* и *E*. Прямолинейность образующих составляет менее микрометра. Угол α составляет 90° , при точности изготовления не хуже 10 угл. с. Поверхность *G* – плоская, изготавливается за один установ с рабочими поверхностями и используется для предварительной юстировки аксикона. При использовании специализированных станков возможно изготовление *W*-аксиконов не с прямолинейными образующими, а имеющими сложную, например, параболическую, форму.

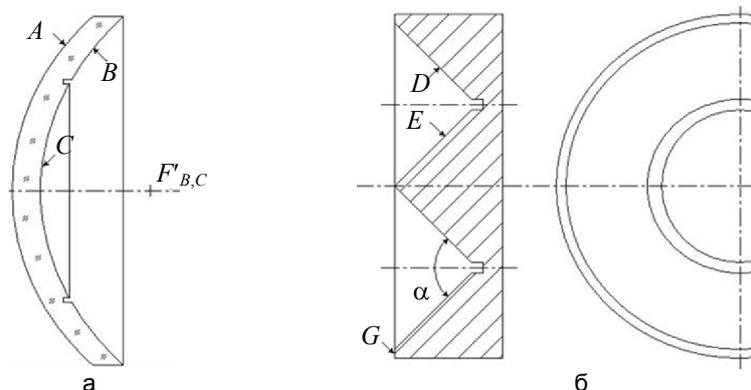


Рис. 1. Конструкции оптических элементов: из ПММК (а); *W*-аксикон из алюминиевого сплава (б)

Из алюминиевого сплава АМг-6 было изготовлено зеркало диаметром 150 мм с внутренним отверстием диаметром 30 мм и толщиной 15 мм. Масса зеркала составляла около 400 г. Рабочая поверхность зеркала – парабола с фокусным расстоянием 150 мм. Формообразование велось на станке токарного типа МО-200. Финишная обработка осуществлялась при глубине резания 0,005 мм, скорости вращения шпинделя станка 700 об/мин и скорости подачи суппорта 0,003 мм на оборот. В качестве смазочно-охлаждающей жидкости использовался спирт технический этиловый марки «А». В качестве режущего инструмента использовался резец из природного монокристаллического алмаза массой 0,5 кар. Передний угол заточки резца отрицательный – 4° , задний – 10° , радиус режущей кромки 5 мм. Среднеквадратичное отклонение полученной поверхности составило $0,15\lambda$ для $\lambda = 0,6328$ мкм, а среднеквадратичная шероховатость Rq не превышала 0,006 мкм. Формообразование вместе с юстировкой станка и проведением контрольно-измерительных операций заняло три рабочих смены (24 ч). Конструкция резцов, используемых для АМТ, приведена на рис. 2.

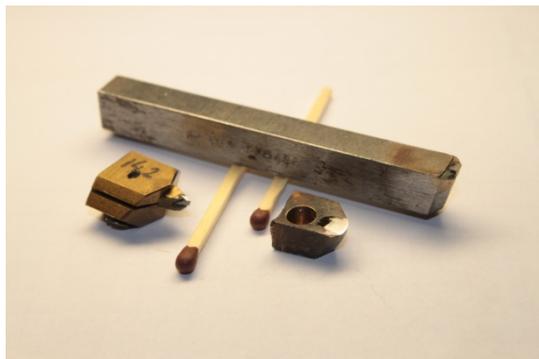


Рис. 2. Резцы из природных монокристаллических алмазов для АМТ

Зеркало предназначено для использования в оптико-электронном приборе космического базирования, рабочий спектральный диапазон которого составлял 8–14 мкм. Данные условия и определили выбор используемых материалов и технологии. Зеркало устанавливалось на платформу, изготовленную также из АМг-6, что делало конструкцию термонерасстраиваемой. Конструкция зеркала предусматривает наличие гладких отверстий для крепления его к платформе и не требует оправы, наличие которой привело бы к увеличению массы конструкции как минимум в 1,5 раза. Использование технологии полирования малоразмерным инструментом или технологии ионной обработки для достижения тех же параметров рабочей поверхности зеркала, по предварительным оценкам, заняло бы не менее 10 рабочих смен. Данный пример показывает существенное преимущество технологии АМТ.

Сравнительные характеристики различных технологических процессов приведены в таблице. Для детали, изготовленной посредством АМТ, приведены измеренные характеристики, для остальных случаев использовался метод экспертных оценок и результаты измерений аналогичных деталей. Шероховатость поверхности после ионной обработки не приведена, так как авторам не известны случаи формообразования металлических поверхностей такого размера с использованием данной технологии.

Характеристика	Алмазное микроточение	Полирование малоразмерным инструментом	Ионная обработка
Продолжительность формообразования и контроля, ч	24–30	80–100	100–120
Стоимость (без учета стоимости амортизации оборудования), тыс. руб.	20–30	60–90	90–120
Коэффициент отражения 8–14 мкм, %	97–98	95–96	–

Таблица. Сравнительные характеристики технологических процессов формообразования зеркала диаметром 150 мм с шероховатостью поверхности Rq 0,006 мкм и точностью формы 0,15λ для $\lambda = 0,6328$ мкм из сплава АМг-6

Заключение

Применение технологии алмазного микроточения позволяет изготавливать оптические элементы с асферическими поверхностями и дифракционными структурами, что делает возможным эффективное исправление аберрации. Это позволяет получать требуемые характеристики оптических систем, используя меньшее количество линз и зеркал [3]. Технологические возможности алмазного микроточения позволяют широко применять облегченные металлооптические элементы с рабочими, базовыми и юстировочными поверхностями. Все это ведет к уменьшению габаритов и массы оптико-механических систем, упрощению конструкции и технологии сборки, повышению коэффициента пропускания оптических систем.

Чтобы обеспечить высокие технические характеристики приборов, необходимо учитывать возможности и ограничения конкретных станков алмазного микроточения уже на стадии выполнения оптического расчета системы и конструирования прибора.

Литература

1. Справочник технолога-оптика / Под ред. М.А. Окатова. 2-е изд., перераб. и доп. СПб: Политехника, 2004. 679 с.
2. Мирошников М.М., Любарский С.В., Любарский Н.Х. Оптические зеркала для космической инфракрасной астрономии // Оптический журнал. 2010. Т. 77. № 4. С. 36–40.
3. Ган М.А. 50 лет киноформной оптики. Итоги и перспективы развития // Оптический журнал. 2006. Т. 73. № 7. С. 9–16.

4. Лукин А.В. Голограммные оптические элементы // Оптический журнал. 2007. Т. 74. № 1. С. 80–87.
5. Сабинин В.Е., Солк С.В. Проблемы проектирования и изготовления оптики из полимерных материалов // Оптический журнал. 2002. Т. 69. № 1. С. 61–64.
6. Солк С.В., Сабинин В.Е. Новые области применения технологии алмазного микроточения // Оптический журнал. 2005. Т. 72. № 11. С. 82–85.
7. Solk S., Shevtsov S., Iakovlev A. Designing of optical elements manufactured by diamond turning // Proc. of SPIE. 2000. V. 4231. P. 181–188.
8. Макин В.С., Макин Р.С. Взаимодействие осесимметрично поляризованного лазерного излучения с конденсированными средами // Оптика и спектроскопия. 2013. Т. 115. № 4. С. 670–675.
9. Солк С.В., Яковлев А.А. Технология изготовления малогабаритных зеркал из магниевых сплавов // Оптический журнал. 2010. Т. 77. № 3. С. 84–85.
10. Добровольский Г.Г., Саксеев П.Ю. Алмазное микроточение кремния и германия (обзор) // Сверхтвердые материалы. 2004. № 1. С. 46–51.
11. Клименко С.А., Манохин А.С. Термодинамический анализ напряженно-деформированного состояния поверхностного слоя изделия, обработанного алмазным микроточением // Процеси механічної обробки в машинобудуванні. 2010. В. 9. С. 74–81.
12. Черезова Л.А., Михайлов А.В. Применение ионной обработки в нанотехнологиях получения высокоточных поверхностей оптических деталей // Оптический журнал. 2010. Т. 77. № 5. С. 51–54.
13. Арутюнян В.В., Шевцов С.Е. Формообразование оптических поверхностей методом алмазного микроточения на материалах, применяемых в ИК диапазоне спектра // Контенант. 2013. Т. 12. № 2. С. 60–63.

- Медунецкий Виктор Михайлович* – доктор технических наук, профессор, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, vm57med@yandex.ru
- Солк Сергей Вольдемарович* – кандидат технических наук, докторант, ОАО «НИИ оптико-электронного приборостроения», Санкт-Петербург, Россия, solk@sbor.net
- Viktor Medunetskiy* – D.Sc., Professor, Saint Petersburg National University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, vm57med@yandex.ru
- Sergei Solk* – PhD, doctoral candidate, PLC «Scientific Research Institute for Optoelectronic Instrument Engineering», Saint Petersburg, Russia, solk@sbor.net

УДК 621.9

ОСОБЕННОСТИ ФОРМИРОВАНИЯ ПРЯМЫХ ЗУБЬЕВ ЦИЛИНДРИЧЕСКИХ КОЛЕС СТУПЕНЧАТЫМ ДОЛБЯКОМ

Н.М. Расулов^а, Г.Р. Гусейнов^а, У.М. Надиров^а

^аАзербайджанский технический университет, Баку, Азербайджан, n_ugurlu@mail.ru

Трудоемкость формирования зубчатых поверхностей традиционными методами достаточно высока. Разработка новых, относительно эффективных методов их обработки имеет важное научно-практическое значение.

Разработан новый способ – скоростное зубодобление прямых зубьев цилиндрических колес косозубым ступенчатым долбяком, зубья которого расположены ступенчато, вдоль собственной оси. После наладки технологической системы инструменту и заготовке сообщается непрерывное станочное зацепление. Инструмент получает одновременно и радиальное перемещение. При этом, несмотря на отсутствие осевого относительного перемещения заготовки и инструмента в целом, режущие кромки последнего совершают осевое перемещение, что обеспечивается конструкцией инструмента и кинематикой нарезания, и снимают материал. После достижения требуемой глубины нарезки производится однократное перемещение инструмента вдоль оси заготовки. Обеспечивается калибровка поверхности, и уменьшаются погрешности обработки, так как каждый зуб долбяка имеет возможность участвовать в формировании любого зуба колеса.

Спроектированы и изготовлены ступенчатые долбяки разной конструкции, было создано модернизированное оборудование на базе вертикально-фрезерного станка мод. 6Р13. Опробование и промышленные испытания разработанного метода проведены на машиностроительном заводе им. Г. Мусабекова.

Установлено, что нарезание зубьев ступенчатым долбяком обеспечивает 8–9 степень точности по [1] при значительном сокращении машинного времени по сравнению с традиционным зубодоблением. На специальном станке можно нарезать зубчатые поверхности, обладающие точностью не ниже 8 степени.

Ключевые слова: зуб, долбяк, обработка, согласованное вращение, скрещивание, угол.

STRAIGHT COGS FORMATION FEATURES FOR CYLINDRICAL SPUR GEARS BY STEPPED GEAR-SHAPED CUTTER

N. Rasulov^b, G. Guseynov^b, U. Nadirov^b

^b Azerbaijan Technical University, Baku, Azerbaijan, n_ugurlu@mail.ru

The complexity of cogged surfaces formation by traditional methods is high enough. Development of new, effective methods of their treatment is an issue of significant scientific and practical importance. A new method is developed: speed gear-cutting of straight cogs of spur gears by skew gear-shaped cutter, which cogs are stepped along its own axis. After the technological system is set up, a continuous machine tool gearing is passed to the tool and workpiece. The tool also gets a radial displacement. Despite the absence of axial displacement of the workpiece and tool in general, cutting edge of the tool commits axial displacement, provided by tool design and the cutting kinematics. After reaching the desired depth of cutting, a single-phase tool moving along the workpiece is done. The surface calibration is supplied and processing errors are reduced. Each cog of the cutter has the opportunity to take part in the formation of any wheel cog. Stepped cutters of different design were created and manufactured; updated equipment was made based on vertical milling machine of 6P13 model. Industrial testing of the developed method was performed on machine-building plant named after G. Musabekov. It was stated that cogs cutting by stepped cutter provides the 8-9th degrees of accuracy according to the state standard specification SSS 1643-81 and a significant reduction in time as compared with the traditional gear shaping. The special machine can cut gearing surfaces with the quality not lower than the 8th degree of accuracy.

Keywords: cog, gear-shaped cutter, treatment, consistent rotation, crossing, angle.

Введение

Одним из важнейших направлений научно-технического прогресса является широкое освоение прогрессивных технологий, способствующее успешному решению задач интенсификации производства. В связи с этим становится необходимым переход к принципиально новым технологиям, обеспечивающим наивысшую эффективность. Непрерывно растут требования к точности современных машин и механизмов, к их надежности и долговечности. Эти требования в большой мере определяют необходимую точность ряда важнейших звеньев машин и механизмов, в том числе зубчатых передач.

Зубчатые поверхности являются одними из наиболее распространенных сложных поверхностей вращения. Трудоемкость их формирования очень высока. Наиболее производительными и эффективными методами формирования зубчатых поверхностей являются зубофрезерование и зубодолбление. Эти методы досконально исследованы и доведены до высокой степени совершенства [2–8], что не исключает возможность разработки новых, относительно эффективных методов обработки на основе управления связями геометрической формы. Однако и они обладают недостатками. При фрезеровании зубьев червячной фрезой величина врезания достигает больших значений, особенно при нарезании зубьев на единичной заготовке: чем больше высота зубьев и диаметр инструмента, тем больше величина врезания. Это приводит к повышению машинного времени и снижению эффективности зубонарезания. Нарезание зубьев долблением осуществляется при наличии большого количества холостых ходов инструмента; отход и возврат инструмента после каждого рабочего хода приводит к повышению «вспомогательной» части машинного времени и снижению производительности зубодолбления.

Неизбежность перехода в производстве зубчатых колес на высокопроизводительную технологию их изготовления, диктуемая необходимостью интенсификации всего машиностроения, является одной из актуальных проблем, имеющей важное научно-практическое значение. Целью исследований является разработка и исследование нового скоростного метода зубодолбления, приводящего к повышению производительности, точности и качества обработанной поверхности.

Разработка метода нарезания зубьев

Новый метод зубонарезания разработан на основе морфологической комбинаторики (предложена швейцарским ученым Ф. Цвикки) существующих методов нарезания огибанием, а именно – зубофрезерование, зубодолбление, зуботочение и шевингование [2–7].

Разработанный метод зубонарезания схематически изображен на рис. 1, он основывается на следующих условиях и предположениях [8–15].

Конструктивная особенность станка. Станок обладает системой обеспечения и наладки скрещивания осей стола и штосселя (шпинделя) в широком диапазоне ($\alpha=0-40^\circ$). Согласованные вращательные движения стола и шпинделя (штосселя) обеспечиваются как в прямом (для нарезания внутренних зубьев), так и в обратном (для нарезания наружных зубьев) направлениях. Станок обладает системой наладок относительных положений инструмента и заготовки и системой подачи как в осевом направлении заготовки, так и в направлении кратчайшего межосевого расстояния инструмента–заготовки.

Конструктивная особенность инструмента. Особенность специального косозубого долбяка – в том, что его зубья расположены ступенчато вдоль собственной оси. Таким образом, плоскость, проходящая через соответствующие точки режущих лезвий зубьев инструмента, в отличие от обычной, не перпендикулярна его оси и составляет некоторый угол наклона $\gamma \neq 90^\circ$ (рис. 1, 2).

- Относительные положения инструмента и заготовки до нарезания зубьев – следующие:
- оси инструмента специального долбяка и заготовки должны скрещиваются под углом α ;
 - для обеспечения наибольшей эффективности нарезания кратчайшее расстояние между осями инструмента и заготовки при наладке должно быть обеспечено примерно в середине длины образующих нарезаемых зубьев.

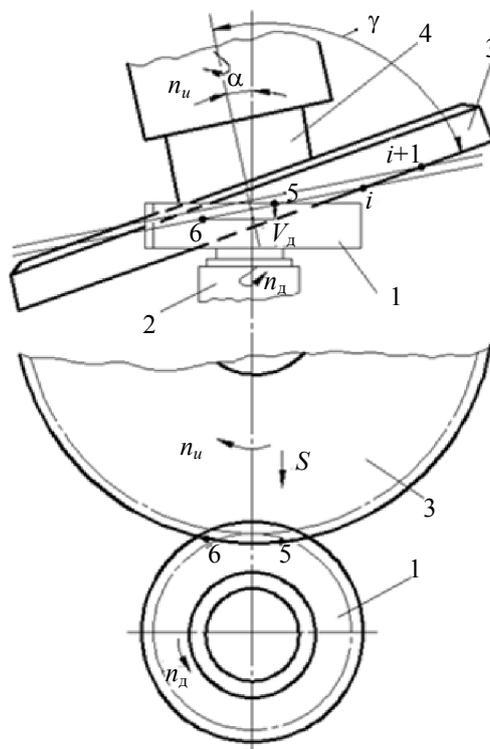


Рис. 1. Схема формирования цилиндрических наружных зубчатых поверхностей специальным долбяком: 1 – заготовка; 2 – оправка; 3 – долбяк; 4 – шпиндель; 5 – позиция i -го зуба при входе на нарезание; 6 – позиция i -го зуба при выходе с нарезания

Необходимые движения при нарезании:

- непрерывное вращательное движение заготовки вокруг своей оси с частотой вращения n_d ;
- непрерывное вращательное движение инструмента вокруг своей оси с частотой вращения n_u , согласованное с частотой вращения заготовки: $n_u = k n_d$. Здесь k – коэффициент, выражающий отношения количества зубьев инструмента и колеса;
- радиальная подача S инструмента (или заготовки) в направлении, перпендикулярном к плоскости, в которой рассматривается угол скрещивания осей инструмента и заготовки α ;
- однократное движение инструмента (или заготовки) вдоль оси заготовки после достижения необходимой высоты нарезанных зубьев для достижения высокой точности формы.

Особенности процесса нарезания. Заготовка 1 (рис. 1) крепится на оправке 2 приспособления, установленного на столе станка, а инструмент-долбяк – 3 на шпинделе 4, причем оси инструмента 3 и заготовки 1 перекрещиваются под острым углом α . Осуществляется наладка технологической системы, сообщается обратное движение инструменту и заготовке (n_u и n_d), и один из них, например, инструмент, получает еще радиальное перемещение (S). При этом, несмотря на отсутствие осевого относительного перемещения заготовки и инструмента в целом, режущие кромки последнего совершают осевое перемещение (рис. 1), например, от точки 5 до точки 6 с некоторой осевой скоростью V_d , что обеспечивается конструкцией инструмента и кинематикой нарезания. На рисунке показаны траектории режущих кромок зубьев инструмента под номером i , $(i+1)$, и т.д. Как показано на схеме, режущий зуб под номером i входит в теоретическое зацепление с заготовкой в точке 5 и выходит из зацепления в точке 6. При этом осуществляется осевое перемещение режущей кромки зуба относительно заготовки, формируется скорость резания V_d . Перемещаясь в осевом направлении со скоростью V_d , режущие кромки зубьев долбяка снимают необходимый слой материала.

Таким образом, при радиальном перемещении инструмента (или заготовки) каждый зуб долбяка, зацепляясь с теоретическим зубом заготовки, перемещается в осевом направлении со скоростью V_d и снимает материал.

При вращении косозубого долбяка обеспечивается возвращение каждого режущего зуба на исходную рабочую позицию. Так как плоскость наклона оси долбяка является разделительной, зуб инструмента, находящийся справа от этой плоскости, оказывается в рабочей позиции, совершает рабочее движение, и часть этого движения используется для нарезания зубьев колес. Когда зуб инструмента находится слева от разделительной плоскости, он оказывается в нерабочей позиции и совершает вспомогательное движение, возвращающее его в рабочую позицию. После достижения требуемой глубины нарезки прекращают радиальную подачу и осуществляют отвод инструмента.

Для повышения точности формирования профилей нарезанных зубьев, т.е. с целью их калибровки, после достижения необходимой глубины нарезки целесообразен незначительный отвод инструмента в радиальном направлении и подача его корпуса вдоль образующих нарезаемых зубьев для однократного осевого перемещения инструмента. При этом наладочный размер технологической системы принимается таким же, что и при нарезании только радиальной подачей. Однократное перемещение инструмента вдоль оси заготовки обеспечивает калибровку поверхности и уменьшение погрешностей обработки по отклонению формы от цилиндричности. Достаточным условием калибровки по всей длине обработки является выход теоретической линии, выражающей кратчайшее расстояние между осями инструмента и заготовки, за пределы торцов зубчатого венца.

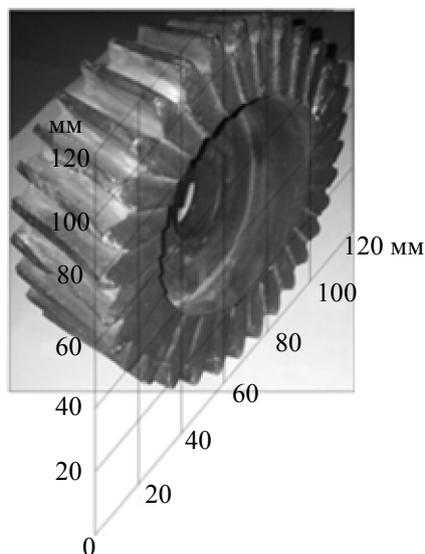


Рис. 2. Косозубый ступенчатый долбяк

Перемещение инструмента вдоль оси заготовки обеспечивает устранение погрешностей обработки в виде отклонений формы от цилиндричности. При обработке внутренних прямых зубьев зубчатых колес специальным долбяком достигается высокая производительность и точность обработки. Это объясняется тем, что каждый нарезаемый зуб колеса формируется всеми зубьями ступенчатого долбяка, что, в свою очередь, обеспечивается определенным отношением количества зубьев инструмента и нарезаемой шестерни.

Высокая производительность и точность обработки могут быть достигнуты тогда, когда каждый нарезаемый зуб обрабатывается, т.е. формируется, всеми зубьями инструмента, что зависит от отношений количества зубьев инструмента и нарезаемой шестерни (рис. 2).

Схема нарезания

Последовательность съема материала осуществляется согласно схеме, представленной на рис. 3. Каждый зуб инструмента формирует определенную часть эвольвентного профиля всех формируемых зубьев (или большинства из них) детали. При этом зуб инструмента, перемещаясь радиально, снимает материал в одном и том же секторе зубьев детали, последовательно углубляясь в радиальном направлении.

На рис. 3 представлены некоторые положения трех зубьев инструмента, обозначенные двузначными числами (11, 12, 21, 22, 23 и 31, 32, 33, 34). Десятичная часть числа обозначает номер зуба инструмента, а единичная часть – его положение. Для удобства нумерация положений зубьев произведена, начиная от последних положений. Необходимо отметить, что при хорошей организации технологической операции общая работа, необходимая для нарезания зубьев колеса, распределяется между зубьями инструмента почти равномерно, так как слабо нагруженный зуб инструмента при образовании одного зуба детали (например, зуб 1 на рис. 3), оказывается сильно нагруженным (подобно зубу 3 на рис. 3) при образовании другого зуба.

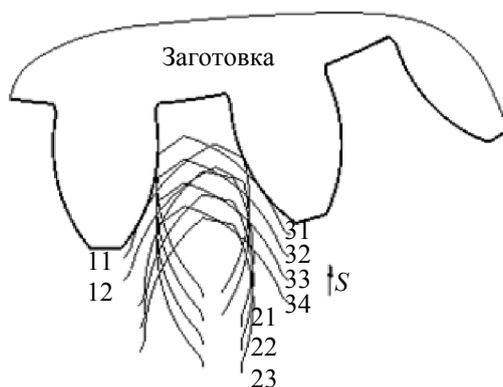


Рис. 3. Схема нарезания зубьев

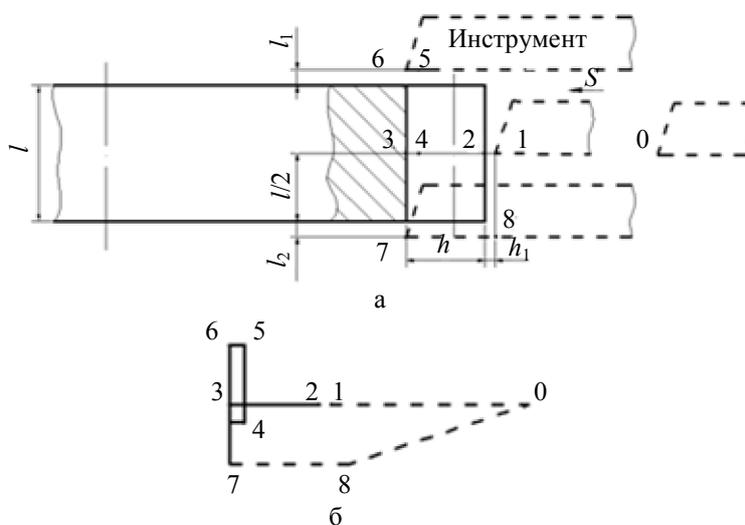


Рис. 4. Схема формирования цикла нарезания зубьев: схема основных положений инструмента (а); цикл обработки (б)

Цикл обработки включает следующие вспомогательные и технологические проходы (рис. 4, б): подача инструмента к заготовке (0–1); радиальная рабочая подача нарезания зубьев на глубину h (1–2–3); отвод инструмента (3–4); его подача в рабочее положение (4–5–6); калибровка зубьев осевой подачей (6–7); отвод инструмента (7–8–0).

Продолжительность нарезания

Машинное время нарезания зубьев складывается из двух частей – время, затрачиваемое на формирование зубьев радиальной подачей и на их калибровку осевым однократным перемещением инструмента – и определяется формулой

$$t_0 = \frac{H_0}{n_n \cdot S} + \frac{L}{V_p} + \left(2 + \frac{L}{2}\right) \cdot \frac{1}{V_x} \approx \frac{H_0}{n_n \cdot S} + \frac{1}{n_0}, \quad (1)$$

где $H_0 = h + h_1$; $L = l + l_1 + l_2$; h и l – высота и длина зубьев; h_1 и l_1 – радиальное и осевое врезания; l_2 – перебег; V_p и V_x – скорость рабочего и холостого перемещений инструментальной головки; n_0 – число двойных ходов в минуту.

В формуле (1) величина отвода инструмента (3–4, рис. 4, а) принята равной 2 мм.

Опробование процесса нарезания зубьев косозубым ступенчатым долбяком

Для опробования и промышленного испытания разработанного метода под руководством одного из авторов было создано модернизированное оборудование на базе вертикально-фрезерного станка мод. 6P13 (существующие зуборезные оборудования не обеспечивают условий, требуемых для реализации данного метода). Для передачи вращательного движения на заготовку и синхронизации частоты вращения инструмента и заготовки оборудование оснащено карданной передачей и коробкой скоростей. Кардан жестко закреплен на боковой поверхности шпиндельной бабки, получает вращение от шпинделя станка с помощью сменных зубчатых пар и передает его на ведущий вал коробки скоростей. Двухступенчатая ко-

робка скоростей закреплена на столе станка, а на ее выходном вале предусмотрена шейка для установки заготовки.

Спроектированы и изготовлены зубонарезные ступенчатые долбяки разной конструкции. Они имеют дисковые конструкции и изготовлены на машиностроительном заводе им. Г. Мусабекова и на Бакинском заводе нефтяного машиностроения (рис. 2). При этом учтены установившиеся на практике особенности проектирования и изготовления долбяков применительно к ступенчатым [16, 17]. Точность параметров рабочих частей инструментов была низкой по сравнению с требованиями [18]. Это связано с низкой культурой производства на указанных заводах по сравнению со специализированными инструментальными заводами, изготавливающими долбяки.

Были изготовлены заготовки из чугуна марки СЧ 18 и из стали марки 40, предназначенные для нарезания цилиндрических наружных прямых зубьев с модулем $m = 3$, количества зубьев $z = 24$ и 34 , длиной зубьев 15 и 24 мм. Стальные заготовки имели конструкцию детали, производимой заводом им. Г. Мусабекова. Эксперименты проведены на машиностроительном заводе им. Г. Мусабекова. Режим нарезания зубьев: частота вращения инструмента $n_n = 315$ об/мин; радиальная подача $S = 0,025$ мм/об; осевая подача калибровки – ручная; смазывающе-охлаждающая жидкость – масло индустриальное.

Измеряли и оценивали отклонение шага, накопленную погрешность шага зубчатого колеса, колебание длины общей нормали, отклонение направления зуба, отклонение толщины зуба. Были использованы средства измерения, предложенные в [2]. Установлено, что нарезание зубьев ступенчатым долбяком обеспечивает 8–9 степени точности по [1]. На специальном станке можно нарезать зубчатые поверхности, обладающие качеством не ниже 8 степени точности.

Таким образом¹, нарезание цилиндрических прямых зубьев разработанным методом оказалось производительнее в 1,5–1,6 раза и эффективнее обычного зубодолбления (с учетом стоимости изготовления инструментов и т.п.) на 17%.

Заключение

1. Разработан новый метод скоростного зубодолбления прямых зубьев цилиндрических колес косозубым ступенчатым долбяком. Метод имеет следующие отличительные особенности:
 - зубья долбяка расположены ступенчато вдоль собственной оси. В результате, плоскость, проходящая через соответствующие точки режущих лезвий зубьев инструмента, составляет некоторый угол наклона ($\gamma \neq 90^\circ$) с осью. Это определяет относительное осевое смещение – высоты ступеней зубьев;
 - нарезание зубьев осуществляется при непрерывном станочном зацеплении долбяк–заготовка, при радиальной подаче инструмента. При этом аналогичные участки профилей всех зубьев или их большинство формируются одним и тем же зубом инструмента. После достижения необходимой высоты нарезанных зубьев осуществляется калибрование – однократное перемещение инструмента вдоль оси заготовки, нарезание профилей всех зубьев реализуется со всеми зубьями инструмента, достигается высокая точность обработки;
 - зуб инструмента, находящийся с одной стороны от плоскости наклона, совершает рабочее движение, на другой стороне разделительной плоскости он оказывается в нерабочей позиции и совершает вспомогательное движение, возвращающее его в рабочую позицию.
2. Опробование и промышленное испытание процесса нарезания зубьев косозубым ступенчатым долбяком осуществлены на модернизированном оборудовании на базе вертикально-фрезерного станка мод. 6Р13. Промышленные испытания показали, что нарезание цилиндрических прямых зубьев разработанным методом оказалось производительнее в 1,5–1,6 раза по сравнению с традиционным зубодолблением, а качество обработанных поверхностей и точность не уступают сравниваемому.

Литература

1. ГОСТ 1643-81. Основные нормы взаимозаменяемости. Передачи зубчатые цилиндрические. Допуски. Введ. 01.07.1981. М.: Изд-во стандартов, 1981. 45 с.
2. Производство зубчатых колес: Справочник/ Под общ. ред. Б.А. Тайца. 3-е изд., перераб. и доп. М.: Машиностроение, 1990. 464 с.
3. Тарапанов А.С., Харламов А.С. Управление процессом зубодолбления. М.: Машиностроение, 1999. 128 с.
4. Калашников А.С. Технология изготовления зубчатых колес. М.: Машиностроение, 2004. 480 с.
5. Технология производства и методы обеспечения качества зубчатых колес и передач / Под общ. ред. В.Е. Старжинского, М.М. Кане. СПб: Профессия, 2007. 832 с.

¹ Результаты проведенных многочисленных теоретических и экспериментальных исследований будут представлены в очередных работах.

6. Артамонов В.Д., Федоров Ю.Н. Анализ эффективности процессов зубонарезания цилиндрических колес. Тула: Изд-во ТулГУ, 2008. 356 с.
7. Артамонов В.Д. Технологические основы повышения эффективности зубонарезания цилиндрических колес с продольной модификацией зубьев. Дис. ... докт. техн. наук: 05.02.07, 05.02.08. Тула, 2011. 456 с.
8. Расулов Н.М., Расулов Р.С. Способ формообразования зубьев зубчатых колес косозубым долбяком. Патент РФ №2064377. Бюл. 1996. № 21.
9. Расулов Н.М. Повышение эффективности обработки деталей со сложными поверхностями вращения. Дис. ... докт. техн. наук: 05.02.08. Баку, 2000. 295 с.
10. Расулов Н.М., Расулов Р.С., Надиров У.М. О нарезании зубьев со специальным долбяком. Проектирование и производство // Материалы 7-й Международной конференции. Анкара, Турция, 1996. С. 399–410 [на турец. яз.]
11. Расулов Н.М., Надиров У.М., Алигулиев Ф.А. Анализ некоторых параметров станочного зацепления при нарезании зубьев специальным долбяком // Материалы докладов 43-й Научно-технической и методической конференции. Баку: АЗТУ, 1995. С. 68–69.
12. Расулов Н.М. Скорость резания при нарезании внутренних зубьев со специальным долбяком // Ученые записки (Азербайджанский технический университет, Баку). 1995. № 1. С. 35–38.
13. Расулов Н.М., Надиров У.М., Амирасланов П.А. О точности по диаметру нарезанных специальным долбяком зубчатых поверхностей // Ученые записки (Азербайджанская государственная нефтяная академия, Баку). 1995. № 1. С. 60–64.
14. Расулов Н.М. Влияние кинематики нарезания зубьев со специальным долбяком на точность обработки // Ученые записки (Азербайджанский технический университет, Баку). 1994. № 3. С. 24–29.
15. Расулов Н.М., Надиров У.М., Гусейнов Г.Р. Анализ точности зубодолбления // Механика-машиностроение (Баку). 2003. № 4. С. 55–57. [на азерб. яз.]
16. Справочник инструментальщика / Под общ. ред. И.А. Ординарцева. Л.: Машиностроение, 1987. 846 с.
17. Фельдштейн Е.Э., Корниевич М.А. Металлорежущие инструменты. Справочник конструктора. Минск: Новое знание, 2009. 1039 с.
18. ГОСТ 9323-79. Долбяки зуборезные чистовые. Технические условия. Введ. 01.07.1981. М.: Изд-во стандартов, 1987. 72 с.

<i>Расулов Нариман Могбил</i>	– доктор технических наук, профессор, зав. кафедрой, Азербайджанский технический университет, Баку, Азербайджан, nariman.rasulov@yahoo.com
<i>Гусейнов Гумбат Рагим</i>	– кандидат технических наук, старший преподаватель, Азербайджанский технический университет, Баку, Азербайджан, rasulzade_office@rambler.ru
<i>Надиров Угурлу Магомед</i>	– кандидат технических наук, доцент, Азербайджанский технический университет, Баку, Азербайджан, n_ugurlu@mail.ru
<i>Nariman Rasulov</i>	D.Sc., Professor, Department head, Azerbaijan Technical University, Baku, Azerbaijan, nariman.rasulov@yahoo.com
<i>Humbat Guseynov</i>	PhD, senior lecturer, Azerbaijan Technical University, Baku, Azerbaijan, rasulzade_office@rambler.ru
<i>Ugurlu Nadirov</i>	PhD, Associate professor, Associate professor, Azerbaijan Technical University, Baku, Azerbaijan, rasulzade_office@rambler.ru

УДК 519.633.6:517.958:536.2

ПРОБЛЕМА УЧЕТА ЗАВИСИМОСТИ КОЭФФИЦИЕНТА ОБЪЕМНОЙ ТЕПЛОЕМКОСТИ ОТ ТЕМПЕРАТУРЫ ПРИ МОДЕЛИРОВАНИИ ЛАЗЕРНО-ДУГОВОЙ НАПЛАВКИ

С.А. Лукьяненко^а, В.А. Третьяк^а^а Национальный технический университет Украины «Киевский политехнический институт», Киев, Украина, valery.tretyak@gmail.com

На примере задачи моделирования динамики температурного поля при лазерно-дуговой наплавке композитных материалов исследуется проблема учета явлений плавления и кристаллизации сталей и сплавов. Математическая модель принадлежит классу задач Стефана и описывается нестационарным нелинейным трехмерным уравнением теплопроводности с граничными условиями. Задача решается конечно-разностным методом сквозного счета. Нелинейность задачи обусловлена зависимостью коэффициента объемной теплоемкости от температуры, которая имеет всплеск в интервале плавления-кристаллизации. Для учета этой зависимости предлагается метод линейно-экспоненциальной аппроксимации. За счет стремительности нагрева в некоторых узлах разностной сетки возникает проблема перехода температуры из диапазона твердой фазы в диапазон жидкой фазы за один шаг интегрирования по времени. Для решения этой проблемы предлагается алгоритмический метод учета всплеска зависимости коэффициента объемной теплоемкости от температуры в интервале плавления-кристаллизации, который основан на идее вложенных сеток. Предложенный метод увеличивает алгоритмическую сложность расчета, однако дает гарантию учета физики процесса плавления-кристаллизации. Приводится сравнение результатов расчетов, полученных при использовании кусочно-линейной и линейно-экспоненциальной аппроксимации зависимости коэффициента объемной теплоемкости от температуры. Показано, что в случае использования линейно-экспоненциальной аппроксимации процесс решения на адаптивной сетке ускоряется за счет уменьшения количества временных шагов интегрирования.

Ключевые слова: уравнение теплопроводности, коэффициент объемной теплоемкости, задача Стефана, метод сквозного счета.

TEMPERATURE DEPENDENCE CONSIDERATION ISSUE FOR COEFFICIENT OF VOLUMETRIC HEAT CAPACITY IN SIMULATION OF LASER-ARC PAD WELD PROCESS

S. Luk'yanenko^b, V. Tretyak^b^b National Technical University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine, valery.tretyak@gmail.com

The paper deals with research of steels and alloys melting and crystallization on the example of temperature field dynamics simulation for the laser-arc pad weld of composite materials. Mathematical model is one of the Stefan problems for nonlinear transient three-dimensional Fourier equation of heat conduction with boundary conditions. The problem is solved by the finite-difference method with the pass-through calculation scheme. Nonlinearity of the problem is caused by the temperature dependence of the volumetric heat capacity coefficient that has a peak in the interval of phase transition (melting – crystallization). The linear-exponential approximation method is offered to consider this dependence. At some nodes of the discrete mesh the temperature changes from solid phase range to liquid phase at one time step on account of high heat velocity. The algorithm of temperature dependence consideration for volumetric heat capacity peak in the interval of melting – crystallization is suggested based on the nested mesh idea. Though the computational complexity is increased by suggested algorithm it provides more accurate solution and ensures account of phase transition phenomenon. Calculation results for linear and linear-exponential approximation of temperature dependence for volumetric heat capacity are compared. It is shown that in case of linear-exponential approximation the solution process on dynamic adaptive mesh is speeded up due to decreasing of time steps number.

Keywords: heat capacity equation, volumetric heat capacity coefficient, Stefan problem, pass-through calculation scheme.

Введение

Использование лазерных технологий упрочнения и восстановления поверхностей позволяет увеличить время работы ключевых деталей машин и механизмов. Лазерная технология дает возможность практически полностью избежать остаточных деформаций, которые нельзя было бы устранить последующей рихтовкой. Высокая стоимость технологических процессов лазерного нагрева привела к возникновению комбинированного лазерно-дугового метода наплавки. Преимуществами этого метода являются универсальность, возможность экономии ресурсов и энергии, экологическая чистота и возможность полной автоматизации [1]. Однако широкое распространение этого метода сдерживается отсутствием эффективных методов исследования. Натурные эксперименты не дают полной информации о динамике температурного поля в процессе наплавки, в то время как она является определяющим параметром качества покрытия [2]. Компьютерное моделирование также сталкивается с рядом проблем [3]. Таким образом,

исследования, направленные на определение адекватных моделей процесса и развитие методов и средств реализации этих моделей, являются актуальными.

Технологический процесс наплавки предполагает, что порошок наносится на поверхность потоком транспортирующего газа, при этом для увеличения эффективности он может подогреваться электрической дугой с распределением плотности мощности теплового потока q_a , создавая плазму. Лазерный луч фокусируется на поверхности детали и завершает процесс плавления, пятно фокусировки луча может находиться перед областью действия дуги (на поверхности подложки) или после (на поверхности слоя наплавки), также они могут пересекаться. Более детально технологический процесс описан в [4]. Он включает множество сложных явлений [5, 6], все их невозможно учесть при моделировании.

Целью настоящей работы является усовершенствование способа учета зависимости коэффициента объемной теплоемкости от температуры в задачах Стефана за счет формирования гладкой аппроксимирующей функции и модификации численного метода сквозного счета.

Математическая постановка задачи

В работе зависимость теплофизических свойств подложки, а также коэффициента теплопроводности покрытия от температуры не учитываются. Считается, что слой напыления в виде идеально спрессованного композитного порошка моментально появляется под электрической дугой, т.е. область определения (ОО) задачи моделирования меняется во времени в соответствии с передвижением дуги (рис. 1). Положим размеры ОО L_x , L_y , L_z по соответствующим координатам, h – высота покрытия, Γ_a – область нагрева дугой, Γ_l – пятно фокусировки лазера, Π_1 – неподвижный цилиндр, который образовался под дугой в первый момент времени, Π_2 – цилиндр, соответствующий текущему положению дуги, который перемещается со скоростью v вдоль оси Oy , оставляя за собой поверхностный слой, Π – параллелепипедная часть покрытия. Таким образом, слой покрытия имеет переменную форму, образованную слиянием Π_1 , Π и Π_2 .

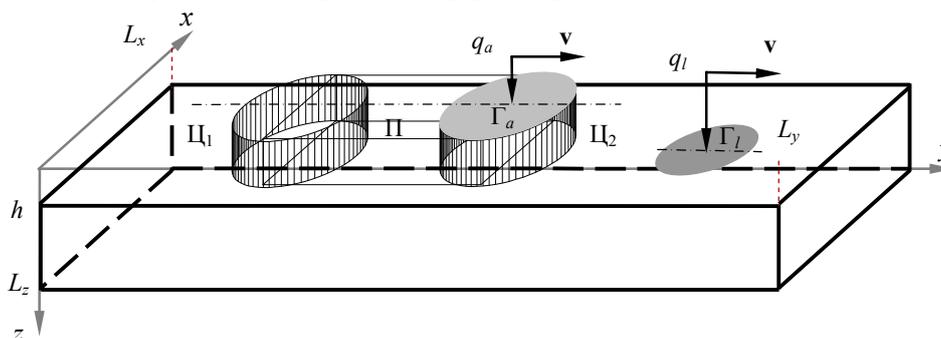


Рис. 1. Схематическое изображение переменной ОО

Распределение плотности мощности (РПМ) лазера по пятну фокусировки может быть различной конфигурации [7], чаще всего для моделирования РПМ лазера применяют функцию Гаусса [8]. В настоящей работе используются РПМ луча, описанные в [9]: гауссово, постоянное, кусочно-линейное и кусочно-постоянное. Хотя регулировать РПМ дуги так же, как и лазера, не представляется возможным, в предложенной модели считается, что конфигурация РПМ дуги может иметь такие же виды, как и луч. В местах пересечения луча и дуги РПМ комбинированного нагрева описывается предложенными в [10] формулами.

Основной математических моделей процесса лазерной наплавки является уравнение теплопроводности, которое, ввиду перечисленных выше допущений, вместе с начальными и граничными условиями формирует граничную задачу:

$$\begin{aligned} \frac{\partial}{\partial t}(U \cdot C(U)) &= \lambda \left(\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} \right), \\ U|_{t=0} &= U_c, \\ \lambda \frac{\partial U(x, y, z_0, t)}{\partial z} + q(x, y, z, t) &= 0, \quad \forall (x, y) \in \Gamma = \Gamma_a \cup \Gamma_l, \\ \lambda \frac{\partial U}{\partial \mathbf{n}} \Big|_{\Omega} + \alpha [U|_{\Omega} - U_c] &= 0, \quad \forall (x, y) \notin \Gamma = \Gamma_a \cup \Gamma_l, \end{aligned} \quad (1)$$

где U – температура; $C(U)$ – коэффициент объемной теплоемкости (КОТ); λ – коэффициент теплопроводности; U_c – температура окружающей среды; z_0 – уровень верхней грани детали, соответствующий либо подложке ($z_0=h$), либо покрытию ($z_0=0$); $q(x, y, z, t)$ – функция РПМ нагревающих элементов; \mathbf{n} – нормаль к поверхности детали; α – коэффициент теплоотдачи; Ω – граница области определения.

При этом λ и $C(U)$ зависят от пространственных координат. Если координаты соответствуют подложке, берутся значения коэффициентов материала подложки, которые считаются константами. Для покрытия берутся эффективные величины λ_{eff} и $C_{eff}(U)$, соответствующие композитному материалу. Эффективный коэффициент теплопроводности, как и ранее [11], рассчитывается по формуле Максвелла:

$$\lambda_{eff} = \lambda_m \frac{1 + 2V_f \left(1 - \frac{\lambda_m}{\lambda_f}\right) / \left(1 + 2\frac{\lambda_m}{\lambda_f}\right)}{1 - V_f \left(1 - \frac{\lambda_m}{\lambda_f}\right) / \left(1 + 2\frac{\lambda_m}{\lambda_f}\right)} = \text{const},$$

где λ_m, λ_f – коэффициенты теплопроводности матрицы и включений композита; V_f – объемная концентрация включений, не превышает 0,3. Определение значений КОТ рассмотрим детальнее.

Зависимость коэффициента объемной теплоемкости от температуры

Задачи, учитывающие явления плавления и кристаллизации, называют задачами Стефана. Для трехмерных задач Стефана с наличием нескольких фронтов расплавленного материала наилучшим образом зарекомендовал себя метод сквозного счета. Согласно этому методу, плавление должно учитываться зависимостью теплофизических параметров материала от температуры, а вид уравнений для всех фаз одинаков. В данной работе фазовый переход учитывается функцией КОТ $C(U)$. Так как материал покрытия – композитный, для него используются эффективные значения КОТ:

$$C_{eff} = (c\rho)_{eff} = \dot{C}_m(1 - V_f) + \dot{C}_f V_f,$$

где \dot{C}_m и \dot{C}_f – эффективные значения КОТ матрицы и включений, которые рассчитываются согласно скрытой теплоте плавления:

$$\dot{C}_m = C_m(U) - L_m \frac{\partial \varphi_m}{\partial U} \quad \text{и} \quad \dot{C}_f = C_f(U) - L_f \frac{\partial \varphi_f}{\partial U},$$

где L_m и L_f – скрытая теплота плавления; φ_m и φ_f – объемная концентрация жидкой фазы в материале матрицы покрытия и включений соответственно. Задача усложняется тем, что достоверной информации о значениях φ_m и φ_f нет [3], φ_m и φ_f зависят от состава материала, температуры и скорости нагрева или охлаждения. Например, в [12] предложена формула для мартенситных превращений. Однако удобнее выполнять аппроксимацию непосредственно зависимости КОТ от температуры. Ранее [11] зависимость КОТ от температуры определялась с помощью кусочно-линейной аппроксимации (КЛА). Очевидно [13], изменение КОТ во время плавления и кристаллизации имеет более плавный характер. В [10] предложено выполнять аппроксимацию квадратичным полиномом. В данной работе предлагается еще один способ аппроксимации КОТ – линейно-экспоненциальный (ЛЭА) (рис. 2), согласно которому КОТ вычисляется по формуле

$$C(U) = L e^{-\beta(U - U_{пл})^2} + kU + Z, \quad (2)$$

где L соответствует скрытой теплоте плавления материала; β определяет ширину интервала плавления-кристаллизации; $U_{пл}$ – температура плавления, для сталей и сплавов считается медианой $[U_s, U_L]$; k – угловой коэффициент линейной составляющей функции; Z – значение аппроксимирующей функции КОТ при температуре 0°C .

Стремительность возрастания $C(U)$ при достижении U_s регулируется параметром γ , который равен значению первого слагаемого формулы (2) при $U = U_s$:

$$L e^{-\beta(U_s - U_{пл})^2} = \gamma.$$

Получаем формулу для расчета β :

$$\beta = \frac{\ln\left(\frac{L}{\gamma}\right)}{(U_s - U_{пл})^2}.$$

Проблема учета фазового перехода на уровне алгоритмического метода

Задача (1) решается методом конечных разностей, при этом используется двуциклическая схема расщепления по координатам, которая имеет второй порядок точности по пространственным и временному шагам. Для экономии машинного времени, а также для получения более точных результатов применяются адаптивные разностные сетки [4].

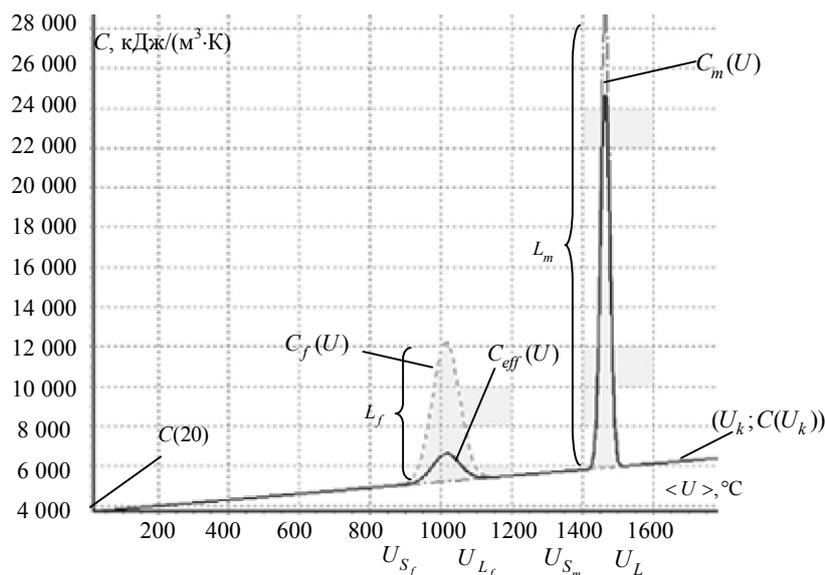


Рис. 2. Линейно-экспоненциальная аппроксимация КОТ: штрихпунктир – материал матрицы покрытия (сталь 9X18), штриховой пунктир – материал включений (бронза ОЦС 3-8-6), сплошная – график эффективного КОТ

При исследовании результатов моделирования было обнаружено, что в некоторых точках за один шаг интегрирования температура возрастает так быстро, что проскакивает всплеск функции КОТ, призванный учитывать фазовый переход. Вследствие этого фазовый переход, хотя и учитывается математической моделью, по факту опускается алгоритмическим методом реализации модели. В литературе этой проблеме уделено немного внимания. В [6] сказано только, что значение шага по времени подбирается, исходя из условия недопустимости изменения фазового состояния вещества за один шаг, однако не описано, каким образом это осуществляется. Если отказаться от результатов расчетов всего временного слоя, и, уменьшив шаг (например, в 2 раза), заново произвести расчет, это приведет к излишнему числу операций.

В работе предлагается по окончании каждого шага интегрирования (перед формированием новой сетки) совершать поиск узлов, где имело место изменение фазы. Затем такие узлы объединяются в подобласти, для которых производится перерасчет с шагом в 2 раза меньше. На границах подобластей необходимо определить граничные условия. Поскольку значения в граничных узлах подобластей уже найдены на k -м и $(k + 1)$ -м шагах, можно использовать граничные условия первого рода. На дробном $(k + 1/2)$ -м шаге значения в граничных узлах неизвестны, поэтому необходимо найти их при помощи интерполяции (используется метод Лагранжа 2-го порядка). Перерасчет предлагается выполнять рекурсивно, пока выделенные подобласти будут содержать узлы. Подобласти выделяются таким образом, что отстоящие друг от друга на 3 и менее пространственных шага узлы с фазовым переходом помещаются в одну подобласть. Более того, размерность подобласти обязательно должна быть не менее трех, что достигается путем добавления узлов по краю подобласти.

Исследование эффективности предложенных методов

Для исследования метода учета фазового перехода в алгоритмическом методе было произведено сравнение (табл. 1) полученных результатов (как максимального значения температуры на поверхности покрытия в каждый момент модельного времени).

Время, с	Результат без пересчета $U_1, ^\circ\text{C}$	Результат с пересчетом $U_2, ^\circ\text{C}$	$ U_1 - U_2 , ^\circ\text{C}$	$\frac{ U_1 - U_2 }{U_1}, \%$
0,300	2300	2027	273	11,870
0,400	2410	2208	202	8,382
0,454	2330	2181	149	6,395
0,500	2311	2119	192	8,308
0,600	2407	2240	167	6,938
0,700	2325	2233	92	3,957
0,810	2430	2359	71	2,922
Средние значения			163,714	6,967

Таблица 1. Анализ метода учета фазового перехода путем пересчета по подобластям

Критерий сравнения	КЛА ($F1$)	ЛЭА ($F2$)	$F1 - F2$	$100 \cdot (F1 - F2) / F1, \%$
Время расчета на равномерной сетке, с	153	187	-34	-22,2
Время расчета адаптивным методом, с	121	98	23	19
Количество временных шагов, шт.	80	62	18	22,5
Количество отказов, шт.	45	17	28	62

Таблица 2. Анализ влияния метода аппроксимации КОТ на характеристики расчета

При использовании перерасчета по подобластям получены меньшие значения температуры, что соответствует теоретическим рассуждениям, поскольку метод позволяет учесть поглощение дополнительной энергии во время плавления.

Также был проведен сравнительный анализ результатов моделирования при условии использования КЛА ($F1$) и ЛЭА ($F2$) аппроксимации КОТ (табл. 2).

При использовании фиксированных сеток КЛА дает меньшее время расчета, поскольку требует меньшего количества операций. Однако на адаптивных сетках расчет производится быстрее для ЛЭА, возможно, это связано с тем, что ЛЭА непрерывно дифференцируема. При этом также уменьшается общее количество временных шагов. В ходе анализа данных численных расчетов выявлено, что на численные параметры процесса решения задачи (1) сильно влияют теплофизические свойства материалов. При уменьшении интервала плавления-кристаллизации наблюдается значительное замедление процесса реализации модели.

Заключение

В работе предложен метод аппроксимации коэффициента объемной теплоемкости линейно-экспоненциальной функцией. Преимуществом этого метода является то, что аппроксимирующая функция непрерывно дифференцируема. Также раскрыта проблема учета фазового перехода в ходе алгоритмического метода решения граничной задачи. Предложен экономичный с точки зрения вычислительной эффективности метод решения этой проблемы, согласно которому на каждом шаге интегрирования необходимо рекурсивно производить перерасчет с меньшим временным шагом не по всей области, а по сформированным подобластям. Хотя этот метод требует много вычислительных ресурсов, он позволяет получить более адекватную реализацию модели распределения тепла в процессе лазерно-дуговой наплавки композитных материалов.

Литература

1. Лазерні технології та комп'ютерне моделювання / Під ред. Л.Ф. Головка, С.О. Лук'яненка. Київ: Вістка, 2009. 296 с.
2. LIA Handbook of Laser Materials Processing / Eds J. F. Ready, D. F. Farson. Orlando, USA: Laser Institute of America, Magnolia publishing, Inc., 2001. 715 p.
3. Князева А.Г., Крюкова О.Н., Букина Н.В., Сорокова С.Н. Проблемы моделирования технологических процессов поверхностной обработки материалов и нанесения покрытий с использованием высокоэнергетических источников // Известия Томского политехнического университета. 2010. Т. 316. № 2. С. 93–101.
4. Третьяк В.А. Усовершенствование адаптивного метода построения сеток для задач теплопроводности с нестационарным источником энергии // Математичне та комп'ютерне моделювання. Сер. Технічні науки. 2012. В. 7. С. 197–206.
5. Picasso M., Marsden C.F., Wagniere J.D., Frenk A., Rappaz M. A simple but realistic model for laser cladding // Metallurgical and Materials Transaction B. 1994. V. 25. N 2. P. 281–291.
6. Масленников А.В., Левин Ю.Ю., Ерофеев В.А. Моделирование нестационарного формирования шва при импульсной лазерной ND:YAG-сварке // Известия Тульского государственного университета. Технические науки. 2010. В. 4. Ч. 1. С. 255–265.
7. Гончарский А.А., Туницкий В.В. Об обратной задаче синтеза оптических элементов для лазерного излучения // Вычислительные методы и программирование. М.: НИВЦ МГУ, 2006. Т. 7. С. 138–162.
8. Qi H., Mazumder J., Ki H. Numerical simulation of heat transfer and fluid flow in coaxial laser cladding process for direct metal deposition // J. Appl. Phys. 2006. V. 100. N 2. P. 024903-1–024903-11 [Электронный ресурс]. Режим доступа: <http://dx.doi.org/10.1063/1.2209807>, свободный. Яз. англ. (дата обращения 05.11.2013).
9. Головка Л.Ф., Лук'яненко С.О., Смаковський Д.С., Михайлова І.Ю., Агеєнко В.А. Моделювання температурного поля при зміцненні матеріалів лазерним випромінюванням // Моделювання та інформаційні технології. 2008. № 45. С. 28–35.

10. Amara E.H., Hamadi F., Achab L., Boumia O. Numerical modelling of the laser cladding process using a dynamic mesh approach // Journal of Achievements in Materials and Manufacturing Engineering. 2006. V. 15. N 1–2. P. 100–106.
11. Головкин Л.Ф., Лукьяненко С.А., Смаковский Д.С., Агеенко В.А., Михайлова И.Ю. Моделирование адаптивным сеточным методом температурного поля при лазерной наплавке порошковых материалов // Электронное моделирование. 2009. Т. 31. № 1. С. 21–32.
12. Mamat M., Tofany N., Kartono A. Numerical analysis of heat conduction and phase transformation in laser transformation hardening: influences of heating duration and laser beam intensity // Applied Mathematical Sciences. 2010. V. 4. N 61–64. P. 3019–3033.
13. Мышковец В.Н., Максименко А.В., Баевич Г.А., Грищенко В.В. Термические циклы в зоне рекристаллизации при импульсной лазерной наплавке среднелегированных высокопрочных сталей // Известия Гомельского государственного университета им. Ф. Скорины. 2011. № 6 (69). С. 105–109.

Лукьяненко Святослав Алексеевич – доктор технических наук, профессор, зав. кафедрой, Национальный технический университет Украины «Киевский политехнический институт», Киев, Украина, lukian@aprodos.kpi.ua

Третьяк Валерия Анатольевна – аспирант, Национальный технический университет Украины «Киевский политехнический институт», Киев, Украина, valery.tretyak@gmail.com

Svyatoslav Luk'yanenko – D.Sc., Professor, Head of Department, National Technical University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine, lukian@aprodos.kpi.ua

Valery Tretyak – postgraduate, National Technical University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine, valery.tretyak@gmail.com

УДК 519.63

МЕТОДЫ МОДЕЛИРОВАНИЯ ТЕМПЕРАТУРНОГО ПОЛЯ ПРИ БЕСКОНТАКТНОЙ ЛАЗЕРНОЙ ДЕФОРМАЦИИ ПЛАСТИНЫ

С.А. Лукьяненко^а, И.Ю. Михайлова^а

^а Национальный технический университет Украины «Киевский политехнический институт», Киев, Украина, lukian@aprodos.kpi.ua

Сущность бесконтактной лазерной деформации состоит в изменении геометрической формы объекта в процессе его нагрева и охлаждения без применения механического воздействия. Среди факторов, влияющих на деформацию, присутствуют нагрев, создающий температурное поле, и скорость изменения температурного поля.

В работе рассмотрен результат компьютерного моделирования температурного поля пластины, возникающего под воздействием перемещающегося лазерного луча. Проведено сравнение расчетов по двум математическим моделям с результатами эксперимента. В первой модели такие параметры, как плотность, удельная теплоемкость и теплопроводность, приняты константами, во второй – зависящими от температуры по закону, полученному путем линейной аппроксимации табличных данных методом наименьших квадратов. В обеих моделях температурное поле определяется из решения трехмерного нестационарного уравнения теплопроводности: линейного – в первом случае, квазилинейного – во втором. Для решения использована шестиступенчатая неявная разностная схема расщепления по координатам, имеющая второй порядок точности по всем координатам. Системы линейных алгебраических уравнений, возникающие в этой разностной схеме, решаются модифицированным методом Гаусса. Для автоматического построения переменной разностной сетки применяется адаптивный метод, который «сгущает» узлы в зонах с большим градиентом температур и располагает их более редко в областях, где температура изменяется плавно. Это позволяет сократить время расчета и получить результат с заранее заданной точностью.

Компьютерное моделирование показало, что учет зависимости параметров материала от температуры дает более точный результат. Однако такой метод связан с большим числом временных шагов и соответственно более длителен.

Ключевые слова: трехмерное нестационарное квазилинейное уравнение теплопроводности, метод разностной аппроксимации.

METHODS OF TEMPERATURE FIELD MODELING FOR CONTACTLESS LASER DEFORMATION OF A PLATE

S. Luk'yanenko^b, I. Mykhailova^b

^b National Technical University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine, lukian@aprodos.kpi.ua

Contactless laser deformation is a process of changing a geometrical form of an object by its heating and cooling without using mechanical forces. One of the factors, influencing deformation, is heating, which creates temperature field and overpatching speed of temperature field.

The article deals with a computer modeling result for temperature field of a plate irradiated by a moving laser beam. Comparison of results, obtained by using two mathematical models and experiments, is conducted. Such parameters as density, thermal capacity and thermal conductivity are constant values in the first model, and they are linearly dependent on temperature in the second one. Linear approximation of table values by least-squares method is used to define this dependence. Both models have the temperature field computed by 3D non-stationary heat equation: it is linear function in first model and quasi-linear function in the second. To solve the equation a six-step implicit finite difference coordinate-wise splitting scheme is

applied, which has the second order of accuracy for all coordinates. Systems of linear algebraic equations, created by this finite difference scheme, are solved by modified Gauss method. Adaptive method is used for automatic design of variable difference mesh. It condenses nodes in high gradient zones and dilutes them in regions where temperature changes gradually. This gives the possibility for computation time shortening and receiving a result with predefined accuracy. Results of computer modeling show that taking into account dependence of material parameters on temperature produces more accurate results. However, it requires more steps and therefore it is more time-consuming.

Keywords: 3D non-stationary quasi-linear heat equation, difference approximation method.

Введение

Одним из направлений лазерной технологии является процесс управляемого деформирования деталей в результате локального нагрева их поверхности сфокусированным лазерным лучом. Неравномерный нагрев детали приводит к бесконтактной деформации в результате возникающего градиента температур. В этой связи становятся актуальными вопросы компьютерного моделирования возникающих температурных полей, дающего наиболее приближенные к эксперименту результаты. В свою очередь, это возможно при условии получения адекватной математической модели и использования корректных методов решения. Особенность поставленной задачи бесконтактной деформации деталей в результате лазерного облучения состоит в том, что получить деформацию можно только при достаточно больших локальных температурах. При этом большинство известных нам моделей не учитывают зависимость физических параметров материала (плотность, теплоемкость, теплопроводность) от температуры [1, 2]. Исходя из этого, актуальной является разработка модели, учитывающей эту особенность процесса бесконтактной деформации. Нами поставлена задача – сопоставить предлагаемую модель с традиционными расчетами, а также с результатами проведенного эксперимента.

В работе предлагается рассмотреть модели распространения тепла в результате локального нагрева металлической пластины под воздействием лазерного луча, перемещающегося с определенной скоростью по ее поверхности. Спроектирована и изготовлена экспериментальная установка, позволяющая оценить изменение температуры в реальном времени. Полученные экспериментальные результаты сопоставлены с расчетами по предложенным моделям.

Постановка задачи

Исследуемый объект – металлическая пластина с геометрическими размерами L_x , L_y , L_z (рис. 1).

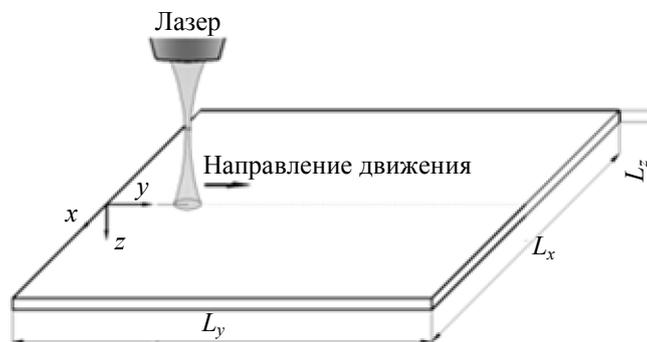


Рис. 1. Схематическое изображение расположения луча лазера относительно пластины

Будем считать, что пластина неподвижна. Все ее поверхности находятся в процессе теплообмена с окружающей средой, температура которой U_c . Для деформации пластины на ее верхнюю грань воздействует луч лазера с плотностью мощности излучения $q(x, y, t)$, который движется со скоростью $V(t)$ параллельно оси ординат в течение времени T_k . Необходимо найти распределение температур в металлической пластине, на которую воздействует луч лазера.

Математические модели

Рассмотрим две математические модели данного процесса.

В первой модели не будем учитывать зависимость физических параметров металла от температуры. Процесс распределения температуры в пластине описывается нестационарным трехмерным уравнением теплопроводности, в котором теплоемкость, плотность и теплопроводность – константы [3]:

$$c\rho \frac{\partial U(x, y, z, t)}{\partial t} = \lambda \left(\frac{\partial^2 U(x, y, z, t)}{\partial x^2} + \frac{\partial^2 U(x, y, z, t)}{\partial y^2} + \frac{\partial^2 U(x, y, z, t)}{\partial z^2} \right), \quad (1)$$

где c – теплоемкость; ρ – плотность; λ – коэффициент теплопроводности материала, $x \in [0; L_x]$, $y \in [0; L_y]$, $z \in [0; L_z]$, $t \in [0; T_k]$.

Краевые условия вне зоны действия лазера моделируют теплообмен с окружающей средой по закону Ньютона (с учетом закона Фурье) [4]:

$$\lambda \frac{\partial U(x, y, z, t)}{\partial \bar{n}} + \alpha[U_c - U(x, y, z, t)] = 0, \quad (2)$$

где \bar{n} – нормаль к поверхности; α – коэффициент теплоотдачи.

Краевое условие в зоне воздействия лазерного луча:

$$\lambda \frac{\partial U(x, y, 0, t)}{\partial z} + q(x, y, t) = 0. \quad (3)$$

Начальная температура пластины равняется температуре окружающей среды:

$$U(x, y, z, 0) = U_c.$$

Во второй модели будем учитывать зависимость физических параметров металла от температуры. Эти зависимости примем линейными, полученными в результате среднеквадратичной аппроксимации экспериментальных табличных данных. Процесс распределения температуры будем описывать с помощью нестационарного квазилинейного трехмерного уравнения теплопроводности [5]:

$$c(U)\rho(U) \frac{\partial U(x, y, z, t)}{\partial t} = \frac{\partial}{\partial x} \left[\lambda(U) \frac{\partial U(x, y, z, t)}{\partial x} \right] + \frac{\partial}{\partial y} \left[\lambda(U) \frac{\partial U(x, y, z, t)}{\partial y} \right] + \frac{\partial}{\partial z} \left[\lambda(U) \frac{\partial U(x, y, z, t)}{\partial z} \right]. \quad (4)$$

Аналогично краевые условия принимают вид

$$\lambda(U) \frac{\partial U(x, y, z, t)}{\partial \bar{n}} + \alpha[U_c - U(x, y, z, t)] = 0, \quad (5)$$

$$\lambda(U) \frac{\partial U(x, y, 0, t)}{\partial z} + q(x, y, t) = 0. \quad (6)$$

Начальное условие остается без изменений.

Методы решения

Для решения задачи будем использовать метод конечных разностей, состоящий из трех этапов:

1. дискретизация расчетной области;
2. замена дифференциального уравнения в частных производных системами алгебраических уравнений;
3. решение этих систем.

Для первого этапа будем использовать метод построения переменной неравномерной разностной сетки [6], в которой узлы сгущаются в зоне влияния лазера и разрежаются вне этой зоны. На втором этапе будем использовать метод покоординатного расщепления [7], который реализует на неравномерной сетке переход с $(k-1)$ -го временного слоя на $(k+1)$ -й и заключается в выполнении шести этапов, на каждом из которых решается одномерная задача, т.е. необходимо решить систему алгебраических уравнений с трехдиагональной матрицей. Для формирования систем уравнений на каждом шаге метода покоординатного расщепления будем использовать метод разностной аппроксимации. Для решения полученных систем будем использовать модифицированный метод Гаусса.

Пусть приближенное решение найдено на $(k-1)$ -м слое на неравномерной сетке в узлах (x_i, y_j, z_m) , где $i = 0, \dots, n_1$, $j = 0, \dots, n_2$, $m = 0, \dots, n_3$, n_1 – количество узлов по оси абсцисс, n_2 – количество узлов по оси ординат, n_3 – количество узлов по оси аппликата. Обозначим величину шагов сетки в направлении оси абсцисс через $h_{1,i} = x_i - x_{i-1}$, в направлении оси ординат – через $h_{2,j} = y_j - y_{j-1}$, в направлении оси аппликата – через $h_{3,m} = z_m - z_{m-1}$. Обозначим также среднее арифметическое двух соседних шагов для узла (x_i, y_j, z_m) в каждом из координатных направлений через $h_{1,c}$, $h_{2,c}$, $h_{3,c}$; шаг по времени через $\tau = \tau_k = t_k - t_{k-1} = t_{k+1} - t_k$; значение приближенного решения в точке (x_i, y_j, z_m, t_k) – через u_{ijm}^k .

Рассмотрим метод разностной аппроксимации [8]. Заменяя производные в уравнении (1) соответствующими разностными схемами, получим операторную форму:

$$\frac{u^{k-\frac{2}{3}} - u^{k-1}}{\tau} = \Lambda_3 \frac{u^{k-\frac{2}{3}} + u^{k-1}}{2}, \quad \frac{u^{k-\frac{1}{3}} - u^{k-\frac{2}{3}}}{\tau} = \Lambda_2 \frac{u^{k-\frac{1}{3}} + u^{k-\frac{2}{3}}}{2}, \quad \frac{u^k - u^{k-\frac{1}{3}}}{\tau} = \Lambda_1 \frac{u^k + u^{k-\frac{1}{3}}}{2},$$

$$\frac{u^{k+\frac{1}{3}} - u^k}{\tau} = \Lambda_1 \frac{u^{k+\frac{1}{3}} + u^k}{2}, \quad \frac{u^{k+\frac{2}{3}} - u^{k+\frac{1}{3}}}{\tau} = \Lambda_2 \frac{u^{k+\frac{2}{3}} + u^{k+\frac{1}{3}}}{2}, \quad \frac{u^{k+1} - u^{k+\frac{2}{3}}}{\tau} = \Lambda_3 \frac{u^{k+1} + u^{k+\frac{2}{3}}}{2},$$

где $\Lambda_1, \Lambda_2, \Lambda_3$ – разностные операторы аппроксимации вторых производных на неравномерной сетке:

$$\Lambda_1 u_{ijm}^k = a \frac{h_{1,i+1} u_{i-1,j,m}^k - 2h_{1,c} u_{ijm}^k + h_{1,i} u_{i+1,j,m}^k}{h_{1,c} h_{1,i} h_{1,i+1}}, \quad \Lambda_2 u_{ijm}^k = a \frac{h_{2,j+1} u_{i,j-1,m}^k - 2h_{2,c} u_{ijm}^k + h_{2,j} u_{i,j+1,m}^k}{h_{2,c} h_{2,j} h_{2,j+1}},$$

$$\Lambda_3 u_{ijm}^k = a \frac{h_{3,m+1} u_{i,j,m-1}^k - 2h_{3,c} u_{ijm}^k + h_{3,m} u_{i,j,m+1}^k}{h_{3,c} h_{3,m} h_{3,m+1}}, \quad (7)$$

где $a = \frac{\lambda}{c\rho}$ – коэффициент температуропроводности.

Запись алгебраических систем на каждом этапе схемы покоординатного расщепления для уравнения с переменными коэффициентами (4) будет отличаться от предыдущей только тем, что физические параметры металла будут линейно зависеть от температуры. Таким образом, для разностных операторов (7) коэффициент температуропроводности примет вид

$$a(u_{ijm}^k) = \frac{\lambda(u_{ijm}^k)}{c(u_{ijm}^k)\rho(u_{ijm}^k)}.$$

К полученным системам алгебраических уравнений добавляются уравнения, соответствующие крайевым условиям (2), (3) в первом случае и (5), (6) во втором.

Результаты моделирования

Моделирование проводилось для тонкой пластины из стали углеродной 65Г [9] толщиной 0,5 мм, шириной 30 мм и длиной 50 мм. Мощность лазера – 0,2 кВт, скорость перемещения – 0,0167 м/сек, диаметр пятна – 1 мм. Температура окружающей среды – 27°C. Значения теплофизических параметров, используемых при моделировании, без учета их зависимости от температуры: $\alpha = 50$ Вт/(м²·К), $c = 705$ Дж/(кг·К), $\rho = 7730$ кг/м³, $\lambda = 28$ Вт/(м·К), что соответствует температуре 800°C. Расчет проводился на компьютере с процессором Intel Core i7-3770 и тактовой частотой 3,4 ГГц под управлением ОС Windows Vista x64.

Поскольку экспериментальное определение температуры непосредственно в зоне обработки крайне затруднительно, измерение проводилось на некотором расстоянии. Для этого к образцу припаивались две термопары (рис. 2): одна на лицевой стороне образца на расстоянии 7 мм от края обрабатываемой зоны, вторая – на обратной стороне, непосредственно на центральной оси прохода. Сигнал с термопар регистрировался осциллографом, скомутированным с самописцем.

Обозначим температуру, измеренную верхней термопарой, через t_1 , а нижней – через t_2 . В результате проведения эксперимента были получены значения $t_1^e = 44^\circ\text{C}$ и $t_2^e = 80^\circ\text{C}$.

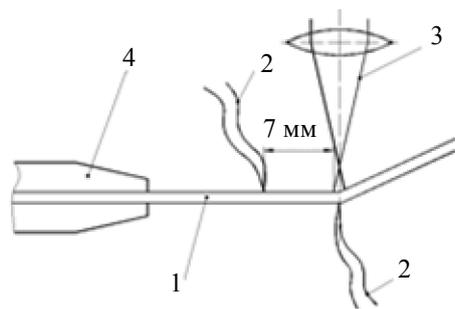


Рис. 2. Схема измерения температуры образца во время обработки: 1 – пластина; 2 – термопары; 3 – лазерный луч; 4 – фиксатор

Результаты для контрольной точки на верхней стороне пластины, полученные при компьютерном моделировании с учетом и без учета зависимости плотности, теплоемкости и коэффициента теплопроводности от температуры, отличаются примерно на 7% (таблица). Модель с переменными параметрами дает результаты на 5% ближе к результатам эксперимента по сравнению с моделью, в которой коэффициенты постоянны. Из таблицы видно, что температура в верхней контрольной точке для второго метода выше по сравнению с первым. Это объясняется прямо пропорциональной зависимостью теплоемкости от температуры и константным значением, соответствующим температуре 800°C.

Полученные значения температуры на нижней грани для обоих методов (рис. 3) расположены в большую и меньшую стороны от экспериментальных данных примерно на одинаковую величину перегрева (таблица), что также объясняется зависимостью теплоемкости от температуры и выбором константного значения теплоемкости. Кроме того, из таблицы видно, что расхождение между t_1^e и t_1 для обоих методов больше, чем между t_2^e и t_2 .

В связи с отсутствием экспериментальных данных на поверхности в зоне действия лазерного луча, сложно говорить о том, какой метод дает более точный результат. Однако если сравнивать результаты, полученные при моделировании процесса разными методами, то лучшим методом следует признать метод разностной аппроксимации с зависимостью физических параметров от температуры, поскольку суммарное отклонение полученных в результате расчета значений от экспериментальных в этом случае меньше. Объяснить расхождение между экспериментальными и расчетными данными можно тем, что подход, использованный для измерения температуры во время эксперимента, не обеспечивает достаточно высокой точности результата.

Использованный метод	Количество шагов	$t_1, ^\circ\text{C}$	$t_2, ^\circ\text{C}$	$ t_1 - t_1^e $	$\frac{ t_1 - t_1^e }{t_1^e} \cdot 100\%$	$ t_2 - t_2^e $	$\frac{ t_2 - t_2^e }{t_2^e} \cdot 100\%$
Метод разностной аппроксимации с постоянными C, ρ, λ	227	29,5	70	14,5	33%	10	12,5%
Метод разностной аппроксимации с переменными C, ρ, λ	388	31,6	90,7	12,4	28%	10,7	13%

Таблица. Сравнение эмпирических и расчетных температур

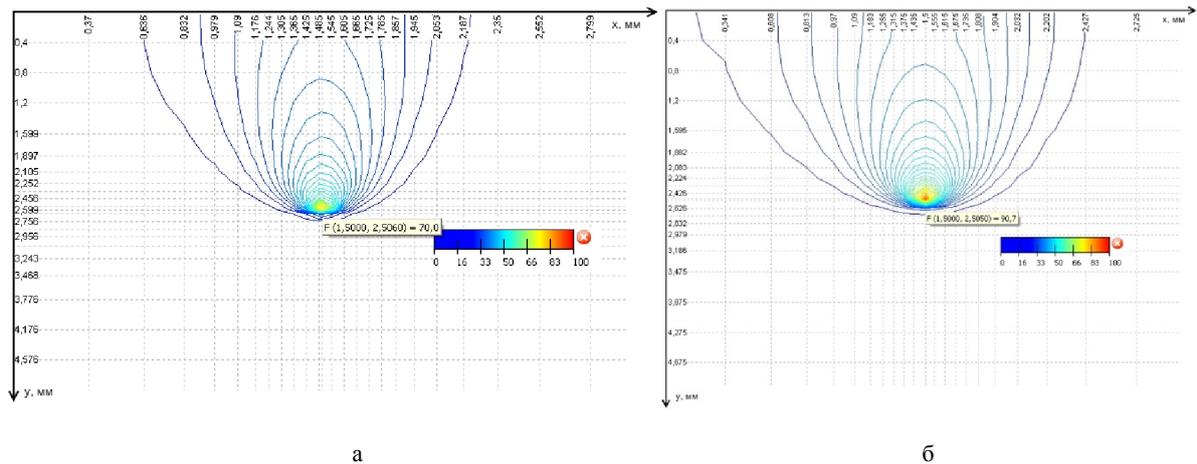


Рис. 3. Изотермы на нижней грани в разрезе xOy : с постоянными C, ρ, λ (а); с переменными C, ρ, λ (б)

Заключение

В работе рассмотрены методы решения линейного и квазилинейного уравнений теплопроводности, используемых для моделирования распределения температуры при лазерном нагревании пластины для последующей бесконтактной деформации.

Результаты компьютерного моделирования показали, что метод разностной аппроксимации с учетом зависимости параметров материала от температуры дает более точный результат. Отметим, что метод более трудоемок, так как требует выполнения большего числа шагов на адаптивной сетке, что увеличивает время расчета примерно на 40%.

Литература

1. Price S. Laser forming // Journal of Manufacturing Science and Engineering. 2007. V. 129. P. 117–124.
2. Shi Y., Shen H., Yao Z., Hu J. Temperature gradient mechanism in laser forming of thin plates // Optics & Laser Technology. 2007. V. 39. N 4. P. 858–863.
3. Головки Л.Ф., Лук'яненко С.О., Смаковський Д.С., Михайлова І.Ю., Агеєнко В.А. Моделювання температурного поля при зміцненні матеріалів лазерним випромінюванням // Моделювання та інформаційні технології. 2008. № 45. С. 28–35.
4. Кутателадзе С.С. Основы теории теплообмена. М.: Атомиздат, 1979. 416 с.

5. Калиткин Н.Н. Численные методы. М.: Наука, 1978. 512 с.
6. Лук'яненко С.О. Адаптивні обчислювальні методи моделювання об'єктів з розподіленими параметрами. Київ: Політехніка, 2004. 236 с.
7. Марчук Г.И. Методы вычислительной математики. М.: Наука, 1977. 456 с.
8. Тихонов А.Н., Самарский А.А. Уравнения математической физики. М.: Наука, 1977. 735 с.
9. 65Г – сталь конструкционная рессорно-пружинная. Марочник стали и сплавов [Электронный ресурс]. Режим доступа: http://www.splav.kharkov.com/mat_start.php?name_id=265, свободный. Яз. рус. (дата обращения 07.11.2013).
10. Михайлова І.Ю. Спосіб функціонального тестування точності результатів, отриманих з використанням адаптивної сітки // Математичне та комп'ютерне моделювання. Сер. Технічні науки. 2012. В. 7. С. 124–132.

Лукьяненко Святослав Алексеевич – доктор технических наук, профессор, Национальный технический университет Украины «Киевский политехнический институт», Киев, Украина, lukian@aprodos.kpi.ua

Михайлова Ирина Юрьевна – ассистент, Национальный технический университет Украины «Киевский политехнический институт», Киев, Украина, imikh@aprodos.kpi.ua

Svyatoslav Luk'yanenko – D.Sc., Professor, National Technical University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine, lukian@aprodos.kpi.ua

Irina Mykhailova – assistant, National Technical University of Ukraine «Kyiv Polytechnic Institute», Kyiv, Ukraine, imikh@aprodos.kpi.ua

УДК 532.542.4:533.6.011.32:612.215.41

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ ТУРБУЛЕНТНОГО ПОТОКА ВОЗДУХА С ИСПОЛЬЗОВАНИЕМ МЕТОДА ОТСОЕДИНЕННЫХ ВИХРЕЙ

А.А. Воронин^а, Г.Н. Лукьянов^а, Е.В. Фролов^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, ale_vor@rambler.ru

Приведен краткий обзор основных математических моделей описания потоков жидкостей и газов, от модели пути смешения Прандтля до одно- и двухпараметрических дифференциальных RANS-моделей, а также нестационарных моделей крупных LES- и отсоединенных DES-вихрей. Последний тип моделей, впервые предложенный П. Спалартом в 1997 г., позволил объединить преимущества моделей LES и RANS, что впоследствии расширило область применения нестационарных математических моделей к описанию потоков жидкостей и газов. Авторы исследования приводят математическую формулировку моделей DES, подробно останавливаясь на принципиальных отличиях данных методов от моделей RANS.

Также приведен вид трехмерной геометрической модели носовой полости человека, полученной на основе данных компьютерной томографии с использованием программного пакета Mercury Amiga. После сегментации данной модели с помощью программного пакета Altair Hypermesh была построена объемная нерегулярная сетка из $1,5 \times 10^7$ конечных элементов, на основе которой произведен нестационарный расчет параметров потока (программный пакет Ansys Fluent). Приведены полученные в результате расчета поля скоростей потока для вдоха и выдоха. Использование подробной расчетной сетки и нестационарной модели DES позволило выделить внутри потока отдельные мелкомасштабные вихревые структуры.

Авторами разработана твердотельная модель носовой полости, на основе которой были проведены экспериментальные исследования процесса дыхания. Измеренные значения перепада давления потока в преддверии носа показали удовлетворительное согласие с соответствующими расчетными данными, и был сделан вывод о возможности применения моделей DES для моделирования течений жидкостей и газов в каналах нерегулярной формы.

Ключевые слова: численное моделирование воздушных потоков, метод отсоединенных вихрей, турбулентность.

DETACHED-EDDY SIMULATION OF TURBULENT AIRFLOW

A. Voronin^а, G. Luk'yanov^а, E. Frolov^а

^а Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, Saint Petersburg, Russia, ale_vor@rambler.ru

A brief survey of the most significant mathematical models of the air and fluid flows from Prandtl's mixing length layer theory to RANS models with one and two differential equations, as well as unsteady LES and DES models is given. Detached-eddy simulation was first proposed by P. Spalart in 1997 and combined the main advantages of LES and RANS methods which gave the researchers the possibility to widen the sphere of such models application. The authors give the basic mathematical description of DES models paying special attention to the most important differences between DES and RANS models. The 3D geometrical model of human nasal cavities obtained from computer-aided tomography data using Mercury Amira program is also given. The 3D unstructured mesh with $1,5 \times 10^7$ finite elements was constructed after the segmentation using Altair Hypermesh software had been finished. The mesh was used to set up an unsteady simulation of airflow inside the obtained geometrical model. Application of DES method on the mesh of a good quality made it possible to distinguish the

small-scale turbulent swirls inside the flow. A solid-state model of human nasal cavities for experimental investigation of the breathing process was designed. The measured values of the pressure drops inside the model during breathing simulation showed good correlation with results obtained from the numerical calculations using Ansys software. The conclusion was made that DES models may be used for flows simulation of liquids and gases in non-regular shaped channels.

Keywords: airflow computational modeling, detached-eddy simulation, turbulence.

Введение

Численное моделирование турбулентных течений жидкостей и газов представляет собой весьма сложную и одновременно актуальную задачу, возникающую в различных областях науки и техники. Разработка и последующее внедрение в научно-техническую практику различных математических моделей турбулентности происходило в ответ на возникновение новых, принципиально более сложных задач расчета турбулентных течений, связанных с моделированием высокоскоростных потоков воздуха в аэрокосмических приложениях, определением параметров атмосферных циркуляционных вихрей и др.

С 1925 г., когда Л. Прандтлем была предложена первая математическая модель, позволяющая осуществлять расчет осредненных уравнений Навье–Стокса для движения жидкости посредством моделирования турбулентной вязкости через параметр пути смешения, различными исследователями было предложено множество моделей турбулентности, широко используемых в различных инженерных приложениях: алгебраические модели, RANS (Reinolds Averaged Navier-Stokes) модели с одним и двумя дифференциальными уравнениями для определения турбулентной вязкости, RSM (Reinolds Stress Modeling) модели рейнольдсовых напряжений, а также более совершенные модели LES (Large Eddy Simulation – модели крупных вихрей) и многие другие. При этом, несмотря на то, что разработка моделей RANS и LES происходила фактически одновременно, использование последних для расчета широкого спектра течений на сегодняшний день не представляется возможным по причине недостаточного уровня развития имеющихся в распоряжении исследователей вычислительных мощностей.

Указанная сложность была во многом преодолена, когда в 1997 г. П. Спаларт [1] предложил метод DES (Detached Eddy Simulation – моделирование отсоединенных вихрей), представляющий собой численное решение трехмерных нестационарных уравнений с использованием единой модели турбулентности, которая действует как подсеточная модель в областях с высоким сеточным разрешением и как модель URANS (Unsteady RANS – формулировка моделей RANS для расчета нестационарных задач) в областях, где сеточное разрешение недостаточно [2]. Основная концепция метода DES состоит в использовании модели LES для области свободного течения и моделей RANS для пограничных слоев, а также в качестве подсеточных функций. Это позволяет избежать чрезмерного дробления расчетной сетки в пристеночных областях и, таким образом, получать достаточно точное нестационарное решение гидродинамической задачи при сохранении приемлемого уровня затрат вычислительных ресурсов.

При этом, если в первоначальной формулировке, данной П. Спалартом, в качестве RANS-модели использовалась однопараметрическая модель SA (Spalart-Allmarès – модель Спаларта–Аллмареса), то в модели DES, представленной в работах [2, 3], предполагалось использование двухпараметрической RANS-модели SST (Shear Stress Transport – модель сдвиговых напряжений Ментера [4]).

В последующие годы были предложены усовершенствованные модификации модели DES – DDES (Delayed Detached Eddy Simulation) [5] и IDDES [6] (Improved Delayed Detached Eddy Simulation), позволившие распространить использование методов DES на моделирование широкого круга безотрывных течений. Такая универсализация открывает ряд новых возможностей для приложения методов DES к нестационарному моделированию турбулентных потоков.

В настоящем исследовании было осуществлено моделирование турбулентных течений в областях нерегулярной формы при помощи метода DES-SST, произведен сравнительный анализ полученных результатов и соответствующих экспериментальных данных.

Математическое описание модели отсоединенных вихрей

Для описания движения жидкости как сплошной среды используется система уравнений Навье–Стокса, включающая в себя уравнение неразрывности, а также уравнения для трех компонент скорости потока [7].

Уравнение неразрывности:

$$\frac{\partial \rho}{\partial \tau} + \frac{\partial \rho u_j}{\partial x_j} = 0. \quad (1)$$

Уравнения движения:

$$\frac{\partial \rho u_i}{\partial \tau} + \frac{\partial \rho u_j u_i}{\partial x_j} = -\frac{\partial P}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j}, \quad (2)$$

где u_i , u_j – компоненты скорости, ρ – плотность, P – давление, а τ_{ij} – компоненты тензора сдвиговых напряжений.

Основное отличие методов отсоединенных вихрей от RANS-моделей турбулентности состоит в том, что уравнения (1) и (2) могут быть отфильтрованы по области пространства с сохранением полного объема информации о высокочастотных пульсациях исследуемых величин. Иными словами, фильтрация уравнений движения по времени заменяется осреднением по области пространства.

Математическая формулировка фильтрации векторного поля скорости $\mathbf{u}(\mathbf{x}, t)$ выглядит следующим образом:

$$\bar{\mathbf{u}}(\mathbf{x}, t) = \int \mathbf{G}(\mathbf{r}, \mathbf{x}) \mathbf{u}(\mathbf{x} - \mathbf{r}, t) d\mathbf{r}, \quad (3)$$

где $\mathbf{G}(\mathbf{r}, \mathbf{x})$ – фильтрующая функция. При этом ширина фильтра, определяющая функцию $\mathbf{G}(\mathbf{r}, \mathbf{x})$, может быть задана априорно либо определенным образом соотнесена с размером расчетной сетки конечных элементов. Отфильтрованные в соответствии с выражением (3) уравнения движения (1) и (2) могут быть приведены к следующему виду:

$$\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial \bar{u}_i \bar{u}_j}{\partial x_j} = - \frac{\partial}{\partial x_i} \left(\bar{p} + \frac{2}{3} \mu \frac{\partial \bar{u}_k}{\partial x_k} \right) + \frac{\partial}{\partial x_i} \left(\mu \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) - \rho \tau_{ij}^r \right), \quad (4)$$

где $\tau_{ij}^r = \bar{u}_j \bar{u}_i - \bar{u}_j \bar{u}_i$ – тензор остаточных напряжений.

Моделирование тензора остаточных напряжений, определение которого оказывается необходимым для замыкания системы уравнений (4), производится с помощью так называемых моделей подсеточной вязкости, наиболее известная из которых – модель Смагоринского – содержит в своей основе гипотезу о пропорциональности тензора остаточных напряжений тензору скоростей деформаций отфильтрованного поля (соотношение, по виду аналогичное гипотезе Буссинеска).

Формулировка метода DES состоит в замене линейного масштаба l_{DES} , входящего в используемую RANS-модель, на следующее соотношение:

$$l_{DES} = \min \{ l_{RANS}, C_{DES} \Delta \},$$

где l_{DES} – линейный масштаб DES, $C_{DES} = \text{const}$, Δ – размер фильтра, определяемый как

$$\Delta = \max \{ \Delta_x, \Delta_y, \Delta_z \},$$

где $\Delta_x, \Delta_y, \Delta_z$ – размеры расчетной сетки в данной точке.

В результате такой замены в случае, если размер фильтра (расчетной сетки) Δ оказывается слишком грубым, и $C_{DES} \Delta > l_{RANS}$, то $l_{DES} = l_{RANS}$, и модель DES функционирует как модель RANS. В ином случае модель DES преобразуется к виду модели LES и выполняется решение отфильтрованных уравнений (4), что позволяет получить существенно нестационарное решение задачи.

Геометрическая модель внутренних носовых каналов

Построение геометрической модели внутренних носовых каналов человека осуществлялось на основе данных компьютерной томографии черепной коробки пациентов отоларингологического отделения медицинского центра «Адмиралтейские верфи» (Санкт-Петербург) при помощи специализированного программного пакета Mercury Amira. Вид трехмерной геометрической модели носовой полости человека приведен на рис. 1.

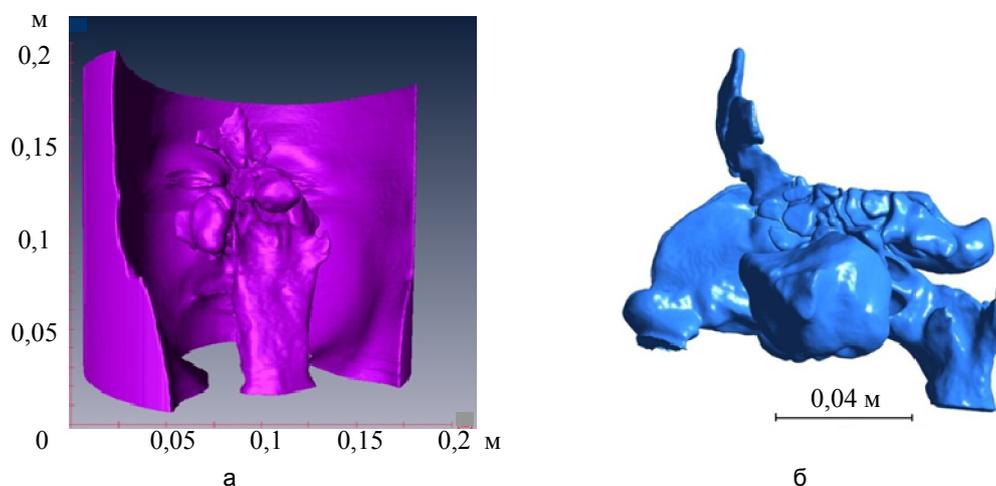


Рис. 1. Геометрическая модель внутренних носовых каналов человека: исходная модель (а); отредактированная модель (б)

Сегментация модели, приведенной на рис. 1 (выделение границ «вход», «выход», «стенка»), осуществлялась при помощи программного пакета Altair Hypermesh, построение расчетной сетки конечных элементов – с использованием программного пакета Ansys Icem CFD.

Численное моделирование воздушного потока в носовой полости человека

Результаты численного расчета нестационарного поля скоростей потока воздуха для геометрической модели внутренних носовых каналов (вдох и выдох), представленной на рис. 1, приведены на рис. 2. Моделирование осуществлялось с использованием метода DES-SST (программный пакет Ansys Fluent) на основе объемной нерегулярной сетки из $1,5 \times 10^7$ тетраэдрических конечных элементов со следующими граничными условиями: $p = 0$ – давление на входе в носовую полость; $u_1 = u_2 = u_3 = 0$ – компоненты скорости потока на стенке; $P = 50 \times \cos(\pi \times (0,869\tau - 0,47828)) + 50$ – давление в носоглотке при вдохе; $P = 50 \times \sin(\pi \times (1,17647\tau + 0,5)) - 50$ – давление в носоглотке при выдохе.

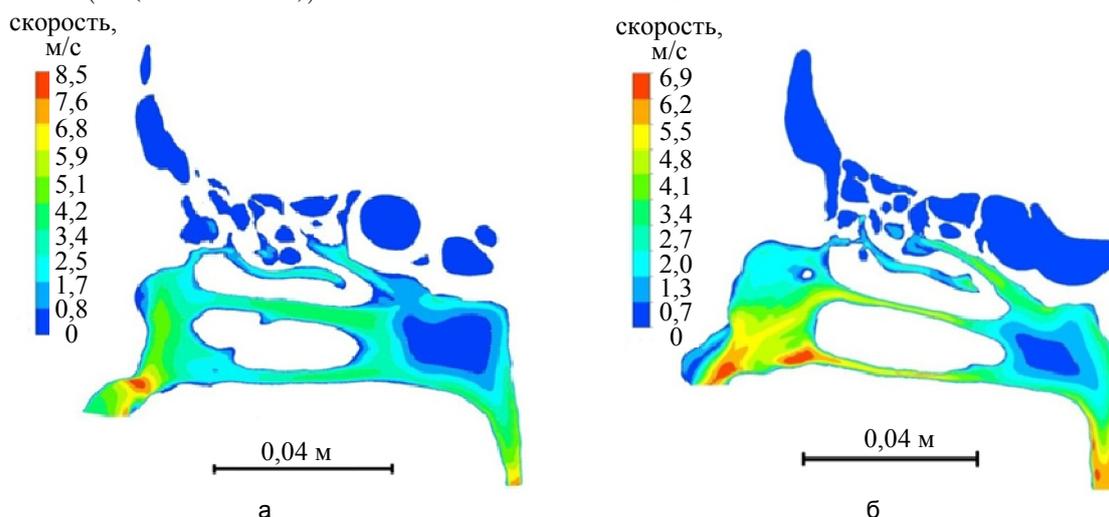


Рис. 2. Моделирование потока воздуха в носовой полости человека. Метод DES-SST. Поле скоростей: вдох (а); выдох (б)

В результате моделирования были получены пространственные поля скоростей и давлений потока, а также траектории движения малых объемов воздушной струи, визуализация которых способствовала выявлению в общей структуре потока воздуха отдельных вихревых структур различных масштабов. На рис. 3 приведен результат моделирования траекторий движения воздушного потока внутри гайморовой пазухи до (рис. 3, а) и после операции (рис. 3, б) по удалению новообразования (кисты). Вентиляция пазухи производится через естественное соустье. Удаление кисты осуществлялось через искусственное соустье, сформированное в ходе операции.

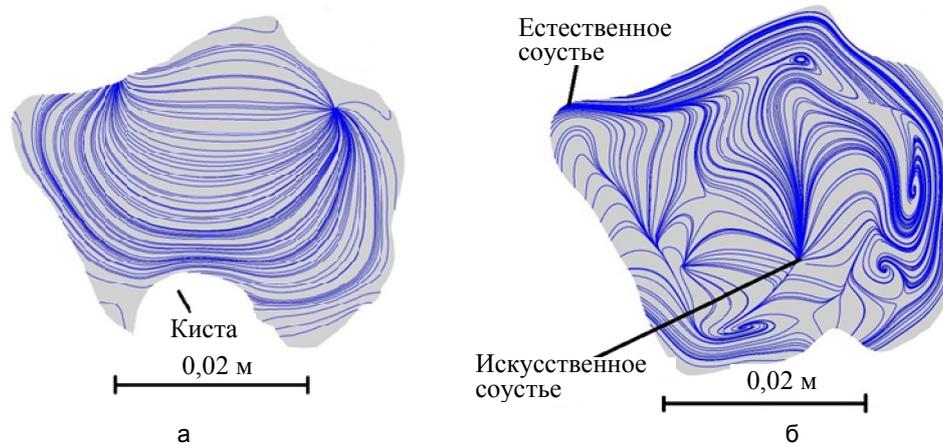


Рис. 3. Траектории потока воздуха в гайморовой пазухе носовой полости человека: до операции по удалению новообразования (кисты) (а); после операции (б)

Использование нестационарной модели отсоединенных вихрей применительно к мелкомасштабной сетке конечных элементов позволило также смоделировать процессы преобразования и перемещения

данных структур в пространстве с течением времени. Возможность моделирования таких вихревых структур малых размеров, формирующих высокочастотные компоненты колебания динамических характеристик турбулентного потока, при сохранении приемлемого уровня затрат вычислительных ресурсов представляет собой важное преимущество использованной в ходе настоящего исследования нестационарной модели отсоединенных вихрей. Моделирование потока воздуха внутри гайморовой пазухи с использованием метода DES для моделей, полученных из данных компьютерной томографии пациентов до и после операции по удалению новообразования, позволило сделать вывод о важности качественного воспроизведения таких вихревых структур, анализ положения которых внутри турбулентного потока может стать важным инструментом адекватной оценки параметров моделируемого течения и рассматриваемой геометрической модели.

Экспериментальное моделирование воздушного потока в каналах нерегулярной формы на примере носовой полости человека

На основе данных компьютерной томографии была создана твердотельная модель носовой полости (рис. 4), с использованием которой было осуществлено экспериментальное моделирование процесса дыхания. Имитация движения воздуха при дыхании производилась через трубку, прикрепленную к модели в области носоглотки. В различных точках модели были установлены пьезорезистивные датчики давления, позволяющие фиксировать изменение перепада давления потока внутри воздушной струи с течением времени.

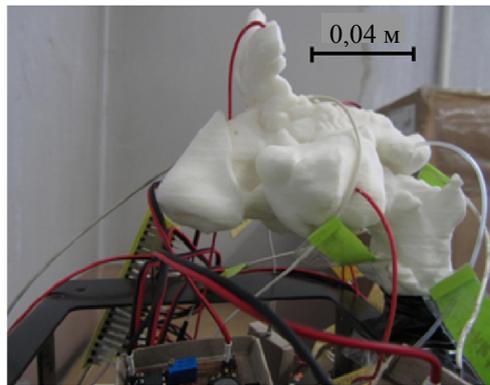


Рис. 4. Твердотельная модель носовой полости человека

Результаты экспериментального измерения перепада давления для твердотельной модели носовой полости, а также численного моделирования для соответствующей компьютерной модели внутренних носовых каналов (преддверие носа) приведены на рис. 5. Из рис. 5 видно, что результаты численного моделирования находятся в удовлетворительном согласии с данными эксперимента. Учитывая, что нестационарные граничные условия для изменения давления потока при численном расчете были заданы для выхода (область носоглотки), установленный факт сходства значений амплитуд колебаний давления на входе (преддверие носа) для твердотельной и компьютерной моделей позволил сделать вывод об адекватности использованной компьютерной модели носовых каналов, а также метода расчета применительно к решению гидродинамической задачи.

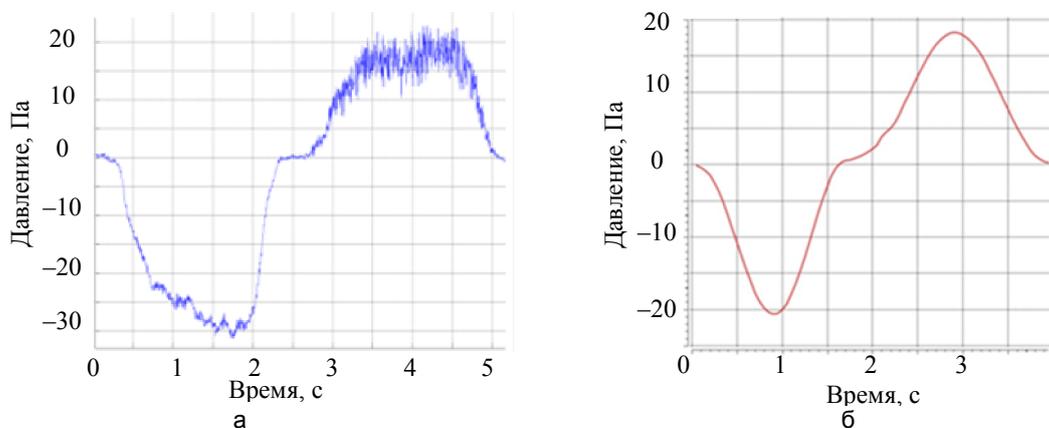


Рис. 5. Результаты моделирования изменения перепада давления в преддверии носа в процессе дыхания: твердотельная модель (а); численный расчет (б)

Заключение

В ходе настоящей работы были исследованы возможности применения нестационарной математической модели отсоединенных вихрей Спаларта к решению задач, связанных с моделированием движения воздушного потока в областях нерегулярной формы на примере внутренних носовых каналов человека. Полученные результаты численного расчета показали удовлетворительное согласие с соответствующими опытными данными.

Использование мелкомасштабной сетки конечных элементов ($1,5 \times 10^7$ тетраэдрических элементов) позволило воспроизвести пространственные распределения скоростей и давлений потока воздуха с достаточно высоким разрешением. Визуализация траекторий движения малых объемов воздуха в общей структуре воздушной струи позволила идентифицировать отдельные вихревые компоненты турбулентного потока. Возможность выявления таких структур является несомненным достоинством методов DES, которые могут быть использованы в задачах моделирования широкого круга течений в областях как регулярной, так и нерегулярной геометрии.

Литература

1. Spalart P.R., Jou W.-H., Strelets M., Allmaras S.R. Comments on the feasibility of LES for wings, and on a hybrid RANS/LES approach // Proc. of First AFOSR International Conference on DNS/LES. Ruston, Louisiana: Greyden Press, 1997. P. 137–147.
2. Travin A., Shur M., Strelets M., Spalart P. Detached-Eddy Simulations Past a Circular Cylinder // Flow, Turbulence and Combustion. 1999. V. 63. N 1–4. P. 293–313.
3. Travin A., Shur M., Strelets M., Spalart P.R. Physical and numerical upgrades in the detached-eddy simulation of complex turbulent flows // 412 Euromech Colloquium on LES of complex transitional and turbulent flows. Abstracts. Munich, Germany, 2000. P. 87–93.
4. Menter F.R. Zonal Two Equation $k-\omega$ Turbulence Models for Aerodynamic Flows // AIAA Paper. 1993. V. 93-2906. 21 p.
5. Spalart P.R., Deck S., Shur M.L., Squires K.D., Strelets M.Kh., Travin A. A new version of detached-eddy simulation, resistant to ambiguous grid densities // Theoretical and Computational Fluid Dynamics. 2006. V. 20. N 3. P. 181–195.
6. Shur M.L., Spalart P.R., Strelets M.Kh., Travin A.K. A hybrid RANS-LES approach with delayed-DES and wall-modelled LES capabilities // International Journal of Heat and Fluid Flow. 2008. V. 29. N 6. P. 1638–1649.
7. Воронин А.А., Лукьянов Г.Н., Неронов Р.В. Моделирование воздушного потока в каналах нерегулярной формы // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 3 (85). С. 113–118.

- Воронин Алексей Анатольевич** – аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, ale_vor@rambler.ru
- Лукьянов Геннадий Николаевич** – доктор технических наук, профессор, зав. кафедрой, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Gen-lukjanow@yandex.ru
- Фролов Евгений Викторович** – ассистент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Frolov_evg.v@mail.ru
- Aleksei Voronin** – postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, ale_vor@rambler.ru
- Gennadiy Luk'yanov** – D.Sc., Professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Gen-lukjanow@yandex.ru
- Evgeny Frolov** – assistant, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Frolov_evg.v@mail.ru

УДК 004.051, 336.76

СТРАТЕГИЯ МАРКЕТ-МЕЙКИНГА В СИСТЕМЕ ВЫСОКОЧАСТОТНОЙ АЛГОРИТМИЧЕСКОЙ ТОРГОВЛИ

А.В. Торопов^{а, б}

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, toropov@rain.ifmo.ru

^б Tbricks AB, Санкт-Петербург, Россия, tav@tbricks.com

Маркет-мейкер – важнейший участник современной биржевой торговли, обеспечивающий повышение ликвидности рынка и уменьшение спреда между ценой покупки и продажи. В работе представлена автоматическая стратегия маркет-мейкера для котирувания опционов и других финансовых инструментов на электронных биржах. Котировки инструментов рассчитываются на основе их теоретической стоимости, что является ресурсоёмким вычислением. Для обеспечения смены котировок в режиме реального времени предложены алгоритмические оптимизации, позволившие добиться четырехкратного ускорения смены котировок в рыночных условиях – механизма кэширования котировок в заданном диапазоне вокруг текущей цены базового актива. Механизм кэширования предварительно рассчитывает котировки в заданном диапазоне вокруг текущей цены базового актива. При изменении цены на рынке в пределах диапазона стратегия отправляет на биржу готовое сообщение о смене котировок вместо выполнения нового трудоемкого расчета. Сглаживание колебаний цены базового актива позволяет избежать постоянного изменения границ диапазона и реагировать только на существенные движения рынка. Экспериментально определен размер кэшируемого диапазона (40 элементов), обеспечивающий оптимальное соотношение скорости смены котировок и потребляемых ресурсов. При котирувании 36 опционов на рынке Eurex Exchange средняя задержка смены котировок при изменении цены базового актива составила 277 мкс. Замеры производились на сервере Sun X4170 M3: CPU(s): 2xXeon 2.9 ГГц RAM: 128 ГБ под управлением операционной системы Solaris 10. Полученные результаты соответствуют современным требованиям маркет-мейкеров. Разработанная стратегия используется крупными европейскими банками и трейдинговыми фирмами.

Ключевые слова: алгоритмическая торговля, маркет-мейкинг, котирувание.

MARKET-MAKING STRATEGY IN THE SYSTEM OF ALGORITHMIC HIGH-FREQUENCY TRADING

А. Торопов^{с, d}

^с Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, toropov@rain.ifmo.ru

^d Tbricks AB, Saint Petersburg, Russia, tav@tbricks.com

Market maker is the most important participant of modern exchange trading, it provides the increase of market liquidity and reduces the difference between bid and ask (spread). The paper presents automatic market-making strategy for quoting of options and other kinds of financial instruments on electronic markets. Quotes are based on theoretical pricing which is a resource-intensive task. Presented algorithmic optimizations, in particular quotes caching and smoothing of underlying asset price oscillation, give the possibility up to four times boost for quote modify scenario on real market data. Mechanism of quotes caching precalculates quotes in certain diapason around current underlying price. If underlying price changes within the diapason, algorithm sends already filled message for quote modification, instead of new complex computation. Smoothing of underlying asset price oscillation prevents permanent moving of the diapason and reacts only on significant market moving. A size of caching diapason which provides optimal correlation between speed of quotes modification and resource consumption has been defined experimentally (40 elements). In case of quoting 36 options on Eurex Exchange an average delay between underlying price change and quote modification is 277 usec. The measurements were carried out on the Sun X4170 M3: CPU(s): 2xXeon 2.9GHz RAM: 128 GB server under Solaris 10 operating system. Obtained results correspond to modern market-making requirements. The developed strategy is used by big European banks and trading firms.

Keywords: algorithmic trading, market-making, quota system.

Введение

Алгоритмическая торговля (algorithmic trading) – процесс совершения сделок на финансовых рынках в автоматическом режиме с помощью специализированных компьютерных программ, реализующих различные торговые стратегии – алгоритмы принятия решений о совершении торговых операций. Такие программы называют автоматическими торговыми стратегиями. Одной из наиболее популярных на сегодняшний день стратегий на рынке является маркет-мейкинг.

Маркет-мейкер (market-maker) – один из важнейших участников биржевой торговли, берущий на себя риск приобретения и хранения ценных бумаг с целью организации их продаж. Маркет-мейкер заключает с биржей контракт, согласно которому он обязан выставлять активные заказы на покупку и продажу (котирувать, от quoting) определенной группы финансовых инструментов, как правило, группы опционов на определенный базовый актив, в течение торгового времени. Такие соглашения оговаривают максимальную разницу между ценой покупки и продажи (спред, от spread), штрафы за невыполнение условий и т.д. От биржи маркет-мейкер получает либо денежную компенсацию, либо уменьшенную комиссию на заключение сделок, что дает ему возможность вести торговую деятельность с меньшими издержками [1]. Пример обязательств маркет-мейкера на рынке Eurex можно найти, например, в [2].

В условиях нестабильной экономической ситуации в мире крупнейшие биржи, такие как NYSE, NASDAQ, EUREX, LSE, CME и многие другие, для повышения лояльности инвесторов привлекают сотни маркет-мейкеров, которые ежедневно котируют десятки тысяч финансовых инструментов [3]. Основной задачей маркет-мейкера является повышение ликвидности котируемых инструментов и уменьшение спреда между ценой покупки и продажи. Большое количество активных маркет-мейкеров способствует приближению цены при больших объемах сделки к лучшей на рынке и понижению волатильности [4].

Задачей настоящей работы являлась разработка программного обеспечения для осуществления деятельности маркет-мейкеров в автоматическом режиме на ведущих мировых электронных финансовых рынках с учетом всех современных требований к такой функциональности. В качестве основы для разработки был взят программный комплекс для алгоритмической торговли компании Tbricks AB [5]. Система Tbricks представляет собой распределенную систему с клиент-серверной архитектурой, разделенную на отдельные сервисы. Ее специфическими особенностями являются: широкие функциональные возможности, покрывающие требования различных финансовых организаций; поддержка протоколов для соединения с ведущими мировыми финансовыми рынками; хорошая масштабируемость; эффективная работа на любом количестве серверов (успешный опыт использования от 1 до 20 серверов с 8–24 CPU каждый); автоматическое перераспределение нагрузки (load-balancing) [5].

Алгоритм выбора котировок

Главным вопросом при создании стратегии маркет-мейкера становится выбор цен для котировок, так как торговля по необоснованным ценам может принести серьезные убытки. Естественное решение здесь – базирование котировок на теоретической стоимости финансового инструмента. Теоретическое ценообразование позволяет получить справедливую стоимость [6, 7], которая определяет рациональную, свободную от арбитражных возможностей оценку потенциальной рыночной стоимости финансового инструмента, принимающую во внимание объективные рыночные факторы (цена базового актива, время жизни инструмента, финансовая ставка, дивиденды и др.). Такая теоретическая оценка позволяет, например, правильно застраховать свои риски путем хеджирования сделок [7].

На современных биржах торгуется множество видов финансовых инструментов: европейские и американские опционы [6, 7], экзотические (exotic) опционы (азиатские, барьерные опционы и турбо варранты) [6] и многие другие. Известны различные подходы к вычислению справедливой стоимости этих финансовых инструментов, приводящие к аналитическим или численным методам расчета [6, 8]. Стоит отметить, что часто теоретическая оценка является ресурсоемким вычислением и требует относительно больших временных затрат [9]. Более подробно о теоретическом ценообразовании можно узнать в [6–8], а о его применении в автоматических торговых стратегиях – в [10].

Схема алгоритма выбора котировок, реализованного нами, проиллюстрирована на рис. 1. Базовый инструмент торгуется на рынке по цене покупки S_b и продажи S_a . Соответствующие теоретические цены покупки $V_b(S_b)$ и продажи $V_a(S_a)$ производного инструмента раздвигаются в меньшую и большую сторону до значений M_b и M_a соответственно, чтобы обеспечить задаваемый маркет-мейкером спред. Разница между теоретическими ценами (V_b или V_a) и соответствующими им котировками (M_b и M_a соответственно) включает в себя всевозможные риски, ожидаемый доход от сделок, учитывает текущую волатильность рынка.

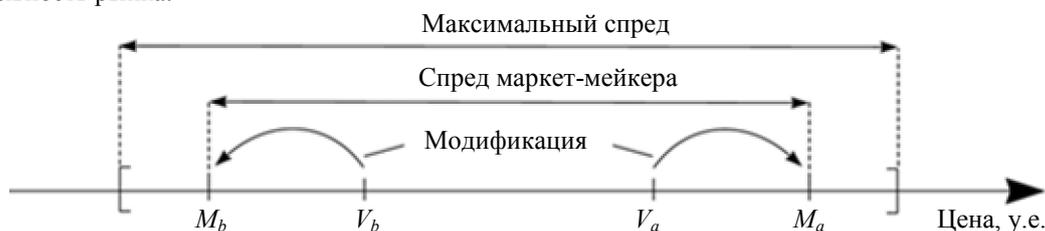


Рис. 1. Схема алгоритма выбора котировок на основе теоретической цены

Детальная настройка стратегии включает более 70 параметров, среди которых величина спреда маркет-мейкера, минимальный и максимальный объем выставляемых котировок, максимальный объем купленных или проданных активов, максимальное количество совершенных сделок в определяемый временной интервал и др.

Алгоритм обновления котировок

Важным требованием при таком подходе к выбору котировок является высокая скорость расчета теоретической стоимости при изменении цены базового актива. Задержки в расчете приводят к котиروанию инструментов по устаревшим ценам и соответственно могут принести убытки маркет-мейкеру.

Часто теоретическая стоимость не может быть посчитана за удовлетворительное время, поэтому нами был предложен алгоритм кэширования котировок.

Алгоритмом предварительно рассчитываются теоретические значения для диапазона $[S_{-k}, S_{-k+1}, \dots, S_{-1}, S_0, S_1, \dots, S_{k-1}, S_k]$ вокруг текущей цены базового актива S_0 , на их основе рассчитываются котировки и кэшируются в памяти. Разница между соседними ценами в диапазоне определяется минимальным шагом, заданным биржей. При изменении цены на базовый актив с S_0 на S_1 стратегии достаточно отправить на рынок уже готовое сообщение об изменении котировок на заранее посчитанные значения. При изменении цены на S_1 весь диапазон сдвигается вправо, так что текущая цена базового актива S_1 становится его центром. При этом котировки, посчитанные для цены актива S_{-k} , стираются из кэша, а котировки для цены S_{k+1} рассчитываются в фоновом режиме и помещаются в кэш (рис. 2). Количество предварительно рассчитанных котировок $N = 2 \cdot k$ называется размером кэша. При существенном движении рынка новая цена S не попадает в пределы кэшируемого диапазона ($S < S_{-k}$ или $S > S_k$). Тогда необходимо вычислить теоретические значения для цены S , рассчитать на их основе котировки, отослать сообщение на биржу и перестроить диапазон вокруг цены S . Смена котировок в этом случае занимает намного больше времени по сравнению с поиском кэшированных котировок. В разделе «Анализ производительности» мы покажем, что при правильно выбранном N доля непопаданий в кэш в реальных рыночных условиях мала.

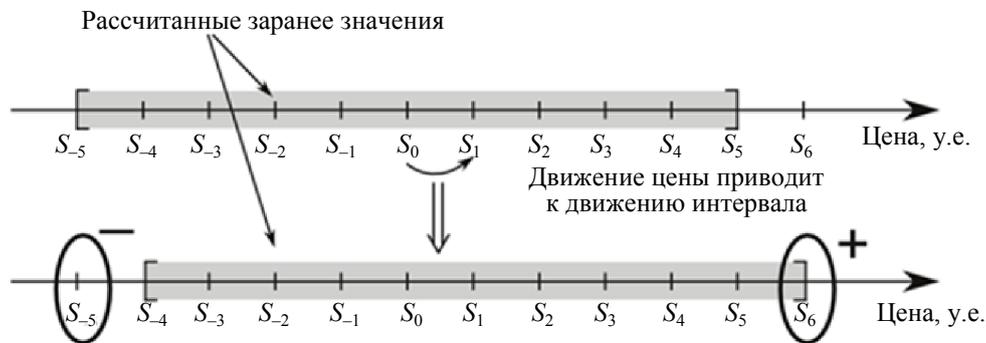


Рис. 2. Алгоритм кэширования котировок при $N = 10$

Чтобы избежать частого изменения границ диапазона в связи с малыми колебаниями цены на рынке, был применен механизм реагирования только на существенные движения цены.

Пусть $S_1, S_2, S_3, \dots, S_m, \dots$ – последовательность цен акции. Определим величину H_m :

$$H_m = \sum_{i=1}^{m-1} (S_{i+1} - S_i) \cdot$$

Когда цена акции осциллирует вокруг некоторого уровня, значение H_m близко к нулю. Если же уровень цены акции существенно движется в большую или меньшую сторону, абсолютное значение $|H_m|$ увеличивается. Когда оно превысит задаваемый порог M , границы диапазона сдвигаются (рис. 2).

Анализ производительности

Анализ производительности реализованных в работе алгоритмов заключался в измерении полного времени, затрачиваемого на смену котировок при изменении цены базового актива. Это время от поступления сообщения об изменении цены базового актива в систему алгоритмической торговли до отправки сообщения на биржу с измененными котировками (рис. 3). Результаты позволяют оценить влияние предлагаемых оптимизаций на работу маркет-мейкера в реальных рыночных условиях.

В качестве источника цен взяты котировки акции компании Volvo AB на рынке NASDAQ OMX Nordic [11] за 6 августа 2012 г. Отправка котировок производилась на рынок Eurex ETI 2.0 [12]. Для анализа использован сервер Sun X4170 M3: CPU(s): 2xXeon 2.9GHz RAM: 128 GB, под управлением операционной системы Solaris 10. На основе фреймворка динамической трассировки DTTrace [13] был создан скрипт, позволяющий измерить минимальное, максимальное и среднее время между поступлением новой цены с биржи и отправкой котировок.

В таблице приведено минимальное, среднее и максимальное время смены котировок на 36 американских опционов (18 опционов типа колл и 18 типа пут [6]) при выключенном алгоритме кэ-

ширования ($N = 0$) и размерах кэша $N = 10, 20, 40, 45$. Колонка «Попадание в кэш» показывает долю измерений, когда после изменения цена на базовый актив осталась в пределах кэшируемого диапазона.

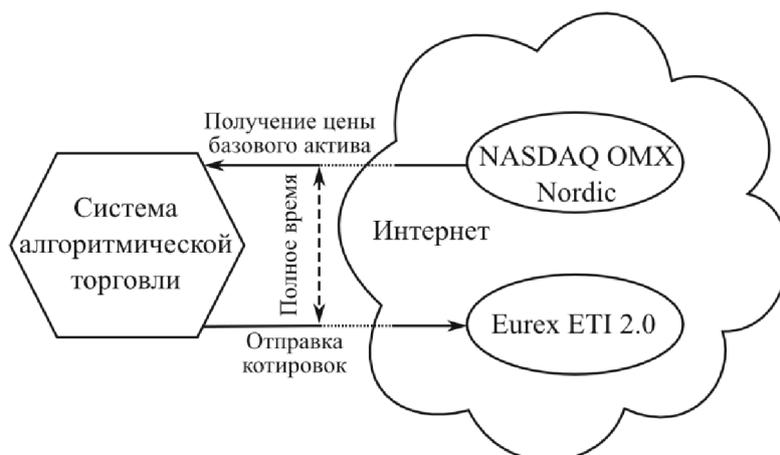


Рис. 3. Схема измерения производительности работы маркет-мейкера

Под выбросами (колонка «Выбросы» в таблице) понимается доля замеров, отработавших 1,5–2 и более миллисекунд в связи с задержками в работе операционной системы (например, задержки при возвращении управления потоку при системных вызовах, загрузка исходного кода в кэш при кросс-библиотечных вызовах и др.), а также за счет сетевых подключений, проб DTtrace и других причин, не зависящих от системы алгоритмической торговли. Данные выбросов не учитывались при вычислении среднего и максимального времени смены котировок.

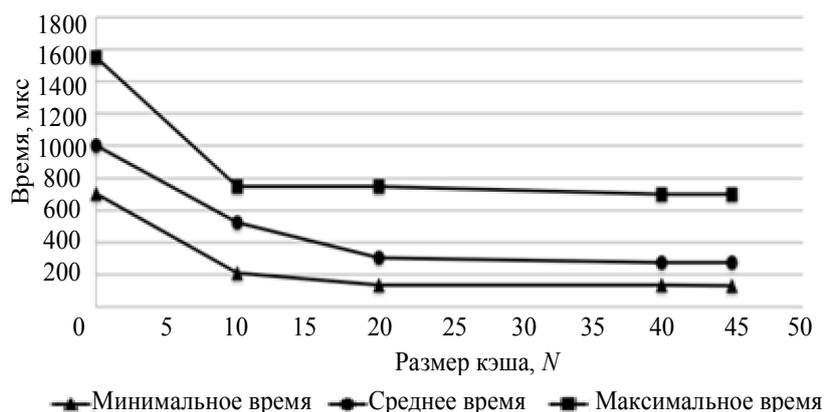


Рис. 4. Зависимость минимального, среднего и максимального времени от размера кэша

Размер кэша	Минимальное время, мкс	Среднее время, мкс	Максимальное время, мкс	Попадание в кэш, %	Выбросы, %
0	702	1001	< 1600	0%	5,83%
10	209	524	< 800	1,54%	0,44%
20	136	308	< 800	77,6%	0,41%
40	134	276	< 800	94,07%	0,14%
45	132	277	< 800	94,43%	0,15%

Таблица. Скоростные характеристики смены котировок

На рис. 4 изображена зависимость минимального, среднего и максимального времени от размера кэша. Из таблицы и графика (рис. 4) видно, что при размере кэша $N \geq 40$ среднее время работы не превышает 280 мкс и практически не изменяется. На рис. 5 изображены зависимости доли попадания в кэш и выбросов от размера кэша. Как видно из графика (рис. 5), при $N \geq 40$ попадание в кэш стремится к 95%, а доля выбросов к 0%. Опираясь на эти данные, нами даны рекомендации по выбору размера кэшируемого диапазона $N = 40$.

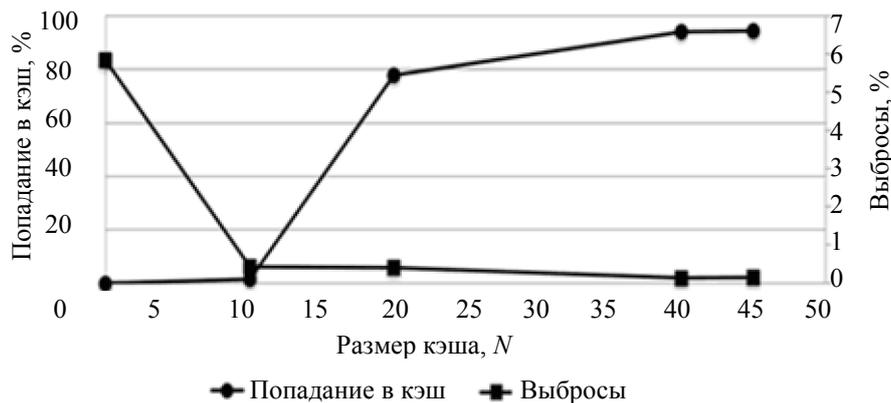


Рис. 5. Зависимость доли попаданий в кэш и доли выбросов от размера кэша

Заключение

На основе инфраструктуры компании Tbricks AB разработана стратегия для осуществления деятельности маркет-мейкеров в автоматическом режиме на ведущих мировых электронных финансовых рынках [14]. Предложенный оптимизированный алгоритм смены котировок позволил обеспечить среднюю задержку менее 280 мкс и пренебрежимо малую долю выбросов при рекомендованном автором размере кэша $N = 40$. Полученные характеристики соответствуют современным требованиям маркет-мейкеров к системам высокочастотной алгоритмической торговли и являются конкурентоспособными на ведущих мировых финансовых рынках. Разработанная стратегия успешно применяется клиентами компании Tbricks AB – крупными европейскими банками и трейдинговыми фирмами [15–17].

Автор признателен научному руководителю д.ф.-м.н. Ю.А. Шполянскому за чтение рукописи статьи и полезные замечания, М.А. Святухиной за помощь в проведении тестирования, коллегам в компании Tbricks AB за плодотворное сотрудничество по тематике работы.

Литература

1. Investopedia. Electronic Trading: The Role of a Market Maker [Электронный ресурс]. Режим доступа: <http://www.investopedia.com/university/electronictrading/trading3.asp>, свободный. Яз. англ. (дата обращения 18.10.2013).
2. Eurex Exchange. Market Makers Obligations at Eurex [Электронный ресурс]. Режим доступа: http://www.eurexexchange.com/blob/exchange-en/62496/650910/2/data/obligations_13_10_01e.pdf, свободный. Яз. англ. (дата обращения 18.10.2013).
3. NYSE Liffe. Current liquidity providers and vacant places [Электронный ресурс]. Режим доступа: https://globalderivatives.nyx.com/sites/globalderivatives.nyx.com/files/current_liquidity_providers_and_vacant_places_website_9.xls, свободный. Яз. англ. (дата обращения 18.10.2013).
4. JSE Securities Exchange. The Role of a Market Maker [Электронный ресурс]. Режим доступа: http://www.jse.co.za/Libraries/Commodity_Derivatives_Products_Quanto_Futures_Options/Market_Making_Quanto_Futures.sflb.ashx, свободный. Яз. англ. (дата обращения 18.10.2013).
5. Tbricks AB. Official website [Электронный ресурс]. Режим доступа: <http://www.tbricks.com>, свободный. Яз. англ. (дата обращения 18.10.2013).
6. Wilmott P. Paul Wilmott on Quantitative Finance. 2nd ed. John Wiley & Sons, Ltd., 2006. V. 1–3. 1500 p.
7. Hull J.C. Options, futures and other derivatives. 7th ed. Prentice-Hall International, Inc, 2009. 814 p.
8. Haug E.G. The complete guide to option pricing formulas. 2nd ed. McGraw-Hill, 2007. 536 p.
9. Косяков М.С., Шинкарук Д.Н., Торопов А.В., Шполянский Ю.А. Применение технологии CUDA для ускорения расчета цен опционов европейского типа сеточным методом // Изв. вузов. Приборостроение. 2012. Т. 55. № 10. С. 13–19.
10. Торопов А.В., Иванов Д.В., Шполянский Ю.А. Применение теоретического ценообразования опционов в автоматических торговых стратегиях // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 2 (84). С. 136–141.
11. NASDAQ OMX Nordic. Official website [Электронный ресурс]. Режим доступа <http://www.nasdaqomxnordic.com>, свободный. Яз. англ. (дата обращения 18.10.2013).
12. Eurex Exchange. Official website [Электронные ресурсы]. Режим доступа <http://www.eurexexchange.com/exchange-en/>, свободный. Яз. англ. (дата обращения 18.10.2013).
13. Oracle. Solaris Dynamic Tracing Guide [Электронные ресурсы]. Режим доступа <http://docs.oracle.com/cd/E19253-01/817-6223/>, свободный. Яз. англ. (дата обращения 18.10.2013).

14. Tbricks AB. Market making solution [Электронный ресурс]. Режим доступа: <http://www.tbricks.com/solutions/market-making>, свободный. Яз. англ. (дата обращения 18.10.2013).
15. Tbricks AB. Tbricks wins market making platform deal from All Options [Электронный ресурс]. Режим доступа: <http://www.tbricks.com/company/news#tbricks-wins-market-making-platform-deal-from-all-options>, свободный. Яз. англ. (дата обращения 18.10.2013).
16. Tbricks AB. Carnegie chooses Tbricks, replaces legacy system for market making [Электронный ресурс]. Режим доступа: <http://www.tbricks.com/company/news#carnegie-chooses-tbricks>, свободный. Яз. англ. (дата обращения 18.10.2013).
17. Tbricks AB. Nino Options chooses Tbricks for market making and trading [Электронный ресурс]. Режим доступа: <http://www.tbricks.com/company/news#nino-options-chooses-tbricks-for-market-making-and-trading>, свободный. Яз. англ. (дата обращения 18.10.2013).

- Торопов Александр Владимирович* – Аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики; инженер-программист, Tbricks AB, Санкт-Петербург, Россия, tav@tbricks.com, toropov@rain.ifmo.ru
- Alexander Toropov* – Postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia; software engineer, Tbricks AB, Saint Petersburg, Russia, tav@tbricks.com, toropov@rain.ifmo.ru

УДК 37:004

ОТ ТРАДИЦИОННОГО ДИСТАНЦИОННОГО ОБУЧЕНИЯ
К МАССОВЫМ ОТКРЫТЫМ ОНЛАЙН-КУРСАМ

В.Н. Васильев^а, С.К. Стафеев^а, Л.С. Лисицына^а, А.В. Ольшевская^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, olshevskaya@mail.ifmo.ru

Поставлена проблема перехода вузов России от традиционного дистанционного обучения к массовому электронному обучению на основе открытых онлайн-курсов, обоснована ее актуальность. Проведен анализ важнейших предпосылок для успешности перехода (востребованность образовательных интернет-ресурсов, используемых с мобильных устройств; наличие большого количества различных электронных ресурсов, успешно используемых вузами на практике в дистанционных образовательных технологиях; наличие опыта ведения электронных журналов успеваемости студентов для планирования и оценивания результатов обучения; существенный рост материальных затрат в мировом рынке онлайн-обучения). Определены ключевые задачи перехода и сформулированы основные принципы разработки электронных онлайн-курсов. Приведена методика разработки результативного электронного онлайн-курса. Методика содержит четыре этапа: планирование ожидаемых результатов обучения, структурирование электронного содержания курса и разработка сценария обучения, разработка плана испытаний и электронных оценочных средств для автоматического контроля запланированных результатов обучения, реализация курса с использованием игровых механик и технологий сетевого общения обучаемых. Определены требования к различным формам контроля запланированных в курсе результатов обучения. Выделены два вида электронных онлайн-курсов (научные и технологические курсы). Приведены примеры их реализации в авторских онлайн-курсах «Волновая оптика», «Теория графов», «Разработка веб-интерфейсов на основе HTML и CSS», созданных и используемых на практике в 2013 году в НИУ ИТМО. Поставлены актуальные задачи развития массового открытого образования в ведущих вузах России.

Ключевые слова: электронное обучение; массовое открытое образование; массовые открытые онлайн-курсы; технологическая платформа: AcademicNT; дистанционные образовательные технологии; информационные технологии; система управления обучением; социальная сеть.

FROM TRADITIONAL DISTANCE LEARNING TO MASS ONLINE
OPEN COURSES

V. Vasiliev^b, S. Stafeev^b, L. Lisitsyna^b, A. Ol'shevskaya^b

^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, olshevskaya@mail.ifmo.ru

The issue of transition for higher education institutions of Russia from traditional distance learning to mass electronic education on the basis of the online open courses is considered, its relevance is proved. Analysis of the major prerequisites for transition success is carried out (a demand for the educational Internet resources from mobile devices; existence of a large number of various electronic resources which are successfully used in practice by higher education institutions in remote educational technologies; maintaining experience for electronic magazines of students' progress for planning and estimation of training results; essential growth of material costs in the world online training market). Key issues of transition are defined and the basic principles of electronic online courses development are formulated. A technique for electronic online course development aimed at the result is given. The technique contains the following four stages: planning of expected training results, course electronic content structuring and training scenarios creation, development of the tests plan and electronic estimated means for automatic control of the planned training results; course realization by means of game mechanics and technologies of network communication between students. Requirements to various forms of control planned in the course of learning results are defined. Two kinds of electronic online courses are assigned (knowledge-intensive and technological courses). Examples of their realization in the authors' online courses "Wave Optics", "Theory of Graphs", "Development of Web Interfaces on the Basis of HTML and CSS" created and practically used in NRU ITMO in 2013 are given. Finally, the actual tasks of mass open education development in the leading higher education institutions of Russia are set forth.

Keywords: e-learning; mass open education, mass online open courses, technological platforms, AcademicNT, distance education technologies, information technologies, learning management system, social network.

Введение

Дистанционные образовательные технологии и электронные ресурсы для их реализации прочно вошли в педагогическую практику на всех уровнях образования, и уже стали традиционными для подготовки студентов в вузах Российской Федерации (РФ). В последние годы в мире произошла настоящая революция, характеризуемая аббревиатурой МООС (Massive Open Online Courses), которая поставила новые задачи по модернизации информационно-образовательных сред и систем электронного обучения

на основе открытых онлайн-курсов. Взрыв популярности МООС вполне объясним. Востребованное образование по актуальным специальностям всегда и везде стоит дорого, а МООС позволяют не только получать знания условно бесплатно, но еще и обучаться у лучших преподавателей мира. Газета «The New York Times» – третья по популярности газета в США, одна из влиятельнейших газет мира – назвала 2012 год годом массовых открытых онлайн-курсов [1]. По данным на июль 2013 года, общее количество студентов на образовательном портале онлайн-курсов Coursera (<https://www.coursera.org/>) превысило 4 млн человек из 195 стран. РФ входит в пятерку стран-лидеров по числу студентов, обучающихся на этом портале, что говорит о весьма высокой востребованности сегмента МООС как новой технологии электронного обучения. В этой связи есть серьезные опасения, что русский язык как язык профессионального общения в некоторых странах СНГ постепенно сдает свои позиции [2]. Понимая это, разработчики группы компаний АйТи создали и подготовили к апробации собственную русскоязычную платформу CourseHub (<http://скомпьютеромнаты.рф>).

В этой статье мы познакомим читателей с предпосылками, созданными для перехода к электронному обучению в вузах РФ, а также осветим некоторые, на наш взгляд, успешные методики и проекты создания и использования онлайн-курсов в нашем университете.

Предпосылки для перехода к электронному обучению

В настоящее время созданы объективные предпосылки для успешного перехода к электронному обучению на основе онлайн-курсов в системе высшего образования России. Отметим наиболее значимые из них.

- Изменение в составе электронных ресурсов, используемых россиянами для получения новых знаний и навыков.** По результатам опроса Фонда общественного мнения (рис. 1), проведенного 15.10.2013 года среди граждан РФ от 18 лет и старше (всего в интервью принимало участие 1500 респондентов из 43 субъектов РФ), тенденции очевидны: для получения новых знаний и навыков россияне отдают все большее предпочтение образовательным интернет-ресурсам, в том числе доступным с мобильных устройств.

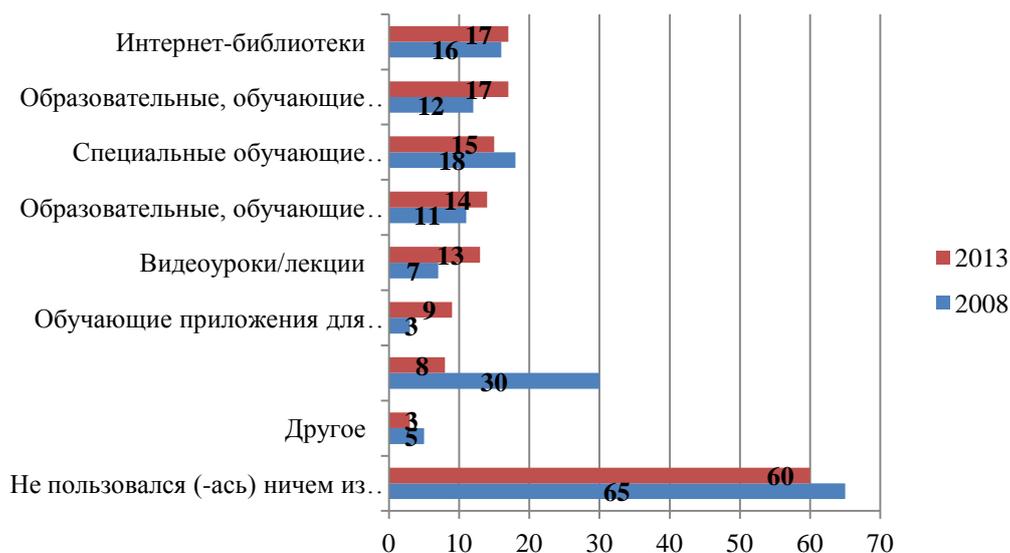


Рис. 1. Распределение россиян (%) по использованию компьютера и Интернета для получения новых знаний и навыков

- Накопление практического опыта разработки и использования электронных ресурсов для дистанционной поддержки учебного процесса в вузах.** На рис. 2 показана динамика роста курсов НИУ ИТМО, загруженных в систему дистанционного обучения второго поколения – AcademicNT (<http://de.ifmo.ru>) [3], с электронными ресурсами для обучения и аттестаций (компьютерные тесты, виртуальные лаборатории для практических занятий, электронные тренажеры и т.п.). Здесь небольшое падение роста таких курсов на 1 и 2 курсах связано с вводом новых образовательных программ, реализующих ФГОС ВПО для подготовки бакалавров. Тем не менее, следует отметить, что в НИУ ИТМО и в других вузах России имеется значительный потенциал методически проработанных и многократно апробированных ресурсов, которые могут послужить хорошим заделом для перехода на электронное обучение.

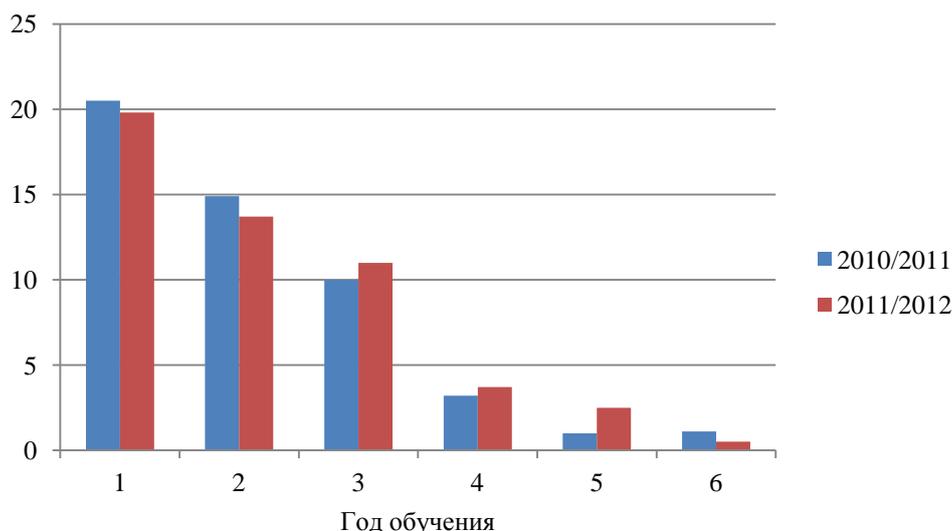


Рис. 2. Доля курсов НИУ ИТМО (%), содержащих электронные ресурсы

3. **Накопление практического опыта разработки и использования электронных журналов успеваемости студентов, обеспечивающих планирование и контроль результатов обучения студентов.** На рис. 3 приведен пример электронного журнала студентов НИУ ИТМО, в котором индивидуально в течение всего семестра накапливаются баллы по всем запланированным результатам обучения (тесты, отчеты за лабораторные работы, контрольные работы и т.д.), а также оценки за личностные качества, проявленные студентами в их испытаниях (сдача в срок, оригинальность предложенного решения и т.п.). Балльно-рейтинговая система успеваемости студентов [4], реализованная в технологии ведения и мониторинга электронных журналов – это еще одна предпосылка для успешного перехода к электронному обучению.

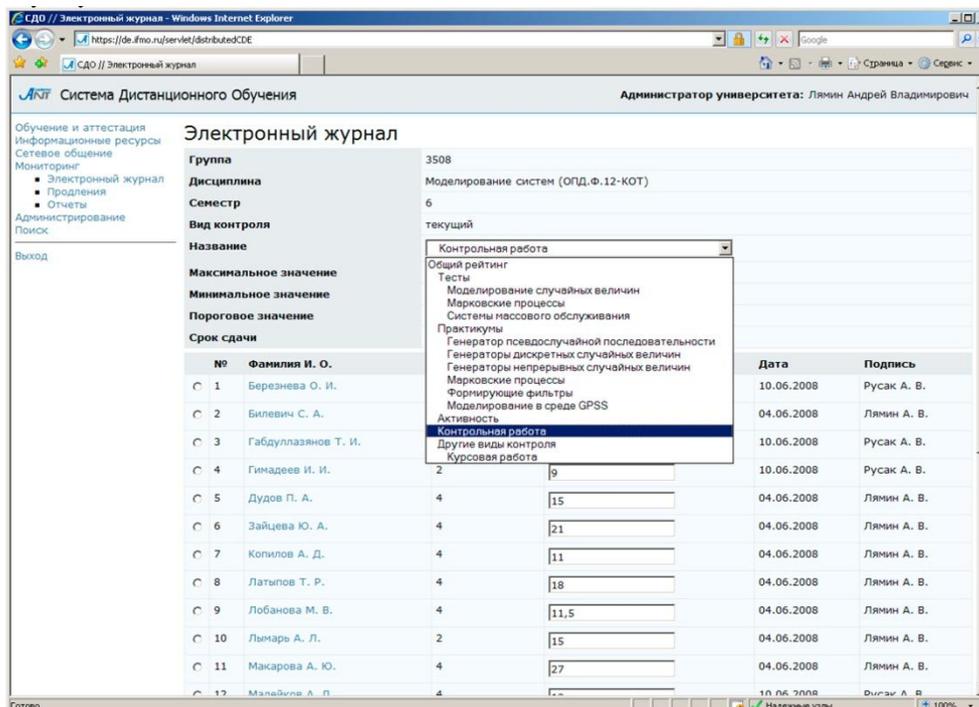


Рис. 3. Пример электронного журнала успеваемости студентов

4. **Существенные материальные ресурсы, направляемые на развитие электронного обучения.** На рис. 4 приведена динамика мирового рынка онлайн-обучения [5] с прогнозом до 2017 года.

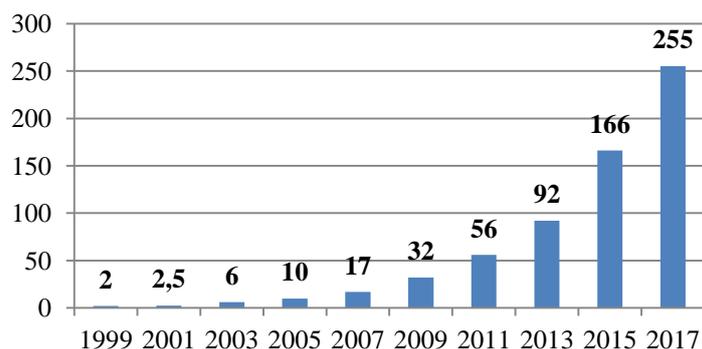


Рис. 4. Динамика мирового рынка онлайн-обучения 1999–2017 г.г. (прогноз), млрд долл.

Методика разработки электронного курса

Электронный курс должен обеспечивать достижение запланированных результатов обучения [6]. Его разработка должна отвечать следующим принципам:

- использование технологии MOOC и требований международного стандарта инженерного образования CDIO;
- результативность и междисциплинарность курса (планирование конкретных результатов освоения компетенций ФГОС ВПО в дисциплине (модуле) основной образовательной программы (ООП), структурирование на их основе содержания и характеристик учебной нагрузки курса);
- автоматический контроль запланированных результатов освоения компетенций (планирование и разработка оценочных средств курса, обеспечивающих контроль умений применять знания на практике в установленные сроки сдачи);
- автоматическое управление целостным образовательным процессом на основе отобранного структурированного содержания курса (весь учебный материал должен быть поделен на небольшие легкие для усвоения законченные фрагменты, завершающиеся автоматическим контролем и оценкой уровня сформированности запланированных результатов обучения);
- прозрачность критериев оценивания результатов обучения в баллах, установленных в соответствии с трудоемкостью их формирования;
- повышение мотивации к обучению за счет использования различных игровых механик;
- совместное обучение за счет сетевых технологий общения и консультирования самими обучаемыми в процессе изучения курса;
- доступность электронных курсов для любого желающего средствами базовых интернет-технологий, в том числе с мобильных устройств.

Методика разработки электронного курса состоит из следующих этапов.

1. Планирование ожидаемых результатов обучения курса и отбор из избыточного содержания предметной области обучения материалов для их формирования.
2. Структурирование отобранного содержания в соответствии с вышеизложенными принципами, разработка сценария обучения.
3. Разработка плана испытаний и оценочных средств для контроля результатов обучения.
4. Реализация курса, в том числе включение игровых механик и технологий сетевого совместного обучения.

Все виды контроля результатов обучения [7] в курсе должны быть автоматическими. Текущий контроль должен предусматривать проверку знаниевого и функционального компонентов формируемых компетенций – проверку усвоения знаний после очередного фрагмента лекций в виде небольшого практического упражнения, проверку сформированности навыков в виде практической работы на виртуальном стенде, тренажере и т.п. Проверку мотивационно-личностного компонента формируемых компетенций в ходе текущего контроля необходимо планировать и оценивать как проявления ответственности (например, прохождение в срок практических упражнений, тестов, виртуальных лабораторных работ и т.п.), а также активности в сетевом обсуждении проблем прохождения курса (например, заданные вопросы и (или) разъяснения по ним, отмеченные другими обучаемыми как лучшие и своевременные).

Рубежный контроль в электронном курсе проводится после завершения законченной по отношению к запланированным результатам обучения части электронного курса и включает в себя задания, выполнение которых позволяет проверить способность применять сформированные знания, умения и навыки для решения практических задач, в том числе междисциплинарных.

Итоговый контроль должен содержать задания, подтверждающих сформированность ожидаемых результатов обучения в курсе.

Реализация электронного онлайн-курса

Каждый электронный курс является авторским, т.е. у него всегда должен быть автор из числа ведущих преподавателей вуза, отвечающий за отбор и структурирование минимального и достаточного содержания, обеспечивающего достижение и контроль ожидаемых результатов обучения. Его использование на практике должно быть синхронизировано для повышения эффективности игровых технологий (организация соревнований, публикация топ-рейтингов и открытых профилей обучаемых и т.п.), для сетевого общения обучаемых в потоке (обмен опытом прохождения курса, совместное обучение в потоке). Несомненно, что такой курс в процессе его использования студентами будет «прирастать» новыми ресурсами, отобранными и рекомендованными самими студентами в процессе совместного обучения. Для автоматизации поиска и связей дополнительных ресурсов с элементами электронного курса перспективным и многообещающим является онтологический подход [8–11].

В зависимости от предметной области обучения и планируемых результатов обучения электронные курсы можно разбить на два вида:

1. электронные курсы для наукоемких предметных областей обучения, при реализации которых необходимы видео-лекции с различными интерактивными элементами для закрепления базовых знаний;
2. технологические электронные курсы, при реализации которых нужна дозированная подача теоретического материала для формирования запланированных практических навыков использования данной технологии, оборудования и т.п. (так называемые Fab lab, от англ. fabrication laboratory).

Рассмотрим далее некоторые приемы в реализации электронных курсов, апробированные в 2013 году в НИУ ИТМО для подачи теоретического материала. На рис. 5 приведены фрагменты видео-лекции из курса «Волновая оптика» (автор – Стафеев С.К.) с иллюстрациями в виде анимированных рисунков, которые по ходу лекции могут появляться на экране, в том числе и как дополненная реальность (при желании обучаемого). На рис. 6 приведен фрагмент из курса «Теория графов» (автор – Лисицына Л.С.), который содержит описание изучаемого алгоритма и пошаговую анимацию его применения для конкретного примера. На рис. 7 представлен интерфейс практического упражнения технологического электронного курса «Разработка веб-интерфейсов на основе HTML и CSS» [12] (автор – Першин А.А.). В верхней части левой колонки расположен блок с теорией, необходимой для выполнения упражнения. Внизу левой колонки расположен блок с перечнем заданий, которые нужно выполнить в упражнении. В центральной колонке расположены онлайн-редакторы кода, которые по функциональности и внешнему виду очень похожи на привычные «офф-лайновые» приложения, используемые разработчиками веб-интерфейсов в своей профессиональной работе. В правой колонке расположен визуализатор (так называемый «мини-браузер»), в котором отображается результат выполнения кода. При выполнении упражнения обучаемый изменяет исходный код таким образом, чтобы выполнить все требования своего задания. Например, в упражнении (рис. 7) необходимо расположить фоны определенным образом, меняя код CSS. Интерфейс является интерактивным, реагирует на изменение кода в режиме реального времени без перезагрузки страницы. Обновление полученного результата в визуализаторе, а также проверка статуса выполнения задач также происходят в режиме реального времени. Возможность на одном экране видеть и исходный код, и полученный результат его выполнения делает каждое практическое упражнение курса очень наглядным.

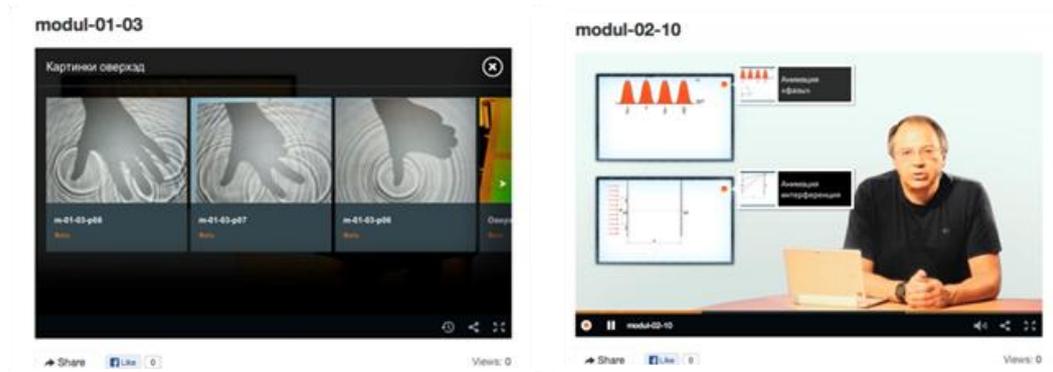


Рис. 5. Фрагменты видео-лекций из электронного курса «Волновая оптика»

Алгоритм Прима для поиска МОД графа

Пример Ход решения Ответ

1			4		6	4		4
2			4					10
3	4	4						
4							6	10
5	6						4	
6	4						4	
7		10		6	4	4		
8	4			10				

Дана матрица весов W связанного взвешенного графа $G(X,U)$ с $n=8$ (рис. сверху). Вес каждого ребра графа определен как расстояние между вершинами его концов в **линейной (ортогональной) метрике**. Для наглядности бесконечно большие веса отсутствующих ребер графа в матрице не указаны (пустые клетки).

Введем следующие обозначения:
 g – индекс вершины графа $G(X,U)$, выбранной для построения очередного поддерева в цепочке,
 S – суммарная стоимость ребер очередного дерева в цепочке. На этапе **построения поддерева T** выполняются следующие действия:

- $g=1$
- $S=0$

На этапе построения **цепочки поддеревьев МОД** в цикле от $i=1$ до $(n-1)$ строится множество ребер МОД – $U^* = \{u_1^*, u_2^*, \dots, u_{n-1}^*\}$. Тело цикла включает в себя следующие действия:

1. В матрице W вычеркиваем g -столбец и выделяем g -строку.
2. В матрице W просматриваем все выделенные строки сверху вниз и слева направо и выбираем

Рис. 6. Фрагмент лекций из электронного курса «Теория графов»

Свойство background-position [4/16]

Свойство `background-position` управляет расположением фонового изображения. Значение свойства состоит из двух частей, разделенных пробелом: `x y`.

- `x` задает расположение по горизонтали, а `y` по вертикали.
- В качестве значения `x` можно использовать ключевые слова `left`, `center`, `right`, значения в процентах и в пикселях.
- В качестве значения `y` можно использовать ключевые слова `top`, `center`, `bottom`, значения в

Задание

Задать блокам расположение фона:

1. 50% 50% — 1 блок.
2. 100% 0 — 2 блок.
3. 0 100% — 3 блок.
4. 100% 100% — 4 блок.

Цели задания:

1. 1 блок
2. 2 блок
3. 3 блок
4. 4 блок

[Что это за кнопка?](#)

```

1 <!DOCTYPE html>
2 <html>
3 <head>
4 <title>Свойство background-position</title>
5 <meta charset="utf-8">
6 </head>
7 <body>
8 <div class="block block1"></div>
9 <div class="block block2"></div>
10 <div class="block block3"></div>
11 <div class="block block4"></div>
12 </body>
13 </html>

```

```

3 padding: 0;
4 padding-top: 20px;
5 margin: 0;
6 text-align: center;
7 }
8 .block{
9 display: inline-block;
10 width: 128px;
11 height: 128px;
12 margin: 5px;
13 border-radius: 5px;
14 box-shadow: 1px 1px 3px #999;
15 }
16 .block1{
17 background-color: #2ecc71;
18 background-image: url("../assets/course13/cat_box.png");
19 background-repeat: no-repeat;
20 background-position: 50% 50%;
21 }

```

В ноябре HTML Academy запускает новый формат курсов: **месячный онлайн-интенсив**. Записаться и зарезервировать место можно уже сейчас.

Рис. 7. Фрагмент технологического электронного курса «Разработка веб-интерфейсов на основе HTML и CSS»

Заключение

В настоящее время уже созданы предпосылки для перехода вузов России от традиционного дистанционного обучения к электронному обучению на основе открытых онлайн-курсов. Практический опыт создания и использования в учебном процессе электронных ресурсов и онлайн-курсов, систем для автоматизации управления и контроля результатов обучения, технологий сетевого общения является основой для решения следующих актуальных задач развития массового открытого образования в вузах: формирование электронных онлайн-курсов для реализации совместных образовательных программ; дальнейшее развитие внутривузовской системы электронного обучения и аттестации; реализация внешних имиджевых (репутационных) проектов по созданию авторских открытых онлайн-курсов; сотрудничество с внешними держателями онлайн-курсов и открытых медиа-коллекций.

Литература

1. Bohle S. Librarians and the Era of the MOOC [Электронный ресурс]. Режим доступа: http://www.scilogs.com/scientific_and_medical_libraries/librarians-and-the-era-of-the-mooc/, свободный. Яз. англ. (дата обращения 01.12.2013).
2. Булгакова Н. Многоточие в сети. Вузы СНГ будут вместе развивать дополнительное образование [Электронный ресурс]. Режим доступа: http://www.poisknews.ru/php/FILES/core/contentfile/contentfile/3b/ad/97/poisk_20131025_017.pdf, свободный. Яз. рус. (дата обращения 01.12.2013).
3. Васильев В.Н., Лямин А.В., Чежин М.С. Система дистанционного обучения второго поколения // Научно-технический вестник СПбГУ ИТМО. 2007. № 11 (45). С. 148–157.
4. Фомина Н.Н., Кузьмина О.В., Лямин А.В. Гуманитарное образование в условиях балльно-рейтинговой системы // Информационные технологии в гуманитарном образовании. Материалы II Международной научно-практической конференции. Пятигорск: ПГЛУ, 2009. С. 492–498.
5. Морозов И. АйТи-трансформация: как технологии позволяют обучать миллионы людей одновременно [Электронный ресурс]. Режим доступа: <http://www.forbes.ru/reklama/246648-aiti-transformatsiya-kak-tekhnologii-rozvoluyayut-obuchat-milliony-lyudei-odnovremenno>, свободный. Яз. рус. (дата обращения 30.11.2013).
6. 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.
7. Васильев В.Н., Лисицына Л.С. Планирование и оценивание ожидаемых результатов освоения компетенций ФГОС ВПО // Научно-технический вестник СПбГУ ИТМО. 2013. № 2 (84). С. 142–148.
8. Боярский К.К., Катков Ю.В., Муромцев Д.И., Ольшевская А.В., Стафеев С.К., Яговкин В.И. Комплексная визуализация предметной онтологии на основе взаимосвязанных конструкций // Компьютерные инструменты в образовании. 2012. № 5. С. 38–45.
9. Гаврилова Т.А., Муромцев Д.И. Интеллектуальные технологии в менеджменте: инструменты и системы: Учебн. пособие. 2-е изд. СПб: Высшая школа менеджмента, 2008. 488 с.
10. Муромцев Д.И., Злобин А.Н., Катков Ю.В., Починок И.Н. Разработка базы знаний по оптике для образовательных веб-приложений // Искусственный интеллект и принятие решений. 2011. № 3. С. 3–10.
11. Муромцев Д.И., Варгин Г.В., Семерханов И.А. Применение онтологий в системе управления интеллектуальными ресурсами // Научно-технический вестник СПбГУ ИТМО. 2011. № 2 (72). С. 170.
12. Лисицына Л.С. Першин А.А. Электронный курс для обучения технологиям разработки веб-интерфейсов на основе HTML и CSS // Сборник материалов I Международной научно-практической конференции «Современные проблемы компьютерных наук». Пенза: Изд-во ПГУ, 2013. С. 83–84.

- Васильев Владимир Николаевич** – доктор технических наук, профессор, член-корреспондент РАН, ректор, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, vasilev@mail.ifmo.ru
- Стафеев Сергей Константинович** – доктор технических наук, профессор, декан, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, stafeevs@yahoo.com
- Лисицына Любовь Сергеевна** – доктор технических наук, профессор, зав. кафедрой, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, lisizina@mail.ifmo.ru
- Ольшевская Анастасия Владимировна** – начальник отдела лицензирования и аккредитации, аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, olshevskaya@mail.ifmo.ru
- Vladimir Vasilev** – D.Sc., Professor, RAS corresponding member, Rector, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, vasilev@mail.ifmo.ru
- Sergei Stafeev** – D.Sc., Professor, Dean, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, stafeevs@yahoo.com
- Luybov' Lisitsyna** – D.Sc., Professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, lisizina@mail.ifmo.ru
- Anastasiya Ol'shevskaya** – Head of Licensing and Accreditation Department, postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, olshevskaya@mail.ifmo.ru

КРАТКИЕ СООБЩЕНИЯ BRIEF PAPERS

УДК 535.21; 681.7.03

ОБРАТИМАЯ ФОТОДЕСТРУКЦИЯ НАНОЧАСТИЦ СЕРЕБРА В ФОТО-ТЕРМО-РЕФРАКТИВНЫХ СТЕКЛАХ

Д.А. Игнатьев^a, А.И. Игнатьев^a, Н.В. Никоноров^a, Д.С. Стародубов^b^a Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, ignatiev_d_a@mail.ru^b Университет Южной Калифорнии, Лос-Анжелес, США, dstarodubov@gmail.ru

Исследован процесс фотодеструкции наночастиц серебра в фото-термо-рефрактивных стеклах. Показано, что при облучении импульсным лазером в фото-термо-рефрактивных стеклах происходит фотофрагментация и фотоионизация наночастиц серебра. Последующая термообработка приводит к росту наночастиц серебра, что говорит об обратимости процесса фотодеструкции наночастиц серебра.

Ключевые слова: фотодеструкция, фото-термо-рефрактивные стекла, серебряные наночастицы.

REVERSIBLE PHOTO DESTRUCTION OF SILVER NANOPARTICLES IN PHOTO-THERMO-REFRACTIVE GLASS

D. Ignatiev^c, A. Ignatiev^c, N. Nikonorov^c, D. Starodubov^d^c Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, ignatiev_d_a@mail.ru^d University of Southern California, CA, USA, dstarodubov@gmail.ru

The paper deals with research of silver nanoparticles photo destruction process in photo-thermo-refractive glass. It is shown, that photo fragmentation and photo ionization of silver nanoparticles in photo-thermo-refractive glass takes place under pulsed laser radiation. Additional thermal treatment results in the growth of silver nanoparticles, which indicates photo destruction process reversibility for silver nanoparticles.

Keywords: photo destruction, photo-thermo-refractive glass, silver nanoparticles.

Процесс исследования обратимости фотодеструкции наночастиц серебра (НЧС) в фото-термо-рефрактивных (ФТР) стеклах состоит из трех этапов. На первом этапе образец ФТР стекла (рисунок, кривая 1) облучался УФ излучением ртутной лампы с последующей термообработкой (ТО) (при 530°C в течение 10 ч). В результате этого происходил рост НЧС с возникновением полосы поглощения с максимумом 450 нм (рисунок, кривая 2).

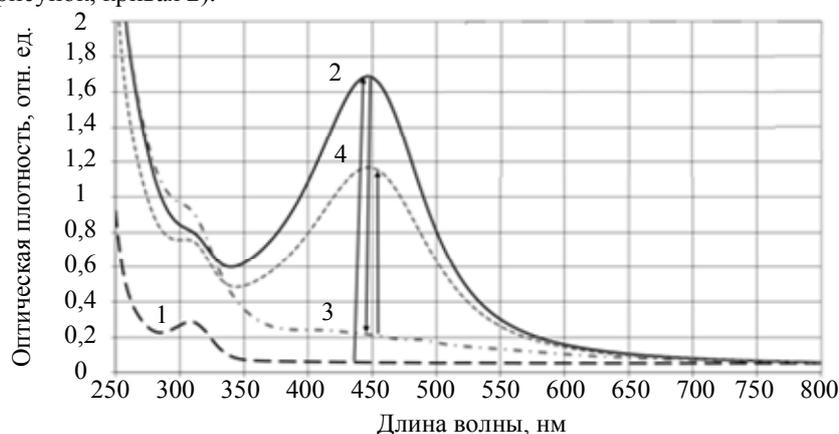


Рисунок. Спектры поглощения ФТР стекла при разных видах обработки

На втором этапе производилось «обесцвечивание» образца под действием импульсного лазерного излучения с длиной волны 532 нм (энергия импульса составляла 68 мДж/см², длительность импульса – 13 нс, частота следования импульсов – 10 Гц). В результате этого процесса полоса поглощения уменьшалась, вплоть до ее полного исчезновения (рисунок, кривая 3). Механизмы этого эффекта были предложены в работе [Л]. На третьем этапе образец подвергался дополнительной ТО, в результате чего наблюдалось возникновение полосы поглощения НЧС (рисунок, кривая 4), что говорит об обратимости фотодеструкции НЧС.

Схематично процесс обесцвечивания можно выразить следующим образом. При облучении импульсным лазерным излучением происходят:

1. фотофрагментация ($Ag_n^0 + h\nu = mAg_x^0 + kAg_0^0$);
2. фотоионизация, включающая генерацию фотоэлектронов ($Ag_n^0 + h\nu = Ag^+ + e^- + Ag_{n-1}^0$) и захват фотоэлектронов сурьмой ($e^- + Sb^{5+} \rightarrow [Sb^{5+}]^-$).

Этот факт вполне согласуется с предложенным нами механизмом: при вторичной термообработке происходит «сброс» электрона от сурьмяного комплекса, захват его ионом серебра с образованием нейтрального атома и присоединение атомов к оставшимся фрагментам (мелким наночастицам, необладающим плазмонным резонансом), что приводит вновь к росту наночастиц. Однако из-за «потерь» электронов количество НЧС несколько меньше, чем в первоначальном облученном состоянии. Схематично процесс образования НЧС при повторной ТО можно выразить следующим образом:

1. «сброс» электрона с сурьмы ($[Sb^{5+}]^- + kT \rightarrow e^- + Sb^{5+}$);
2. захват освободившихся термоэлектронов ($e^- + Ag^+ \rightarrow Ago$) и 3) рост НЧС ($Ag_n0 + kAg0 = Ag_n + k0$).

[Л]. Игнатъев Д.А., Игнатъев А.И., Никоноров Н.В. Фотодеструкция наночастиц серебра в фото-термо-рефрактивных стеклах // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 3 (85). С. 158–159.

- Игнатъев Дмитрий Александрович** – инженер, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, ignatiev_d_a@mail.ru
- Игнатъев Александр Иванович** – зав. лабораторией, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, ignatiev@oi.ifmo.ru
- Никоноров Николай Валентинович** – доктор физ.-мат. наук, профессор, зав. кафедрой, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Nikonorov@oi.ifmo.ru
- Стародубов Дмитрий Сергеевич** – кандидат физ.-мат. наук, научный сотрудник, Университет Южной Калифорнии, Лос-Анджелес, США, dstarodubov@gmail.ru
- Dmitry Ignatiev** – engineer, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, ignatiev_d_a@mail.ru
- Alexander Ignatiev** – Head of laboratory, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, ignatiev@oi.ifmo.ru
- Nicolai Nikonorov** – D.Sc., Professor, Department head, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Nikonorov@oi.ifmo.ru
- Dmitry Starodubov** – PhD, research scientist, University of Southern California, CA, USA, dstarodubov@gmail.ru

УДК 535.3+519.642.7

УРАВНЕНИЯ ПЕРЕНОСА ИЗЛУЧЕНИЯ В ИНФРАКРАСНОЙ ТОМОГРАФИИ В СЛУЧАЕ АКТИВНО-ПАССИВНОЙ ДИАГНОСТИКИ И ВЕЕРНОГО СКАНИРОВАНИЯ¹

А.А. Макарова^а

^а Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, alena.etalon@gmail.com

Сформулирована схема веерного сканирования горячего газа в задаче инфракрасной томографии. Используются два режима диагностики: активный (ON) – с включенным источником, пассивный (OFF) – без источника. Выведены два интегральных уравнения относительно коэффициента абсорбции k и функции Планка B среды (по которой можно рассчитать температурный профиль среды T).

Ключевые слова: ИК томография, интегральные уравнения переноса излучения, активный и пассивный режимы диагностики, веерное сканирование, коэффициент абсорбции, температурный профиль.

EQUATIONS OF RADIATION TRANSFER IN INFRARED TOMOGRAPHY IN THE CASE OF ACTIVE-PASSIVE DIAGNOSIS AND SWEEPING SCANNING²

A. Makarova^б

^б Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, alena.etalon@gmail.com

Sweeping scanning scheme of a hot gas in the task of infrared tomography is formulated. Two diagnosis regimes are used: the active one (ON) – with included source and the passive one (OFF) – without it. Two integral equations are deduced concerning the absorption coefficient k and the Planck function B of a medium (by which it is possible to calculate the temperature profile of a medium T).

Keywords: IR tomography, integral equations of radiation transfer, active and passive diagnosis regimes, sweeping scanning, absorption coefficient, temperature profile.

¹ Работа выполнена при поддержке РФФИ (грант № 13-08-00442).

² The work was done with support from the Russian Foundation for Basic Research (grant № 13-08-00442)

В дополнение к обзорной статье [Л] рассмотрим случай веерного сканирования некоторого z -сечения среды в задаче инфракрасной (ИК) томографии. На рисунке представлено два варианта такого сканирования. На рисунке, а, несколько сканеров посылают в направлении одного детектора лучи с интенсивностью I_0 , и детектор измеряет интенсивности $I_R(\theta)$ с включенным источником (активный режим (ON)), а также интенсивности $I_g(\theta)$ без источника (пассивный режим (OFF)). На рисунке, б, один сканер посылает в направлении нескольких детекторов лучи с интенсивностью I_0 , и детекторы фиксируют интенсивности $I_R(\theta)$, а также $I_g(\theta)$ без источника, где θ – угол сканирования.

Математическое описание обоих вариантов одинаковое. Рассмотрим для определенности схему, представленную на рисунке, б. В режиме ON интенсивность на детекторе в функции θ запишется как

$$I_R(\theta) = B(T_0) \exp\left(-\int_0^{\xi_m} k(\theta, \xi) d\xi\right) + \exp\left(-\int_0^{\xi_m} k(\theta, \xi) d\xi\right) \cdot \left\{ \int_0^{\xi_m} k(\theta, \xi) B(T_g(\theta, \xi)) \exp\left(\int_0^{\xi} k(\theta, \xi') d\xi'\right) d\xi \right\}, \quad (1)$$

а в режиме OFF

$$I_g(\theta) = \int_0^{\xi_m} k(\theta, \xi) B(T_g(\theta, \xi)) \exp\left(-\int_{\xi}^{\xi_m} k(\theta, \xi') d\xi'\right) d\xi, \quad (2)$$

где ξ – координата вдоль луча, $\xi_m = \xi_{\max}(\theta)$, $B(T_0)$ – функция Планка источника. Разность функций $I_R(\theta)$ и $I_g(\theta)$ равна

$$I_T(\theta) = I_R(\theta) - I_g(\theta) = B(T_0) \exp\left(-\int_0^{\xi_m} k(\theta, \xi) d\xi\right). \quad (3)$$

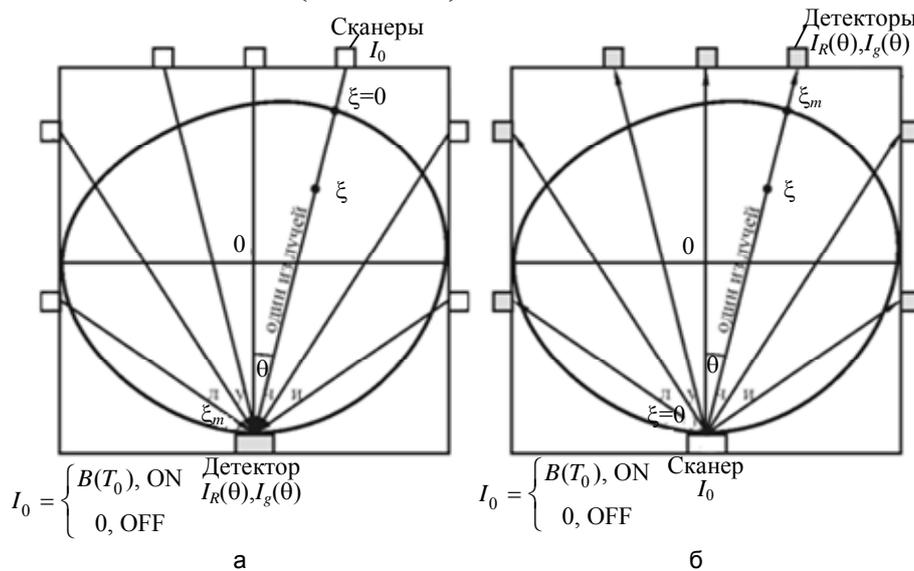


Рисунок. Два варианта веерного сканирования при некотором одном ракурсе φ : несколько сканеров генерируют излучение в направлении одного детектора (а); один сканер генерирует излучение в направлении нескольких детекторов (б)

Интегральные уравнения (1)–(3) позволяют определить коэффициент абсорбции $k(\theta, \xi)$ и функцию Планка среды $B(T_g(\theta, \xi))$, а также температурный профиль $T_g(\theta, \xi)$ при условии, что экспериментальные функции I_R , I_g и I_T получены для ряда ракурсов φ , т.е. получены $I_R(\theta, \varphi)$, $I_g(\theta, \varphi)$ и $I_T(\theta, \varphi)$.

Л. Сизиков В.С. Инфракрасная томография горячего газа: математическая модель активно-пассивной диагностики // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 6 (88). С. 3–17.

Макарова Елена Алексеевна – студент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, alena.etalon@gmail.com

Alena Makarova – student, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, alena.etalon@gmail.com

УДК 615.47, 616-71, 004. 383.3

**ПОЛУЧЕНИЕ СПЕКТРОВ ЭЛЕКТРОФИЗИОЛОГИЧЕСКИХ СИГНАЛОВ
В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ¹****Д.В. Толкович^a, Е.С. Андрианова^a**

^a Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, PYBY@inbox.ru

Рассматриваются проблемы получения спектров электрофизиологических сигналов, зависимости результата диагностики от зашумленности данных и квалификации эксперта. Разработан модуль оценки амплитудно-частотных спектров биосигналов в реальном времени. Приводится общая схема подключения измерительного модуля к спектроанализатору в составе многофункционального диагностического комплекса. Модуль реализован на базе миникомпьютера Cubieboard A20. Осуществляется непрерывное разложение сигналов (до 32 каналов) на фрагменты равной продолжительности и их независимая математическая обработка по каждому измерительному каналу. Предложенное устройство позволяет анализировать динамику изменений спектра в реальном времени.

Ключевые слова: спектральный анализ, энцефалограф, физиологические сигналы, нестационарные сигналы, диагностический комплекс.

REAL TIME REGISTRATION OF THE ELECTROPHYSIOLOGICAL SIGNALS SPECTRA²**D. Tolkovich^b, E. Andrianova^b**

^b Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, PYBY@inbox.ru

The paper deals with the issues of spectra acquisition for the electrophysiological signals, diagnostics result dependence on the noise level and expert's qualification. A real-time module for amplitude-frequency spectra estimation of electrophysiological signals is designed. A general diagram of connection between measuring module and spectrum analyzer (as a part of multifunctional diagnostic system) is shown. The module is realized on the base of Cubieboard A20 minicomputer. It provides continuous decomposition of signals (up to 32 channels) into fragments of equal length and their independent mathematical processing for each measuring channel. The suggested device makes it possible to analyze the real-time dynamics of spectrum changes.

Keywords: spectral analysis, EEG, physiological signals, non-stationary signals, diagnostic complex.

Для оценки физиологического состояния человека используют различные методы регистрации биологических сигналов. К электрофизиологическим сигналам относятся электроэнцефалограмма, электрокардиограмма, электромиограмма и т.д. Современные приборы способны сохранять и визуализировать зарегистрированные сигналы, обрабатывать их в режиме реального времени (накопление, усиление, фильтрация и т.д.). Ряд приборов обеспечивает возможность математической постобработки сигнала (в частности, спектральный анализ), однако результат ее использования существенно зависит от квалификации персонала. Основной причиной отсутствия в приборах данного класса визуализации спектров в режиме реального времени является сложная природа биосигнала, характеризующегося наличием помех (шумов) физиологического и инструментального происхождения, затрудняющих качественный анализ сигнала.

Спектральный анализ в физиологических исследованиях выполняется следующим образом: исследователь с учетом своего опыта выделяет фрагменты, имеющие видимые изменения зарегистрированного прибором сигнала, для последующего анализа. Этот процесс во многом зависит от мастерства эксперта, так как физиологический биосигнал имеет сложную форму и является композицией нескольких сигналов с различными амплитудно-частотными характеристиками [1, 2]. Спектральный анализ позволяет оценить изменение амплитудно-частотных характеристик физиологического сигнала при изменении внешних воздействий в различные моменты времени [3–5].

Особенностью физиологических сигналов является то, они могут иметь значительную продолжительность и располагаться в низкочастотной области спектра (от 0,01 Гц до 150 Гц). Это вызывает сложности при их регистрации и последующей математической обработке [1]. Данная проблема может быть решена различными методами, например, методом «вызванных потенциалов», в котором для последующего анализа исследователем выделяются фрагменты сигнала определенной длительности (например, 500 мс). Фрагменты сигнала, оказавшиеся за границей выделения, анализу не подвергаются.

Использование спектроанализатора для обработки электрофизиологических сигналов в режиме реального времени позволит наблюдать динамику изменения спектров, визуализировать полученные данные о спектре, а также сохранять и в дальнейшем воспроизводить их при последующих исследованиях. В настоящей работе предлагается использовать спектроанализатор в качестве компонента информационно-измерительного комплекса [6] для регистрации биосигналов при исследовании напряженной нервной деятельности. Модуль обработки спектров представляет собой миникомпьютер Cubieboard 20 со

¹ Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы по теме «Исследование нестационарных физиологических сигналов и изображений при интеллектуальной деятельности с помощью многофункционального диагностического измерительного комплекса».

² Done in the framework of Federal Target Program «Research and scientific-pedagogical personnel of innovative Russia» for 2009-2013 entitled «Study of non-stationary physiological signals and images at intellectual activity by the multifunction diagnostic measuring complex»

специализированным программным обеспечением. Выбранный компьютер реализован на открытой архитектуре ARM (Advanced RISC Machine – усовершенствованная RISC-машина) на базе двоядерного процессора a20 Cortex™-A7. К порту материнской платы подключена видеокарта Mali400 MP1 с поддержкой OpenGL ES 2.0/1.1 и аппаратным декодером H.264, также 1GB DDR3 480 МГц оперативной памяти. Использование ARM-процессора позволяет ускорить однопоточные вычисления благодаря увеличению «плотности» обрабатываемых данных. Используемый набор команд полезен для организации итерационных вычислительных процессов, в том числе при разложении сигналов в спектр.

Зарегистрированные электрофизиологические сигналы после оцифровки поступают на модуль спектроанализатора, где подвергаются математической обработке по разложению их в спектр; полученные результаты сохраняются в виде ряда серий на отдельном носителе и выводятся на специализированный монитор. Предложенная схема подключения спектроанализатора (рисунок) позволяет без модернизации используемого медицинского прибора перехватывать и обрабатывать данные независимо от алгоритмов, заложенных в регистрирующем приборе.

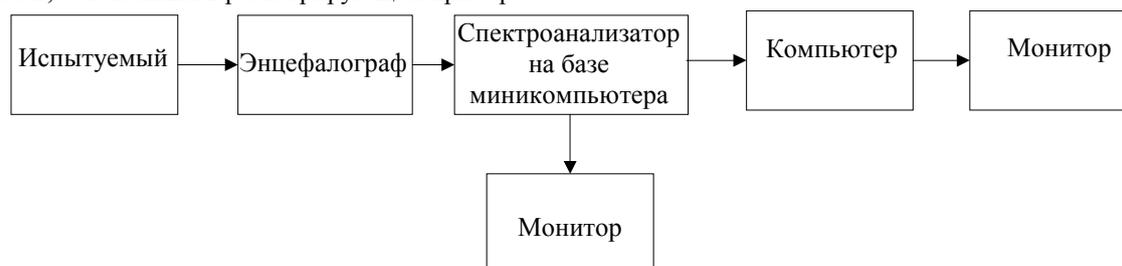


Рисунок. Блок-схема взаимодействия компонентов диагностического комплекса

В модуле спектроанализатора реализована возможность параллельной независимой спектральной обработки каждого канала в отдельности (от одного до 32 каналов). Спектроанализатор позволяет применять различные виды математических преобразований для выделения спектральных компонентов исследуемых сигналов [7]. Модуль может использоваться совместно с любыми электрофизиологическими приборами (миографы, кардиографы, энцефалографы с количеством каналов до 32), позволяя получить более полное представление о составляющих электрофизиологического сигнала по зарегистрированному спектру, следовательно, обеспечивая больше возможностей для анализа данных и медицинской диагностики.

Программное обеспечение спектроанализатора обеспечивает непрерывное разложение регистрируемого сигнала на фрагменты равной длительности и последующее применение к этим фрагментам математических преобразований. Длительность фрагмента выбирается до начала исследования и в процессе исследования не меняется. Дальнейшее усовершенствование устройства связано с реализацией записи результатов спектрального анализа в соответствии со стандартом медицинских данных DICOM 3.0.

Разработанный спектроанализатор в виде отдельного модуля входит в состав информационного измерительного комплекса для исследования нестационарных физиологических сигналов [6, 8], разработанного совместно сотрудниками кафедры ИТиКТ НИУ ИТМО и отдела экологической физиологии Научно-исследовательского института экспериментальной медицины Северо-Западного отделения Российской академии медицинских наук (НИИЭМ СЗО РАМН). В настоящее время спектроанализатор проходит испытания в лаборатории нейробиологии НИИЭМ СЗО РАМН.

1. Гнездицкий В.В. Вызванные потенциалы мозга в клинической практике. М.: МЕДпресс-информ, 2003. 264 с.
2. Рутман Э.М. Вызванные потенциалы в психологии и психофизиологии. М.: Оникс, 2012. 301 с.
3. Курова Н.С., Черемушкин Е.Л. Спектральные характеристики ЭЭГ при усложнении контекста когнитивной деятельности // Журнал высшей нервной деятельности им. И.П. Павлова. 2006. Т. 56. № 2. С. 211–218.
4. Нидеккер И.Г., Антонов А.А. Спектральный анализ длительных записей электроэнцефалограммы // Физиология человека. 2003. Т. 29. № 3. С. 129–135.
5. Павлыгина Р.А., Давыдов В.И., Сахаров Д.С., Тутушкина М.В., Прямоносова А.А. ЭЭГ при решении математических логических задач // Журнал высшей нервной деятельности им. И.П. Павлова. 2010. Т. 60. № 5. С. 534–542.
6. Марусина М.Я., Суворов Н.Б., Козаченко А.В., Толкович Д.В. Синхронизация физиологических сигналов интеллектуальной деятельности человека с помощью многофункционального измерительного комплекса // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 4 (86). С. 49–55.
7. Туровский Я.А., Кургалин С.Д., Максимов А.В. Выбор анализирующих вейвлетов для системы с параллельной обработкой биомедицинских данных // Вестник Воронежского государственного университета. Сер. Системный анализ и информационные технологии. 2011. № 2. С. 74–79.

8. Анодина-Андреевская Е.М., Божокин С.В., Полонский Ю.З., Суворов Н.Б., Марусина М.Я. Перспективные подходы к анализу информативности физиологических сигналов и медицинских изображений человека при интеллектуальной деятельности // Изв. вузов. Приборостроение. 2011. Т. 54. № 7. С. 27–35.

Толкович Дмитрий Владиславович – аспирант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, RYBY@inbox.ru

Андрянова Евгения Сергеевна – студент, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия, Zeneka44@gmail.com

Dmitry Tolkovich – postgraduate, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, RYBY@inbox.ru

Evgenia Andrianova – student, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia, Zeneka44@gmail.com

ПРАВИЛА ПРЕДСТАВЛЕНИЯ РУКОПИСЕЙ

Статьи принимаются в электронном виде (электронный адрес karmanov@mail.ifmo.ru), либо на DVD. В распечатанном виде с подписями авторов материалы представляются в редакцию по запросу после прохождения процедуры рецензирования.

Комплект документов должен включать:

- текст статьи с завершенными рисунками и таблицами;
- форму сведений о статье (рус., англ.);
- формы сведений об авторах (на каждого автора); допускается указание нескольких мест работы автора;
- файлы с рисунками к статье в оригинальном формате (предпочтительно JPEG) с максимальным разрешением; допускается представление цветных рисунков, если в черно-белом варианте теряется полезная информация.

Объем статьи, включая иллюстрации, таблицы и список литературы, не должен превышать 8 страниц машинописного текста. Оригинал-макет статьи выполняется в редакторе Word for Windows (версия не ниже 2003 года).

Основной текст набирается шрифтом Times New Roman, размер 12 pt обычный, межстрочный интервал одинарный, поля сверху 20 мм, остальные поля 25 мм. Абзацный отступ 10 мм. Аннотация (объем 150–250 слов) печатается шрифтом Times New Roman 10 pt обычный, межстрочный интервал одинарный, выравнивание по ширине страницы.

Название статьи печатается шрифтом Times New Roman 12 pt полужирный, межстрочный интервал одинарный.

Название подраздела печатается шрифтом Times New Roman 12 pt полужирный, межстрочный интервал одинарный, пустыми строками не отделяется. Формулы набираются в редакторе Microsoft Equation Ed. 3.0.

Параметры стиля:

текста – Times New Roman, прямой;

функций, переменных и цифр – шрифт латинских букв Times New Roman курсив, греческих букв – Symbol прямой; русских букв, цифр, скобок – Times New Roman прямой. Обозначения функций (sin, cos, tg, ctg, min, max, exp, extr) – прямой шрифт. Те же правила написания применяются и в отношении верхних и нижних индексов: обычный – 10 pt, крупный индекс – 12 pt, мелкий индекс – 8 pt, крупный символ – 16 pt, мелкий символ – 12 pt.

матриц и векторов – Times New Roman, прямой, полужирный (знак вектора «стрелка» не ставится).

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 рекомендуется не менее 15 позиций для полнотекстовой статьи и не менее 8 – для краткого сообщения. Оформляется в соответствии с ГОСТ 7.0.5-2008. Нумерация позиций – согласно очередности ссылок в тексте.

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