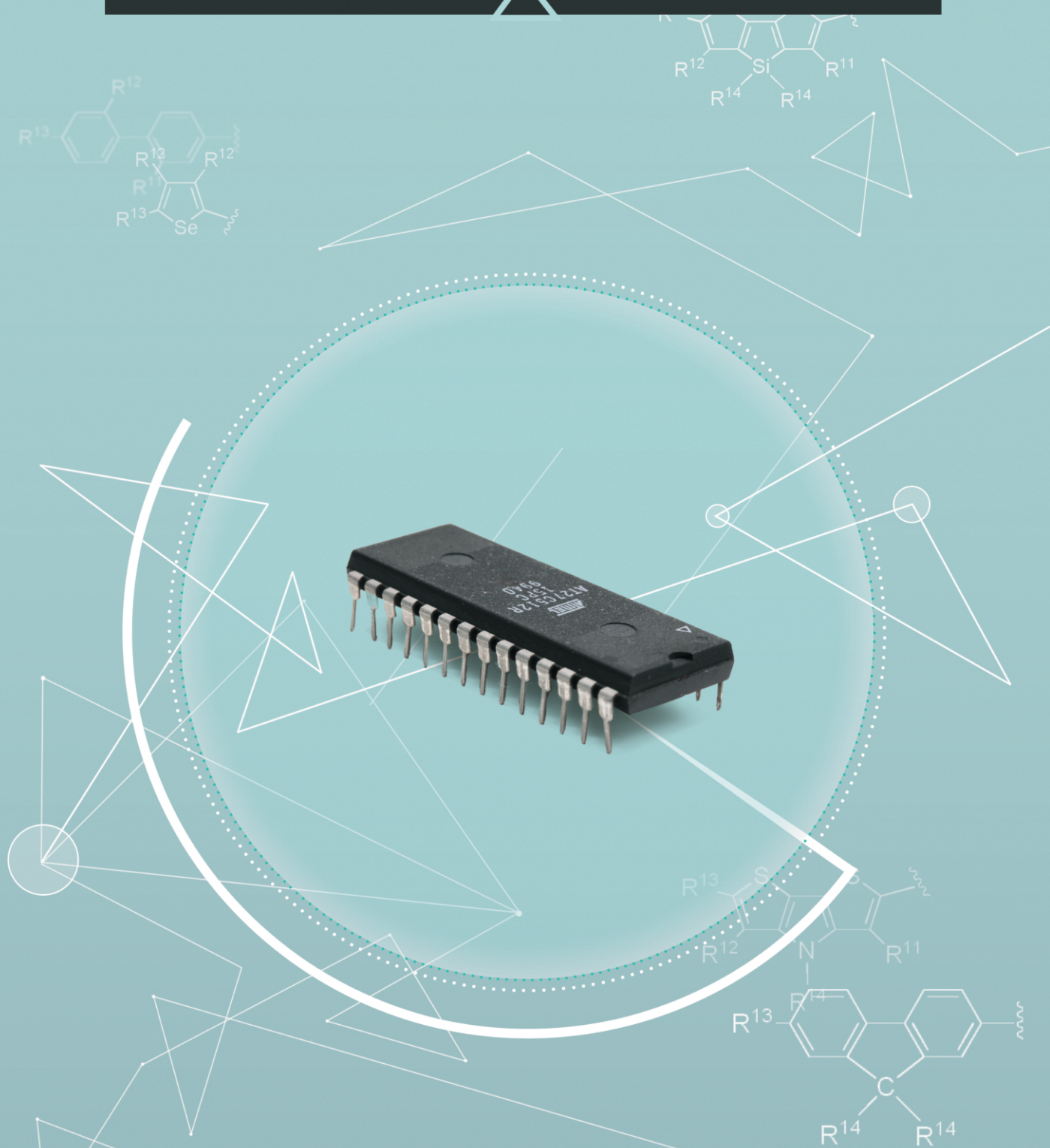


СБОРНИК РАБОТ СТУДЕНТОВ УНИВЕРСИТЕТА ИТМО,
ПОБЕДИТЕЛЕЙ КОНКУРСА ГРАНТОВ ПРАВИТЕЛЬСТВА САНКТ-ПЕТЕРБУРГА



Санкт-Петербург
2014

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ,
МЕХАНИКИ И ОПТИКИ

Сборник работ студентов Университета ИТМО,
победителей конкурса грантов
Правительства Санкт-Петербурга

Санкт-Петербург

2014 г.

Сборник работ студентов Университета ИТМО, победителей конкурса грантов правительства Санкт-Петербурга / Под. ред. В.О. Никифоров – СПб.: Университет ИТМО, 2014 – 230 с.

Сборник содержит работы студентов Университета ИТМО, удостоенных премии правительства Санкт-Петербурга, победителей конкурса грантов для студентов и аспирантов вузов, отраслевых и академических институтов, расположенных на территории Санкт-Петербурга. Научные отчеты по результатам выполнения грантов издаются с целью демонстрации достижений студентов университета, стимулирования их участия в научных исследованиях, повышения качества подготовки специалистов с высшим образованием и формирования резерва для кадров высшей квалификации.

ISBN 978-5-7577-0469-2

Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики, 2014

Авторы, 2014

ОГЛАВЛЕНИЕ

Планирование и стабилизация движений двуногих роботов <i>Д.Н. Базылев</i>	6
Мультиплексирование классических и квантовых каналов в оптоволоконных линиях связи <i>Д.Н. Вавулин</i>	15
Трудовой потенциал как фактор развития национальной экономики <i>О.В. Волкова</i>	21
Особенности проведения измерений органических и неорганических объектов методом атомно-силовой микроскопии в жидких средах <i>М.В. Жуков</i>	29
Математические модели для анализа напряженно-деформированного бесстыкового пути при статических и динамических нагрузках <i>Т.И. Зайцева</i> ..	39
Методы обработки сигналов электромиографии в системах управления реабилитационных устройств <i>К.А. Зименко</i>	46
Разработка алгоритма упрощения графа перекрытий при сборке геномных последовательностей <i>С.В. Казаков</i>	55
Моделирование и идентификация нелинейных систем с приложением к системам гидропривода <i>А.А. Лосенков</i>	66
Анализ и синтез регуляторов антропоморфного манипулятора <i>А.А. Маргун</i>	68
Применение метода Хартри-Фока в задаче хранения частиц <i>И.Ф. Мелихов</i>	78
Спектральная задача для Y-разветвленной цепочки слабо связанных шарообразных резонаторов <i>А.С. Мелихова (Аникевич)</i>	83
Расчет времени когерентности фемтосекундного спектрального суперконтинуума в кварцевом оптическом волокне и выявление оптимальных условий его генерации <i>М.В. Мельник</i>	89
Разработка схемных решений для регистрации высоких пространственных частот для безлинзовых методов формирования изображений <i>Т.Ю. Николаева</i>	100
Исследование и разработка принципов обработки данных в видеоинформационных системах цветового анализа минералов <i>Д.Б. Петухова</i>	107
Разработка проекта Глоссария по информационной безопасности для Парламентской Ассамблеи ОДКБ <i>В.Е. Подкидышева</i>	116
Разработка гибридного микрофлюидного чипа для полимеразной цепной реакции <i>Я.С. Посмитная</i>	128
Микропроекторные системы смешанной и дополненной реальности на плоских волноводных экранах <i>М.С. Рудакова</i>	143
Лазерно-индуцированная локальная модификация пористых стекол <i>М.М. Сергеев</i>	155

Влияние структуры, химического состава и способа получения пленок на электрохромные окна <i>А.О. Трофимов</i>	166
Применение методов решения задачи о выполнимости булевой формулы для построения управляющих конечных автоматов по сценариям работы <i>В.И. Ульянов</i>	177
Синтез системы управления роботом через Интернет <i>А.А. Хасанов</i>	185
Разработка методов генерации управляющих конечных автоматов на основе муравьиных алгоритмов <i>Д.С. Чивилихин</i>	192
Исследование и разработка методики настройки цветопередачи видеоинформационных систем цветового анализа <i>Д.Д. Шитов</i>	200
Разработка и обоснование требований к построению системы, реализующей защиту от несанкционированного доступа к создаваемым файловым объектам, основанной на исключении сущности «объект доступа» из разграничительной политики <i>К.А. Щеглов</i>	213



Награждение победителей конкурса грантов 2013 года для студентов и аспирантов вузов

В 2009 г. нашему вузу по итогам всероссийского конкурсного отбора программ развития университетов была присвоена категория национального исследовательского Университет (НИУ). Это, прежде всего, означает, что Университет одинаково эффективно осуществляет образовательную и научную деятельность на основе принципов интеграции науки и образования.

На протяжении многих лет Университет ИТМО участвует в конкурсах на получение грантов для студентов, аспирантов из вузов и академических институтов, расположенных на территории Санкт-Петербурга, проводимых Комитетом по науке и высшей школе при правительстве Санкт-Петербурга. Университет входит в лидирующую тройку среди вузов Санкт-Петербурга по количеству поданных и выигранных заявок. В 2013 году студентами нашего Университета было подано более 150 заявок, из них выиграно 28.

В настоящий сборник вошли работы студентов - победителей конкурса грантов 2013 г. для студентов и аспирантов вузов, отраслевых и академических институтов Санкт-Петербурга. Представленные работы демонстрируют высокий уровень и широкий диапазон научных исследований, проводимых студентами Университета ИТМО под руководством ведущих ученых вуза.

Итогом развития научной деятельности Университета ИТМО явилось включение нашего вуза в число 15 вузов страны, вошедших в программу повышения конкурентоспособности среди ведущих международных научных и образовательных центров.

Проректор, д.т.н., проф. В.О. Никифоров



Базылев Дмитрий Николаевич
Факультет компьютерных технологий и управления
Кафедра систем управления и информатики
E-mail: bazylevd@mail.ru

ПЛАНИРОВАНИЕ И СТАБИЛИЗАЦИЯ ДВИЖЕНИЙ ДВУНОГИХ РОБОТОВ

Д.Н. Базылев

Научный руководитель – профессор А.А. Бобцов

Цели работы – построение математических моделей двуногих шагающих роботов; планирование траекторий, соответствующих их устойчивым движениям; разработка систем управления шагающими роботами, а также апробация полученных алгоритмов управления на практике. Исследования проводились в области стабилизации двуногих роботов, стоящих в положении стоя на поверхности, угол наклона которой изменяется произвольным образом.

В ходе работы была определена кинематическая схема, соответствующая рассматриваемому устройству. В качестве модели выбран двуногий шагающий робот, собранный из набора BioloidPremiumKit, производимого корейской компанией Robotis. Общий вид сконструированного робота на базе этого робототехнического набора представлен на рис. 1. Он имеет восемь степеней свободы.



Рис. 1. Общий вид двуножного робота Bioloid

Набор BioloidPremiumKit включает в себя небольшие сервоприводы, которые называются Dynamixel и представляют собой отдельные самостоятельные модули. С их помощью могут быть собраны роботы различной конструкции, например колесные или шагающие. В набор входит микроконтроллер CM-510, специально разработанный на базе ATmega 2561. Помимо двигателей и микроконтроллера, BioloidKit включает в себя различные датчики и устройства (ИК-дальномер, двухосный гироскоп, ИК-приемник/передатчик, модуль ZigBee, пульт дистанционного управления и др.), а также ряд соединительных крепежей, болтов, гаек, которые и обеспечивают разнообразие моделей роботов.

Микроконтроллер CM-510 имеет порты для работы с датчиками и сервомоторами Dynamixel, порты проводной и беспроводной связи (рис. 2).

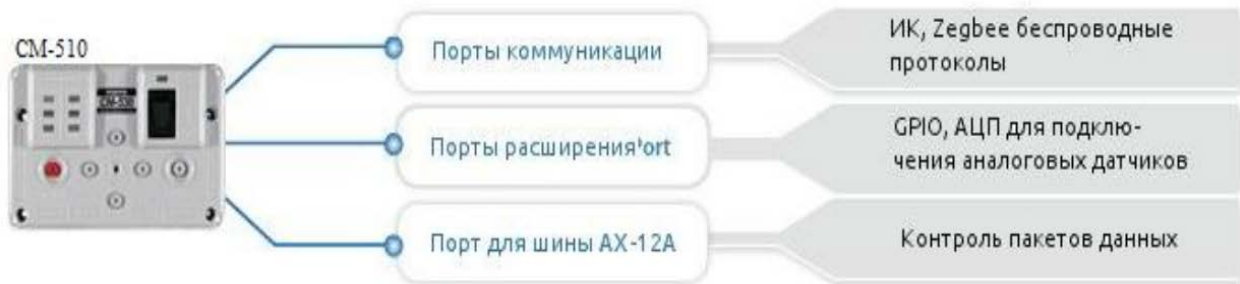


Рис. 2. Возможности микроконтроллера CM-510

Программирование робототехнического комплекса Bioid может осуществляться как на языке C/C++, так и на с-подобном языке RoboPlus, разработанном компанией Robotis. В случае написания программ на языке C все необходимые библиотеки для работы с микроконтроллером ATmega 2561 можно найти в AtmelStudio или AVRStudio. Для использования сервомоторов Dynamixel и организации последовательной связи необходимо подключить специальные библиотеки: serial.h и dynamixel.h. При использовании передачи данных по Zigbee-протоколу необходимо подключить библиотеку libzgbbee.h.

Карта портов микроконтроллера CM-510 и выполняемые портами функции представлены на рис. 3 и в Табл. 1.

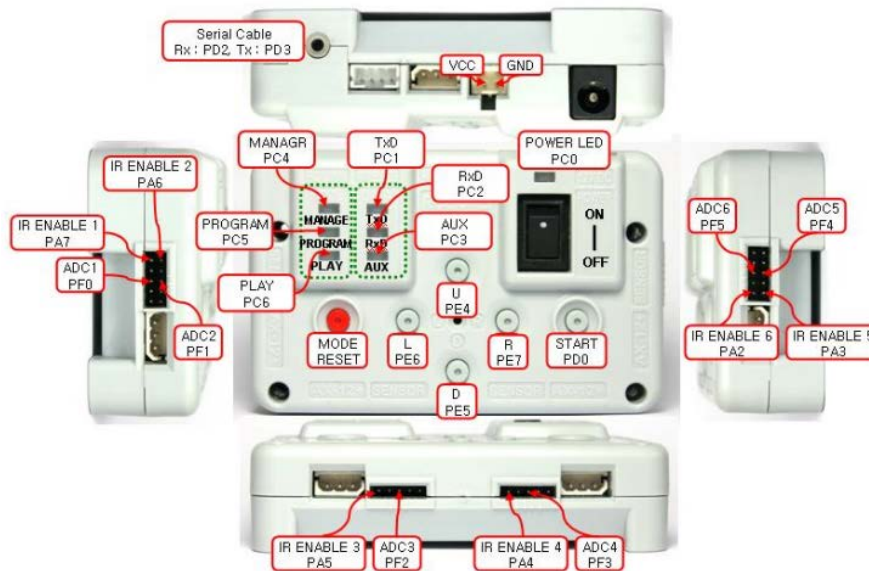


Рис. 3. Карта портов микроконтроллера CM-510

Таблица 1

Функции портов микроконтроллера CM-510

Название порта	Функция
PORTF1 ~ PORTF6	АЦП
PORTD0	Кнопка старта
PORTD1 ~ PORTD2	Запись (Tx) и чтение (Rx) данных
PORTA2 ~ PORTA7	Выходы (порт с 5 выводами (рисунок 2.19))
PORTC0 ~ PORTC6	Светодиоды (статус, питание)
PORTB5	Порт управления бипером
PORTE4 ~ PORTE6	Кнопки направления (U, D, L, R)
PORTD4 ~ PORTD6	Порты управления подключениями

Сервоприводы Dynamixel являются довольно мощным и удобным средством для построения робототехнических систем. Они имеют встроенные микроконтроллеры, используемые для управления, и целый ряд энкодеров; позволяют измерять температуру и угловое положение вала двигателя, скорость его вращения, напряжение, нагрузку на валу

двигателя и т. д. Все данные хранятся в памяти микроконтроллера, некоторые из них доступны только для чтения, другие – можно перезаписывать. Параметры встроенного регулятора могут быть изменены перезаписыванием значений соответствующих байтов в памяти. Управление двигателем может осуществляться путем задания желаемого углового положения вала двигателя, угловой скорости или вращающего момента. Также можно регулировать максимальный момент двигателя.

Ниже приведены технические характеристики серводвигателей Dynamixel.

Масса: 54,6 г.

Размеры: 32×50×40 мм.

Разрешение: 0,29°.

Передаточное отношение: 254:1.

Момент опрокидывания: 1,5 Нм (12 В, 1,5 А).

Скорость без нагрузки: 59 рад/мин (при 12 В).

Степень вращения: 1) 0–300°; 2) бесконечные вращения.

Рабочая температура: -5 – +70 °С.

Напряжение: 9–12 В (рекомендуемое напряжение: 11,1 В).

Управляющий сигнал: цифровой пакет.

Тип протокола: полудуплексный асинхронный последовательный протокол (8 бит, стоп-бит, без проверки четности).

ID: 254 ID (0–254).

Скорость передачи данных: 7343bps – 1Mbps.

Обратная связь: положение, температура, нагрузка, входное напряжение и т. д.

Материал: технический пластик.

На рис. 4 приведен пример подключения сервомоторов Dynamixel к CM-510 по порту шины AX-12A.

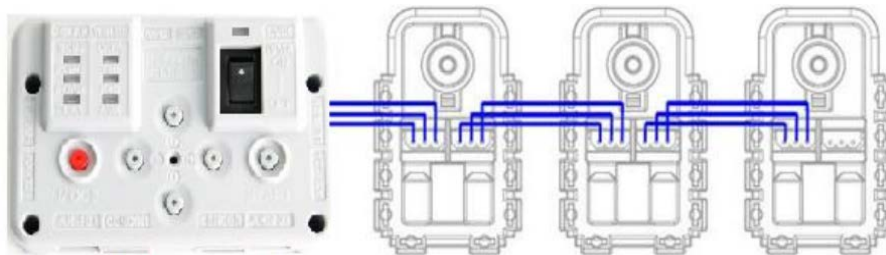


Рис. 4. Микроконтроллер CM-510 с подключенными сервоприводами Dynamixel

Более подробное описание набора BioloidPremiumKit, а также технические характеристики входящих в этот набор устройств можно найти на официальном сайте компании Robotis [1].

Для построения кинематической схемы была получена упрощенная модель рассматриваемого робота, приведенная на рис. 5. Учитывая то, что сервоприводы робота имеют встроенные регуляторы, задачу управления можно свести к определению желаемых траекторий робота и угла наклона поверхности.

Поскольку рис. 5 представляет собой вид сбоку, на нем проиллюстрирована модель только одной ноги робота. Параметры для второй – такие же. Модель состоит из четырех звеньев и трех вращательных сочленений. Углы q_1 , q_2 и q_3 могут быть измерены энкодерами серводвигателей, режим управления которых переключен в «joint». Массы звеньев обозначены как m_1 , m_2 , m_3 и M . Вращение звеньев происходит только в сочленениях робота, обозначенных как O_1 , O_2 и O_3 . Все остальные соединения считаются жесткими. Также было сделано допущение, что стопы робота не отрываются от наклонной поверхности, а при скольжении по ней его ускорение мало. Таким образом, модель имеет шесть степеней свободы.

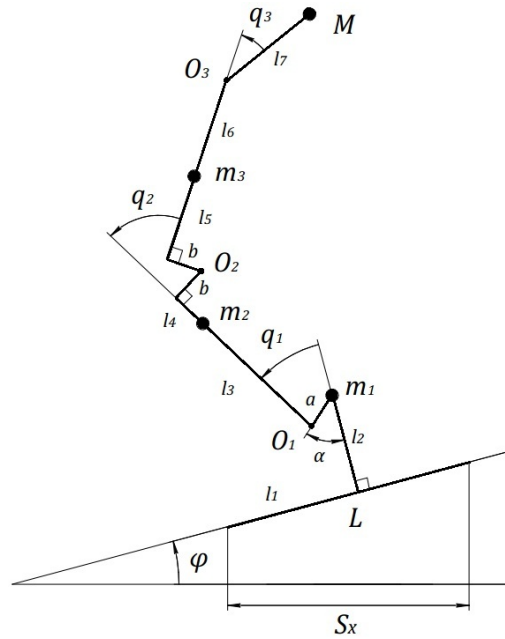


Рис. 5. Упрощенная плоская модель двуногого робота *Bioid*

В соответствии с обозначениями на рис. 5 были определены следующие параметры модели робота:

$$\begin{aligned} m_1 &= 165 \text{ г}, m_2 = 77,5 \text{ г}, m_3 = 29 \text{ г}, M = 322,3 \text{ г}, \\ L &= 100 \text{ мм}, l_1 = 54 \text{ мм}, l_2 = 40 \text{ мм}, l_3 = 66 \text{ мм}, l_4 = 9 \text{ мм}, \\ l_5 &= 35 \text{ мм}, l_6 = 40 \text{ мм}, l_7 = 42 \text{ мм}, \\ b &= 15 \text{ мм}, a = 10 \text{ мм}, \alpha = \text{const} = 37^\circ. \end{aligned}$$

Известно, что тело, находящееся на горизонтальной плоскости, начинает падать, когда положение его центра масс выходит за площадь опоры. В связи с этим цель управления роботом была сформулирована как поддержание координат центра масс над центром линии опоры S_x .

$$x_c \rightarrow \frac{1}{2} S_x, \quad (1)$$

где $x_c = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3 + M x_4}{m_1 + m_2 + m_3 + M}$, $S_x = L \cos \varphi$,

$$x_1 = l_1 \cos \varphi - l_2 \sin \varphi,$$

$$x_2 = x_1 - a \sin(\alpha - \varphi) - l_3 \sin(q_1 + \varphi),$$

$$x_3 = x_2 - l_4 \sin(q_1 + \varphi) + b \cos(q_1 + \varphi) - b \cos(q_2 - q_1 - \varphi) + l_5 \sin(q_2 - q_1 - \varphi),$$

$$x_4 = x_3 + l_6 \sin(q_2 - q_1 - \varphi) + l_7 \sin(q_3 + q_2 - q_1 - \varphi).$$

Если учесть, что угол наклона поверхности φ измерен или известен, то x_1 также является известным. Тогда уравнение (1) содержит три неизвестных величины $q = (q_1, q_2, q_3)^T$. Таким образом, имеется бесконечное число подходящих траекторий q_1^*, q_2^*, q_3^* и для того, чтобы получить определенные, необходимо дополнить уравнение (1) некоторыми связями между обобщенными координатами q_1, q_2 и q_3 .

Для сужения поиска желаемых траекторий робота q^* к уравнению (1) были добавлены следующие два уравнения связей

$$q_3 + q_2 - q_1 - 90^\circ = 0, \quad (2)$$

$$q_2 - k q_1 = 0, k < 1. \quad (3)$$

Уравнение (2) обеспечивает горизонтальное положение торса робота вне зависимости от угла наклона поверхности. Связь (3) делает его движения более похожими на человеческие

и говорит о том, что угловые перемещения в стопе должны быть большими по сравнению с перемещениями в колене.

Подставляя уравнения связей (2) и (3) в (1) получаем желаемые траектории q_1^* , q_2^* , q_3^* для каждого звена одной из ног робота в зависимости от угла наклона φ . Для второй ноги желаемые траектории те же.

Угол φ наклонной поверхности может быть измерен различными способами. Например, если установить на борту робота видеокамеру и снимать нарисованную на себе (сбоку от робота) вертикальную линию. По углу отклонения вертикали снимка видеокамеры от нарисованной вертикали можно определить угол отклонения робота и наклон поверхности.

Одним из основных отличий данной работы от публикаций [1, 2] является повышение качества измерения угла наклонной поверхности. В указанных работах угол наклона измерялся косвенным методом, исходя из ошибок углов поворота серводвигателей и гироскопа. В данном проекте угол φ наклонной поверхности измеряется на основе данных с двухосного гироскопа компании Robotis и трехосного акселерометра, производимого фирмой Pololu. Общий вид гироскопа и акселерометра приведены на рис. 6а и 6б соответственно, а в Табл. 2 и 3 – их технические характеристики.

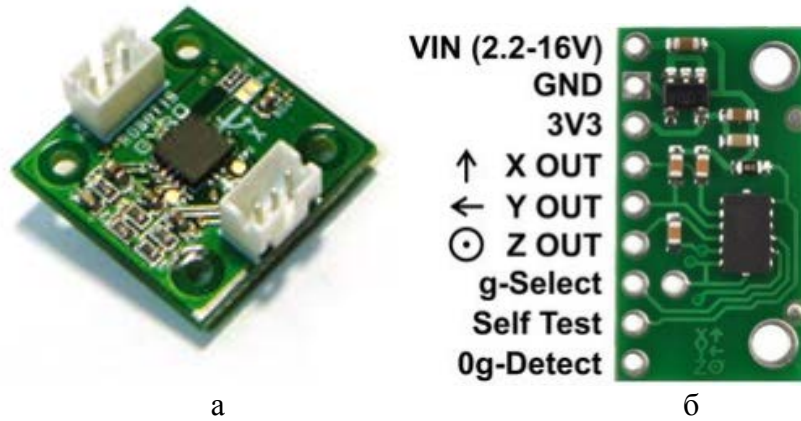


Рис. 6. Общий вид: а – двухосного гироскопа Robotis; б – трехосного акселерометра PololuMMA7341L

Таблица 2

Технические характеристики гироскопа фирмы Robotis

Параметр	Значение
Масса	2,8 г
Размеры	23×23×10 мм
Рабочая температура	-40–85 °С
Диапазон измерения угловой скорости	-300–300 °/с
Пропускная способность	140 Гц
Чувствительность	3,33 мВ/град/с
Напряжение питания	4,5–5,5 В

Таблица 3

Технические характеристики акселерометра MMA7341L

Параметр	Значение
Масса	0,7 г
Размеры	1,2×2,3×0,23 см
Рабочая температура	-40–85 °С
Диапазон измерения	±3 g или ±11 g
Пропускная способность	400 Гц
Чувствительность	440 мВ/g или 118 мВ/g
Напряжение питания	2,2–16 В

Аналоговые сигналы гироскопа и акселерометра измерялись с помощью встроенного в микроконтроллер ATmega 2561 аналого-цифрового конвертера. Для получения значений угла наклона с гироскопа и акселерометра использовались соответственно выражения (4) и (5).

$$\varphi_{gyroX} = \varphi_{gyroX} + Gx \cdot dt, \quad (4)$$

где φ_{gyroX} – угол, а Gx – угловая скорость вокруг оси X, измеренные с помощью гироскопа, dt – интервал времени между измерениями.

$$\varphi_{accX} = \arctg \frac{A_x}{\sqrt{A_y^2 + A_z^2}}, \quad (5)$$

где φ_{accX} – угол вокруг оси X, измеренный с помощью акселерометра, A_x , A_y , A_z – проекции вектора ускорения свободного падения на оси X, Y, Z соответственно.

Для корректировки значений угла наклона, полученных с гироскопа, существует несколько путей решения, например использование ПИД-регулятора или комплементарного фильтра, также фильтра Калмана [2]. Самым популярным ввиду его простоты является комплементарный фильтр. Его использование высокоэффективно, при этом не затрачивается много машинных ресурсов.

$$\varphi = (1 - K)\varphi_{gyro} + K\varphi_{acc}, \quad (6)$$

где φ_{gyro} – значение угла, полученное с помощью гироскопа, φ_{acc} – значение угла, полученное с помощью акселерометра, K – коэффициент фильтра.

Как видно, итоговая величина угла наклона представляет собой сумму интегрированного значения показаний с гироскопа и мгновенного значения с акселерометра. По сути, главная задача комплементарного фильтра заключается в том, чтобы нивелировать дрейф нуля гироскопа и ошибки интегрирования. Вес корректирующих значений определяется коэффициентом фильтра K .

Выбор коэффициента K зависит от величины дрейфа нуля гироскопа и от скорости накопления ошибок вычисления. Так, слишком большое значение коэффициента приведет к тому, что на результат работы фильтра будут сильно влиять внешние вибрации корпуса робота (например, при ходьбе). Слишком же малое может оказаться недостаточным для ликвидации дрейфа нуля гироскопа. Как правило, коэффициент фильтрации K подбирается при настройке для каждого инклинометра исходя из вышеуказанных условий. Типичное его значение лежит в диапазоне от 0,05 до 0,15. В данном проекте был задан коэффициент фильтрации K комплементарного фильтра равным 0,1. Результаты измерений угла наклона поверхности показаны на рис. 7–10.

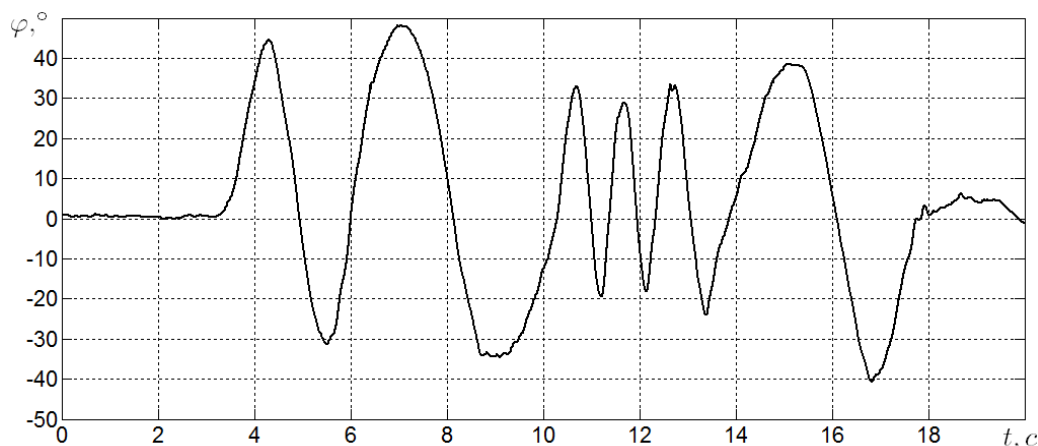


Рис. 7. График изменения угла наклона φ нестационарной поверхности, полученный с использованием комплементарного фильтра

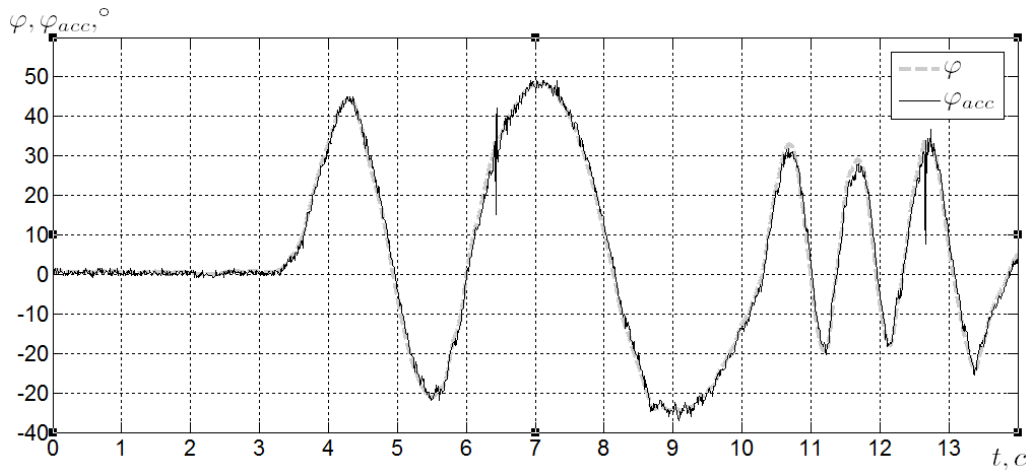


Рис. 8. График изменения угла наклона нестационарной поверхности:
 φ – угол, полученный с использованием комплементарного фильтра;
 $\varphi_{асс}$ – угол, полученный на основе измерений акселерометра

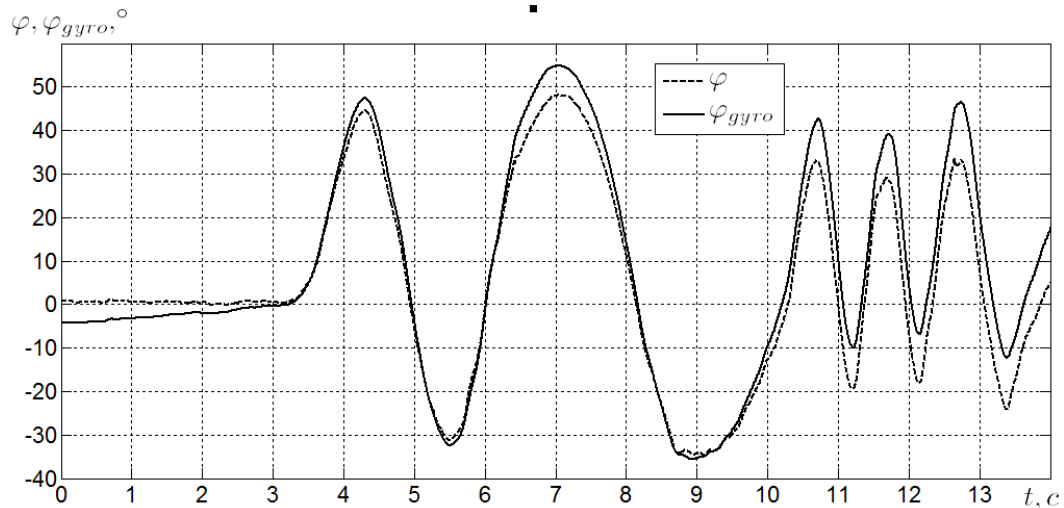


Рис. 9. График изменения угла наклона нестационарной поверхности:
 φ – угол, полученный с использованием комплементарного фильтра;
 $\varphi_{гиро}$ – угол, полученный на основе измерений гироскопа

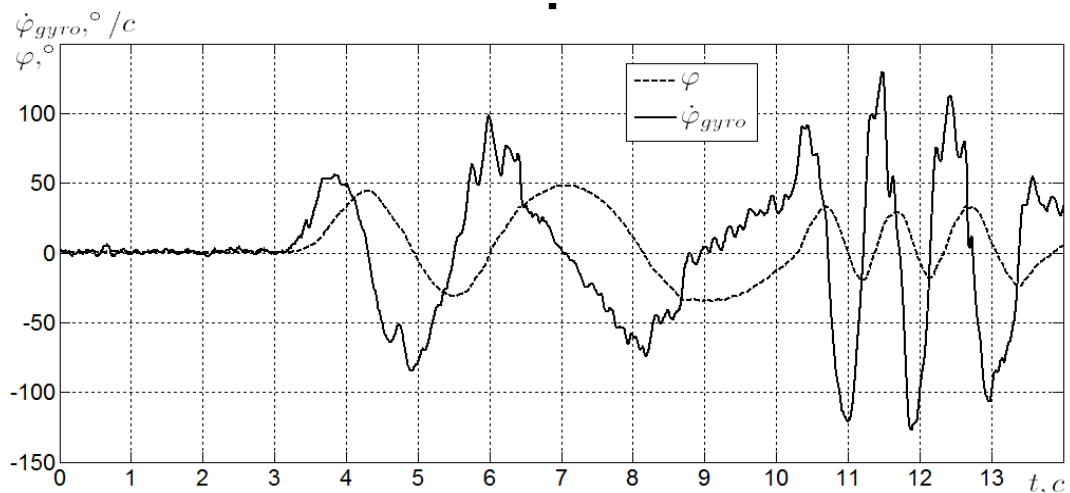


Рис. 10. График изменения угла φ наклона нестационарной поверхности,
 полученного с использованием комплементарного фильтра,
 и угловой скорости $\dot{\varphi}_{гиро}$, измеренной гироскопом

Для управления роботом был использован ПД-регулятор [3] с коэффициентами пропорциональной и дифференциальной составляющей соответственно равными $K_p = K_d = 1$.

$$u = K_p e + K_d \dot{e}, \quad (7)$$

где $u = (u_1 \ u_2 \ u_3)^T$, $e = (e_1 \ e_2 \ e_3)^T$, $e_i = q_i - q_i^*$.

График переходных процессов для одного из двигателей расположенного в стопе, приведен ниже на рис. 11. Для двигателя, расположенного во второй стопе, графики будут выглядеть аналогично.

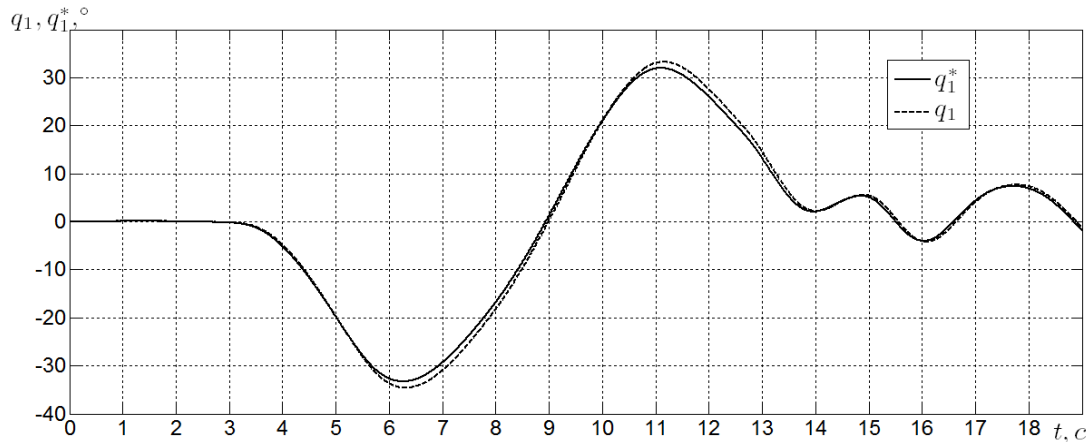


Рис. 11. Графики текущей и желаемой траекторий для двигателя в стопе

На рис. 12 приведен пример работы разработанной системы управления двуногим роботом.

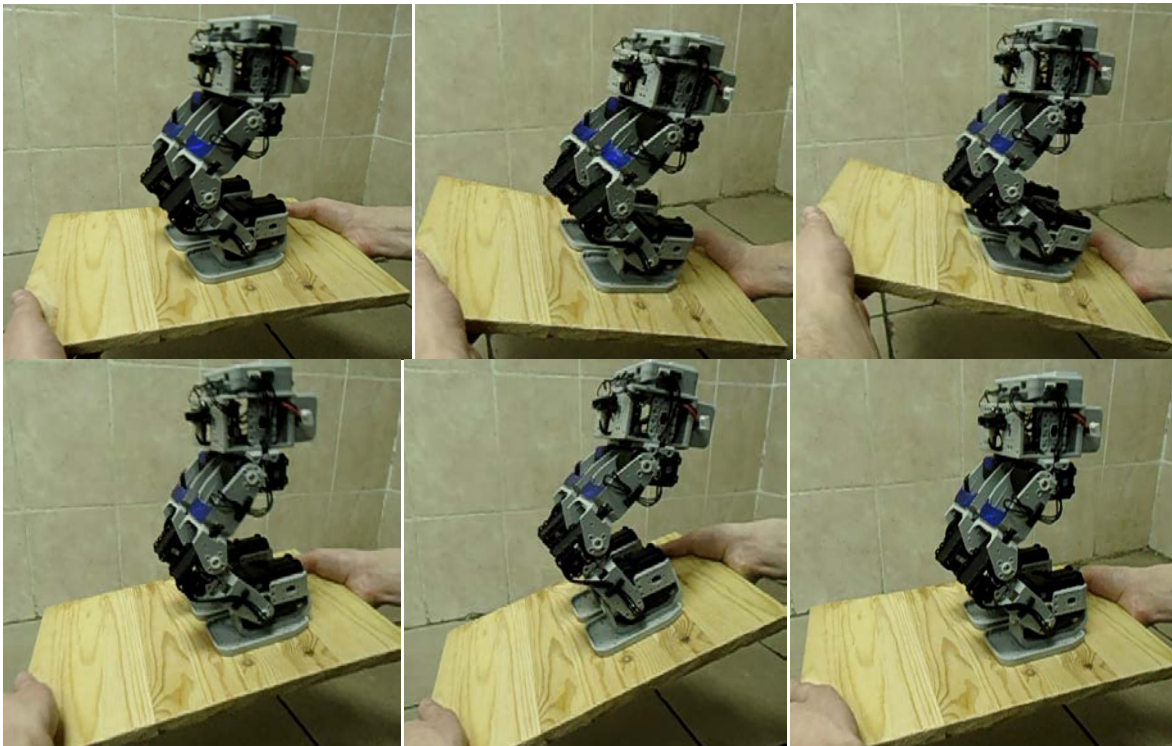


Рис. 12. Эксперимент

Язык RoboPlus является более простым и удобным, чем C, однако он не обладает необходимым функционалом для построения сложных систем управления роботом. Поэтому его целесообразно использовать на начальных и ознакомительных этапах работы с робототехническим набором BioloidKit. Ввиду этого полученный алгоритм управления рассматриваемым двуногим роботом был реализован на языке C в среде AtmelStudio.

Заключение

В ходе данной работы была решена задача стабилизации двуногого шагающего робота в положении стоя на нестационарной поверхности.

Была построена и исследована кинематическая схема робота, реализованного на практике и имеющего шесть степеней свободы. В качестве цели управления, обеспечивающей стабилизацию робота, было выбрано условие поддержания координат центра масс робота над центром площади опоры. Исходя из этого условия, были получены желаемые траектории движения робота в зависимости от угла наклона нестационарной поверхности. Для управления каждым сервоприводом робота использован ПД-регулятор.

При балансировании робота угол наклонной поверхности менялся произвольным образом в пределах от 0 до 50°.

Для измерения угла наклона поверхности, на которой располагался робот, были использованы двухосный гироскоп и трехосный акселерометр. Вычисленные углы, полученные исходя из показаний данных датчиков, обрабатывались комплементарным фильтром с коэффициентом $K=0,1$. Использование этого фильтра позволило получить более точные значения угла наклона поверхности по сравнению с использованием только одного из датчиков, а также позволило исключить дрейф гироскопа.

Предложенные в статье алгоритмы могут быть использованы для стабилизации подобных роботов, а также служить основой при разработке регуляторов и систем управления для более сложных роботов.

Публикации

1. *Bazylev D.N., Pyrkin A.A.* Stabilization of biped robot standing on nonstationary plane // 18th IEEE International Conference on Methods and Models in Automation & Robotics. Miedzyzdroje, 2013. P. 459–463.
2. *Базылев Д.Н., Громов В.С., Борисов О.И.* Стабилизация двуногого робота в положении стоя // Сб. тезисов докладов конгресса молодых ученых. СПб., 2013. Вып. 1: Интеллектуальные системы управления и обработки информации. С. 170–171.
3. *Базылев Д.Н.* Планирование и стабилизация движений шагающего робота // Аннотированный сборник научно-исследовательских выпускных квалификационных работ бакалавров УНИВЕРСИТЕТ ИТМО. СПб., 2012. С. 87–89.

Литература

1. Официальный сайт компании Robotis – <http://www.robotis.com>.
2. Микросистемы ориентации беспилотных летательных аппаратов / Под ред. В.Я. Распопова. М., 2011. – 180 с.
3. *Мирошник И.В.* Теория автоматического управления. Линейные системы. СПб., 2005. – 336 с.



Вавулин Дмитрий Николаевич
Факультет фотоники и оптоинформатики
Кафедра фотоники и оптоинформатики
E-mail: dima-vavulin@mail.ru

МУЛЬТИПЛЕКСИРОВАНИЕ КЛАССИЧЕСКИХ И КВАНТОВЫХ КАНАЛОВ В ОПТОВОЛОКОННЫХ ЛИНИЯХ СВЯЗИ

Д.Н. Вавулин

Научный руководитель – доцент С.А. Чивилихин

Введение

Системы квантовой криптографии, существующие на данный момент, зачастую используют не только слабый квантовый сигнал как носитель защищенной информации, но и классический (опорный), предназначенный для синхронизации временных отсчетов принимающей и передающей стороны [1–3]. По наличию классического сигнала также можно определить попытку передачи информации, так как квантовый может полностью поглотиться в волокне. Однако для этих целей используются два оптических волокна: одно для квантового канала и второе – для классического. Необходимость прокладывать отдельное волокно для квантового канала даже в пределах одного города может вызвать большие финансовые трудности.

Данная работа позволила решить проблему необходимости прокладки дополнительного оптоволоконного канала путем мультиплексирования обоих каналов связи в одном волокне.

Методы исследования

Стандартные методы квантовой криптографии

Стандартные телекоммуникационные системы связи используют оптические волокна в качестве среды, по которой передается информация в виде большого числа квазимонохроматических классических сигналов, расположенных в небольшом частотном диапазоне вблизи длины волны 1550 нм (WDM-мультиплексирование). Если квантовый сигнал пропускать на частоте, близкой к частоте классических сигналов, то будет наблюдаться множество нелинейных эффектов. Они значительно искажат качество квантового сигнала, так как он приблизительно на восемь порядков слабее классического. Основными двумя нелинейными эффектами являются эффекты четырехволнового смешения (ЧВС) и спонтанного комбинационного рассеяния света [4–8].

В данной работе рассматривалось и анализировалось влияние эффекта ЧВС для различных мощностей классического сигнала. При этом использовалась длина волны квантового сигнала в 1300 нм, а для классического сигнала – 1550 нм. Данные длины волн обеспечивают не только малое влияние эффекта ЧВС, но и ослабленное значение Рамановского рассеяния света, так как они сильно удалены друг от друга. Однако поглощение квантового сигнала в волокне при этом будет больше по сравнению со случаем, когда оба сигнала располагаются вблизи длины волны 1550 нм.

Исходные уравнения, значения используемых параметров, начальные условия

Исходные уравнения для решения поставленной задачи были взяты из статьи [9]. Они описывают изменения комплексных амплитуд x - и y -компонент поляризации полей накачки (p – поле в классическом канале, волна синхронизации), сигнальной волны (s – поле в квантовом канале) и холостой (i – паразитная волна, возникающая за счет ЧВС).

Использовались уравнения для x -компоненты накачки, а также x - и y -компонент сигнальной волны. Они были дополнены еще тремя для y -компоненты накачки, а также x - и y -компонент холостой волны. Все уравнения были упрощены, так как в этом случае нет необходимости учитывать двулучепреломление волокна. Далее они были дополнены членами, ответственными за затухание полей в волокне. Полученная система из шести уравнений выглядит следующим образом:

$$i \frac{\partial \phi_{1p}}{\partial z} + \frac{8}{9} (2\phi_{1i} \phi_{1p} \phi_{1i}^* + \phi_{1p}^2 \phi_{1p}^* + 2\phi_{1i} \phi_{1s} \phi_{1p}^* e^{i\Delta k z} + 2\phi_{1p} \phi_{1s} \phi_{1s}^* + \phi_{1p} \phi_{2i} \phi_{2i}^* + \phi_{1i} \phi_{2p} \phi_{2i}^* + \phi_{1s} \phi_{2i} \phi_{2p}^* e^{i\Delta k z} + \phi_{1p} \phi_{2p} \phi_{2p}^* + \phi_{1i} \phi_{2s} \phi_{2p}^* e^{i\Delta k z} + \phi_{1s} \phi_{2p} \phi_{2s}^* + \phi_{1p} \phi_{2s} \phi_{2s}^*) + i\alpha(\omega_p) \phi_{1p} = 0 \quad (1)$$

$$i \frac{\partial \phi_{2p}}{\partial z} + \frac{8}{9} (2\phi_{2i} \phi_{2p} \phi_{2i}^* + \phi_{2p}^2 \phi_{2p}^* + 2\phi_{2i} \phi_{2s} \phi_{2p}^* e^{i\Delta k z} + 2\phi_{2p} \phi_{2s} \phi_{2s}^* + \phi_{2p} \phi_{1i} \phi_{1i}^* + \phi_{2i} \phi_{1p} \phi_{1i}^* + \phi_{2s} \phi_{1i} \phi_{1p}^* e^{i\Delta k z} + \phi_{2p} \phi_{1p} \phi_{1p}^* + \phi_{2i} \phi_{1s} \phi_{1p}^* e^{i\Delta k z} + \phi_{2s} \phi_{1p} \phi_{1s}^* + \phi_{2p} \phi_{1s} \phi_{1s}^*) + i\alpha(\omega_p) \phi_{2p} = 0 \quad (2)$$

$$i \frac{\partial \phi_{1s}}{\partial z} + \frac{8}{9} (\phi_{1p}^2 \phi_{1s}^* e^{-i\Delta k z} + 2\phi_{1i} \phi_{1s} \phi_{1i}^* + 2\phi_{1p} \phi_{1s} \phi_{1p}^* + \phi_{1s}^2 \phi_{1s}^* + \phi_{1s} \phi_{2i} \phi_{2i}^* + \phi_{1p} \phi_{2p} \phi_{2i}^* e^{-i\Delta k z} + \phi_{1i} \phi_{2s} \phi_{2i}^* + \phi_{1s} \phi_{2p} \phi_{2p}^* + \phi_{1p} \phi_{2s} \phi_{2p}^* + \phi_{1s} \phi_{2s} \phi_{2s}^*) + i\alpha(\omega_s) \phi_{1s} = 0 \quad (3)$$

$$i \frac{\partial \phi_{2s}}{\partial z} + \frac{8}{9} (\phi_{2p}^2 \phi_{2s}^* e^{-i\Delta k z} + 2\phi_{2i} \phi_{2s} \phi_{2i}^* + 2\phi_{2p} \phi_{2s} \phi_{2p}^* + \phi_{2s}^2 \phi_{2s}^* + \phi_{2s} \phi_{1i} \phi_{1i}^* + \phi_{2p} \phi_{1p} \phi_{1i}^* e^{-i\Delta k z} + \phi_{2i} \phi_{1s} \phi_{1i}^* + \phi_{2s} \phi_{1p} \phi_{1p}^* + \phi_{2p} \phi_{1s} \phi_{1p}^* + \phi_{2s} \phi_{1s} \phi_{1s}^*) + i\alpha(\omega_s) \phi_{2s} = 0 \quad (4)$$

$$i \frac{\partial \phi_{1i}}{\partial z} + \frac{8}{9} (\phi_{1p}^2 \phi_{1i}^* e^{-i\Delta k z} + 2\phi_{1s} \phi_{1i} \phi_{1s}^* + 2\phi_{1p} \phi_{1i} \phi_{1p}^* + \phi_{1i}^2 \phi_{1i}^* + \phi_{1i} \phi_{2s} \phi_{2s}^* + \phi_{1p} \phi_{2p} \phi_{2s}^* e^{-i\Delta k z} + \phi_{1s} \phi_{2i} \phi_{2s}^* + \phi_{1i} \phi_{2p} \phi_{2p}^* + \phi_{1p} \phi_{2i} \phi_{2p}^* + \phi_{1i} \phi_{2i} \phi_{2i}^*) + i\alpha(\omega_i) \phi_{1i} = 0 \quad (5)$$

$$i \frac{\partial \phi_{2i}}{\partial z} + \frac{8}{9} (\phi_{2p}^2 \phi_{2i}^* e^{-i\Delta k z} + 2\phi_{2s} \phi_{2i} \phi_{2s}^* + 2\phi_{2p} \phi_{2i} \phi_{2p}^* + \phi_{2i}^2 \phi_{2i}^* + \phi_{2i} \phi_{1s} \phi_{1s}^* + \phi_{2p} \phi_{1p} \phi_{1s}^* e^{-i\Delta k z} + \phi_{2s} \phi_{1i} \phi_{1s}^* + \phi_{2i} \phi_{1p} \phi_{1p}^* + \phi_{2p} \phi_{1i} \phi_{1p}^* + \phi_{2i} \phi_{1i} \phi_{1i}^*) + i\alpha(\omega_i) \phi_{2i} = 0 \quad (6)$$

Здесь поля ϕ_{1p} , ϕ_{1s} , ϕ_{1i} – комплексные амплитуды x -компонент поляризации полей накачки (p), сигнальной (s) и холостой волны (i). Соответственно, ϕ_{2p} , ϕ_{2s} , ϕ_{2i} – это амплитуды y -компонент поляризации полей накачки (p), сигнала (s) и холостой волны (i). Здесь амплитуды измеряются в единицах Вт^{1/2}, то есть величина $|\phi_{1p}|^2$ – это пиковая мощность x -компоненты поляризации поля накачки, измеряемая в Ваттах. Под продольной координатой z понимается произведение физической длины волокна Z , измеряемой в километрах, на коэффициент керровской нелинейности γ , измеряемой в (Вт·км)⁻¹ [10–13]. Он принимался равным $\gamma = 1$ (Вт·км)⁻¹ (типичное значение для волокна); Δk – величина, определяемая из теории ЧВС. Частоты для полей накачки, сигнальной и холостой волн не зависят от поляризации волн: $\omega_{1p} = \omega_{2p} = \omega_p$, $\omega_{1s} = \omega_{2s} = \omega_s$, $\omega_{1i} = \omega_{2i} = \omega_i$. Далее, $\alpha_p(\omega_p)$, $\alpha_s(\omega_s)$ и $\alpha_i(\omega_i)$ – коэффициенты поглощения для полей накачки, сигнальной и холостой волн соответственно.

Расчет дополнительных параметров

Величина Δk рассчитывалась по выражению:

$$\Delta k = k_i + k_s - 2k_p.$$

Здесь $k_i = n_i \omega_i / c$ – постоянная распространения холостой волны на частоте ω_i , c – скорость света в вакууме, n_i – показатель преломления волокна на частоте ω_i (аналогично для k_s и k_p). Частоты волн связаны соотношением $2\omega_p = \omega_s + \omega_i$. Здесь ω_p соответствует длине волны $\lambda_p = 1,55$ мкм, ω_s – длине волны $\lambda_s = 1,3$ мкм. Частота холостой волны ω_i была найдена из условия $\omega_i = 2\omega_p - \omega_s$. Она соответствует длине волны $1/\lambda_i = 2/\lambda_p - 1/\lambda_s$.

$$\lambda_i = (\lambda_p \lambda_s) / (2\lambda_s - \lambda_p),$$

$$\lambda_i = (1,3 \cdot 1,55) / (1,3 \cdot 2 - 1,55) = 1,919 \text{ мкм} \sim 1,92 \text{ мкм}.$$

Показатели преломления на всех трех длинах волн: λ_p , λ_s , λ_i были рассчитаны по формуле Сельмейера:

$$n^2(\omega) = 1 + \sum_{j=1}^m \frac{B_j \omega_j^2}{\omega_j^2 - \omega^2}$$

Здесь $B_1=0,696163$, $B_2=0,4079426$, $B_3=0,8974794$, $\lambda_1=0,0684043$ мкм, $\lambda_2=0,1162414$ мкм, $\lambda_3=9,896161$ мкм, где $\lambda_j=2\pi c/\omega_j$.

Результаты расчета показателя преломления представлены на рис. 1.

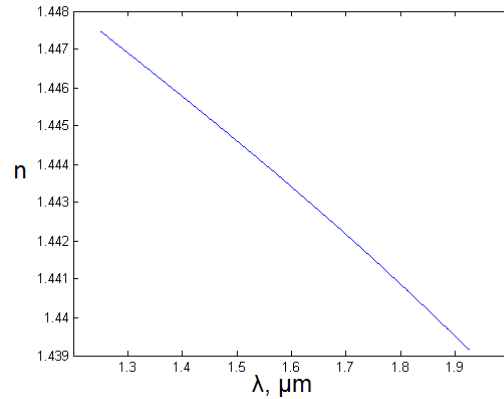


Рис. 1. Результаты расчета показателя преломления волокна по формуле Сельмейера

Показатели преломления на всех трех длинах волн ($\lambda_s = 1,3$ мкм, $\lambda_p = 1,55$ мкм, $\lambda_i = 1,92$ мкм) были рассчитаны по формуле Сельмейера и получились равными: $n_s (1,3 \text{ мкм}) = 1,4469$, $n_p (1,55 \text{ мкм}) = 1,4440$, $n_i (1,92 \text{ мкм}) = 1,4392$.

Постоянные распространения k_s , k_p , k_i рассчитывались по формуле (с соответствующими n и λ): $k = 2\pi n / \lambda$. Они получились равными: $k_s (1,3 \text{ мкм}) = 6,9932$ (1/мкм), $k_p (1,55 \text{ мкм}) = 5,8535$ (1/мкм), $k_i (1,92 \text{ мкм}) = 4,7122$ (1/мкм).

Величина Δk рассчитывалась по формуле $\Delta k = k_i + k_s - 2k_p$ и получилась равной: $\Delta k (1,92) = -0,0016$ (1/мкм).

Были использованы стандартные коэффициенты поглощения телекоммуникационного волокна для различных длин волн: $\alpha_s (1,3 \text{ мкм}) = 0,34$ дБ/км, $\alpha_p (1,55 \text{ мкм}) = 0,2$ дБ/км, $\alpha_i (1,92 \text{ мкм}) = 6,5$ дБ/км.

Коэффициенты поглощения были переведены из дБ/км в 1/км по формуле:

$$\alpha(1/\text{км}) = 0,5 \cdot 0,1 \cdot \alpha(\text{дБ/км}) \ln 10.$$

Они получились равными: $\alpha_s (1,3 \text{ мкм}) = 0,03914$ (1/км), $\alpha_p (1,55 \text{ мкм}) = 0,02303$ (1/км), $\alpha_i (1,92 \text{ мкм}) = 0,74834$ (1/км).

Начальные условия

Поля классического (p) и квантового (s) каналов заданы на входе ($z = 0$) своими комплексными амплитудами [14]:

$$\begin{aligned} \phi_{1p} &= (10^{-3})^{1/2} \text{ Вт}^{1/2}, \\ \phi_{2p} &= 0 \text{ Вт}^{1/2}, \\ \phi_{1s} &= (10^{-11})^{1/2} \text{ Вт}^{1/2}, \\ \phi_{2s} &= 0 \text{ Вт}^{1/2}. \end{aligned}$$

Холостая волна на входе не возбуждена:

$$\begin{aligned} \phi_{1i} &= 0 \text{ Вт}^{1/2}, \\ \phi_{2i} &= 0 \text{ Вт}^{1/2}. \end{aligned}$$

Упрощение модели

В нашем случае возбуждались только x -компоненты всех трех полей, а y -компоненты оставались нулевыми, поэтому уравнения (1)–(6) были упрощены до следующего вида:

$$\phi_{2p} = 0 \text{ Вт}^{1/2}, \quad (7)$$

$$\phi_{2s} = 0 \text{ Вт}^{1/2}, \quad (8)$$

$$\phi_{2i} = 0 \text{ Вт}^{1/2}, \quad (9)$$

$$i \frac{\partial \phi_{1p}}{\partial z} + \frac{8}{9} (2\phi_{1i} \phi_{1p} \phi_{1i}^* + \phi_{1p}^2 \phi_{1p}^* + 2\phi_{1i} \phi_{1s} \phi_{1p}^* e^{i\Delta k z} + 2\phi_{1p} \phi_{1s} \phi_{1s}^*) + i\alpha(\omega_p) \phi_{1p} = 0, \quad (10)$$

$$i \frac{\partial \phi_{1s}}{\partial z} + \frac{8}{9} (\phi_{1p}^2 \phi_{1s}^* e^{-i\Delta k z} + 2\phi_{1i} \phi_{1s} \phi_{1i}^* + 2\phi_{1p} \phi_{1s} \phi_{1p}^* + \phi_{1s}^2 \phi_{1s}^*) + i\alpha(\omega_s) \phi_{1s} = 0, \quad (11)$$

$$i \frac{\partial \phi_{1i}}{\partial z} + \frac{8}{9} (\phi_{1p}^2 \phi_{1s}^* e^{-i\Delta k z} + 2\phi_{1s} \phi_{1i} \phi_{1s}^* + 2\phi_{1p} \phi_{1i} \phi_{1p}^* + \phi_{1i}^2 \phi_{1i}^*) + i\alpha(\omega_i) \phi_{1i} = 0. \quad (12)$$

Численная реализация модели

Численное моделирование процесса выполнялось в программе Matlab. Был использован метод Рунге-Кутты четвертого порядка.

При точном расчете на длине линии 100 км уложилось бы $N = 8 \cdot 10^9$ шагов. Поэтому шаг сетки для ускорения вычислений выбирался равным Δz приближенный $= 0,2(\Delta k)^{-1}$, а не Δz точный $= 0,02(\Delta k)^{-1}$ (это обеспечивало хорошую точность). Величина Δk при вычислениях изменялась для различных длин линии связи по формуле: Δk приближенная $= \Delta k \cdot (10^2)/L$, где L – длина линии связи, выраженная в километрах. То есть, например, для линии длиной 10 км величина Δk была увеличена в тысячу раз по сравнению с истинным значением.

Данные приближения обеспечивали достаточно точный вид графика функции и позволяли проследить ее изменение с расстоянием. Существенно менялось лишь количество осцилляций волн на единицу длины линии. Точный расчет был проведен только для линии связи длиной 1 м, что, конечно, не представляет практического интереса, но позволяет проследить плотность осцилляций холостой волны на единицу длины линии в реальном случае.

Расчет мощностей накачки, сигнальной и холостой волн

На рис. 2 представлены результаты расчета для линии связи в 100 км для волн накачки (красная, классический канал), сигнала (зеленая, квантовый канал) и холостой волны (синяя). Для наглядности значения интенсивностей квантового канала и холостой волны были увеличены соответственно в 10^8 и в 10^{17} раз.

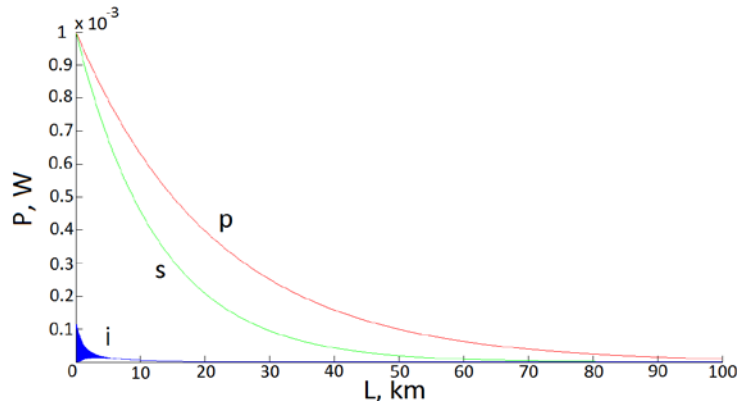


Рис. 2. Результаты расчетов для линии связи в 100 км: p – волна накачки, s – квантовая волна, i – холостая волна

На рис. 3а представлена разность в интенсивностях сигнальных волн ΔP_s в присутствии накачки (1 мВт) и без нее, на фоне интенсивности сигнальной волны, полученной в присутствии накачки. Начальная интенсивность сигнальной волны равна 10^{-11} Вт, а максимальная разность составляет порядка $1,6 \cdot 10^{-21}$ Вт, то есть на 10 порядков меньше. Отсюда видно, что наличие классического сигнала практически не будет влиять на квантовый. На рис. 3б представлена та же разность, но вычисленная для сигнала накачки, увеличенного

в тысячу раз (1 Вт). Видно, что графики имеют схожий вид, но максимальная разность теперь составляет порядка $1,6 \cdot 10^{-15}$ Вт, то есть на шесть порядков больше. Также видно, что даже такая интенсивность классического канала будет слабо влиять на квантовый.

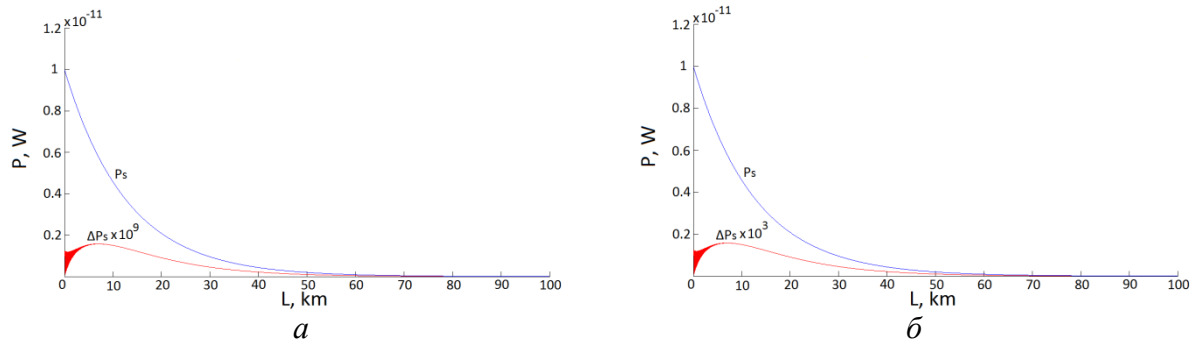


Рис. 3. Разность в интенсивностях сигнальных волн в присутствии накачки и без нее на фоне сигнальной волны: а – мощность сигнала накачки составляет 1 мВт (для наглядности значение разности увеличено в 10^9 раз); б – мощность сигнала накачки составляет 1 Вт (для наглядности значение разности увеличено в 10^3 раз)

Расчет эффективности ЧВС

Для накачки интенсивностью 1 мВт, была рассчитана зависимость эффективности ЧВС от длины линии оптического волокна. Эффективность ЧВС рассчитывалась как разность сигнальных волн в присутствии накачки и без нее, деленная на интенсивность сигнальной волны в присутствии накачки (рис. 4). Видно, что эффективность ЧВС с увеличением длины линии стремится к некоторому пределу.

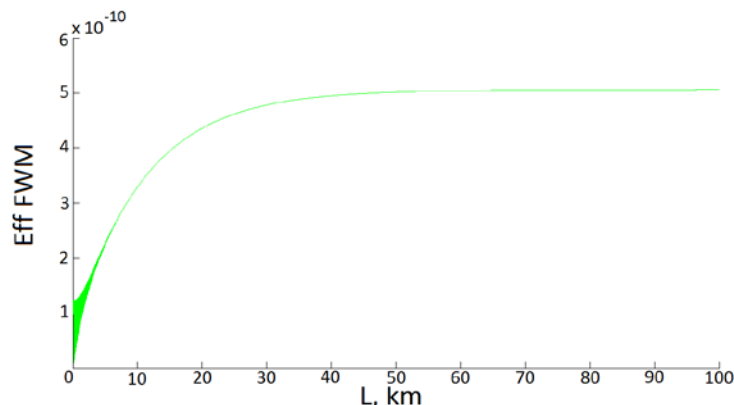


Рис. 4. Зависимость эффективности ЧВС от длины линии оптического волокна

Заключение

Получена возможность оптимальным образом подбирать параметры для квантовой рассылки криптографического ключа, мультиплексированного с классическим сигналом. Это позволяет достаточно просто осуществить абсолютно секретную передачу информации.

Благодаря использованию только одного волокна возможно осуществлять непрерывную синхронизацию, а не останавливать систему каждый раз для синхронизации отправителя и получателя. Это экономически удобно, так как квантовый сигнал пропускается по уже имеющимся оптическим волокнам.

Области возможного использования

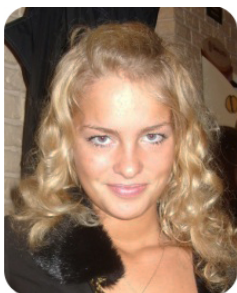
Основное применение результатов исследования – системы квантовой криптографии. Потенциальные потребители – военные и оборонные организации, банковские системы, системы связи и пр. Использование результатов также возможно для следующих приложений: квантовые коммуникации (квантовый компьютеринг, квантовые алгоритмы), выделение одиночных фотонов из телекоммуникационных каналов. Потребителями здесь могут стать научно-исследовательские лаборатории и группы, а также университеты (в учебных целях).

Публикации

1. Egorov V.I., Vavulin D.N., Latypov I.Z., Gleim A.V., Rupasov A.V. Analysis of a sidebands based quantum cryptography system with different detector types // *Nanosystems: physics, chemistry, mathematics*. 2013. Vol. 4 (2). P. 190–195.
2. Vavulin D.N., Egorov V.I., Gleim A.V. Investigation of four-wave mixing effect on quantum key distribution // *Nanosystems: physics, chemistry, mathematics*. 2013 (in print).

Литература

1. Patel K.A., Dynes J.F., Choi I. et al. Coexistence of high-bit-rate quantum key distribution and data on optical fiber // *Physical Review X*. 2012. Vol. 2. P. 041010.
2. Chapuran T.E., Toliver P., Peters N.A. et al. Optical networking for quantum key distribution and quantum communications // *New Journal of Physics*. 2009. Vol. 11. P. 105001.
3. Townsend P.D. Simultaneous quantum cryptographic key distribution and conventional data transmission over installed fibre using wavelength-division multiplexing // *Electronics Letters*. 1997. Vol. 33. P. 188–190.
4. Choi I., Young R.J., Townsend P.D. Quantum information to the home // *New Journal of Physics*. 2011. Vol. 13. P. 063039.
5. Scarani V., Bechmann-Pasquinucci H., Cerf N. et al. The security of practical quantum key distribution // *Reviews of Modern Physics*. 2009. Vol. 81. P. 1301–1350.
6. Boyd W.R. *Nonlinear Optics*. New York, 2003. – 614 p.
7. Eraerds P., Walenta N., Legré M. et al. Quantum key distribution and 1 Gbps data encryption over a single fibre // *New Journal of Physics*. 2010. Vol. 12. P. 063027.
8. Lancho D., Martinez J., Elkouss D. et al. QKD in standard optical telecommunication networks // *Lecture Notes of the Institute for Computer Sciences. Social Informatics and Telecommunications Engineering*. 2010. Vol. 36. P. 142–149.
9. Guasoni M., Kozlov V.V., Wabnitz S. Theory of polarization attraction in parametric amplifiers based on telecommunication fibers // *Journal of the Optical Society of America B*. 2012. Vol. 29. No 10. P. 2710–2720.
10. Paul D. Townsend experimental investigation of the performance limits for first telecommunications-window quantum cryptography systems // *IEEE Photonics Technology Letters*. 1998. Vol. 10. No 7. P. 1048–1050.
11. Peev M., Pacher C., Alléaume R. The SECOQC quantum key distribution network in Vienna // *New Journal of Physics*. 2009. Vol. 11. P. 075001.
12. Qi B., Zhu W., Qian L., Lo H.K. Feasibility of quantum key distribution through a dense wavelength division multiplexing network // *New Journal of Physics*. 2010. Vol. 12. P. 103042.
13. Stucki D., Legré M., Buntschu F. et al. Long-term performance of the SwissQuantum quantum key distribution network in a field environment // *New Journal of Physics*. 2011. Vol. 13. P. 123001.
14. Egorov V.I., Vavulin D.N., Latypov I.Z., Gleim A.V., Rupasov A.V. Analysis of a sidebands based quantum cryptography system with different detector types // *Nanosystems: physics, chemistry, mathematics*. 2013. Vol. 4 (2). P. 190–195.



Волкова Ольга Владимировна
Гуманитарный факультет
Кафедра экономической теории и бизнеса
E-mail: 15cool15@mail.ru

ТРУДОВОЙ ПОТЕНЦИАЛ КАК ФАКТОР РАЗВИТИЯ НАЦИОНАЛЬНОЙ ЭКОНОМИКИ

О.В. Волкова

Научный руководитель – доцент Л.А. Пятлина

Теоретической основой исследования явились работы отечественных и зарубежных ученых, изучающих трудовой потенциал, таких как Кознова А.Б., Гузакова О.Л., Озова А.А., Морозова А.А., Радько С.Г., Усачева В.И., Меньшикова О.И. и Шлендера П.Э.

Для анализа трудового потенциала использовались теоретические и эмпирические методы. Эмпирической основой для получения прогнозных значений служили эконометрические расчеты, которые выполнялись с помощью: косвенного метода наименьших квадратов, двухшагового метода наименьших квадратов и различных экономических моделей. Для сравнения показателей использовались формулы роста и прироста, корреляционно-регрессионного анализа, анализ трендов и остатков.

Результаты данного проекта дают характеристику трудового потенциала России в наше время, полученные прогнозные значения помогут решить уникальную, не имеющую исторических аналогов задачу по поддержанию высоких темпов экономического роста в условиях сокращающегося населения и, прежде всего, его экономически активной части.

Понятие трудового потенциала и его структура

Обобщающим показателем процесса становления и развития человека в трудовой деятельности является трудовой потенциал общества. Словом «потенциал» обычно обозначают средства, запасы, источники, которые могут быть использованы, а также возможности отдельного лица, группы лиц, общества в конкретной обстановке.

Целесообразность процесса труда предполагает его конечный результат – продукт труда, который «есть потребительная стоимость, вещество природы, приспособленное к человеческим потребностям посредством изменения формы».

Процесс труда как систему можно представить следующим образом (рис. 1).

Понятие «рабочая сила» соответствует понятию «экономически активное население», или «функционирующая рабочая сила», состоящая из занятых и безработных.

В трудовом потенциале предприятия можно выделить следующие компоненты:

- кадровая составляющая включает: квалификационный потенциал (профессиональные знания, умения и навыки) и образовательный потенциал (познавательные способности);
- профессиональная структура коллектива связана с изменениями в содержании труда (появление новых и отмирание старых профессий, усложнение трудовых операций);
- квалификационная структура определяется качественными изменениями в трудовом потенциале (рост умений, знаний, навыков) и отражает прежде всего изменения в его личностной составляющей;

Общественный трудовой потенциал, аккумулирующий и синтезирующий совокупные способности к общественно полезной деятельности трудоспособного населения, представляет собой конкретную форму материализации человеческого фактора, показатель уровня развития и границы созидательной активности трудящихся масс.

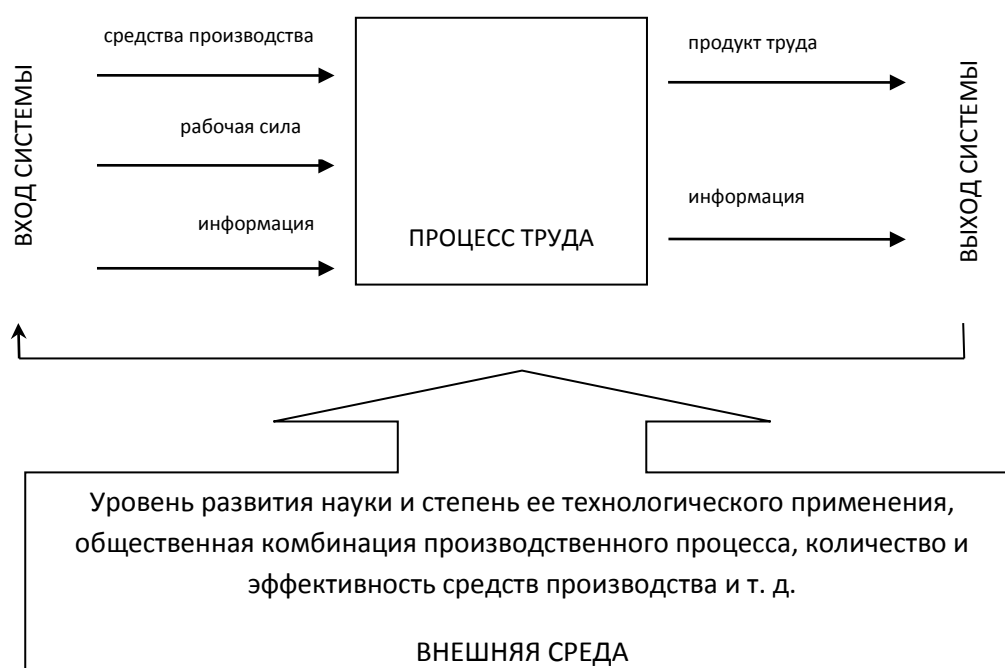


Рис. 1. Процесс труда как система

В качественном отношении трудовой потенциал общества – это его реальные возможности по реализации через участие в общественно полезном труде всего многообразия личных способностей и качеств: знаний, умений и навыков, приобретаемых людьми в процессе воспитания, подготовки к труду, непосредственно трудовой деятельности, переобучения и повышения квалификации. Он зависит от качественных характеристик совокупной рабочей силы, определяющих уровень развития способностей к труду (профессионального мастерства, интеллекта, творческой активности, мобильности) и специфику важнейших социально формирующих черт членов общества (ответственности, исполнительности, инициативности, дисциплинированности, организаторских способностей и др.).

Исследование трудового потенциала и его прогноз

От демографических характеристик трудоспособного населения (половозрастной структуры, здоровья, уровня занятости) и показателей демографического развития (динамики рождаемости и смертности населения) зависит величина совокупного общественного и национального дохода. Это также связано с тем, что демографический фактор параллельно с другими социально-экономическими факторами (род занятий или деятельности, вид производства, условия труда и т. д.) воздействует на уровень трудовой активности населения.

Таблица 1

Численность мужчин и женщин за 2003–2013 гг. (данные на начало года) [5]

Годы	Все население, млн человек	в том числе		В общей численности населения, %		На 1000 мужчин приходится женщин
		мужчины	женщины	мужчины	женщины	
2003	146,9	68,7	78,2	46	53	1138
2004	146,3	68,3	78,0	47	53	1154
2006	145,0	67,5	77,5	47	53	1160
2007	144,2	67,0	77,2	46	54	1161
2008	143,5	66,6	76,9	46	54	1162
2009	142,8	66,2	76,6	46	54	1162
2010	142,2	65,8	76,4	46	54	1162
2011	142,0	65,7	76,3	46	54	1162
2012	141,9	65,6	76,3	46	54	1162
2013	141,9	65,6	76,3	46	54	1162

По данным Табл. 1 можно сказать, что в общей численности населения России женщины, как и прежде, имеют перевес: превышение численности женщин над численностью мужчин на 2013 г. составляло 10,7 млн человек против 9,5 млн в 2003 г. По сравнению с 2003 г. в 2013 г. общая численность населения сократилась на 5 млн человек (3,4 %), в том числе мужчин на 3,1 млн (4,5 %), а женщин – на 1,9 (2,4 %) млн, в результате чего диспропорция полового состава населения несколько увеличилась. На 1 000 мужчин в 2013 г. приходилось 1 162 женщины, а в 2003 г. – 1 138.

Преобладание численности женщин над численностью мужчин отмечается с 30-летнего возраста. Перевес числа женщин над числом мужчин складывается в средних возрастах в результате разной скорости вымирания женского и мужского населения (для отдельных территорий определенное значение могут иметь также миграционные процессы) и нарастает при переходе к старшим возрастным группам.

Мужчины меньше заботятся о своем здоровье, чем женщины, они больше подвержены воздействию стрессов и вредных привычек, мужчины больше гибнут в автомобильных катастрофах, по причине употребления наркотических веществ и алкоголя. Численность населения непосредственно связана с естественным приростом.

Проанализировав Табл. 2, необходимо отметить увеличение количества родившихся на период с 2003 по 2013 гг. на 41,2 % и снижение числа умерших на 8,8 %. На 2013 г. естественный прирост составляет минус 240 тыс. человек. Так, количество умерших выше числа родившихся на 14,1 %.

Таблица 2

Естественный прирост населения на протяжении 2003–2013 гг. [5]

Годы	Всего, тыс. человек			На 1000 человек населения, чел.		
	родившихся	умерших	естественный прирост	родившихся	умерших	естественный прирост
	Все население					
2003	1266,8	2225,3	-958,5	8,7	15,3	-6,6
2004	1311,6	2254,9	-943,3	9,0	15,6	-6,6
2005	1397	2332,3	-935,3	9,7	16,2	-6,5
2006	1477,3	2365,8	-888,5	10,2	16,4	-6,2
2007	1502,5	2295,4	-792,9	10,4	16,0	-5,6
2008	1457,4	2303,9	-846,5	10,2	16,1	-5,9
2009	1479,6	2166,7	-687,1	10,4	15,2	-4,8
2010	1610,1	2080,4	-470,3	11,3	14,6	-3,3
2011	1713,9	2075,9	-362	12,1	14,6	-2,5
2012	1761,7	2010,5	-248,8	12,4	14,2	-1,8
2013	1789	2028,5	-240	12,5	14,2	-1,7

В расчете на 1 000 человек количество родившихся также увеличилось с 2003 по 2013 г. на 43,6 %, число умерших на 1 000 человек сократилось на 7,2 %. Естественный прирост составил на 2013 г. минус 1,7 %.

Причиной увеличения рождаемости является улучшение благосостояния населения, желание создавать семьи и рожать детей, государственные программы по улучшению демографической ситуации в России, льготные условия покупки квартир, надбавки и социальные льготы многодетным семьям.

Все же социально-демографический фон остается не совсем благоприятным. Такая ситуация связана с динамикой ожидаемой продолжительности жизни при рождении (Табл. 3). Она определяется как число лет, которое в среднем предстояло бы прожить человеку из поколения родившихся при условии, что на протяжении всей жизни этого поколения по возрастной смертности останется на уровне того года, для которого исчислен показатель.

Таблица 3

Ожидаемая продолжительность жизни при рождении (количество лет) [5]

Годы	Все население			Городское население			Сельское население		
	всего	мужчины	женщины	всего	мужчины	женщины	всего	мужчины	Женщины
2003	65,34	59,03	72,26	65,69	59,35	72,46	64,34	58,14	71,66
2004	65,23	58,92	72,17	65,57	59,23	72,37	64,25	58,07	71,57
2005	64,95	58,68	71,9	65,4	59,1	72,2	63,7	57,5	71,09
2006	64,85	58,55	71,8	65,4	59,0	72,2	63,4	57,3	70,86
2007	65,27	58,89	72,3	65,8	59,4	72,7	63,8	57,6	71,27
2008	65,3	58,87	72,4	66,0	59,5	72,9	63,4	57,2	71,07
2009	66,6	60,37	73,2	67,3	61,0	73,7	64,7	58,7	71,89
2010	67,51	61,39	73,9	68,2	62,1	74,3	65,6	59,6	72,62
2011	67,88	61,83	74,2	68,6	62,5	74,6	66,0	60,0	72,86
2012	68,67	62,77	74,7	69,4	63,5	75,1	66,7	60,9	73,38

Ожидаемая продолжительность жизни всего населения с 2003 по 2012 гг. увеличилась на 4,6 % и в конце периода составила 68,67 лет. Произошло увеличение данного показателя у мужчин на 6,3 %, у женщин на 3,3 %. Для городского населения данный показатель увеличился на 5,6 %, для женщин – на 3,6%, для мужчин – на 6,9 %. Ожидаемая продолжительность жизни для сельского населения увеличилась на 3,7 %, у женщин данный показатель возрос на 2,4 %, у мужчин – на 4,6 %.

В целом стоит заметить, что продолжительность жизни у мужчин имеет более интенсивную тенденцию к увеличению, чем у женщин, как в городской, так и в сельской местности.

Абсолютная численность экономически активного населения в 2013 г. по сравнению с 2003 г. увеличилась на 4,2 % главным образом за счет роста экономической активности женского населения (на 5,6 %) (Табл. 4). За этот период наблюдался рост абсолютной численности занятых – на 6,9 %, в том числе женщин – на 8,8 % и мужчин – на 5,1 %. Общая численность безработных, зарегистрированных в органах госслужбы занятости сократилась на 20 %, в т. ч. мужчин – на 17,8 % и женщин – на 22,6 %.

Таблица 4

Численность экономически активного населения России за 2003–2013 гг. (тыс. человек) [5]

	2003 г.	2005 г.	2007 г.	2010 г.	2013 г.
Численность экономически активного населения	72332	72835	73811	75060	75440
Мужчины	37499	37206	37511	37975	38575
Женщины	34833	35629	36300	37085	36865
В том числе занятые в экономике – всего	65273	67152	68603	70814	69804
Мужчины	33754	34199	34710	35704	35500
Женщины	31519	32953	33893	35110	34304
Безработные – всего	7059	5683	5208	4246	5636
Мужчины	3745	3007	2801	2271	3075
Женщины	3314	2676	2407	1975	2562

Все больше женщин, расставляя жизненные приоритеты, хотят быть финансово независимыми от мужчин. Прежде чем создавать семью, представительницы прекрасного пола стараются сначала подняться по карьерной лестнице и самореализовать себя.

Несмотря на перевес женщин в общей численности населения в 2013 г., мужчин было занято больше: из общей численности занятого населения (69,8 млн человек) мужчин – 35,5 млн, а женщин – 34,3 млн.

Основная обязанность по уходу за членами семьи и выполнению домашней работы ложится на женские плечи, поэтому, по сравнению с мужчинами, женщины сталкиваются с большими ограничениями по отношению объема времени и усилий, которые они могут посвятить оплачиваемой занятости и производительной деятельности.

Безработица за последние десять лет имеет тенденцию к снижению, примерно на 12,2 %, это видно из Табл. 5.

Таблица 5

Основные показатели безработицы в России за 2001–2012 гг. [5]

Год	Численность безработных, тыс. человек	Уровень безработицы, %	Среднее время поиска работы, месяцев
2001	6 423,7	10	8,8
2002	5 698,3	8,3	9,1
2003	5 959,2	12,1	9,7
2004	5 674,8	9	9,1
2005	5 262,8	9	8,4
2006	5 311,9	8,8	8,6
2007	4 588,5	6,5	8,2
2009	4 791,5	8,4	8,5
2012	5 636,3	7,5	8,2

Но при этом за анализируемый промежуток времени с 2001 по 2012 гг. средний возраст безработных практически не изменился. Стоит также сказать о снижении уровня безработицы с 10 до 7,5 %, что говорит об улучшении стабильности в стране. Своего минимального значения в России безработица достигла в 2007 г. на пике подъема отечественной экономики. Наблюдается достаточно высокое среднее время поиска работы, хоть при этом оно сократилось с 8,8 до 8,2 месяца.

Реализованы программы по сдерживанию безработицы, предусматривающие дополнительные мероприятия, направленные на снижение напряженности на рынке труда субъектов РФ, а именно: предотвращение социальных рисков, связанных с возможным сокращением и вынужденным простоем работников.

Целевым показателем программ является сдерживание зарегистрированной безработицы по каждому субъекту РФ. Для реализации программ выделены средства из бюджета Российской Федерации, составляющие 95 % финансирования, а также средства региональных бюджетов, размер которых равен соответственно 5 %.

Безработица во многих странах, ведет к поиску работы и, соответственно, к миграции населения. По данным Табл. 6, миграционный прирост на период с 2003 по 2013 гг. увеличился примерно в 3,3 раза.

Таблица 6

Миграционная динамика в России за 2003–2013 гг. [5]

Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Прибыло всего (тыс. человек)	2662,3	2334	2201,9	2168,1	2117,4	2088,6	2122	2284,9	2215,9	1987,5	2102,3
Выбыло всего (тыс. человек)	2420,5	2252,2	2114,7	2124,2	2076,1	1981,2	1989,7	2044,9	1973,8	1740,1	1944,2
Миграционный прирост (тыс. человек)	241755	81781	87149	43884	41275	107432	132319	239943	242106	247449	158078

Причиной сокращения безработицы также послужило прибытие населения в страну и динамичное сокращение покидающих пределы России. За 2013 г. число мигрантов внутри России увеличилось на 5,7 % по сравнению с предыдущим. Миграционный прирост населения России снизился за последние десять лет на 36,1 % в результате уменьшившегося числа зарегистрированных по месту жительства прибывших в Российскую Федерацию (на 88,3 тыс. человек, или на 31,5 %), в том числе за счет иммигрантов из стран СНГ – на 89,6 тыс. человек, или на 34,2 %.

Наряду с этим отмечено увеличение числа выбывших за пределы России на 11,7 %, в том числе в страны СНГ – на 0,9 тыс. человек, или на 4,3 %.

Большую часть легальных трудовых иммигрантов приняла Россия, благодаря быстрому развитию экономики и емкому рынку труда. Согласно данным Федеральной миграционной службы России (ФМС России), одним из крупнейших поставщиков, который обеспечивает около 1/3 всего притока легальной рабочей силы в Российскую Федерацию, является Украина. В 2009 г. появился новый лидер – Узбекистан, на который пришлось более 20 % всех трудовых иммигрантов. Также поднялся в рейтинге и Таджикистан.

Важнейшими характеристиками качественного состава населения, и особенно его трудоспособной части, является образовательный уровень и профессиональное мастерство.

Наибольший выпуск студентов среднего профессионального образования с 2004 по 2013 гг. приходится на специальности: экономика и управление (произошло динамичное увеличение количества выпускников практически в 12 раз); гуманитарные науки имеют отрицательный прирост в выпуске студентов на 27,8 %; специальности: образование и педагогика, здравоохранение также претерпевают спад выпускников на 33,4 % и 64,6 % соответственно. Количество выпускников технических специальностей осталось практически неизменным (прирост 4 %), архитектура и строительство имеет отрицательную динамику (прирост – 86,6 %).

Целями демографической политики РФ на период до 2025 г. являются стабилизация численности населения к 2015 г. на уровне 142–143 млн человек и создание условий для ее роста к 2025 г. до 145 млн человек, а также повышение качества жизни и увеличение ожидаемой продолжительности жизни к 2015 г. до 70 лет, а к 2025 г. – до 75 .

Достижение целей демографической политики Российской Федерации в значительной степени зависит от успешного решения широкого круга задач социально-экономического развития. Они включают обеспечение стабильного экономического роста и рост благосостояния населения, снижение уровня бедности и уменьшение дифференциации по доходам, интенсивное развитие человеческого капитала и создание эффективной социальной инфраструктуры (здравоохранение, образование, социальная защита населения), рынка доступного жилья, гибкого рынка труда, улучшение санитарно-эпидемиологической обстановки.

К 2025 г. предполагается:

- обеспечить постепенное увеличение численности населения (в том числе за счет замещающей миграции) до 145 млн человек;
- увеличить ожидаемую продолжительность жизни до 75 лет;
- увеличить в 1,5 раза суммарный коэффициент рождаемости, снизить уровень смертности в 1,6 раза;
- обеспечить миграционный прирост на уровне более 300 тыс. человек ежегодно.

Одной из важнейших долгосрочных целей экономической политики правительства РФ является стимулирование экономического роста, поддержание его темпов на стабильном и оптимальном уровне.

Выведение экономики на пути устойчивого экономического роста возможно на основе повышения производительности труда и преобразований в структуре занятости экономически активного населения. Необходимо сочетать высокую эффективность производства с предоставлением гарантий занятости, включая поддержку предприятий и создание новых рабочих мест.

Основной характеристикой экономического роста является валовой внутренний продукт (Табл. 7).

Таблица 7

Валовой региональный продукт в текущих ценах, млрд руб. [5]

Год	2005	2006	2007	2008	2009	2010	2011	2012	2013
ВВП млрд руб.	10 819,2	13 208,2	17 027,2	21 609,8	26 917,2	33 247,5	41 276,8	38 807,2	45 172,7

За период с 2005 по 2013 гг. заметен рост данного показателя в 4,17 раза, что свидетельствует об улучшении экономической ситуации в стране.

Для рассмотрения прогнозных значений ВВП автором была исследована модель зависимости ВВП от численности экономически активного населения (Табл. 8).

Таблица 8

Изменения ВВП и численности экономически активного населения [5]

Год	ВВП (млрд руб.)	Численность экономически активного населения (тыс. человек)
2005	10 819,20	72357,1
2006	13 208,20	72391,4
2007	17 027,20	72949,6
2008	21 609,80	73431,7
2009	26 917,20	74166,9
2010	33 247,50	75159
2011	41 276,80	75756,6
2012	38 807,20	75657,7
2013	45 172,70	75439,9

Для проверки корректности данных был проведен множественный корреляционно-регрессионный анализ, а именно: определена теснота связи между признаками, отобраны факторные признаки в модель, определен вид уравнения регрессии, построена регрессионная модель и оценены ее параметры, проверена значимость параметров связи, их интервальная оценка. Полученные данные показали, что модель является верной и адекватной, следовательно, прогнозные значения будут рассчитаны с наибольшим процентом точности.

Для дальнейшего построения прогноза были определены прогнозные значения численности экономически активного населения с 2014 по 2016 гг., которые представлены в Табл. 9. Доверительный интервал изображен на рис. 2.

Таблица 9

Прогнозные значения численности экономически активного населения

Год	Прогнозная численность (тыс. человек)	Границы	
		нижняя граница (тыс. человек)	верхняя граница (тыс. человек)
2014	77 100,48	76 284,6	77 916,3
2015	77 630,32	76 789,0	78 471,7
2016	77 892,43	76 996,5	78 788,4

С помощью корреляционно-регрессионных вычислений были получены все необходимые данные для построения прогноза ВВП, зависящего от численности экономически активного населения. Результаты представлены в Табл. 10 и на рис. 3.

Прогнозные значения находятся в границах доверительного интервала, а значит, найденные значения имеют смысл.

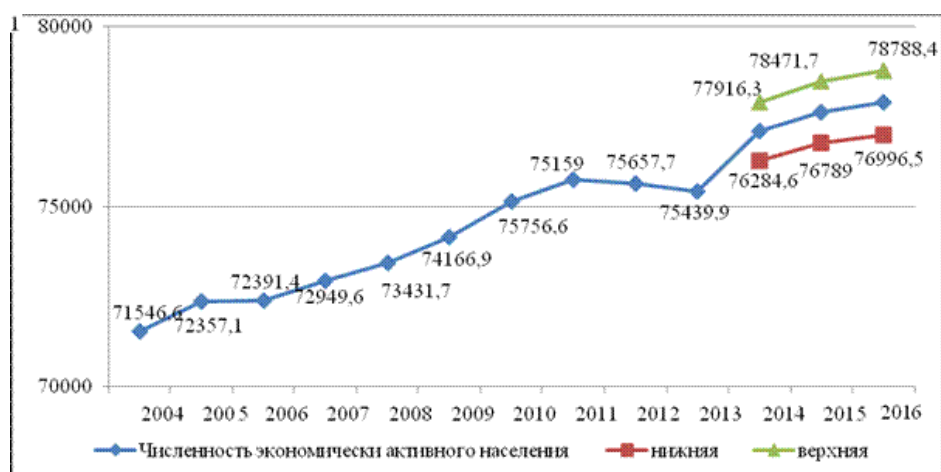


Рис. 2. Прогнозная численность экономически активного населения (тыс. человек)

Таблица 10

Прогнозные значения ВВП и численности экономически активного населения

Год	ВВП (млрд руб.)	Экономически активное население (тыс. человек)
2014	54 374,87	77 100,48
2015	58 356,59	77 630,32
2016	60 672,63	77 892,43

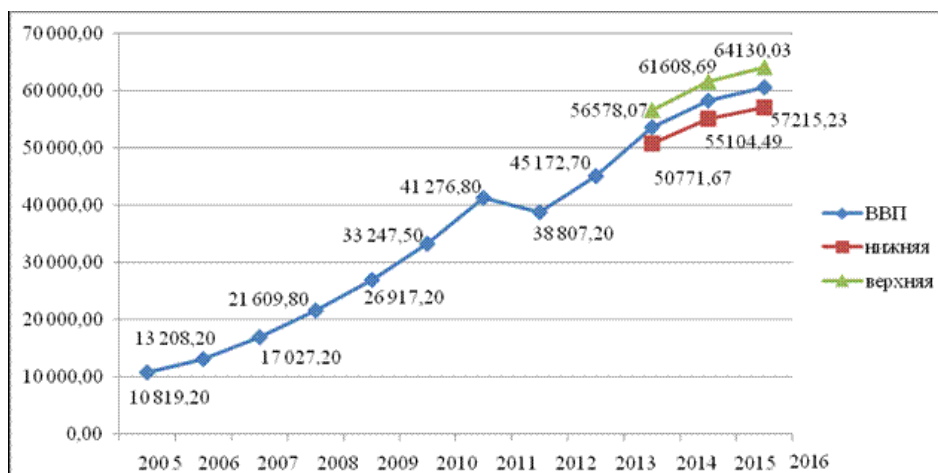


Рис. 3. Прогнозные значения объема ВВП (млрд руб.)

По результатам прогнозирования можно сделать вывод, что за период с 2013 по 2016 гг. численность экономически активного населения будет увеличиваться на 3,25 %. Благодаря этому на 34,5 % вырастет ВВП а следовательно, экономическая обстановка в России станет улучшаться.

Заключение

В рамках данного проекта была проанализирована динамика показателей трудового потенциала, их взаимосвязь. С помощью теоретических и эмпирических моделей осуществлен прогноз и сделаны выводы по развитию трудового потенциала в России на ближайшее время.

Результаты проекта определяют характеристику трудового потенциала России в настоящее время. Полученные прогнозные значения помогут решить уникальную, не имеющую исторических аналогов задачу по поддержанию высоких темпов экономического роста в условиях сокращающегося населения и, прежде всего, его экономически активной части.

Публикации

1. Волкова О.В. Трудовой потенциал национальной экономики // Межвузовский сб. научных трудов. Вып. 17. СПб., 2012. С. 60–62.

Литература

1. Аксенова Н.В. Роль и место организационных знаний в структуре задач управления компанией // Философские дескрипты: Сб. статей. Барнаул, 2006. С. 11–17.
2. Балдин К.В., Быстров О.Ф., Соколов М.М. Эконометрика: Учеб. пособие для вузов. М., 2004. – 254 с.
3. Булгаков М., Кузьмина Н. Оценка трудового потенциала как основа его эффективного использования // Человек и труд. 2005. № 2. С. 34–67.
4. Шлендер П.Э. Экономика трудовых ресурсов: Учеб. пособие. М., 2008. – 302 с.
5. Доклад Министра здравоохранения и социального развития Российской Федерации Татьяны Голиковой на тему «Итоги реализации приоритетного национального проекта „Здоровье“ в 2006–2010 годах» (28.02.2011): <http://www.minzdravsoc.ru/health/prior/75>.
6. Федеральная служба государственной статистики: <http://www.gks.ru>



Жуков Михаил Валерьевич
Факультет точной механики и технологий
Кафедра нанотехнологий и материаловедения
E-mail: cloudjyk@yandex.ru

ОСОБЕННОСТИ ПРОВЕДЕНИЯ ИЗМЕРЕНИЙ ОРГАНИЧЕСКИХ И НЕОРГАНИЧЕСКИХ ОБЪЕКТОВ МЕТОДОМ АТОМНО-СИЛОВОЙ МИКРОСКОПИИ В ЖИДКИХ СРЕДАХ

М.В. Жуков

Научный руководитель – доцент В.В. Левичев

Введение

Одной из актуальных и мало изученных областей науки и техники является сканирование неорганических и биологических объектов (клеток, бактерий и пр.) в жидкостных растворах методом атомно-силовой микроскопии (АСМ) [1, 2]. Данные работы крайне актуальны, так как позволяют проводить прецизионные высокоточные *in situ* исследования биологических объектов в нативном состоянии [3].

Однако применение метода АСМ связано с рядом трудностей, которые возникают при проведении измерений. К ним следует отнести сложность закрепления биологических объектов на подложке в нативном состоянии, а также ухудшение разрешения и возможность появления паразитных шумов при работе с жидкостными средами [4], обусловленных капиллярными силами, поверхностным натяжением, микротечениями и свойствами жидкости (вязкость, плотность и др.).

Одним из основных элементов, отвечающих за разрешающую способность метода АСМ, являются особенности зонда, так как получаемое изображение всегда представляет собой конволюцию (свертку) геометрических параметров зонда и поверхности изучаемого образца [5]. В последнее время для улучшения их параметров применяются различные методы выращивания на кончиках зондов нановискерных (НВ) структур, позволяющих значительно повысить разрешающую способность метода АСМ [6, 7].

Данный проект направлен на решение перечисленных проблем, возникающих при работе в жидкости в ходе измерений объектов различной природы, а также рассмотрению возможности применения специализированных зондов с НВ в жидкостных средах для получения более высокоточной и достоверной информации при проведении исследований методом АСМ.

Цель и задачи проекта

Цель проекта: выявление особенностей измерения неорганических и органических объектов в различных буферных растворах с использованием стандартных зондов и зондов с НВ методом АСМ.

Для решения поставленной цели были обозначены следующие задачи:

1. Провести измерения эталонной тестовой решетки в различных буферных растворах в полуконтактном и контактном режимах сканирования.
2. Провести сканирование биологических объектов методом АСМ в жидкости.
3. Выполнить анализ полученных данных и выявить особенности, возникающие при сканировании объектов в буферных растворах методом АСМ.
4. Сравнить изображения, полученные с помощью стандартных зондов и зондов с НВ, проанализировать полученные результаты для обоих случаев.

Методы исследования

Для создания и модификации зондов с НВ использовался двухлучевой электронно-ионный микроскоп CrossBeam Neon 40 (Carl Zeiss, Германия) (рис. 1) [8]. Формирование НВ производилось с помощью газа-прекурсора на вершине зонда под фокусированным электронным пучком в высоком вакууме с контролем параметров выращиваемого НВ (длины, диаметра, наклона и состава) [9].

Для непосредственного тестирования и выявления особенностей проведения измерения в жидких средах использовался сканирующий зондовый микроскоп (СЗМ) Ntegra Aura (NT-MDT, Россия) (рис. 2) в полуконтактной и контактной силовых модах [10].

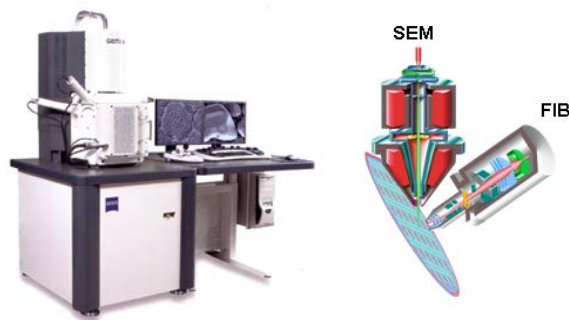


Рис. 1. Общий вид установки Carl Zeiss CrossBeam Neon 40 и расположение электронной (SEM) и ионной (FIB) колонн



Рис. 2. Зондовый микроскоп Ntegra Aura:
1 – базовый блок, 2 – измерительная головка, 3 – система виброизоляции,
4 – система видеонаблюдения, 5 – система управления

Специализированные зонды с нановискерами

В работе использовались зонды с НВ длиной 400–600 нм, диаметром 30–40 нм и радиусом скругления кончика менее 10 нм. Зонды с НВ сохраняют свои параметры после 2-3 циклов сканирования, что сопоставимо с таковыми значениями для стандартных зондов (рис. 3). НВ изготавливались на двухлучевом электронно-ионном микроскопе методом осаждения сфокусированным электронным пучком в вакууме в присутствии газа-прекурсора (Pt).

Состав НВ: 30 % платина, 70 % остатки газа-прекурсора и частицы углерода. Зонды изготавливались под углом наклона НВ относительно острия в $\sim 22\text{--}25^\circ$ для уменьшения его колебаний при двойном касании за счет компенсации угла держателя чипа с кантилевером и зондом относительно поверхности образца [11].

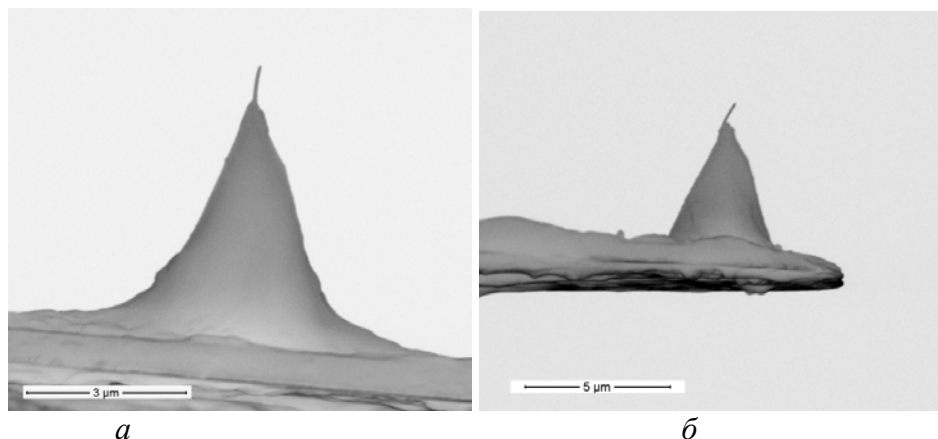
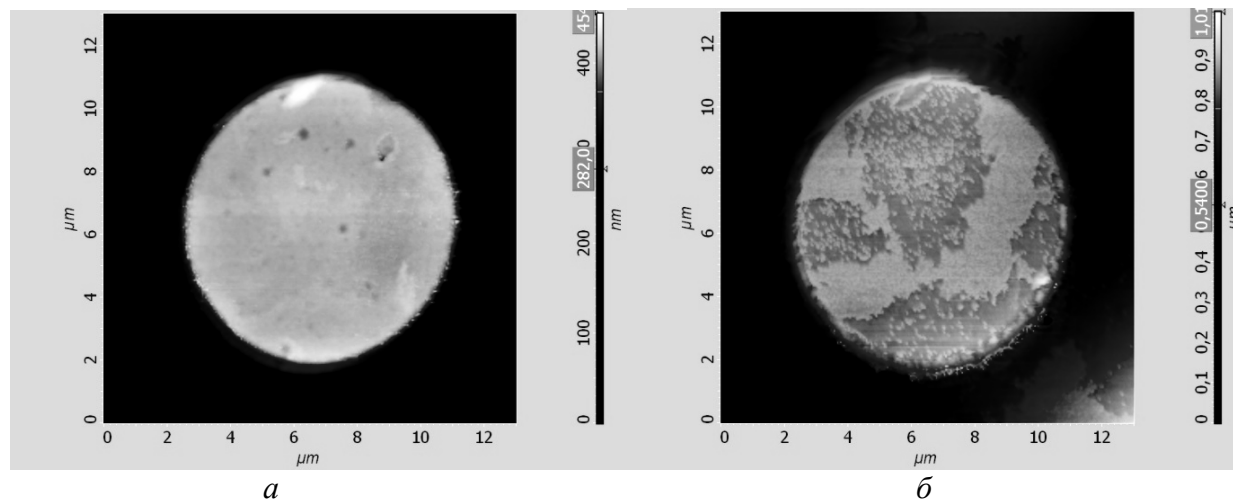


Рис. 3. СЗМ-изображения кантилеверов с НВ на кончиках зондов, полученные после проведения сканирования: а – вид спереди, б – вид сбоку

В качестве основы для создания зондов с НВ использовались нитридные зонды SNL-10 фирмы Bruker [12], предназначенные для проведения измерений в жидкостной среде. Для исследований в воздушной среде использовались зонды фирмы NT-MDT серии NSG01 [13]. Работа зондами Bruker проводилась как на короткой (120 мкм), так и на длинной (205 мкм) балках. Лучший подвод и качество отображения исследуемых объектов было обнаружено на кантилеверах с большими коэффициентами жесткости.

Исследование биологических объектов

Изучение кровяных телец (эритроцитов) велось методом АСМ как в воздушной среде стандартными зондами и зондами с НВ в полуконтактной силовой моде, так и в жидкости с применением гелеобразующих веществ (желатина и агар-агара) для фиксации бактерий *E. Coli* в нативном состоянии. Для сравнения был исследован один и тот же эритроцит, что видно по дефекту в верхней части и линиям перепада высот на его поверхности (рис. 4а, 4б).



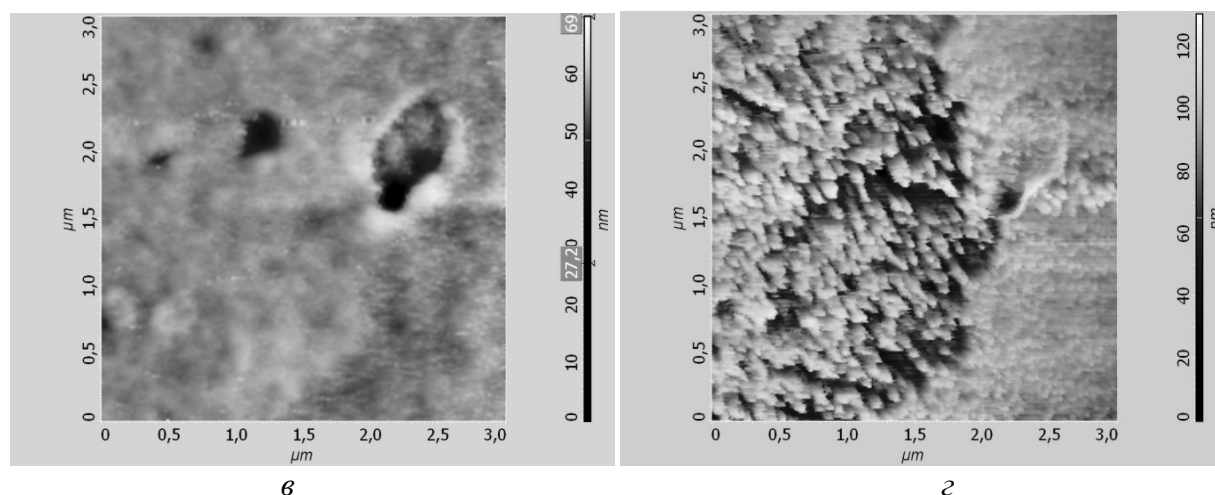


Рис. 4. СЗМ-изображения эритроцита 12х12 мкм, полученные: а – стандартным зондом; б – зондом с НВ. Изображения поверхности мембраны эритроцита 3х3 мкм, полученные: в – стандартным зондом; г – зондом с НВ

Обнаружено, что при использовании зондов с Pt/C НВ существенно повышается качество изображения и разрешающая способность СЗМ. Зонды с НВ отображают поверхность значительно детальней и четче стандартных зондов за счет высокого аспектного отношения и малого радиуса скругления кончика. Особенно сильно эти различия видны при увеличении масштаба области сканирования (рис. 4в, 4г). При сканировании мембран эритроцитов можно увидеть наноразмерные объекты, которые стандартным зондом обнаружить не удалось.

Для иммобилизации бактерий *E.Coli*, предоставленных Институтом цитологии РАН (Санкт-Петербург), на поверхности пластиковых чашек Петри использовались гелеобразующие вещества желатин и агар-агар. Желатин (агар-агар) массовой долей 0,5 % и 1,8 % разводился в дистиллированной воде в термоустойчивой стеклянной посуде, а затем доводился до кипения. Остывший до $\sim 60^\circ\text{C}$ раствор разливался в заранее подготовленные стерильные чашки Петри, после чего при комнатной температуре оставался до полного затвердевания [4].

Для исследования бактерий *E.Coli* методом АСМ было решено использовать агар-агар массовой долей 1,8 % в исходном растворе, так как именно в этом случае была выявлена наилучшая иммобилизация этих бактерий [4].

В ходе измерений методом АСМ удалось визуализировать отдельную бактерию *E.Coli* в жидкости (дистиллированной воде), зафиксированную на пленке агар-агара (рис. 5а). Однако оказалось, что из-за наличия микротечений в жидкости, где находился образец, и высокой эластичности пленки агар-агара в ряде случаев при попадании жидкости под пленку наблюдалось ее набухание и деформация. Поэтому не удавалось измерить с высоким пространственным разрешением относительно протяженные участки поверхности и визуализировать группы бактерий *E.Coli*.

Для устранения данного эффекта было решено нанести агар-агар на пластинку слюды таким образом, чтобы на ее поверхности образовалась тонкая пленка. Так удалось визуализировать не только одиночные бактерии, но и их группы (рис. 5б). Результаты исследований позволили получить оценки размеров бактерий *E.Coli*, соответствующие литературным данным (диаметр $\sim 0,5$ мкм, длина $\sim 2-3$ мкм) [14].

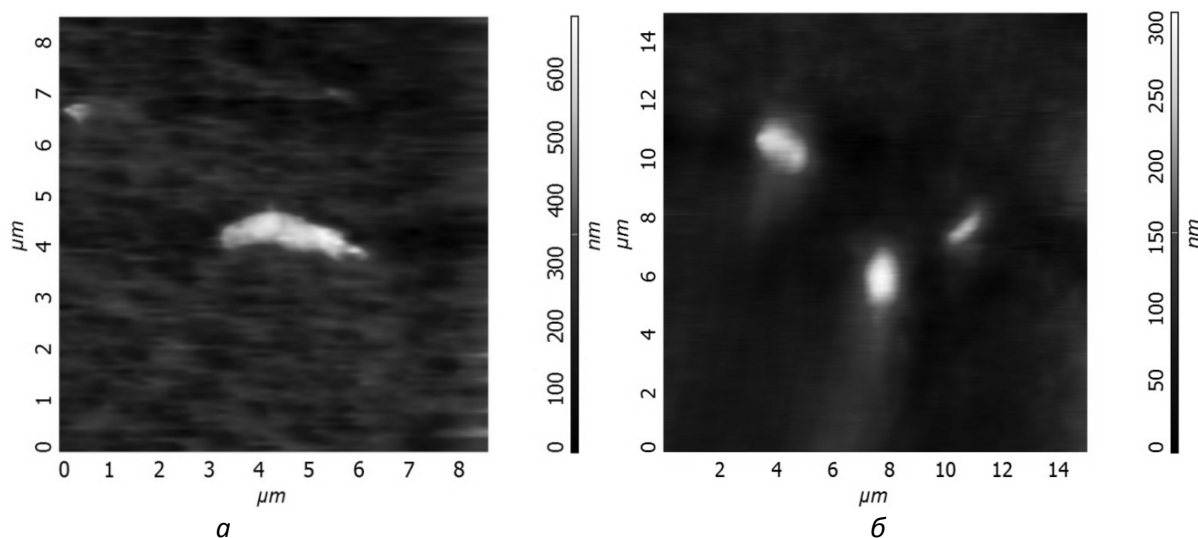


Рис. 5. СЗМ-изображения бактерий *E. Coli*, зафиксированных на пленке агар-агара (массовая доля 1,8 %) в жидкости (дистиллированной воде): а – без слюды; б – на слюде

Измерения эталонной тестовой решетки TGQ01

Тестовым объектом для изучения особенностей сканирования объектов исследования методом АСМ в различных средах служила калибровочная решетка TGQ01 (НТ-МДТ, Россия) (рис. 6) [15].

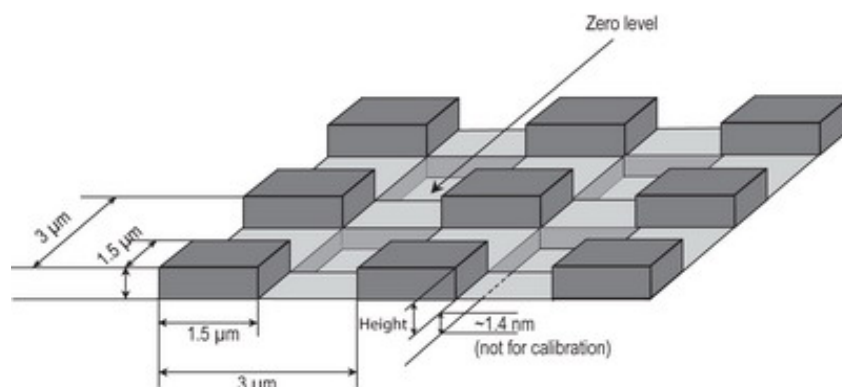


Рис. 6. Схематичное изображение калибровочной решетки TGQ01

Средние значения параметров калибровочной решетки TGQ01, которые можно считать эталонными, приведены в Табл. 1.

В качестве тестовых жидкостных сред были выбраны: дистиллированная вода, натрий-фосфатный буфер (PBS, Phosphate buffered saline), боратный буфер (рН 9,18), растворы NaOH 0,1 М и NaOH 0,5 М. Исследования проводились в контактом и полуконтактном режимах АСМ.

Таблица 1.

Параметры калибровочной решетки TGQ01

Высота калибровочной площадки, нм	Период между калибровочными площадками по X, мкм	Период между калибровочными площадками по Y, мкм	Размер калибровочной площадки по X, мкм	Размер калибровочной площадки по Y, мкм
20±1,5	3±0,05	3±0,05	1,5±0,35	1,5±0,35

В процессе исследований было выяснено, что при сканировании тестовой решетки в буферных растворах наблюдается уменьшение линейных размеров периода между площадками по осям x и y на ~8,3 % (с 3 до ~2,75 мкм). Значения по высоте варьируются в достаточно широких пределах относительно эталона. Для полуконтакта они составляют от

18 до 28 нм, что может объясняться наличием частотных шумов при сканировании, а для контакта – от 13 до 22 нм (небольшое уменьшение высоты объясняется отсутствием влияния частот на изображение и иным характером взаимодействия зонда с образцом в жидкости при прямом «ощупывании» поверхности).

Лучший результат по соответствию линейным размерам при работе в буферных растворах был получен с использованием длинного кантилевера в полуконтакте. Отклонения от эталона составили 3–12,25 % в дистиллированной воде, 3–14,22 % в PBS, 1,33–13,44 % в боратном буфере, 3,22–20,22 % в NaOH 0,1М и 5,73–16,18 в NaOH 0,5М растворах.

Наилучшие по качеству изображения получены при работе с длинными кантилеверами в контакте. При работе в полуконтакте выявлено смазывание изображения (за исключением буфера PBS и NaOH 0,1М) и зашумленность при работе на воздухе.

Лучшими при работе в жидких средах следует признать зонды на длинных балках, причем для получения наилучшего соответствия размерам стоит использовать полуконтактные режимы сканирования, а для получения наилучшего качества – контактную методику измерения. Лучшие буферы – PBS и натриевые гидроксиды, причем самый легкий подвод был обнаружен именно при работе в PBS, а лучшее качество замечено в натриевых гидроксидах. Буфер PBS проявил себя хорошо как в качестве среды для получения качественных СЗМ-изображений, так и в плане соответствия размеров эталонным значениям. Поэтому сравнение стандартных и специализированных зондов с НВ начинается с его использования.

Для сравнения результатов проведено тестирование в воздушной среде и рассмотрены растворы натриевых гидроксидов, также проявивших себя как годные среды для сканирования в жидкости методом АСМ. Зонды с НВ для повышения стабильности работы при силовых нагрузках при сканировании и обеспечения неразрушающего взаимодействия с поверхностью преимущественно использовались в полуконтактном режиме.

Лучшие результаты по качеству отображения калибровочной решетки TGQ01 выявлены при использовании специализированных зондов с НВ в PBS-буфере (рис. 7, 8). Замечено общее снижение уширения фронтов перепадов высот ячеек решетки при использовании зондов с НВ и увеличение шероховатости поверхности, что может говорить о лучшей проникающей способности данного типа зондов (Табл. 2, рис. 9–11). Однако на границе перепадов высот в верхней и нижней областях замечены колебания, обусловленные высоким аспектным отношением НВ (наличие двойных касаний при сканировании). При сканировании зондом в воздушной среде также отсутствовали вибрации, выявленные при сканировании стандартным зондом SNL-10 фирмы Bruker.

Изображения в растворе NaOH 0,1М оказались более смазанными, чем в случае сканирования стандартными зондами, однако результаты в растворе NaOH 0,5М гораздо четче при сканировании зондом с НВ, но более зашумлены. В последних двух средах был замечен артефакт «проваливания» изображения в областях перепадов высот. Это может быть связано с химическими свойствами сред, так как щелочные буферные растворы довольно «агрессивны» и могут разъесть углеродную часть НВ структур в процессе сканирования.

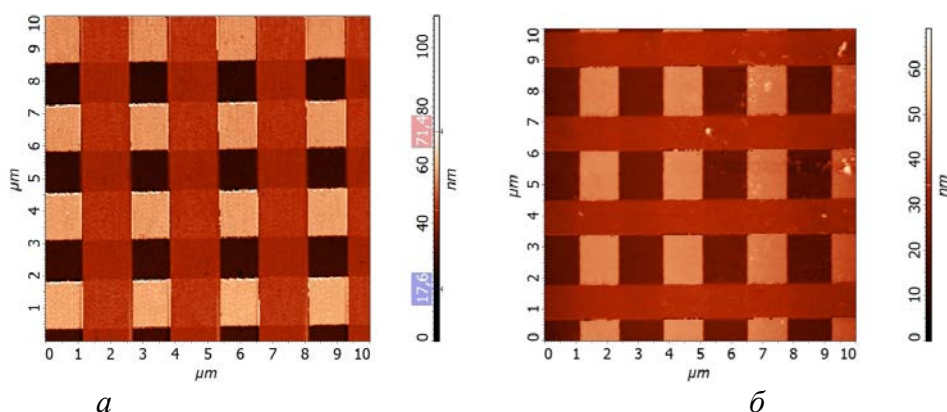


Рис. 7. СЗМ-изображения решетки TGQ01, полученные в PBS-растворе:
а – зондом с Pt-C НВ; б – стандартным зондом

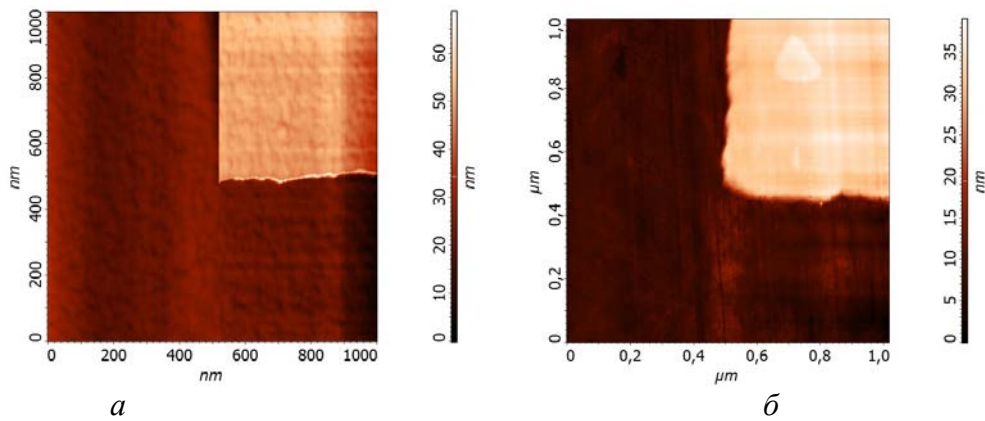


Рис. 8. СЗМ-изображения края ячейки калибровочной решетки TGQ01, полученные в PBS-растворе: а – зондом с Pt-C НВ; б – стандартным зондом

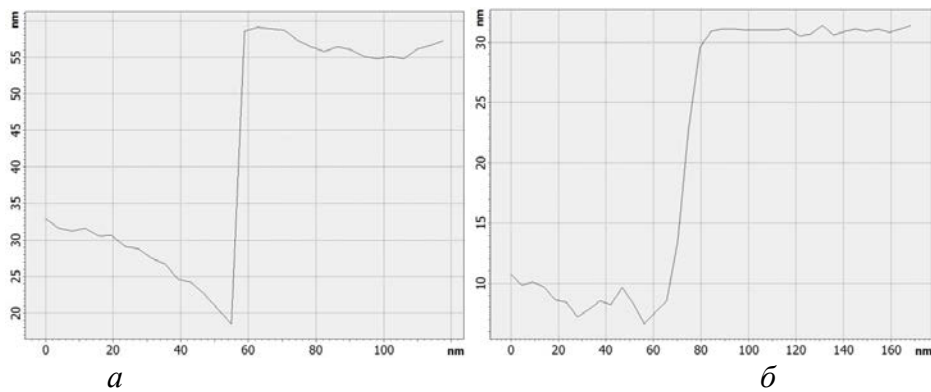


Рис. 9. Сечения ячейки по оси X, полученные при работе: а – зондом с Pt-C НВ; б – стандартным зондом

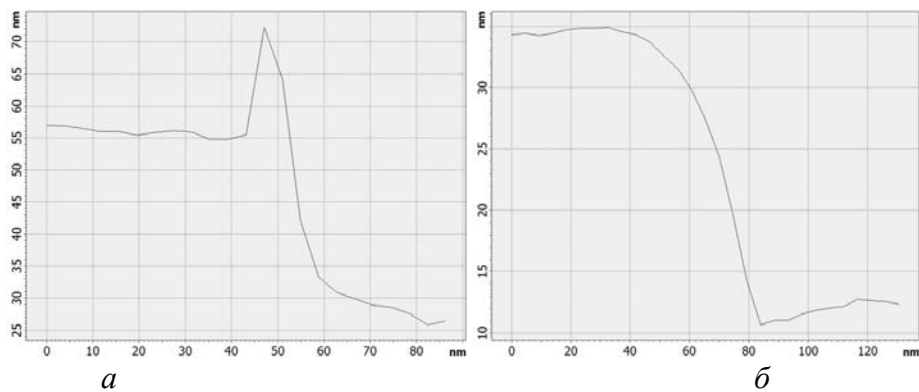


Рис. 10. Сечения ячейки по оси Y, полученные при работе: а – зондом с Pt-C НВ; б – стандартным зондом

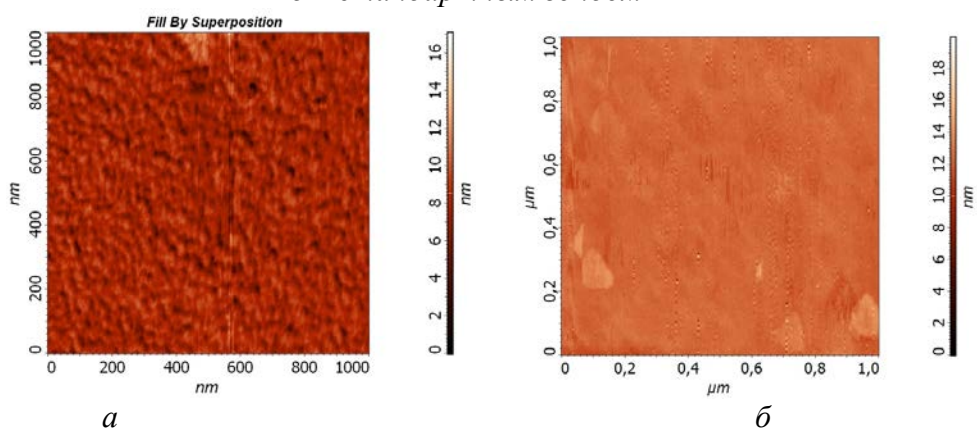


Рис. 11. СЗМ-изображение поверхности ячейки калибровочной решетки TGQ01, полученное в PBS-растворе: а – зондом с Pt-C НВ; б – стандартным зондом

Таблица 2

Средние значения перепадов высот и шероховатостей, полученные в ходе измерений калибровочной решетки TGQ01 в различных средах с использованием короткого кантилевера АСМ-зондов в полуконтактном режиме сканирования

Параметр	Уширение по X, нм		Уширение по Y, нм		Шероховатость, нм	
	Стандартный зонд	Зонд с НВ	Стандартный зонд	Зонд с НВ	Стандартный зонд	Зонд с НВ
1. Воздух	59	39	59	11	0,15	0,17
2. PBS	19	4	37	12	0,32	0,89
3. NaOH 0,1M	31	4–12	63	55	7,23 (шумы)	0,78
4. NaOH 0,5M	16–140	16–100	170–180	80–110	0,37	0,75

Увеличение погрешности по высоте замечено при отображении решетки на воздухе с использованием зондов с НВ. Связано это с колебаниями при двойном касании на границах перепадов высот. Однако размеры по периоду отображаются более достоверно. Длины ячеек калибровочной решетки при сканировании зондом с НВ уменьшились, что может быть связано с более четкой прорисовкой границ ячеек при сканировании. Что касается разрешения отображения поверхности, то оно существенно улучшилось, исчезли паразитные колебания и шумы, обнаруженные на решетке при сканировании стандартным зондом.

Таблица 3

Значения погрешностей, полученные в ходе измерений калибровочной решетки TGQ01 в различных средах с использованием короткого кантилевера АСМ-зондов в полуконтактном режиме сканирования

Среда	Зонд	Высота, % от эталона	Период по Y, % от эталона	Период по X, % от эталона	Размер по Y, % от эталона	Размер по X, % от эталона
1. Воздух	Стандартный зонд	1,80	0,61	5,16	9,11	4,22
	Зонд с НВ	12,25	0,20	0,19	13,12	22,23
2. PBS	Стандартный зонд	39,14	9,72	9,89	5,56	15,22
	Зонд с НВ	4,30	9,30	9,50	1,10	19,60
3. NaOH 0,1M	Стандартный зонд	30,90	9,89	9,39	6,67	13,67
	Зонд с НВ	12,90	9,33	9,50	4,47	18,00
4. NaOH 0,5M	Стандартный зонд	24,88	10,83	10,56	3,22	17,00
	Зонд с НВ	3,22	8,27	8,65	0,33	17,36

Как видно из Табл.3, зонд с НВ в PBS-растворе дает уменьшение погрешности по высоте примерно в 10 раз. Однако на нижнем краю по оси Y заметны небольшие колебания, что связано с формой НВ, его большим аспектным отношением. Размеры по периоду решетки также отображаются более четко, особенно это видно на изображениях и сечениях перепадов высот краев ячеек.

При работе в растворах NaOH зонд с НВ дает более точное отображение по высотам. Выявлено немного лучшее соответствие по периоду (особенно для NaOH 0,5M), а также общее улучшение точности отображения длины ячейки по Y, но ухудшение по X. Следует учитывать, что размеры по длине ячейки предоставляются изготовителем с большой степенью погрешности, поэтому данный параметр изначально имеет большой разброс значений.

В целом видно улучшение разрешающей способности и соответствия эталонным размерам при использовании в исследуемых средах зондов с НВ по сравнению со стандартными зондами.

Области возможного использования результатов проекта

Методы атомно-силовой микроскопии позволяют проводить прецизионные высокоточные *in situ* исследования биологических объектов в нативном состоянии при наличии питательных сред (буферов).

С помощью разрабатываемых специализированных зондов возможно высокоточное исследование наноразмерных структур с высоким разрешением, изучение топологии различных биологических и неорганических объектов, нано- и микроструктур, контролировать качество продукции, а также производить модификацию поверхности методами высокоточной литографии.

В отличие от стандартных зондов, зонды с НВ гораздо лучше прорисовывают области резких перепадов высот за счет более высокого аспектного отношения НВ. Эти зонды обладают лучшей проникающей способностью, а также дают более контрастное и достоверное изображение при сканировании нанорельефных поверхностей в области 2–50 нм (мезоразмеры).

Используемые в НВ материалы (30 % платина, 70 % углерод) приводят к их более высокой износостойкости и прочности. НВ обладают гидрофобными свойствами, способствующими исследованию поверхности, покрытой мономолекулярной водяной пленкой.

Таким образом, данная модификация зондов может использоваться для высокоточных исследований объектов различной природы, лабораторных работ по модификации и анализу поверхностей, а также в промышленности для создания и контроля нано- и микроразмерных структур.

Заключение

В работе выявлены особенности измерения неорганических и органических объектов в различных буферных растворах с использованием стандартных зондов и зондов с НВ методом АСМ.

В процессе выполнения работы были выполнены следующие этапы:

1. Проведены измерения эталонной тестовой решетки в различных буферных растворах в полуконтактной и контактной силовых модах.

2. Проведены измерения объектов биологической природы в различных средах методом АСМ.

3. Выполнен анализ полученных данных и выявлены особенности сканирования объектов в буферных растворах методом АСМ.

4. Проведено тестирование стандартных зондов и зондов с НВ с последующим сравнением и анализом полученных результатов, причем:

– выявлено общее улучшение проникающей способности и контраста отображения резких перепадов высот зондами с НВ;

– обнаружено улучшение отображения высот и периода зондами с НВ относительно стандартных зондов по эталонным значениям;

– для получения наилучшего пространственного разрешения и соответствия размерам по осям X , Y , Z рекомендуется использовать PBS-растворы.

Таким образом, все достигнутые результаты в рамках реализации данной работы соответствуют поставленным целям и задачам.

Публикации

1. Кухтевич И.В., Жуков М.В., Чубинский-Надеждин В.И., Букатин А.С., Евстапов А.А. Фиксация бактерий *E. Coli* на подложке для измерений в жидкости методом атомно-силовой микроскопии // Научное приборостроение. 2012. Т. 22. № 4. С. 56–61.

2. Белоусова О.Д., Гайдаш А.А., Толмачев И.А., Ивченко Е.В., Голубок А.О., Левичев В.В., Мухин И.С., Жуков М.В., Белоусов И.С., Ткачук И.В. Структура цитоплазматических мембран и химический состав эритроцитов в раннем периоде раневых повреждений по данным сканирующей зондовой микроскопии // Военно-медицинский журнал. 2013. Т. СССXXXIV. № 1. С. 19–24.
3. Левичев В.В., Жуков М.В., Мухин И.С., Денисюк А.И., Голубок А.О. Об устойчивости работы сканирующего силового микроскопа с нановискером на вершине зонда // Журнал технической физики. 2013. Т. 83. Вып. 7. С. 115–120.
4. Белоусова О.Д., Толмачев И.А., Гайдаш А.А., Левичев В.В., Жуков М.В., Белоусов И.С. Особенности подготовки проб крови для исследования морфологических параметров и структуры мембран эритроцитов методом атомно-силовой микроскопии // Клиническая медицина. Судебная медицина. Электронный биомедицинский журнал Medline.ru. 2012. Т. 13. С. 954–966 (<http://medline.ru/public/art/tom13/art80.html>).
5. Жуков М.В., Левичев В.В. Исследование влияния нановискеров на контраст и пространственное разрешение СЗМ изображений // Сб. тезисов докладов конгресса молодых ученых. СПб, 2012. С. 185–187.
6. Жуков М.В., Левичев В.В. Модифицированные зонды с нановискерами для сканирующей зондовой микроскопии // Студенты, аспиранты и молодые ученые – малому наукоемкому бизнесу: Мат-лы всероссийской молодежной научной школы. СПб, 2012. С. 30–32.
7. Жуков М.В., Левичев В.В. Модифицированные зонды с металлорганическими нановискерами для сканирующей зондовой микроскопии // Наука и инновации в технических университетах: Мат-лы Шестого всероссийского форума студентов, аспирантов и молодых ученых. СПб, 2012. С. 83–84.
8. Жуков М.В., Левичев В.В. Исследование влияния нановискеров на контраст и пространственное разрешение СЗМ изображений // Аннотированный сборник научно-исследовательских выпускных квалификационных работ бакалавров УНИВЕРСИТЕТ ИТМО. СПб., 2012. С. 6–7.
9. Жуков М.В., Кухтевич И.В., Чубинский-Надеждин В.И., Букатин А.С., Евстратов А.А. Фиксация бактерий *E.Coli* в жидкости во время измерений методом атомно-силовой микроскопии // Сб. статей Четвертой научно-педагогической конференции «Высокие технологии, фундаментальные и прикладные исследования в физиологии, медицине, фармакологии». СПб, 2012. С. 213–216.
10. Жуков М.В., Кухтевич И.В., Левичев В.В. Выявление особенностей проведения измерений методом атомно-силовой микроскопии в жидких средах // Сб. тезисов докладов конгресса молодых ученых. СПб, 2013. С. 325–326.
11. Кухтевич И.В., Жуков М.В., Евстратов А.А. Изучение бактерий *E.Coli* в жидкости методом атомно-силовой микроскопии // Сб. тезисов докладов конгресса молодых ученых. СПб, 2013. С. 331–332.

Литература

1. Touhami A., Jericho M.H., Beveridge T.J. Atomic force microscopy of cell growth and division in *Staphylococcus aureus* // Journal of bacteriology. 2004. Vol. 186. No 11. P. 3286–3295.
2. Lyubchenko Y.L. Preparation of DNA and nucleoprotein samples for AFM imaging // Micron. 2011. Vol. 42. No 2. P. 196–206.
3. Kailas L., Ratcliffe E.C., Hayhurst E.J. et al. Immobilizing live bacteria for AFM imaging of cellular processes // Ultramicroscop. 2009. Vol. 109. No 7. P. 775–780.
4. Кухтевич И.В., Жуков М.В., Чубинский-Надеждин В.И. Фиксация бактерий *E.Coli* на подложке для измерений в жидкости методом атомно-силовой микроскопии // Научное приборостроение. 2012. Т. 22. № 4. С. 56–61.
5. Миронов В.Л. Основы сканирующей зондовой микроскопии: Учеб. пособие для студентов старших курсов высших учебных заведений. Нижний Новгород, 2004. – 114 с.
6. Kado H., Yokoyama K., Tohda T. Atomic force microscopy using ZnO whisker tip // Review of Scientific Instruments. 1992. Vol. 63. No 6. P. 3330–3332.

7. Kado H., Yokoyama K., Tohda T. A novel ZnO whisker tip for atomic force microscopy // Ultramicroscopy. 1992. Vol. 42–44. P. 1659–1663.
8. Сайт фирмы Carl Zeiss / Сканирующие электронные микроскопы серии CrossBeam: <http://optec.zeiss.ru/electro/?n=23431403>.
9. Мухин И.С., Мухин М.С., Феклистов А.В., Голубок А.О. Специализированные СЗМ-зонды на основе каркасных висцерных структур // Научное приборостроение. 2011. Т. 21. № 3. С. 23–29.
10. Сайт фирмы НТ-МДТ / Модельный ряд: ИНТЕГРА Аура: <http://www.ntmdt.ru/modular-afm/aura>.
11. Левичев В.В., Жуков М.В., Мухин И.С. и др. Об устойчивости работы сканирующего силового микроскопа с нановискером на вершине зонда // Журнал технической физики. 2013. Т. 83. Вып. 7. С. 115–120.
12. Bruker AFM probes / SNL-10. Product Description: <https://www.brukerafmprobes.com/Product.aspx?ProductID=3693>.
13. Сайт фирмы НТ-МДТ / NSG01: <http://www.ntmdt-tips.com/products/view/nsg01>.
14. Prescott L.M., Harley J.P., Klein D. Microbiology. New York, 2008. – 1147 p. Сайт фирмы НТ-МДТ / TGQ1: <http://www.ntmdt-tips.com/products/view/tgq1>.



Зайцева Татьяна Игоревна
Факультет естественнонаучный
Кафедра высшей математики
E-mail: tania.zai1010@gmail.com.

АНАЛИЗ ПРОДОЛЬНЫХ И ПОПЕРЕЧНЫХ ПЕРЕМЕЩЕНИЙ РЕЛЬСОВОЙ ПЛЕТИ БЕССТЫКОВОГО ПУТИ ПРИ СТАТИЧЕСКИХ И ДИНАМИЧЕСКИХ НАГРУЗКАХ

Т.И. Зайцева

Научный руководитель – профессор А.М. Уздин
(ПГУПС, Санкт-Петербург)

Цель и задачи проекта

Цель работы – построение математических моделей для оценки напряженно-деформированного состояния (НДС) рельсовых плетей при изменении температуры и определение величины смещения рельса в месте его случайного излома.

Для этого рассмотрим две задачи, одна из которых предполагает составление уравнения для оценки НДС плети при статической температурной нагрузке и получение его решения. При этом необходимо учитывать неоднородность подрельсового основания по длине плети.

Другая задача связана с построением математической модели и решением соответствующих уравнений колебаний плети при разрыве с учетом жесткости и демпфирования в подрельсовом основании [1]. Известно, что реальные свойства балласта меняются в широких пределах в зависимости от того, является ли он свежеложенным или слежавшимся; происходит ли разрыв летом или в зимнее время, когда балласт смерзся.

В данной статье поводится анализ поперечных деформаций плети при поперечных ударах подвижного состава. Это важно при оценке поведения пути на мостах, особенно при наличии податливых сейсмоизолирующих опорных частей. В этом случае из-за резкого

изменения жесткости основания значительная часть поперечной нагрузки передается на него через рельс, что ведет к возникновению в рельсе опасных поперечных усилий.

Методы исследования

Для построения математических моделей и решения уравнений будут использованы классические методы строительной механики и теории дифференциальных уравнений.

Рассмотрим напряженно-деформированное состояние рельсового пути на подходах к однопролетному мосту с сейсмоизолирующими опорными частями при продольных нагрузках (рис. 1). Наиболее распространенными являются резиновые опорные части (РОЧ), однако их применение на железнодорожных мостах существенно ограничено. В процессе продольных сейсмических колебаний и гибкости сейсмоизолирующих опорных элементов пролетные строения могут смещаться относительно насыпи подходов на величину $u_0 = 15\text{--}30$ см в зависимости от длины моста и жесткости опорных частей [3–6]. Указанные смещения передаются на насыпь подходов к мосту. На одном подходе к мосту рельсы будут растянуты, а на другом сжаты, что может привести к потере прочности или устойчивости рельсовой плети в этих местах.

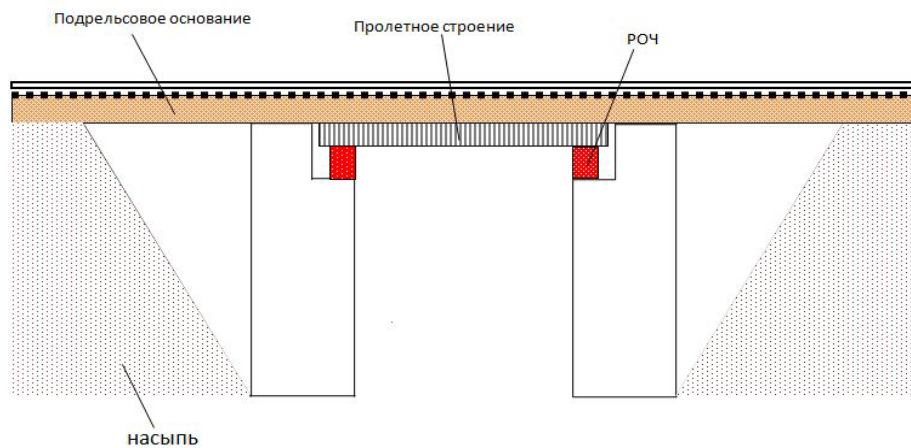


Рис. 1. Схема однопролетного моста с сейсмоизолирующими опорными частями

На участке, примыкающем к мосту, рельсовая плеть будет проскальзывать по подкладкам и смещение рельса w описывается уравнением:

$$w''(x) = q, \quad (1)$$

где q – погонное сопротивление рельсовой плети, равное 20 кН/м.

Уравнение (1) решается при условиях:

$$W(0) = u_0, \quad w|_{x=L} = w_{el}(0), \quad w'|_{x=L} = w'_{el}(0),$$

где L – длина участка плети, на котором будет происходить ее проскальзывание; $w_{el}(0)$ – значение смещений в месте заклинивания плети (начала ее упругой работы);

$w'_{el}(0)$ – значение продольной деформации в месте заклинивания плети (начала ее упругой работы), определяющее напряжения и усилия в рельсе;

Решение уравнения (1) имеет вид:

$$w(x) = \frac{qx^2}{2 \cdot E \cdot F} + C_{slide} \cdot x + u_0,$$

где E – модуль упругости рельса, F – площадь поперечного сечения рельса, q – погонное сопротивление рельсовой плети, u_0 – смещения пролетных строений относительно насыпи, C_{slide} – постоянная интегрирования на участке смещения.

Для определения постоянной интегрирования C_{slide} и значений $w_{el}(0)$ и $w'_{el}(0)$ получаем уравнения:

$$w_{el}^{(0)} = \frac{qL^2}{2 \cdot E \cdot F} + C_{slide}L + u_0, \quad (2)$$

$$w'_{el}{}^{(0)} = \frac{qL}{E \cdot F} + C_{slide}. \quad (3)$$

На участке отсутствия проскальзывания рельсовой плети уравнение ее деформации выглядит следующим образом [2]:

$$w'' - \kappa \cdot w = 0. \quad (4)$$

Принимая во внимание, что $w(\infty)=0$, общее решение уравнения (4) имеет вид:

$$w(x) = C_{el} \cdot e^{-\lambda \cdot (x-L)},$$

где $\lambda = \sqrt{\kappa}$ и C_{el} – постоянные интегрирования.

Постоянная C_{el} должна обеспечить равенство смещений и деформаций (усилий) на границе перехода от проскальзывания к упругой работе плети. Если принять для участка упругой работы рельса свою систему координат, то эти условия имеют вид:

$$C_{el} = w_{el}^{(0)} \quad \text{и} \quad -\lambda C_{el} = w'_{el}{}^{(0)} \quad (5)$$

Из уравнений (2)–(5) получим следующую систему из двух уравнений с тремя неизвестными:

$$\begin{cases} \frac{qL^2}{2EF} - C_{slide} \cdot L + u_0 = C_{el} \\ \frac{qL}{EF} - C_{slide} = -C_{el} \cdot \lambda \end{cases}. \quad (6)$$

К рассмотренным условиям необходимо добавить условие равновесия рельсовой плети:

$$\int_0^L qdx + \int_0^\infty kw(x-L)dx = N_0 = EFw'(0) = EFC_{slide} \quad \text{или} \quad qL + C_{el} \cdot \frac{k}{\lambda} = EFC_{slide} \quad (7)$$

Система уравнений (6) и условие (7) образуют единую систему уравнений третьего порядка для определения трех переменных: C_{slide} , C_{el} и L . Решение ее имеет вид:

$$C_{el} = \frac{\left(-\frac{qL^2}{2EF} + u_0\right)}{(1 + \lambda L)}, \quad C_{slide} = \frac{\left(\frac{qL(\lambda L + 2)}{2EF} + u_0 \cdot \lambda\right)}{(L \cdot \lambda + 1)}.$$

При подстановке их в (5) получим квадратное уравнение относительно L , из которого можно определить:

$$L = \sqrt{\frac{2u_0 \cdot EF}{q}}.$$

В результате для перемещений плети на подходе к мосту имеем:

$$W(x) = \begin{cases} \frac{q \cdot x^2}{2 \cdot E \cdot F} - C_{slide} + U_0, & \text{если } x \leq L \\ C_{el} \cdot e^{-\lambda(x-L)}, & \text{если } x > L \end{cases}.$$

При этом напряжения $\sigma(x)$ можно записать следующим образом:

$$\sigma(x) = \begin{cases} E \cdot \left(\frac{q \cdot x}{E \cdot F} - C_{slide} + U_0\right), & \text{если } x \leq L \\ -E \cdot \lambda \cdot \left(C_{el} \cdot e^{-\lambda(x-L)}\right), & \text{если } x > L \end{cases}.$$

На рис. 2 представлены смещения рельсовой плети, вызванные смещением сейсмоизолированного пролетного строения. Они затухают на расстоянии около 150 м от границы моста. При этом напряжения в рельсе достигают $40\,000\text{ кН/м}^2$. Для обеспечения целостности плети в этом случае необходимо использовать термоупрочненные рельсы с допустимыми напряжениями $49\,000\text{ кН/м}^2$. Из рисунка также видно, что напряжение 25 кН/м^2 (оно традиционно используется на насыпях при закреплении рельса) приводит к заметному увеличению напряжения в рельсовой плети с $447\,500$ до $400\,000\text{ кН/м}^2$. Если учесть температурные напряжения в рельсе, то в первом случае условия прочности будут выполнены при использовании термоупрочненных рельсов с допустимыми напряжениями $490\,000\text{ кН/м}^2$.

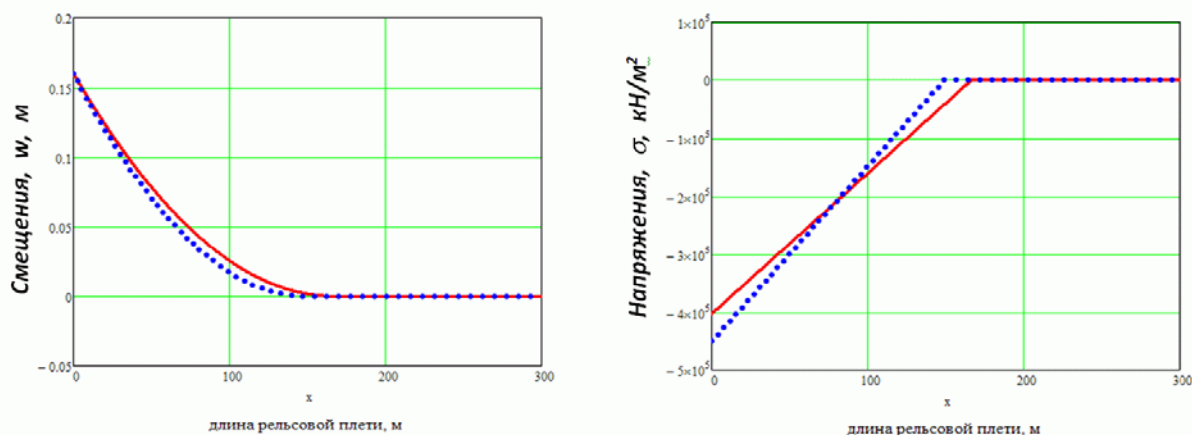


Рис. 2. Напряжения и смещения в рельсовой плети по ее длине x от границы моста.
Сплошная линия – $q=20\text{ кН/м}^2$, точечная – $q=25\text{ кН/м}^2$

Следующая группа задач связана с напряженно-деформированным состоянием рельсового пути на подходах к однопролетному мосту со сейсмоизолирующими опорными частями при поперечных нагрузках. Пролетное строение упруго закреплено на опорах, а мостовой брус укладывается на балласт. В результате поперечных ударов подвижного состава пролетное строение может перемещаться в поперечном направлении на 3–5 см. Это может привести к разрыву рельсовой плети на стыке моста и насыпи подходов. Рассмотрим уравнение поперечной деформации рельсового пути [6]:

$$V = \frac{q}{\kappa} + u_0 + \varphi_0 x + y_0 K_0 + \theta_0 K_1 - \frac{M_0}{EJ} K_2 - \frac{Q_0}{EJ} K_3, \quad (8)$$

где y_0 – смещение балки в середине пролета, θ_0 – угол поворота балки в середине пролета, M_0 – момент в середине пролета, Q_0 – перерезывающая сила в середине балки, u_0 – смещение пролетного строения, φ_0 – угол поворота пролетного строения, q – распределенная нагрузка на рельс, K_i – функции Крылова. Они имеют вид:

$$K_0(\beta x) = ch(\beta x) \cos(\beta x),$$

$$K_1(\beta x) = \frac{ch(\beta x) \sin(\beta x) + sh(\beta x) \cos(\beta x)}{2\beta},$$

$$K_2(\beta x) = \frac{sh(\beta x) \sin(\beta x)}{2\beta^2},$$

$$K_3(\beta x) = \frac{ch(\beta x) \sin(\beta x) - sh(\beta x) \cos(\beta x)}{4\beta^3}.$$

Величины y_0 , θ_0 , M_0 , Q_0 , u_0 , φ_0 называют начальными параметрами.

Особенностью уравнения (8), по сравнению с известными из литературы [3–6], является необходимость учета смещений и поворота пролетного строения от поперечной нагрузки, передаваемой на него через балласт и верхнее строение пути. Рассматривается оно при следующих граничных условиях:

$$Kx \cdot V\left(\pm \frac{L}{2}\right) = V'''\left(\pm \frac{L}{2}\right) EJ,$$

$$K\phi \cdot V'\left(\pm \frac{L}{2}\right) = V''\left(\pm \frac{L}{2}\right) EJ,$$

$$C_1(u_0 + \phi_0 \frac{L}{2}) + C_2(u_0 - \phi_0 \frac{L}{2}) = \int_{-\frac{L}{2}}^{\frac{L}{2}} \kappa(V - u_0 - \phi_0 x) dx,$$

$$C_1(u_0 + \phi_0 \frac{L}{2}) \cdot \frac{L}{2} - C_2(u_0 - \phi_0 \frac{L}{2}) \cdot \frac{L}{2} = \int_{-\frac{L}{2}}^{\frac{L}{2}} \kappa x(V - u_0 - \phi_0 x) dx.$$

где Kx – поперечная жесткость рельса на сдвиг, $K\phi$ – жесткость рельса на поворот.

Известно [6], что

$$EJV''' = -Q, \quad EJV'' = -M, \quad V' = \theta,$$

$$\theta = V' = \varphi_0 + y_0 K'_0 + \theta_0 K'_1 - \frac{M_0}{EJ} K'_2 - \frac{Q_0}{EJ} K'_3 = \varphi_0 + y_0(-4\beta^4 K_3) + \theta_0 K_0 - \frac{M_0}{EJ} K_1 - \frac{Q_0}{EJ} K_2,$$

$$\frac{M}{EJ} = -V'' = -y_0 K''_0 - \theta_0 K''_1 + \frac{M_0}{EJ} K''_2 + \frac{Q_0}{EJ} K''_3 = -y_0(-4\beta^4 K_2) - \theta_0(-4\beta^4 K_3) + \frac{M_0}{EJ} K_0 + \frac{Q_0}{EJ} K_1,$$

$$\frac{Q}{EJ} = -V''' = -y_0 K'''_0 - \theta_0 K'''_1 + \frac{M_0}{EJ} K'''_2 + \frac{Q_0}{EJ} K'''_3 = -y_0(-4\beta^4 K_1) - \theta_0(-4\beta^4 K_2) + \frac{M_0}{EJ}(-4\beta^4 K_3)_0 + \frac{Q_0}{EJ} K_0.$$

Выразим отсюда значение момента и поперечной силы:

$$M = -y_0(-4\beta^4 K_2) \cdot EJ - \theta_0(-4\beta^4 K_3) \cdot EJ + M_0 K_0 + Q_0 K_1,$$

$$Q = -y_0(-4\beta^4 K_1) \cdot EJ - \theta_0(-4\beta^4 K_2) \cdot EJ + M_0(-4\beta^4 K_3)_0 + Q_0 K_0.$$

Распишем граничные условия. Первое граничное условие:

$$Kx \cdot V\left(\pm \frac{L}{2}\right) = V'''\left(\pm \frac{L}{2}\right) EF,$$

$$y_0 \cdot \left[K_0\left(\pm \frac{L}{2}\right) \cdot \frac{Kx}{EJ} + 4\beta^4 K_1\left(\pm \frac{L}{2}\right) \right] + \theta_0 \cdot \left[K_1\left(\pm \frac{L}{2}\right) \cdot \frac{Kx}{EJ} + 4\beta^4 K_2\left(\pm \frac{L}{2}\right) \right] -$$

$$\frac{M_0}{EJ} \cdot \left[K_2\left(\pm \frac{L}{2}\right) \cdot \frac{Kx}{EJ} + 4\beta^4 K_3\left(\pm \frac{L}{2}\right) \right] - \frac{Q_0}{EJ} \cdot \left[K_3\left(\pm \frac{L}{2}\right) \cdot \frac{Kx}{EJ} - K_0\left(\pm \frac{L}{2}\right) \right] + u_0 \frac{Kx}{EJ} + \varphi_0 x \frac{Kx}{EJ} = -\frac{q}{\kappa} \frac{Kx}{EJ}.$$

Второе граничное условие:

$$K\phi \cdot V' \left(\pm \frac{L}{2} \right) = V'' \left(\pm \frac{L}{2} \right) EJ$$

$$y_0 \cdot \left[-4\beta^4 K_3 \left(\pm \frac{L}{2} \right) \cdot \frac{K\phi}{EJ} + 4\beta^4 K_2 \left(\pm \frac{L}{2} \right) \right] + \theta_0 \cdot \left[K_0 \left(\pm \frac{L}{2} \right) \cdot \frac{K\phi}{EJ} + 4\beta^4 K_3 \left(\pm \frac{L}{2} \right) \right] -$$

$$\frac{M_0}{EJ} \cdot \left[K_1 \left(\pm \frac{L}{2} \right) \cdot \frac{K\phi}{EJ} - K_0 \left(\pm \frac{L}{2} \right) \right] - \frac{Q_0}{EJ} \cdot \left[K_2 \left(\pm \frac{L}{2} \right) \cdot \frac{K\phi}{EJ} - K_1 \left(\pm \frac{L}{2} \right) \right] + \frac{K\phi}{EJ} \cdot \phi_0 = 0$$

Третье граничное условие:

$$C_1(u_0 + \phi_0 \frac{L}{2}) + C_2(u_0 - \phi_0 \frac{L}{2}) = \int_{-\frac{L}{2}}^{\frac{L}{2}} \kappa(V - u_0 - \phi_0 x) dx,$$

$$C_1(u_0 + \phi_0 \frac{L}{2}) + C_2(u_0 - \phi_0 \frac{L}{2}) = qL + \kappa y_0 (K_1(\frac{L}{2}) - K_1(-\frac{L}{2})) +$$

$$+ \kappa \theta_0 (K_2(\frac{L}{2}) - K_2(-\frac{L}{2})) - \kappa \frac{M_0}{EJ} (K_3(\frac{L}{2}) - K_3(-\frac{L}{2})) + \kappa \frac{Q_0}{4\beta^4 EJ} (K_0(\frac{L}{2}) - K_0(-\frac{L}{2}))$$

Четвертое граничное условие:

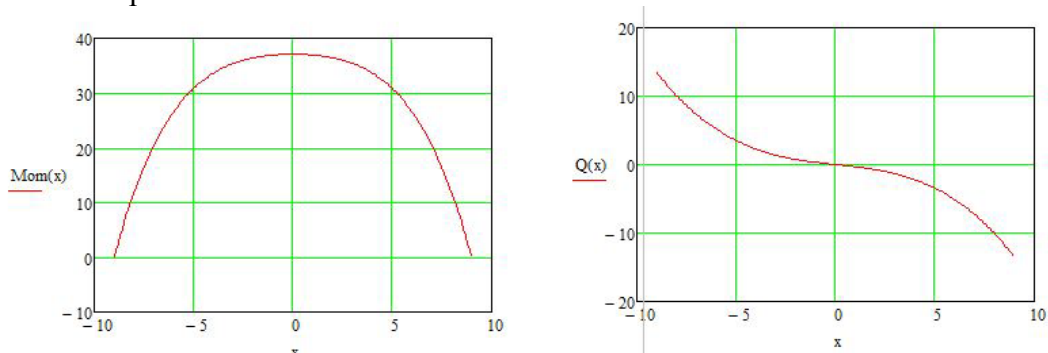
$$C_1(u_0 + \phi_0 \frac{L}{2}) \cdot \frac{L}{2} - C_2(u_0 - \phi_0 \frac{L}{2}) \cdot \frac{L}{2} = \int_{-\frac{L}{2}}^{\frac{L}{2}} \kappa x (V - u_0 - \phi_0 x) dx,$$

$$C_1(u_0 + \phi_0 \frac{L}{2}) \cdot \frac{L}{2} - C_2(u_0 - \phi_0 \frac{L}{2}) \cdot \frac{L}{2} = y_0 \kappa \left[\begin{array}{l} \frac{L}{2} (K_1(\frac{L}{2}) + K_1(-\frac{L}{2})) - \\ - (K_2(\frac{L}{2}) - K_2(-\frac{L}{2})) \end{array} \right] +$$

$$+ \theta_0 \kappa \left[\begin{array}{l} \frac{L}{2} (K_2(\frac{L}{2}) + K_2(-\frac{L}{2})) - \\ - (K_3(\frac{L}{2}) - K_3(-\frac{L}{2})) \end{array} \right] - \frac{M_0}{EJ} \kappa \left[\begin{array}{l} \frac{L}{2} (K_3(\frac{L}{2}) + K_3(-\frac{L}{2})) + \\ + \frac{1}{4\beta^4} (K_0(\frac{L}{2}) - K_0(-\frac{L}{2})) \end{array} \right] - \frac{Q_0}{EJ} \kappa \left[\begin{array}{l} -\frac{1}{4\beta^4} \frac{L}{2} (K_0(\frac{L}{2}) + K_0(-\frac{L}{2})) + \\ + \frac{1}{4\beta^4} (K_1(\frac{L}{2}) - K_1(-\frac{L}{2})) \end{array} \right]$$

Решив систему из шести уравнений, найдем шесть неизвестных констант:

y_0 , θ_0 , M_0 , Q_0 , u_0 , ϕ_0 , с помощью которых получим значения момента M и поперечной силы Q (рис. 3). Из рисунка видно, что для жестких опорных частей (металлических литых $C = 10^{12}$ кН/м) усилия в рельсе составляют $M = 15$ кН·м, $Q = 5$ кН. Они допустимы для рельсовых плетей. При укладке пути на сейсмоизолируемый мост с РОЧ ($C = 100$ кН/м) усилия в рельсе возрастают примерно в три раза, достигая величин $M = 35$ кН·м, и $Q = 15$ кН, и могут приводить к разрыву рельсовой плети. Отсюда следует, что нельзя сочетать сейсмоизоляцию с традиционной конструкцией верхнего строения пути. Как минимум следует использовать термоупрочненные рельсы.



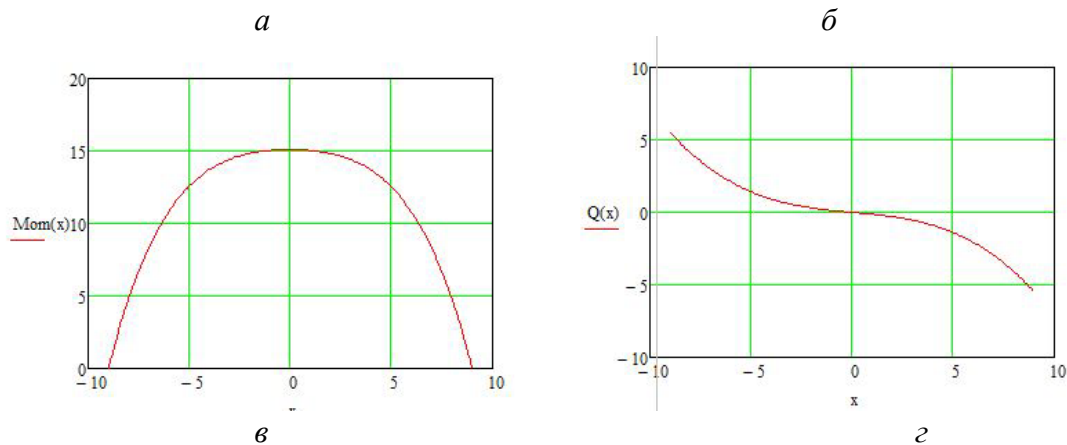


Рис. 3. Зависимость моментов и поперечной силы от жесткости опорных частей:
а, б – при $C_1 = C_2 = 100$ кН/м; в, г – при $C_1 = 10^{12}$ кН/м

Достигнутые результаты

Уменьшение жесткости опорных частей моста ведет к заметному увеличению поперечных усилий на стыке моста с насыпью, что ограничивает применение сейсмоизоляции опор и бесстыкового пути. Возможные способы решения этого вопроса связаны с переходом на термоупрочненные рельсы большой мощности, включением контррельсов в поперечную работу рельса и оптимизацией жесткости подрельсового основания в местах деформационных швов на пролетных строениях и в зоне контакта моста с насыпью подходов.

Разработанные программные средства могут быть использованы при проектировании высокоскоростных магистралей, в частности Хельсинки – Санкт-Петербург – Москва.

Публикации

1. Зайцева Т.И., Жгутова Т.В., Самойлова А.В., Уздин А.М. Оценка безопасности бесстыкового пути по величине зазора при его разрыве // Природные и техногенные риски. Безопасность сооружений. 2011. № 4. С. 53–55.
2. Зайцева Т.И. Влияние изменения жесткости подрельсового основания на напряженно-деформированное состояние рельсовой плети // Сб. статей: Труды студенческого центра прикладных математических исследований. Санкт-Петербург, 2012. С. 47–51.
3. Зайцева Т.И., Уздин А.М. Влияние изменения жесткости подрельсового основания на напряженно-деформированное состояние рельсовой плети // Природные и техногенные риски. Безопасность сооружений. 2012. № 3. С. 53–55.
4. Зайцева Т.И. Математические модели для анализа напряженно-деформированного бесстыкового пути при статических и динамических нагрузках // Сб. работ студентов УНИВЕРСИТЕТ ИТМО, победителей конкурса грантов Правительства Санкт-Петербурга. СПб., 2013. С. 54–61.

Литература

1. Боченков М.С. Зазоры при изломе рельсовой плети // Труды ЦНИИ МПС. Вып. 244. М., 1962. С. 164–172.
2. Зайцева Т.И. Математические модели для анализа напряженно-деформированного бесстыкового пути при статических и динамических нагрузках // Сб. работ студентов УНИВЕРСИТЕТ ИТМО, победителей конкурса грантов Правительства Санкт-Петербурга. СПб., 2013. С. 54–61.
3. Di Sarno L., Del Principe B. Base isolation system for railway bridges: Beneficial or detrimental // Proceedings of workshop “Bridges seismic isolation and large-scale modeling” (Saint Petersburg, 29.06–03.07, 2010). Saint Petersburg, 2010. P. 16–28.
4. Kuznetsova I.O., Uzdin A.M., Zhgutova T.V., Shulman S.A. Seismic protection of railway bridges in Sochi // Proceedings of workshop “Bridges seismic isolation and large-scale modeling” (Saint Petersburg, 29.06–03.07, 2010). Saint Petersburg, 2010. P. 28–37.

5. Huber P. Realized projects of isolation systems for railway bridges in Spain, Hungary and Greece // Proceedings of workshop "Bridges seismic isolation and large-scale modeling" (Saint Petersburg, 29.06–03.07, 2010). Saint Petersburg, 2010. P. 37–51.
6. Филин А.П. Прикладная механика твердого деформируемого тела. Т. II. М., 1978. – 616 с.



Зименко Константин Александрович
Факультет компьютерных технологий и управления
Кафедра систем управления и информатики
E-mail: kostyazimenko@gmail.com

МЕТОДЫ ОБРАБОТКИ СИГНАЛОВ ЭЛЕКТРОМИОГРАФИИ В СИСТЕМАХ УПРАВЛЕНИЯ РЕАБИЛИТАЦИОННЫХ УСТРОЙСТВ

К.А. Зименко

Научный руководитель – профессор А.А. Бобцов

Введение

Задача социальной адаптации инвалидов и лиц, страдающих нарушениями опорно-двигательного аппарата на сегодняшний день весьма актуальна. Ведущим направлением социальной политики в отношении них определена реабилитация, восстановление способностей человека в бытовой, общественной и профессиональной деятельности, соответствующей его физическому и социальному потенциалу. Важнейшим условием реабилитации, возможности возвращения к активному образу жизни, является обеспечение инвалидов специальными средствами передвижения.

Для оценки функционального состояния двигательного аппарата при восстановлении нарушенной двигательной функции в ортопедии и протезировании используется электромиография. Это метод исследования биоэлектрических потенциалов, возникающих в скелетных мышцах животных и человека при возбуждении мышечных волокон. Информация, содержащаяся в сигнале электромиограммы (ЭМГ), может использоваться при создании и использовании активных протезов, экзоскелетов и других устройств, управление которыми происходит посредством изменения биопотенциалов мышц. Анализ сигналов ЭМГ позволяет определить планируемое человеком движение, необходимое для приведения в действие исполнительного механизма (например, электродвигателя протеза или экзоскелета).

Построение классификатора для определения типа движения

Успешность реализации предложенных идей во многом определяется степенью достоверности расшифровки биопотенциалов мышц при планируемом движении. Уровень сигнала в диапазоне частот до 500 Гц лежит в пределах от 0 до 10 мВ [1]. Сегодня точной классификации типа движения препятствует низкое отношение сигнал/шум в измерительной системе. Большинство шумов, присутствующих в сигнале, высокочастотны: шумы от электронных устройств, окружающего электромагнитного излучения, взаимодействия кабелей электродов. Однако помимо них в сигнале могут присутствовать и другие шумы: побочное влияние сигнала кардиограммы, смещение электродов относительно назначенного места, искажения сигнала при изменении биопотенциалов на референте при использовании референтной схемы. Поэтому наименее искаженный и достоверный сигнал ЭМГ находится в диапазоне частот 15–150 Гц. Для получения такого сигнала был применен фильтр низких частот, выполненный на аппаратном уровне.

Для успешной классификации типа движения рассмотрена возможность применения метода слепого разделения источников, нечеткой логики, генетических алгоритмов и др. Благодаря малому времени обучения из существующего многообразия был выбран метод искусственных нейронных сетей [2]. Наилучший результат показал однослойный перцептрон. Использование многослойных сетей затруднено из-за влияния на работу сети высокого уровня шумов сигнала. Для выбранной нейронной сети применена регуляризируемая функция издержек, при нахождении минимума которой использовался метод сопряженных градиентов с алгоритмом Полака-Рибери определяющим направления поиска. Суть алгоритма Полака-Рибери состоит в следующем: генерируется последовательность так называемых рабочих векторов g и направлений поиска h таких, что h взаимно сопряжены. Причем доказано, что если функция издержек, которую необходимо минимизировать, может быть выражена квадратичной формой, то минимизация вдоль первых h -направлений приведет к точному минимуму. Совместно с методом поиска начального размера шага использовался критерий остановки Вульф-Пауэлла.

Для подтверждения работоспособности предложенного алгоритма проведена серия экспериментов. На мышцы предплечья операторов устанавливались четыре электрода, пятый крепился на плечо и выступал в качестве референта (рис. 1а). Данные с электродов снимались регистратором биопотенциалов *Kardi3/9* (рис. 1б) и передавались на персональный компьютер. Операторами выступали пять здоровых мужчин в возрасте от 19 до 24 лет. Они выполняли четыре типа движений: сгибание кисти вперед/назад и вращение кистью по/против часовой стрелки. Каждый из операторов выполнял по пять раз каждое из движений, после чего нейронная сеть обучалась. Теперь классификатор должен был распознавать аналогичные движения оператора с точностью не менее 90 %.

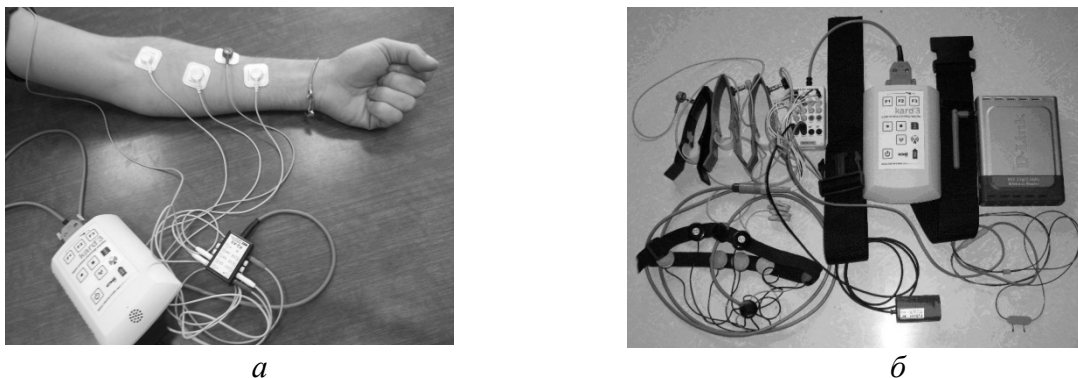


Рис. 1. Экспериментальные исследования: а – схема измерения биопотенциалов; б – внешний вид регистратора биопотенциалов *Kardi3/9*

На первом этапе точность определения типа движения составила 87 %. Этого недостаточно для управления устройствами реабилитации. Ее повышению может способствовать предварительная обработка сигнала с использованием Фурье- или вейвлет-преобразования. В связи с тем, что время на обучение нейронной сети и, соответственно, определение типа движения должно быть минимальным, был выбран вейвлет-анализ. В отличие от преобразования Фурье, анализирующая функция которого покрывает всю временную ось, двухпараметрическая анализирующая функция одномерного вейвлет-преобразования хорошо локализована и во времени, и по частоте. Помимо этого при декомпозиции сигнала использование детализирующей функции позволяет уменьшить размерность сигнала вдвое, а также избавиться от низкочастотных шумов, вызванных изменением потенциала на референте и побочным влиянием сигнала кардиограммы. Понижение размерности сигнала позволяет уменьшить размеры нейронной сети и время ее обучения, что является важным фактором при управлении устройствами реабилитации. При использовании вейвлет-преобразования применялся вейвлет Хаара. Тип планируемого движения определялся по данным, полученным за последние 200 мс, что при частоте дискретизации сигнала в 1000 Гц с учетом уменьшения размерности после вейвлет-преобразования составляет 100 отсчетов для каждого из четырех электродов. На рис. 2

представлен сигнал электромиограммы до и после обработки с использованием вейвлет-преобразования.

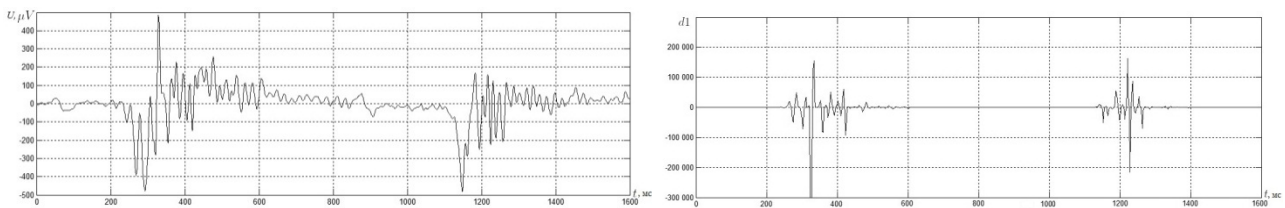


Рис. 2. Сигнал электромиограммы по одному из каналов до и после обработки

Комбинация фильтрации сигнала, его последующего вейвлет-преобразования и его классификатора на основе искусственной нейронной сети позволили повысить точность распознавания движений до 94 %. Предложенный метод обработки сигнала миограммы не вносит значительных временных задержек, что актуально для систем, используемых в реабилитации. Алгоритм обладает высокой скоростью обучения, сравнительно прост в применении и позволяет достичь требуемой точности классификации движений оператора.

Для повышения точности распознавания движений можно использовать дополнительную информацию, например идентифицированные частоты основных гармоник сигнала ЭМГ.

Идентификация частот сигнала ЭМГ в режиме реального времени

Использование сигнала ЭМГ в качестве исходного для диагностики или управления устройствами реабилитации приводит к необходимости осуществления анализа и классификации его в режиме реального времени. Особенности сигнала могут быть представлены в нескольких областях, таких как временная, частотная и частотно-временная. В случае оценки переходных свойств сигнала частотно-временные характеристики наиболее предпочтительны. Центральная концепция большинства методов заключается в разложении сигнала на частотно-временные составляющие:

$$x(t) = \sum_{i=1}^N c_i \beta_i(t), \quad (1)$$

где $\beta_i(t)$ – базисная функция, c_i – соответствующий коэффициент. Однако следует отметить, что большинство методов не работают в режиме реального времени и используют оконные техники, а информация, получаемая с помощью них, ограничена размерами анализирующего окна.

Далее опишем метод идентификации частот сигнала ЭМГ, адаптированный для использования в системах управления робототехническими устройствами для медицинских приложений. Он позволяет идентифицировать частоты мультигармонического сигнала ЭМГ в режиме реального времени. Данные, полученные с помощью этого метода, можно использовать как в качестве новой особенности, так и в диагностических целях.

Основной результат

Перепишем уравнение (1) в следующей форме:

$$y(t) = \sum_{i=1}^k \sigma_i \sin(\omega_i t + \phi_i), \quad (2)$$

где σ_i – амплитуда, ϕ_i – начальная фаза для i -ой гармоники. Каждое движение описывается уникальным набором гармоник. Поэтому, определив частоты основных гармоник сигнала, можно выяснить какой тип движения был совершен человеком. Далее представим краткое описание метода идентификации частот мультигармонического сигнала, представленного в [3].

Известно, что для генерации сигнала $y(t)$ можно использовать уравнение:

$$(p^2 - \theta_1)(p^2 - \theta_2) \cdots (p^2 - \theta_k)y(t) = 0, \quad (3)$$

где $\theta_i = -\omega_i^2$, $i = \overline{1, k}$ – константы, p – оператор Лапласа. Перепишем (3) в виде:

$$p^{2k}y(t) = \bar{\theta}_1 p^{2k-2}y(t) + \cdots + \bar{\theta}_{k-1} p^2 y(t) + \bar{\theta}_k y(t), \quad (4)$$

$$\text{где } \begin{cases} \bar{\theta}_1 = \theta_1 + \theta_2 + \dots + \theta_k, \\ \bar{\theta}_2 = -\theta_1\theta_2 - \theta_1\theta_3 - \dots - \theta_{k-1}\theta_k, \\ \vdots \\ \bar{\theta}_k = (-1)^{k+1}\theta_1\theta_2 \dots \theta_k. \end{cases} \quad (5)$$

Переходя к изображению Лапласа в уравнении (4) получим:

$$s^{2k}\bar{\Delta}(s) = \bar{\theta}_1 s^{2k-2}\bar{\Delta}(s) + \dots + \bar{\theta}_{k-1} s^2\bar{\Delta}(s) + \bar{\theta}_k \bar{\Delta}(s) + D(s), \quad (6)$$

где s – комплексная переменная, $\bar{\Delta}(s) = \mathcal{L}\{y(t)\}$ – изображение Лапласа функции $y(t)$, $D(s)$ – сумма всех членов, содержащих начальные условия.

Помножим уравнение (6) на $\frac{\lambda^{2k}}{(s+\lambda)^{2k}}$, где λ – константа, и рассмотрим вспомогательный фильтр:

$$\xi(s) = \frac{\lambda^{2k}}{(s+\lambda)^{2k}} \bar{\Delta}(s). \quad (7)$$

После обратного преобразования Лапласа для (6) получим:

$$\xi^{(2k)}(t) = \bar{\theta}_1 \xi^{(2k-2)}(t) + \dots + \bar{\theta}_k \xi(t) + \varepsilon(t), \quad (8)$$

где $\varepsilon(t) = \mathcal{L}^{-1}\left\{\frac{D(s)}{(s+\lambda)^{2k}}\right\}$ – экспоненциально затухающая составляющая, вызванная ненулевыми начальными условиями.

Член $\varepsilon(t) = \mathcal{L}^{-1}\left\{\frac{D(s)}{(s+\lambda)^{2k}}\right\}$ зависит от параметра λ , и при его увеличении увеличивается скорость сходимости $\varepsilon(t)$ к нулю.

Пренебрегая $\varepsilon(t)$, получим:

$$\xi^{(2k)} = \Omega^T(t)\bar{\Theta}, \quad (9)$$

где $\Omega^T(t) = [\xi^{(2k-2)}(t) \dots \xi^{(2)}(t) \xi(t)]$ – регрессор, $\bar{\Theta}^T = [\bar{\theta}_1 \dots \bar{\theta}_{k-1} \bar{\theta}_k]$ – вектор неизвестных параметров.

Аналогично [3] и беря в расчет, что $\Omega(t)\xi^{(2k)} = \Omega(t)\Omega^T(t)\bar{\Theta}(t)$, оценка вектора параметров $\bar{\Theta}$ осуществляется следующим образом:

$$\hat{\bar{\Theta}}(t) = K\Omega(t)(\xi^{(2k)} - \Omega^T(t)\hat{\bar{\Theta}}(t)), \quad (10)$$

где $K = \text{diag}\{k_i > 0\}, i = \overline{1, k}$, и k_i – некоторый параметр, увеличение которого может ускорить сходимость $\hat{\bar{\theta}}_i$ к $\bar{\theta}_i$.

Так как частотный спектр сигнала ЭМГ резко изменяется при совершении какого-либо движения, для определения основных частот сигнала используются большие значения параметров k и λ (например, $\lambda > 5, k > 1000$).

Рассмотрим систему (5). Ее уравнения представляют собой формулы Виета. Поэтому параметры $\theta_i, i = \overline{1, k}$ являются корнями полинома $p^{2k} + \bar{\theta}_1 p^{2k-2} + \dots + \bar{\theta}_{k-1} p^2 + \bar{\theta}_k$.

Таким образом, мы получили гибридную схему идентификации параметров $\theta_i, i = \overline{1, k}$. Частоты мультигармонического сигнала мы получим из уравнения (3):

$$\hat{\omega}_i(t) = \sqrt{|\hat{\theta}_i(t)|}, i = \overline{1, k}. \quad (11)$$

Основываясь на исследованиях Пирсона [4], можно заключить, что сигнал ЭМГ состоит в основном из низкочастотных гармоник. Поэтому вместе с вышеописанным методом предполагается использовать низкочастотный или полосовой фильтры, которые позволяют:

- выбрать области частотного спектра, используемые при идентификации движений и в системе управления;
- уменьшить зашумленность сигнала (в том числе от высокочастотных шумов и шумов в 50 Гц, исходящих от промышленной сети).

Выбор полосового фильтра при идентификации движений основывается на предварительном спектральном анализе.

Эксперимент

Чтобы продемонстрировать эффективность предложенного метода, рассмотрим задачу определения вида совершенного движения по сигналу ЭМГ в режиме реального времени. Для подтверждения работоспособности метода был проведен ряд экспериментов на пяти мужчинах в возрасте от 19 до 24 лет. Данные снимались с двух электродов: один из них устанавливался на предплечье, а второй размещался на плече и выступал в качестве референта. В ходе эксперимента выполнялись три типа движений: сгибание кисти и вращение ее вправо/влево.

Основываясь на спектральных характеристиках (рис. 3) можно сделать заключение, что большая часть энергии сигнала содержится в двух-трех гармониках. Одна из них свойственна для всех трех типов движений, остальные зависят от типа произведенного человеком движения. Поэтому был выбран полосовой фильтр с пропускной полосой от 12 до 20 Гц. Таким образом, в дальнейшей обработке сигнала задействована область с одной или двумя наиболее выделяющимися гармониками для каждого типа движения: 15,6 Гц для сгибания кисти; 13,7 Гц и 17,57 Гц для вращения по часовой стрелке; 13,7 Гц для вращения против часовой стрелки.

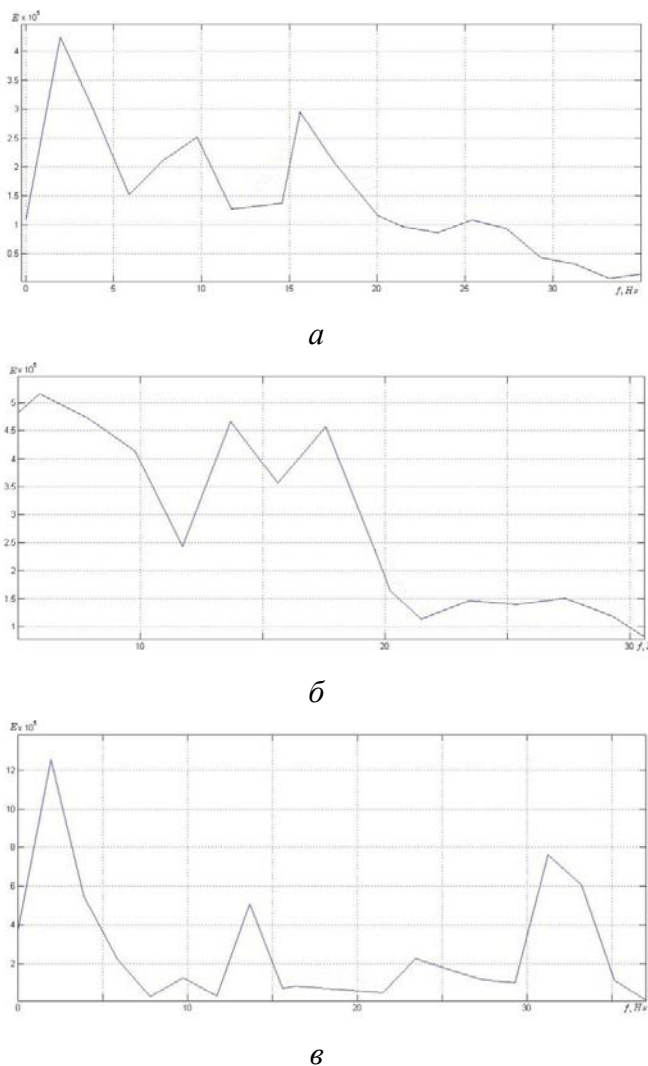


Рис. 3. Спектральная плотность сигнала ЭМГ: а – при сгибании кисти; б – при вращении кисти по часовой стрелке; в – при вращении кисти против часовой стрелки

Таким образом, мы можем переписать уравнение (2) как сумму двух гармоник:

$$y(t) = A_1 \sin(\omega_1 t + \varphi_1) + A_2 \sin(\omega_2 t + \varphi_2), \quad (12)$$

где A_1 , ω_1 , φ_1 , A_2 , ω_2 , φ_2 – неизвестные константы.

В соответствии с (4), (5)

$$p^4 y(t) = \bar{\theta}_1 p^2 y(t) + \bar{\theta}_2 y(t), \quad (13)$$

где $\bar{\theta}_1 = -\omega_1^2 - \omega_2^2$, $\bar{\theta}_2 = \omega_1^2 \omega_2^2$ – неизвестные параметры, которые необходимо идентифицировать. После введения вспомогательного фильтра $\frac{\lambda^4}{(s+\lambda)^4}$ получим:

$$z(t) = \xi^{(4)}(t) = \frac{\lambda^4 s^4}{(s+\lambda)^4} y(t), \zeta_1(t) = \xi^{(2)}(t) = \frac{\lambda^4 s^2}{(s+\lambda)^4} y(t), \zeta_2 = \xi(t) = \frac{\lambda^4}{(s+\lambda)^4} y(t). \quad (14)$$

Используя уравнения (13) и (14) для модели (12) перепишем уравнение (9) в виде:

$$z(t) = \bar{\theta}_1 \zeta_1(t) + \bar{\theta}_2 \zeta_2(t) + \varepsilon(t) = \zeta^T \bar{\theta}, \quad (15)$$

где $\bar{\theta} = \text{col}\{\bar{\theta}_1, \bar{\theta}_2\}$, $\zeta = \text{col}\{\zeta_1, \zeta_2\}$.

В соответствии с уравнением (10) и пренебрегая $\varepsilon(t)$ можно использовать следующий алгоритм идентификации:

$$\dot{\hat{\theta}} = -k \zeta \zeta^T \hat{\theta} + k \zeta z, \quad (16)$$

где $\hat{\theta}$ – оценка вектора $\bar{\theta}$ и $k > 0$ – некоторый коэффициент. Увеличение его в скалярном варианте увеличивает скорость сходимости, но в остальных случаях это не так. Решение этой проблемы может быть найдено при использовании гибридной схемы выбора параметров, основанной на методе каскадной редукции [5]. Изменим уравнение (15) с использованием данного метода. Умножим для этого (15) на ζ_1 и проинтегрируем его:

$$\zeta_1 = \bar{\theta}_1 \zeta_1^2 + \bar{\theta}_2 \zeta_2 \zeta_1, \int_0^t z \zeta_1 d\tau = \bar{\theta}_1 \int_0^t \zeta_1^2 d\tau + \bar{\theta}_2 \int_0^t \zeta_2 \zeta_1 d\tau. \quad (17)$$

Обозначим $\gamma_1 = \int_0^t z \zeta_1 d\tau$, $\gamma_2 = \int_0^t \zeta_1^2 d\tau$ и $\gamma_3 = \int_0^t \zeta_2 \zeta_1 d\tau$. Последовательно разделим последнее отношение на ζ_2 и затем проинтегрируем. Получим следующее выражение:

$$\dot{\gamma}_1 \gamma_2^{-1} - \gamma_1 \dot{\gamma}_2 \gamma_2^{-2} = \bar{\theta}_2 (\dot{\gamma}_3 \gamma_2^{-1} - \gamma_3 \dot{\gamma}_2 \gamma_2^{-2}), \quad (18)$$

или

$$\dot{\gamma}_1 \gamma_2 - \gamma_1 \dot{\gamma}_2 = \bar{\theta}_2 (\dot{\gamma}_3 \gamma_2 - \gamma_3 \dot{\gamma}_2). \quad (19)$$

Обозначим $\bar{z}(t) = \dot{\gamma}_1 \gamma_2 - \gamma_1 \dot{\gamma}_2$ и $\bar{\zeta}_2 = \dot{\gamma}_3 \gamma_2 - \gamma_3 \dot{\gamma}_2$. Тогда уравнение (19) выглядит следующим образом:

$$\bar{z}(t) = \bar{\theta}_2 \bar{\zeta}_2. \quad (20)$$

Из уравнения (20) легко получить идентификационный алгоритм для параметра $\bar{\theta}_2$:

$$\dot{\hat{\theta}}_2 = -k_2 \bar{\zeta}_2^2 \hat{\theta}_2 + k_2 \bar{\zeta}_2 \bar{z}, \quad (21)$$

где $k_2 > 0$ – некоторый параметр, увеличение которого приводит к увеличению скорости сходимости $\hat{\theta}_2$ к $\bar{\theta}_2$. Пренебрегая членом $\varepsilon(t)$ в уравнении (15), для идентификации параметра $\bar{\theta}_1$ можем использовать следующий алгоритм:

$$\dot{\hat{\theta}}_1 = -k_1 \zeta_1^2 \hat{\theta}_1 + k_1 \zeta_1 z', \quad (22)$$

где $z' = z - \hat{\theta}_2 \zeta_2$ и $k_1 > 0$ – некоторый параметр, увеличение которого, как и в предыдущем случае, приводит к увеличению скорости сходимости $\hat{\theta}_1$ к $\bar{\theta}_1$.

Таким образом, мы получили гибридную схему идентификации параметров $\bar{\theta}_1 = -\omega_1^2 - \omega_2^2$ и $\bar{\theta}_2 = \omega_1^2 \omega_2^2$. Решая квадратичное уравнение, легко получить из значений параметров $\bar{\theta}_1$ и $\bar{\theta}_2$ оценку частот.

Запишем уравнения для определения частот сигнала:

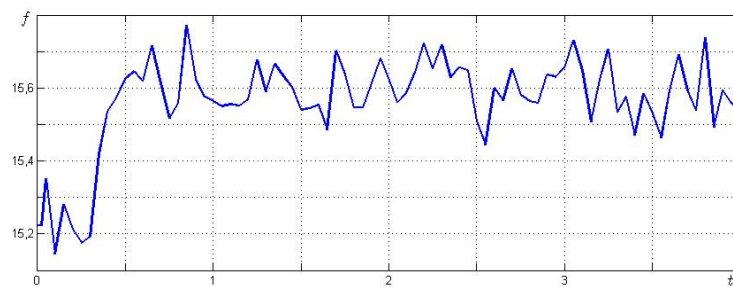
$$\omega_1 = \sqrt{\frac{\bar{\theta}_1 + \sqrt{|\bar{\theta}_1^2 - 4\bar{\theta}_2^2|}}{2}}, \quad (23)$$

$$\omega_2 = \sqrt{|\bar{\theta}_1 - \omega_1^2|}. \quad (24)$$

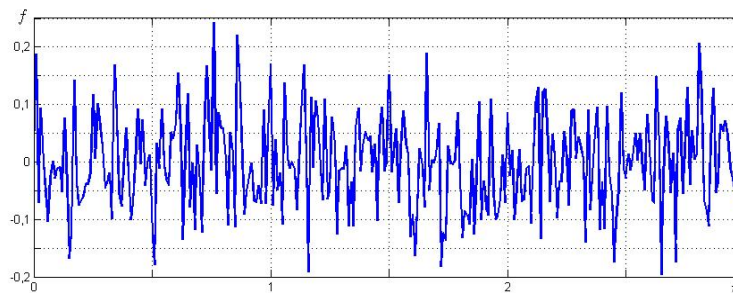
Взяв в расчет, что $f = \frac{\omega}{2\pi}$, получим результаты, которые представлены на рис. 4–6.

Таким образом, для идентификации частот основных гармоник сигнала ЭМГ в режиме реального времени был применен метод каскадной редукции. Полученные результаты совпадают со спектральным анализом с высокой степенью точности.

Получаемые с помощью представленного метода данные могут быть использованы в качестве дополнительной информации к классификаторам, описанным ранее, для идентификации типов совершенных движений на основе сигналов ЭМГ в режиме реального времени.

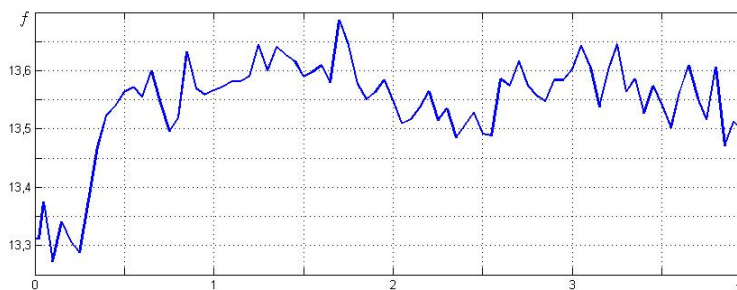


а

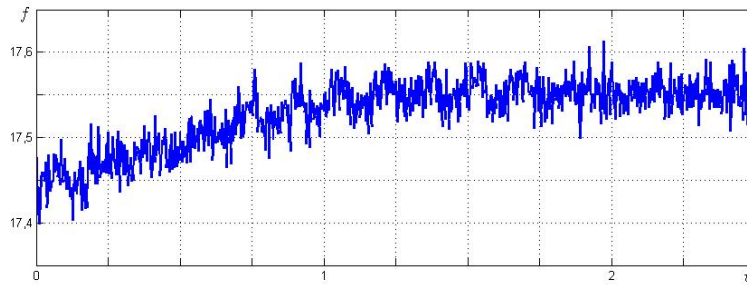


б

Рис. 4. Переходной процесс для оценки сигнала ЭМГ при сгибании кисти:
а – частота ω_1 ; б – частота ω_2



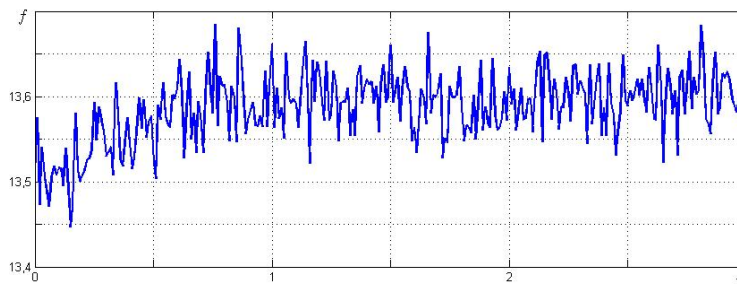
а



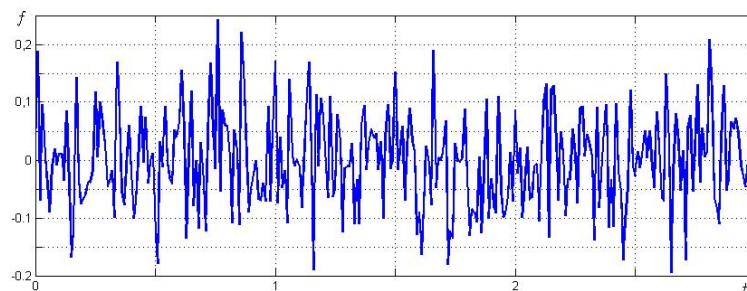
б

Рис. 5. Переходной процесс для оценки сигнала ЭМГ при вращении кисти по часовой стрелке:

а – частота ω_1 ; б – частота ω_2



а



б

Рис. 6. Переходной процесс для оценки сигнала ЭМГ при вращении кисти против часовой стрелки: а – частота ω_1 ; б – частота ω_2

Заключение

В данной работе были рассмотрены способы анализа сигнала ЭМГ для системы управления реабилитационным устройством. Был выбран наиболее подходящий метод по оценке состояния биопотенциалов на мышцах человека. Наряду с теоретическими обоснованиями были также рассмотрены практические аспекты применения результатов анализа сигнала ЭМГ для использования их в системах управления устройствами. Проведен ряд экспериментов по определению частот основных гармоник сигнала в режиме реального времени. Использование предлагаемых методов позволяет классифицировать тип движений человека с высокой точностью. Полученные результаты основываются на использовании метода каскадной редукции в режиме реального времени для определения частот мультигармонического сигнала.

Предложенные методы обработки сигналов ЭМГ могут применяться в системах управления передовыми робототехническими и мехатронными устройствами.

Публикации

1. *Zimenko K.A., Margun A.A., Kremlev A.S.* EMG real-time classification for robotics and HMI // 18th international conference on Methods and Models in Automation and Robotics (MMAR – 2013). Miedzydroje, 2013. P. 340–343.

2. *Zimenko K.A., Borgul A.S., Margun A.A., Kremlev A.S., Krasnov A.Y.* Intuitive control for robotic rehabilitation devices by human-machine interface with EMG and EEG Signals // 17th international conference on Methods and Models in Automation and Robotics (MMAR – 2012). Miedzyzdroje, 2013. P. 308–311.
3. *Bobtsov A.A., Kolyubin S.A., Pyrkin A.A., Borgul A.S., Zimenko K.A., Rabysh E.Y.* Mechatronic and robotic setups for modern control theory workshops // Preprints of the 9th IFAC Symposium on Advances in Control Education (ACE – 2012). Nizhny Novgorod, 2012. P. 348–353.
4. *Боргуль А.С., Зименко К.А., Маргун А.А.* Биоинформационные технологии в управлении техническими системами // Сб. тезисов докладов XIX Всероссийской научно-методической конференции «Телематика – 2012». СПб., 2012. С. 147.
5. *Боргуль А.С., Зименко К.А., Маргун А.А.* Биотехническая система управления устройством реабилитации // Сб. тезисов докладов I Всероссийского конгресса молодых ученых. СПб., 2012. С. 197.

Литература

1. Rohtash Dhiman et al. Detecting the useful electromyogram signals-extracting, conditioning and classification // Indian Journal of Computer Science and Engineering. 2011. Vol. 2. No 4. P. 634–637.
2. *Zimenko K.A., Borgul A.S., Margun A.A. et al.* Intuitive control for robotic rehabilitation devices by human-machine interface with EMG and EEG Signals // 17th international conference on Methods and Models in Automation and Robotics (MMAR – 2012). Miedzyzdroje, 2013. P. 308–311.
3. *Aranovskiy S., Bobtsov A., Kremlev A. et al.* Identification of frequency of biased harmonic signal // European Journal of Control. 2010. Vol. 16. No 2. P. 129–139.
4. *Pearson R.* Theoretical basis of the interpretation of EMG // Human Physiology. 1987. Vol. 13. No 4. P. 65–67.
5. *Pyrkin A., Kremlev A., Bobtsov A., Kolyubin S.* An iterative algorithm of adaptive output control with complete compensation for unknown sinusoidal disturbance // Automation and Remote Control. 2012. Vol. 72. No 8. P. 1327–1336.
6. Статьи о теории и практике цифровой обработки сигналов: <http://www.dsplib.ru/content.html>.
7. *Бернштейн Н.А.* Физиология движений и активность. М., 1990. – 496 с.
8. *Oskoei M., Hu H.* Myoelectric control systems – A survey // Biomedical Signal Processing and Control. 2007. Vol. 2. No 4. P. 275–294.
9. *Subasi A., Yilmaz M., Ozcalik H.R.* Classification of EMG signals using wavelet neural network // Journal of Neuroscience Methods. 2006. Vol. 156. No 1–2. P. 360–367.
10. *Parker P., Englehart K., Hudgins B.* Myoelectric signal processing for control of powered limb prostheses // Journal of Electromyography and Kinesiology. 2006. Vol. 16. No 6. P. 541–548.



Казаков Сергей Владимирович
Факультет информационных технологий и программирования
Кафедра компьютерных технологий
E-mail: svkazakov@rain.ifmo.ru

РАЗРАБОТКА АЛГОРИТМА УПРОЩЕНИЯ ГРАФА ПЕРЕКРЫТИЙ ПРИ СБОРКЕ ГЕНОМНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

С.В. Казаков

Научный руководитель – доцент Ф.Н. Царев

Введение

Многие современные задачи биологии и медицины требуют знания генома живых организмов, который состоит из нескольких нуклеотидных последовательностей ДНК.

При изучении генома живого существа обычно выделяют три основных этапа:

- секвенирование молекул ДНК, содержащих информацию о геноме (выполняется с использованием специальных устройств – секвенаторов);
- сборка геномной последовательности (или коротко – сборка генома, выполняется с использованием компьютеров);
- анализ и сравнение геномов (выполняется с использованием компьютеров).

Изучение генома человека и других живых существ имеет важное прикладное значение. На основании результатов сборки генома конкретного человека возможна реализация персонализированной медицины – определения его предрасположенности к различным болезням, создание индивидуальных лекарств и пр. Кроме этого, на основе результатов исследования геномов растений и животных с применением методов биоинженерии могут быть выведены новые их виды, обладающие определенными свойствами.

Еще одно важное приложение исследования ДНК – генетические заболевания. У особей, зараженных одним генетическим заболеванием, наблюдаются одинаковые изменения в ДНК, что может быть использовано в медицине как для теоретического исследования заболевания, так и для его лечения.

Задача разработки методов сборки геномных последовательностей является в определенном смысле центральной среди всех задач биоинформатики. Это объясняется тем, что без ее решения нельзя приступить к детальному изучению генома живого существа и его анализу с применением других алгоритмов биоинформатики.

При решении этой задачи возникает большое число трудностей. Цель работы – улучшить существующие решения одной из них, а именно – проблемы запутанности графа перекрытий при сборке контигов.

Сборка генома

Сборка геномной последовательности – процесс получения больших фрагментов генома – выполняется с использованием компьютеров. Обычно решением этой задачи является набор контигов, объединенных в скэффолды.

Контигом называется непрерывная последовательность нуклеотидов. Скэффолдом – последовательность контигов с оценкой на расстояния между ними. При этом предполагается, что такие контиги и скэффолды действительно присутствуют в геноме.

В рамках данной работы будет рассматриваться так называемая задача *de novo* сборки генома – сборки генома живого существа, для которого геном еще не известен. Сложность ее обусловлена следующими факторами:

- большой объем входных данных;
- сложность структуры генома – наличие в нем повторов и полиморфизмов;

- наличие ошибок в исходных данных, полученных с секвенаторов.

Для решения указанных проблем сборку геномной последовательности обычно разбивают на этапы:

- исправление ошибок в исходных данных;
- восстановление фрагментов генома по парным чтениям (такие фрагменты называются квазиконтигами);
- сборка контигов (расширенных фрагментов) из квазиконтигов;
- построение скэффолдов – упорядоченных контигов с оценкой на расстояния между ними.

В работе рассматривается сборщик генома *ITMO Genome Assembler* [1–3], который разрабатывается в УНИВЕРСИТЕТ ИТМО. Данный сборщик хорошо зарекомендовал себя в рамках проектов *de novo Genome Assembly Project* [4] и *Assemblathon 2* [5].

На его основе будут тестироваться предлагаемые алгоритмы. Архитектура сборщика *ITMO Genome Assembler* показана на рис. 1.

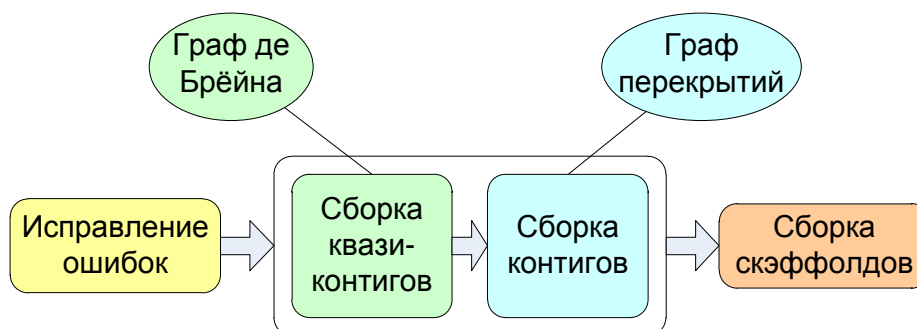


Рис. 1. Архитектура сборщика

Сборка контигов, граф перекрытий

Интересующий нас этап – сборка контигов. В простом понимании, этап сборки контигов из квазиконтигов нужен, для того чтобы расширить квазиконтиги в обоих направлениях настолько, насколько это возможно.

Сборка контигов из квазиконтигов основана на подходе *overlap-layout-consensus* и состоит из нескольких этапов:

- поиск перекрытий;
- удаление транзитивных перекрытий;
- построение графа перекрытий и его упрощение;
- вывод первого приближения контигов;
- микросборка.

Одним из этапов сборки является этап построения графа перекрытий и его упрощение.

Графом перекрытий называется граф, вершины которого – квазиконтиги, а ребра между ними проводятся, если соответствующие квазиконтиги перекрываются. При этом на ребре пишется длина соответствующего перекрытия. Пример графа перекрытий изображен на рис. 2.

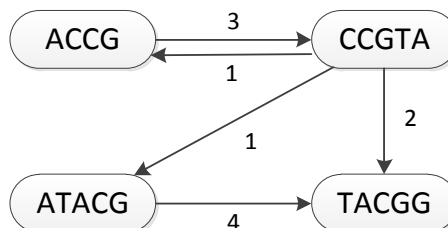


Рис. 2. Граф перекрытий

Граф перекрытий довольно легко построить, зная все перекрытия между квазиконтигами. Он очень удобен для последующей работы над перекрытиями и квазиконтигами, благодаря

тому, что наглядно отображает ситуацию. При построении графа обычно выясняется, что он очень запутан. Одна из маленьких частей реального графа перекрытий изображена на рис. 3.

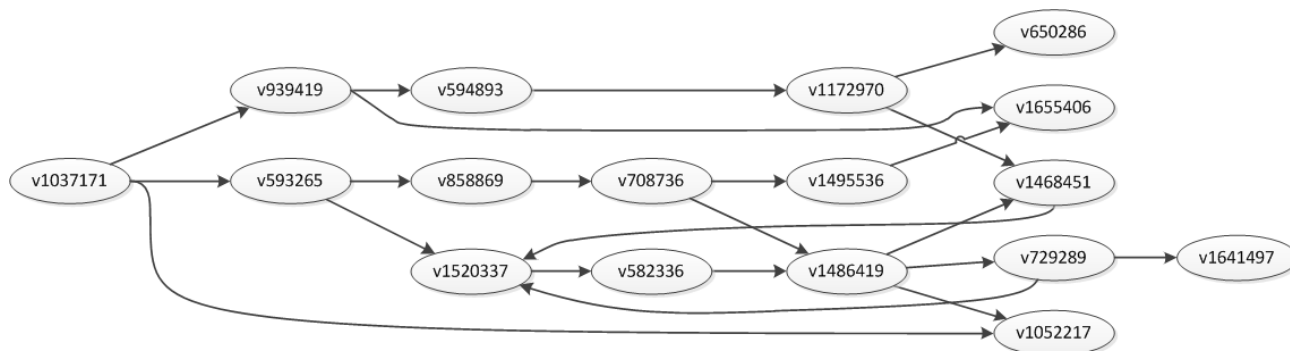


Рис. 3. Часть графа перекрытий

Уже на следующем этапе – стадии *Layout* (вывод первого приближения контигов) – происходит вывод контигов за счет использования простых путей в графе перекрытий. Один контиг получается из одного простого пути в графе перекрытий.

Таким образом, наиболее интересная и трудная задача в процессе сборки контигов из квазиконтигов – не просто расширить квазиконтиги, а распутать граф перекрытий, то есть выделить в нем простые пути.

Для части графа, изображенного на рис. 3, часть графа, которая должна получиться после упрощения, изображена на рис. 4. Из этого рисунка видно, что из полученной части графа получается два контига, один из которых – объединение вершин (квазиконтигов) в первом пути, второй – во втором.



Рис. 4. Распутанный граф перекрытий

Традиционно при сборке контигов выделяют приоритеты:

- уменьшить число ошибочных контигов, то есть тех, которые не встречаются в геноме;
- увеличить среднюю длину собранных контигов (при неизменной суммарной длине всех собранных контигов).

Ошибочные контиги возникают в большинстве случаев из-за того, что путь, который им соответствует, содержал ошибочное ребро, то есть такое, которое соединяло два квазиконтига, которые на самом деле были расположены в разных частях генома. Проходя по нему, мы соединяем две разные несмежные части генома. Ошибочные ребра возникают чаще всего из-за повторяющихся фрагментов, которые в большом количестве присутствуют в любом геноме.

Объясним почему. Предположим у нас есть две разные части одного генома. Пусть квазиконтиг *A* находится в первой части, а квазиконтиг *B* – во второй. Допустим, в этих частях есть небольшой одинаковый фрагмент, будем называть его «повтор». Поясняющая схема изображена на рис. 5.

Если квазиконтиг *A* заканчивается на этом повторе (не покрывая его полностью), а квазиконтиг *B* начинается на нем (тоже не покрывая его полностью), то между квазиконтигом *A* и квазиконтигом *B* будет найдено перекрытие (или ребро). Оно будет ошибочным, так как соединяет два квазиконтига из разных частей генома.

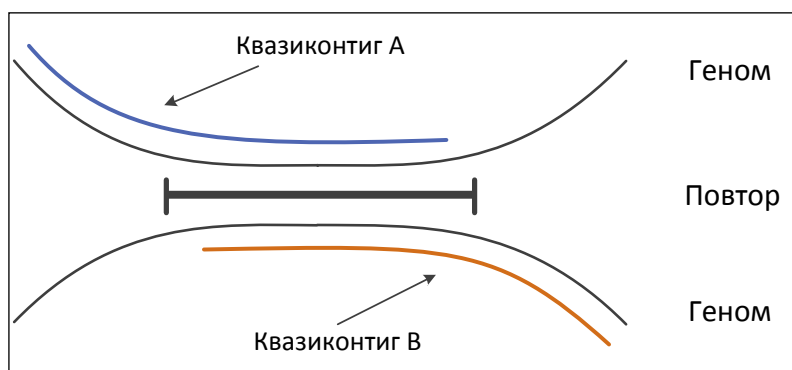


Рис. 5. Ошибочное ребро

На первый взгляд может показаться, что такое случается крайне редко. Однако это не так.

Можно взять для примера распространенный в биоинформатике геном кишечной палочки (*E. coli*, длина генома – 4,6 млн нуклеотидов). Если построить график зависимости числа повторяющихся фрагментов в геноме от размера этого фрагмента (рис. 6), то можно заметить, что число повторов в геноме очень велико вне зависимости от их длины, и они встречаются «на каждом шагу». Причем, чем меньше длина повтора, тем чаще повтор такой длины встречается в геноме.

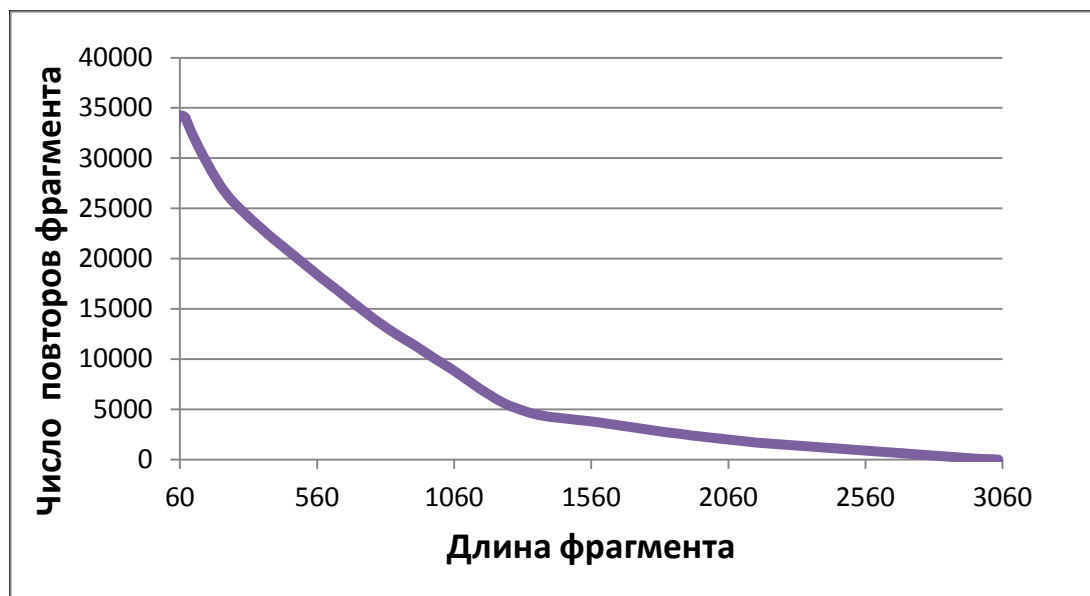


Рис. 6. Зависимость числа повторяющихся фрагментов генома от его длины

Также с помощью генома *E. coli* было экспериментально подсчитано, что, если покрыть его равномерно чтениями (среднее покрытие каждого нуклеотида генома – 40, это значит, что каждый нуклеотид генома встречается в чтениях в среднем 40 раз), то процент ошибочных ребер перед этапом упрощения графа перекрытий будет равен 0,5 % от всех ребер. Эта величина относительно небольшая, но именно из-за таких ребер качество сборки контигов заметно ухудшается.

Решения существующие и предлагаемое

Существующие решения для упрощения графа перекрытий так или иначе используют следующие подэтапы:

- объединение схожих путей;
- удаление отростков;
- анализ развилок и их упрощение.

Сборщик генома *ITMO Genome Assembler* тоже использует эти три этапа. Однако процесс упрощения графа остается малоэффективным из-за ошибочных ребер. Такие ребра, во-первых,

участвуют в создании ошибочных контигов, а во-вторых, мешают всем другим этапам упрощения, так как превращают граф в «запутанный клубок».

Существующие решения так или иначе пытаются бороться с ошибочными ребрами, однако исключить их полностью практически невозможно. Для борьбы с такими ребрами предлагается изменить этап упрощения графа перекрытий: добавить этап вычисления правдоподобия каждого ребра, этап удаления малоправдоподобных ребер, а также использовать числовую характеристику правдоподобия ребра в других подэтапах упрощения.

Предлагаемый этап упрощения графа перекрытий будет выглядеть следующим образом:

- вычисление правдоподобия каждого ребра;
- удаление малоправдоподобных ребер;
- объединение схожих путей;
- удаление отростков;
- анализ развилки и их упрощение (с использованием характеристики правдоподобности ребер).

Благодаря этапу удаления малоправдоподобных ребер становится возможным удалять большую часть ошибочных ребер. Таким образом, кроме непосредственного распутывания графа перекрытий, такое удаление делает работу следующих этапов более эффективной.

Правдоподобие ребра предлагается вычислять на основе двух оценок:

- теоретической, которая вычисляется с использованием длины перекрытия и модели распределения повторов в геноме в зависимости от длины повтора. Такая модель строится благодаря анализу уже собранных геномов, а для несобранных является хорошим приближением к настоящей модели распределения повторов;

- экспериментальной, вычисляемой на основании анализа изменения покрытия строками небольшой длины перекрывающихся квазиконтигов. Она полностью зависит от имеющихся исходных данных, и для ее вычисления не нужно знать никакой другой информации о геноме.

Правдоподобность ребра вычисляется по следующей формуле:

$$P(e) = 1 - P_w(e),$$

где $P(e)$ – правдоподобность ребра e , $P_w(e)$ – оценочная вероятность ребра оказаться ошибочным.

$P_w(e)$ вычисляется на основе двух оценок – теоретической и экспериментальной:

$$P_w(e) = \alpha P_t(e) + \beta P_{exp}(e),$$

где $P_t(e)$ – теоретическая оценка вероятности ребра оказаться ошибочным, $P_{exp}(e)$ – экспериментальная, α и β – коэффициенты учета соответствующих оценок ($\alpha + \beta = 1$).

Теоретическая оценка

Теоретическая оценка вероятности ребра оказаться ошибочным $P_t(e)$ вычисляется с использованием данных о длине перекрытия и о модели распределения повторов в зависимости от длины повтора.

Более формально $P_t(e)$ – вероятность того, что существует повтор, который полностью покрывает перекрывающуюся часть квазиконтигов A и B . Пример такой ситуации изображен на рис. 7.

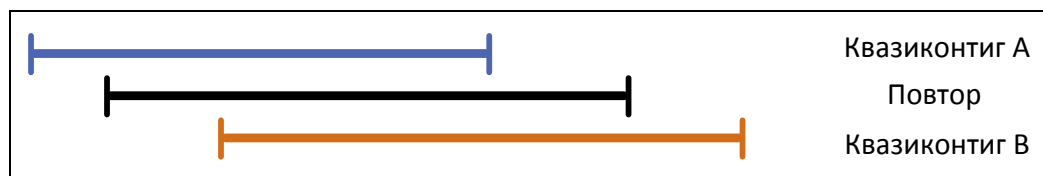


Рис. 7. Схема ошибочного перекрытия квазиконтигов

Экспериментальная оценка

Экспериментальная оценка вероятности ребра оказаться ошибочным $P_{exp}(e)$ вычисляется с использованием статистики о покрытии перекрывающихся квазиконтигов строками небольшой длины (k -мерами). Такая оценка полностью зависит только от исходных данных и не нуждается ни в какой другой информации о геноме.

Например, пусть имеются квазиконтиги A и B , которые перекрываются. Для каждого из них можно подсчитать, сколько раз каждый k -мер в квазиконтиге встречается в исходных парных чтениях. Тогда можно будет построить график зависимости частоты появления k -мера в исходных данных от позиции в соответствующем квазиконтиге.

Заметим, что квазиконтиги A и B перекрываются. Это означает, что можно построить объединенный квазиконтиг $C = A_u \cup OV_{A,B} \cup B_u$, где A_u и B_u – уникальные части квазиконтигов A и B , $OV_{A,B}$ – перекрывающаяся часть квазиконтигов A и B . Поясняющая схема представлена на рис. 8.

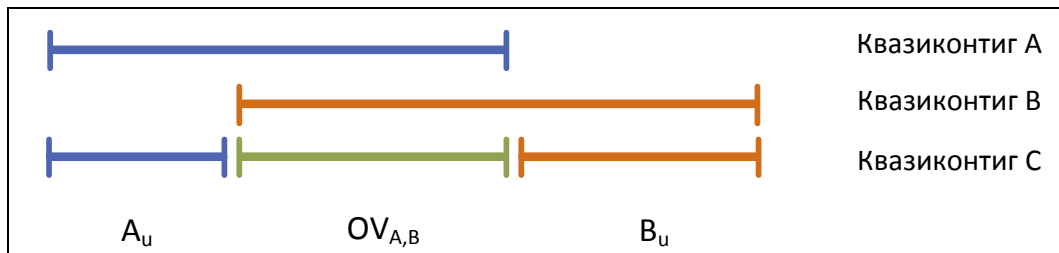


Рис. 8. Схема объединенного квазиконтига

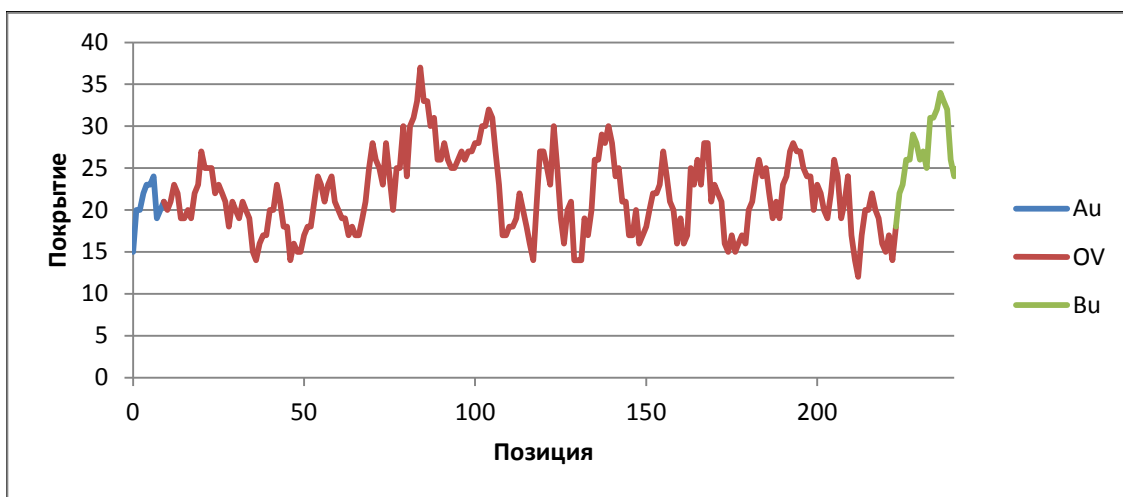
Для объединенного квазиконтига C тоже можно построить график зависимости частоты появления k -мера в исходных данных от позиции в нем. Такой график будет объединением двух графиков для квазиконтигов A и B .

В идеальной ситуации, когда квазиконтиги перекрываются по части $OV_{A,B}$, которая полностью совпадает с соответствующим суффиксом квазиконтига A , а также с соответствующим префиксом квазиконтига B , конец графика зависимости частоты k -мера от позиции для квазиконтига A будет совпадать с началом графика зависимости для квазиконтига B . По такому совпадению они и объединяются. Однако, если квазиконтиги A и B перекрываются с ошибками, то при получении графика для квазиконтига C соответствующие части графиков A и B усредняются.

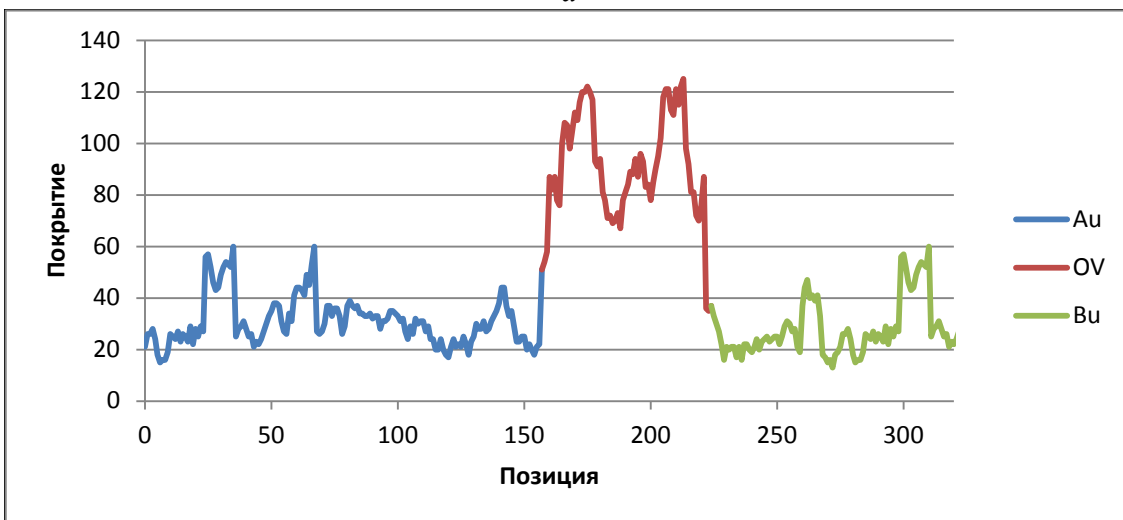
Некоторые возможные зависимости частоты появления k -мера в исходных данных от позиции в квазиконтиге C для реальных данных изображены на рис. 9а, 9б. На них синим цветом показана зависимость появления k -мера в исходных данных для префикса квазиконтига A до перекрывающейся части, красным – для перекрывающейся части и зеленым – для суффикса квазиконтига B после перекрывающейся части.

По этим графикам можно заметить, что их характер поведения отличается. Попытаемся это объяснить.

Предположим, что квазиконтиги A и B действительно находятся рядом и перекрываются как раз из-за этого. Тогда можно допустить, что k -меры из объединенного квазиконтига C не встречаются в других местах генома (то есть нет повторов, которые бы совпадали с частью квазиконтига C и появлялись бы в других местах генома). Такое предположение верно с большой вероятностью, так как суммарная длина всех повторов намного меньше длины генома. Тогда (если k -меры из квазиконтига C не встречаются в других местах генома) с большой вероятностью частота появления этих k -меров в исходных данных будет примерно одинаковой, так как геном равномерно покрыт квазиконтигами. Такое поведение хорошо наблюдается на рис. 9а.



а



б

Рис. 9. Графики зависимости частоты появления k -мера в исходных данных от его позиции в квазиконтиге.

Предположим теперь, что квазиконтиги A и B находятся в разных частях генома и перекрываются только из-за повтора (как изображено на рис. 5). Тогда, ввиду того что повтор встречается как минимум два раза в геноме, k -меры, полученные из него, будут появляться в исходных данных намного больше раз, чем k -меры из других частей (например, чем k -меры из префикса квазиконтига A или из суффикса квазиконтига B). Такую зависимость можно наблюдать на рис. 9б).

Экспериментальная оценка вероятности ребра оказаться ошибочным как раз ориентируется на такую особенность. Для выявления этой ситуации перебираются возможные границы повтора ($R.begin$, $R.end$), такие что:

$$\begin{cases} A.begin < R.begin \wedge R.begin \leq B.begin \\ A.end \leq R.end \wedge R.end < B.end \end{cases}$$

Поясняющая схема изображена на рис. 10.

Далее предполагаем, что повтор начинается именно в этих границах. После чего пытаемся полученный график зависимости частоты появления k -мера в исходных данных от его позиции в квазиконтиге аппроксимировать схемой, изображенной на рис. 11.

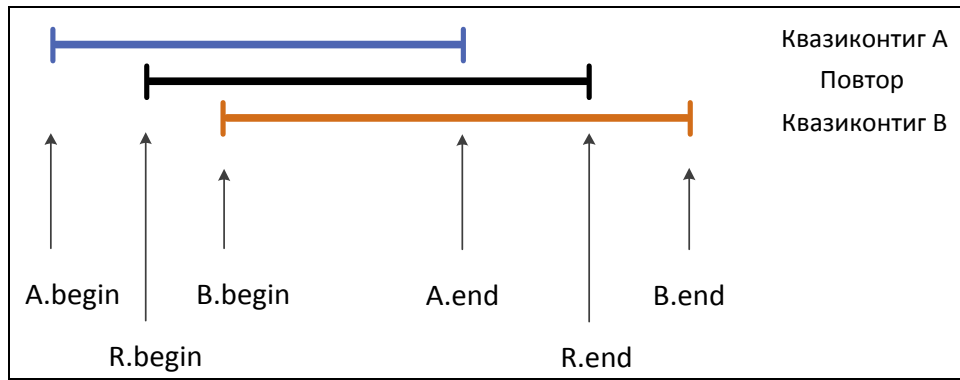


Рис. 10. Схема ошибочного перекрытия квазиконтигов

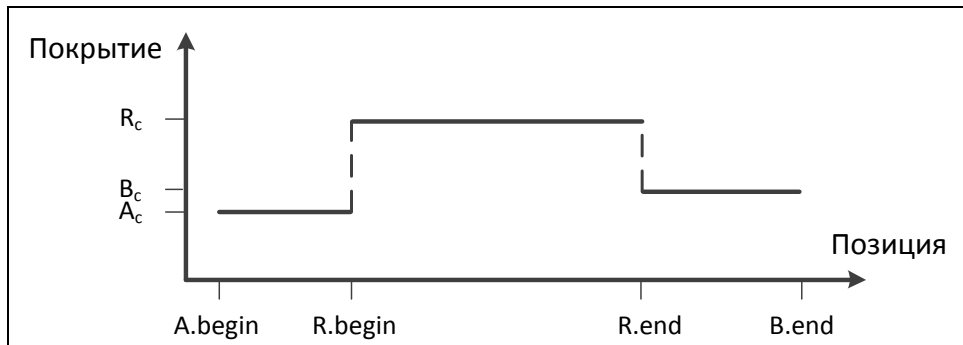


Рис. 11. Зависимость покрытия от позиции в случае, если есть повтор

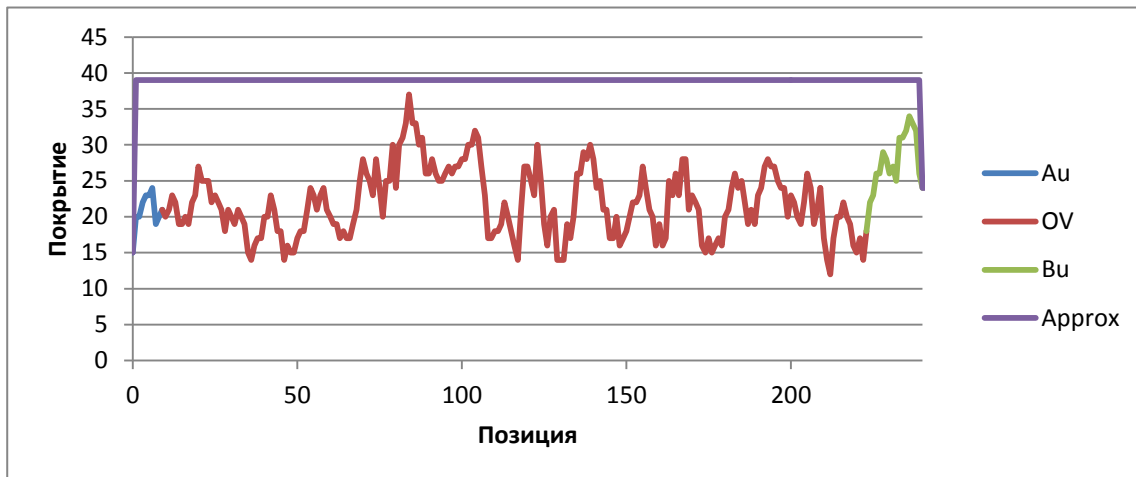
При этом A_c вычисляется как среднее значение на участке $[A.begin, R.begin)$, B_c – на участке $(R.end, B.end]$, R_c – на участке $[R.begin, R.end]$. Отметим, что R_c не должно быть меньше величины A_c+B_c , иначе R_c увеличивается до этого значения.

Далее вычисляем коэффициент несоответствия реального графика и аппроксимации как отношение площади «ошибочной части» (той части, которая отличает один график от другого) к общей площади реального графика.

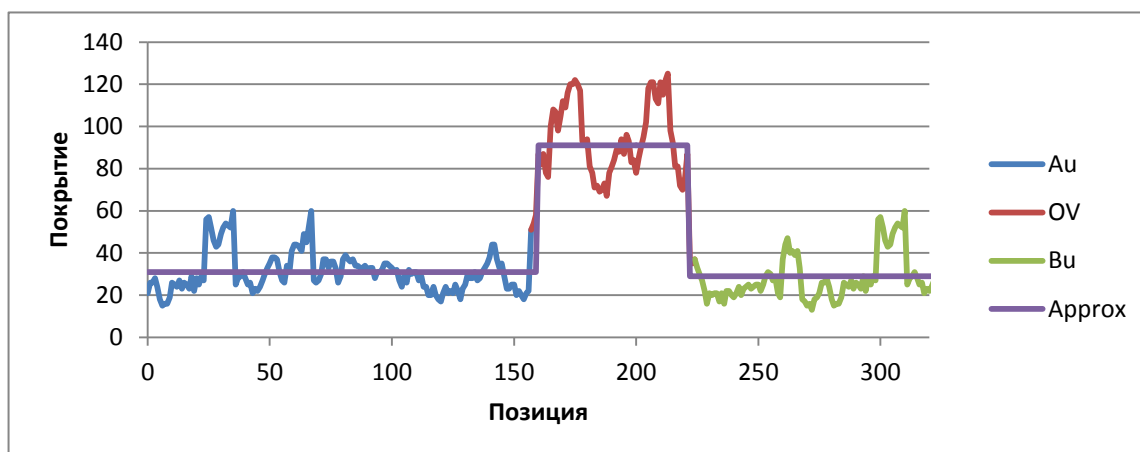
Перебирая всевозможные границы повтора ($R.begin, R.end$), можно выбрать такие, при которых коэффициент несоответствия будет минимальным. При этом график зависимости частоты появления k -мера в исходных данных от его позиции в квазиконтиге будет наиболее хорошо аппроксимироваться схемой, изображенной на рис. 11. Например, для графиков, изображенных на рис. 9, полученная аппроксимация будет выглядеть так, как показано на рис. 12а, 12б.

Таким образом, экспериментальную оценку ребра оказаться ошибочным можно записать

$$P_{exp}(e) = 1 - \min_{\substack{A.begin < R.begin \wedge R.begin \leq B.begin \\ A.end \leq R.end \wedge R.end < B.end}} \frac{S_{wrong}}{S}$$



а



б

Рис. 12. Графики зависимости частоты появления k -мера в исходных данных от его позиции в квазиконтиге.

Коэффициент несоответствия реального графика и аппроксимации для случая а) $k = 0,766$ (то есть ошибочная площадь занимает 76,6 % от площади реального графика), а для случая б) $k = 0,224$. Для варианта а) $P_{exp}(e) = 0,234$, а для б) $P_{exp}(e) = 0,776$.

Экспериментальные исследования

Экспериментальные исследования проводились над разными исходными данными трех бактерий с разными размерами генома. Информация об этих бактериях приведена в Табл. 1.

Таблица 1

Данные исследуемых бактерий

Название (англ.)	Сокращенное название	Название (рус.)	Длина собранного генома
<i>Buchnera aphidicola</i>	<i>Buchnera</i>	-	641'895
<i>Haemophilus influenzae</i>	<i>H. influenzae</i>	Гемофильная палочка, палочка Пфейфера	1'813'033
<i>Escherichia coli</i>	<i>E. coli</i>	Кишечная палочка	4'639'675

Каждая из этих бактерий имеет собранный геном, что помогало посчитать число ошибочных контигов, а также сгенерировать наборы исходных данных для экспериментов, которые были получены следующими способами:

- безошибочные квазиконтиги (сгенерированы из собранного генома);
- квазиконтиги с ошибками (сгенерированы из собранного генома, ошибки сгенерированы программой);
- реальные квазиконтиги (с ошибками, собраны сборщиком *ITMO Genome Assembler* из реальных парных чтений).

Для каждой бактерии было использовано два набора исходных квазиконтигов: безошибочные и реальные. Однако из-за отсутствия подходящих реальных парных чтений для бактерий *Buchnera* и *H. influenzae* были использованы сгенерированные квазиконтиги с ошибками.

Для каждого набора исходных данных для каждой бактерии запускался сборщик *ITMO Genome Assembler* без дополнительных предлагаемых этапов, а также с ними.

Приведем результаты запусков для бактерии *Buchnera* (длина генома – 0,6 млн нуклеотидов). Они показаны в Табл. 2 для безошибочных квазиконтигов и в Табл. 3 для квазиконтигов с ошибками. Как видно из этих таблиц, при добавлении предлагаемых дополнительных этапов результаты сборки не изменились.

Таблица 2

Результаты запусков для бактерии *Vichnera*; исходные данные – безошибочные квазиконтиги

Метод	Число контигов	Суммарная длина контигов	Средняя длина контига	Ошибочных контигов
<i>ITMO Genome Assembler</i> без предлагаемых этапов	33	642'561	19'472	1
<i>ITMO Genome Assembler</i> с предлагаемыми этапами	33	642'561	19'472	1

Таблица 3

Результаты запусков для бактерии *Vichnera*; исходные данные – квазиконтиги с ошибками

Метод	Число контигов	Суммарная длина контигов	Средняя длина контига	Ошибочных контигов
<i>ITMO Genome Assembler</i> без предлагаемых этапов	39	642'816	16'482	1
<i>ITMO Genome Assembler</i> с предлагаемыми этапами	39	642'816	16'482	1

Приведем результаты запусков для бактерии *H. influenzae* (длина генома – 1,8 млн нуклеотидов). Они показаны в Табл. 4 для безошибочных квазиконтигов и в Табл. 5 для квазиконтигов с ошибками.

Таблица 4

Результаты запусков для бактерии *H. influenzae*, исходные данные – безошибочные квазиконтиги

Метод	Число контигов	Суммарная длина контигов	Средняя длина контига	Ошибочных контигов
<i>ITMO Genome Assembler</i> без предлагаемых этапов	71	1'789'421	25'203	1
<i>ITMO Genome Assembler</i> с предлагаемыми этапами	62	1'788'234	28'842	0

Таблица 5

Результаты запусков для бактерии *H. influenzae*, исходные данные – квазиконтиги с ошибками

Метод	Число контигов	Суммарная длина контигов	Средняя длина контига	Ошибочных контигов
<i>ITMO Genome Assembler</i> без предлагаемых этапов	99	1'793'292	18'114	3
<i>ITMO Genome Assembler</i> с предлагаемыми этапами	74	1'789'284	24'180	1

Как видно из представленных таблиц, после добавления предлагаемых этапов результаты сборки значительно улучшились как по характеристике «средняя длина контига» (она увеличилась), так и по характеристике «число ошибочных контигов» (их число уменьшилось).

Результаты запусков для широко распространенного генома кишечной палочки *E. coli* (длина генома – 4,6 млн нуклеотидов) приведены в Табл. 6 (для безошибочных квазиконтигов) и Табл. 7 (для реальных квазиконтигов). Из таблиц видно, что после добавления предлагаемых этапов результаты сборки улучшились. В первом случае улучшения произошли по обоим характеристикам (средняя длина контига и число ошибочных контигов), во втором случае улучшения – только по числу ошибочных контигов (качество при этом тоже улучшилось, виду того что уменьшение числа ошибочных контигов – более приоритетная задача, чем увеличение средней длины).

Таблица 6

Результаты запусков для бактерии *E. coli*, исходные данные – безошибочные квазиконтиги

Метод	Число контигов	Суммарная длина контигов	Средняя длина контига	Ошибочных контигов
<i>ITMO Genome Assembler</i> без предлагаемых этапов	160	4'599'061	28'744	2
<i>ITMO Genome Assembler</i> с предлагаемыми этапами	140	4'598'439	32'846	1

Таблица 7

Результаты запусков для бактерии *E. coli*, исходные данные – реальные квазиконтиги

Метод	Число контигов	Суммарная длина контигов	Средняя длина контига	Ошибочных контигов
<i>ITMO Genome Assembler</i> без предлагаемых этапов	405	4'623'885	11'417	15
<i>ITMO Genome Assembler</i> с предлагаемыми этапами	418	4'629'350	11'075	9

Из представленных результатов экспериментальных исследований можно сделать следующие выводы:

- для разных геномов дополнительные этапы по-разному изменяют качество итоговой сборки. В большинстве случаев качество улучшается, однако оно также может оставаться без изменений. Случаев ухудшения качества из-за предложенных дополнительных этапов замечено не было;

- дополнительные этапы могут уменьшать среднюю длину контигов. Это может происходить из-за того, что ребро, находящееся в центре простого пути, может быть отмечено как маловероятное и впоследствии удалено.

Таким образом, предложенный подход упрощения графа перекрытий показал свою эффективность и применимость в процессе сборки геномных данных.

Области возможного использования результатов

Данный подход можно применять для улучшения качества сборки генома. На настоящий момент он уже используется в сборщике *ITMO Genome Assembler* и показывает хорошие результаты. Автор работы рекомендует внедрять предложенный алгоритм в другие сборщики геномных последовательностей и настраивать его для получения более качественных результатов сборки.

Заключение

В работе предложен метод упрощения графа перекрытий, основанный на вычислении правдоподобности ребер графа. Он вычисляет правдоподобности ребер с использованием как теоретических моделей, так и экспериментальных данных.

Метод был протестирован на нескольких наборах исходных данных разных бактерий. Результаты экспериментов показали применимость и эффективность предлагаемого подхода.

Таким образом, цели работы были достигнуты – существующие решения проблемы запутанности графа перекрытий были улучшены, проведены экспериментальные исследования предложенного подхода и выявлены сильные и слабые его стороны.

Публикации

1. Царев Ф.Н., Александров А.В., Казаков С.В. и др. Метод исправления ошибок в наборе чтений нуклеотидной последовательности // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. 2011. № 5. С. 81–84.

2. Александров А.В., Исенбаев В.В., Казаков С.В. и др. Метод сборки генома с помощью восстановления его фрагментов по парным чтениям // Сб. тезисов докладов конференции молодых ученых. Вып. 1. Труды молодых ученых. Санкт-Петербург, 2011. С. 220.
3. Александров А.В., Казаков С.В., Мельников С.В. и др. Метод сборки контигов геномных последовательностей на основе совместного применения графов де Брюина и графов перекрытий // Научно-технический вестник информационных технологий, механики и оптики. 2012. Вып. 6 (82). С. 93–98.

Литература

1. Царев Ф.Н., Александров А.В., Казаков С.В. и др. Метод исправления ошибок в наборе чтений нуклеотидной последовательности // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. 2011. № 5. С. 81–84.
2. Александров А.В., Исенбаев В.В., Казаков С.В. и др. Метод сборки генома с помощью восстановления его фрагментов по парным чтениям // Сб. тезисов докладов конференции молодых ученых. Вып. 1. Труды молодых ученых. Санкт-Петербург, 2011. С. 220.
3. Александров А.В., Казаков С.В., Мельников С.В. и др. Метод сборки контигов геномных последовательностей на основе совместного применения графов де Брюина и графов перекрытий // Научно-технический вестник информационных технологий, механики и оптики. 2012. Вып. 6 (82). С. 93–98.
4. De novo Genome Assembly Assesment Project: <http://cnag.bsc.es>.
5. Assemblathon 2: <http://assemblathon.org/assemblathon2>.



Лосенков Андрей Андреевич
Факультет компьютерных технологий и управления
Кафедра систем управления и информатики
 E-mail: alosenkov@gmail.com

МОДЕЛИРОВАНИЕ И ИДЕНТИФИКАЦИЯ НЕЛИНЕЙНЫХ СИСТЕМ С ПРИЛОЖЕНИЕМ К СИСТЕМАМ ГИДРОПРИВОДА

А.А. Лосенков

Научный руководитель – профессор А.А. Бобцов

Цель и задачи проекта

Гидравлические машины, системы гидропривода и устройства на их основе широко применяются практически во всех отраслях промышленности и народного хозяйства. Например, в строительстве, дорожном хозяйстве, лесозаготовках. В горнодобывающей промышленности активно применяются тяжелые гидрофицированные машины. Однако в системах автоматического управления гидропривод распространен не столь широко из-за существенной нелинейности протекающих в нем процессов, сложности составления точных математических моделей элементов гидропривода и синтеза законов управления на их основе.

На практике управление рабочими органами гидрофицированной машины и ее перемещением производится оператором в ручном режиме при помощи различных джойстиков и рычагов, что является весьма непростой задачей для человека. Подготовка квалифицированных кадров занимает несколько лет. К тому же точность и скорость работы сильно зависят от навыков, опыта и самочувствия оператора, что ограничивает достижимые сроки, объемы и качество выполнения работ.

Автоматизация гидравлических приводов позволила бы существенно сократить удельные затраты времени и топлива, а также исключить зависимость от квалификации и состояния оператора. Следует отметить, что построение качественных систем управления гидроприводами рабочих органов гидрофицированных машин позволило бы в дальнейшем полностью автоматизировать машину, и она могла бы качественно выполнять полезные функции без участия человека.

Однако важность разработки моделей и методов идентификации параметров сложных нелинейных систем гидропривода имеет не только практическое, но и научное значение. Теория идентификации нелинейных систем – это активно развивающееся и актуальное направление, содержащее большое число нерешенных задач. Разработка нового метода идентификации параметров системы, содержащей степенные, мультипликативные и разрывные нелинейности, каковой является система гидропривода, может внести вклад в развитие этой области знаний.

В существующих работах, как правило, оперируют упрощенными моделями, область применения которых в большей степени ограничивается лабораторными малогабаритными и маломощными установками. При попытках же применить эти модели и методы идентификации для решения задач управления промышленного оборудования оказывается, что они недостаточно точно описывают наблюдаемые в гидроприводе процессы. Обусловлено это неидеальностями элементов гидропривода, которыми возможно пренебречь в лабораторных условиях, но в полевых – их влияние оказывается существенным. Например, в промышленной гидрофицированной технике распространено применение сложных регенеративных схем гидрораспределителя, в которых полости гидроцилиндров соединены между собой. За счет этого в регенеративном режиме обеспечивается увеличение скорости при уменьшении развиваемого усилия. Несмотря на это, в научной литературе подобное промышленное решение практически не освещено: отсутствуют модели, способные качественно описывать процессы, протекающие в гидроприводах с регенеративным распределением.

Итак, целью проекта является разработка новых математических моделей гидроприводов и новых методов идентификации, учитывающих особенности, присущие промышленным системам, что позволило бы внести вклад в заполнение пробела между научными исследованиями и промышленностью. Для достижения поставленной цели предлагается решить следующие задачи: выделение в составе промышленного гидропривода отдельных элементов, не рассматриваемых в большинстве упрощенных моделей; составление математического описания их работы, разработка методов идентификации параметров полученных моделей.

Методы исследования

В ходе выполнения работы были использованы методы нелинейной теории управления, теории идентификации, оптимизации, компьютерного моделирования, адаптации, а также оригинальные методы идентификации.

Основные достигнутые результаты

В ходе исследований были рассмотрены математические модели следующих элементов: гидрораспределителя с нелинейной статической характеристикой; гидравлических цилиндров с регенеративной схемой подключения; соответствующих потоков рабочей жидкости и динамики давлений при различных положениях гидрораспределителя. Предложен метод идентификации модели гидропривода с регенеративным гидрораспределением, в котором, в отличие от большинства известных методов, неизвестные параметры подразделяются на группы и идентифицируются отдельно. При идентификации большей части параметров используются данные о статическом равновесии в системе, что позволяет избежать вычисления производных давления. Кроме того, метод не требует установки дополнительных датчиков в систему гидропривода или применения каких-либо стенов для определения неизвестных параметров модели, что позволяет производить идентификацию непосредственно

на индустриальной гидрофицированной машине и использовать модель для построения системы автоматического управления.

Области возможного использования результатов проекта

Результаты научных исследований вносят вклад в области моделирования и идентификации нелинейных систем: разработаны новые математические модели и новые методы идентификации параметров. Помимо этого они имеют самостоятельное практическое применение при разработке систем управления и автоматизации гидрофицированной техники, например – лесозаготовительной. Разработка и внедрение синтезированных на базе предложенных моделей систем автоматического управления привело бы к повышению эффективности лесозаготовительной и деревообрабатывающей промышленности, что актуально для Северо-Западного региона Российской Федерации.

Публикации

1. Лосенков А.А., Арановский С.В. Моделирование и идентификация гидравлической системы с регенеративным гидрораспределением // Подготовлен макет статьи.



Маргун Алексей Анатольевич
Факультет компьютерных технологий и управления
Кафедра систем управления и информатики
E-mail: alexeimargun@gmail.com

АНАЛИЗ И СИНТЕЗ РЕГУЛЯТОРОВ АНТРОПОМОРФНОГО МАНИПУЛЯТОРА

А.А. Маргун

Научный руководитель – профессор А.А. Бобцов

Цель работы – создание математической модели и синтез регуляторов систем интеллектуального управления роботами и мехатронными комплексами для реабилитации людей с ограниченными возможностями при условии того, что кинематика рассматриваемых устройств подобна подвижности верхней конечности человека. На первом этапе выполнения проекта была определена кинематическая схема, соответствующая рассматриваемому устройству. В качестве модели выбрана механическая схема устройства, повторяющая механику руки человека. Конструкция такого устройства практически совпадает с устройством манипулятора, имеющего три вращательных степени свободы (рис. 1).

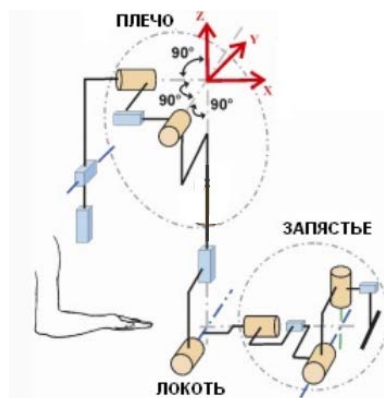


Рис. 1. Кинематическая схема устройства

Для данного класса устройств были решены кинематические задачи. Предметом кинематики манипулятора является аналитическое описание геометрии его движения относительно некоторой заданной абсолютной системы координат без учета сил и моментов, порождающих это движение. Из этого следует, что кинематические задачи являются математическим описанием пространственного расположения манипулятора в зависимости от времени. Данные соотношения также устанавливают связи между значениями присоединенных координат манипулятора, положением и ориентацией его схвата в декартовом пространстве. Обобщенные координаты были получены на основании алгоритма Денавита-Хартенберга (рис. 2).

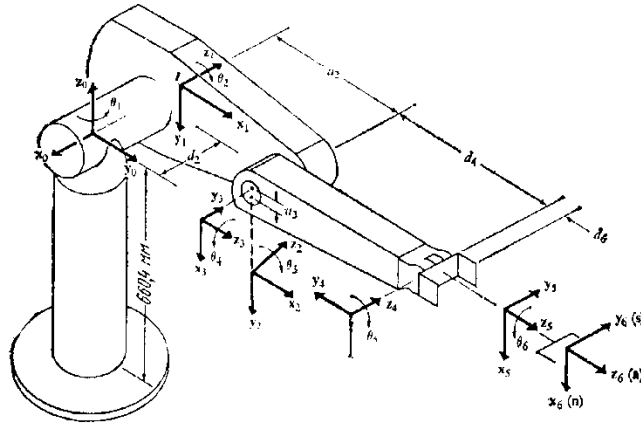


Рис. 2. Локальные системы координат звеньев манипулятора

После введения обобщенных и локальных систем координат для всех звеньев манипулятора были построены однородные матрицы преобразования, связывающие i -ю и $(i-1)$ -ю системы координат:

$${}^{i-1}A_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos Q_i & -\sin Q_i & 0 & 0 \\ \sin Q_i & \cos Q_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & a_i \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times$$

$$\times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha_i & -\sin \alpha_i & 0 \\ 0 & \sin \alpha_i & \cos \alpha_i & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos Q_i & -\cos \alpha_i \sin Q_i & \sin \alpha_i \sin Q_i & a_i \cos Q_i \\ \sin Q_i & \cos \alpha_i \cos Q_i & -\sin \alpha_i \cos Q_i & a_i \sin Q_i \\ 0 & \sin \alpha_i & \cos \alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

где α_i, a_i, d_i – параметры Денавита-Хартенберга (в рассматриваемой системе – константы), а θ_i – обобщенная координата вращательного сочленения [1].

Используя матрицу ${}^{i-1}A_i$, были получены обобщенные координаты точки p относительно i -й системы координат (точка p покоится в i -й системе координат) относительно $(i-1)$ -й локальной системы отсчета, связанной с $(i-1)$ -м звеном. Эта связь устанавливается равенством:

$$p_{i-1} = {}^{i-1}A_i p_i. \quad (2)$$

Для рассматриваемой системы также была решена обратная задача кинематики. Решение ее было получено за счет геометрических связей реабилитационного устройства. Однако зачастую обратная задача кинематики имеет множество решений. Для определения единственности решения в систему были введены переключатели, определяющие конфигурацию устройства [2].

$$\begin{aligned} \text{РУКА} &= \begin{cases} +1, & \text{для ПРАВОЙ руки} \\ -1, & \text{для ЛЕВОЙ руки} \end{cases} & \text{ЛОКОТЬ} &= \begin{cases} +1, & \text{для ВЕРХНЕЙ руки} \\ -1, & \text{для НИЖНЕЙ руки} \end{cases} \\ & & \text{ЗАПЯСТЬЕ} &= \begin{cases} +1, & \text{если КИСТЬ ВНИЗ} \\ -1, & \text{если КИСТЬ ВВЕРХ} \end{cases} \end{aligned}$$

Далее было получено уравнение динамики манипулятора с шестью степенями свободы как в общем виде для манипулятора с произвольной конфигурацией, так и для случая биологического манипулятора, имеющего шесть вращательных степеней свободы:

$$\tau(t) = d(q(t)) \cdot \ddot{q}(t) + c(q(t), \dot{q}(t)) + g(q(t)). \quad (3)$$

Для получения уравнения динамики использован метод Эйлера-Лагранжа, определены выражения для кинетической и потенциальной энергий устройства. Используя их, были получены матрицы инерции d , описывающие взаимовлияние звеньев друг на друга, матрица кариолисовых и центробежных сил – c , матрица g , описывающая гравитационное воздействие на биологический манипулятор. С помощью этих уравнений можно рассчитать динамику любого антропоморфного манипулятора.

Далее была решена задача управления данным классом устройств. Она сводится к расчету управляющих сигналов, подаваемых на сочленения и обеспечивающих положение схвата, необходимое для выполнения желаемого движения. Входными сигналами могут быть как силы и моменты, действующие в сочленениях, так и входные параметры исполнительных устройств, такие как сила тока или напряжение, подаваемые на двигатель в зависимости от используемой модели регулятора. Желаемое движение манипулятора может задаваться последовательностью конечных положений схвата или в виде непрерывного движения.

На сегодняшний день существует множество методов управления манипуляторами. Выбор определенного метода, а также его реализация, могут оказать существенное влияние на производительность, а следовательно, и на область возможного применения. Например, построение регулятора при непрерывном отслеживании положения манипулятора существенно отличается от управления при переходе от точки к точке. Кроме того, на выбор алгоритма управления влияет и конструкция манипулятора. Например, разработчики системы управления картезианского манипулятора сталкиваются с совершенно другими проблемами по сравнению с манипулятором, состоящим из вращательных сочленений. Это создает необходимость поиска компромисса между аппаратной/программной архитектурами манипулятора и системы управления.

Механическая конструкция роботов постоянно совершенствуется, что расширяет возможности их применения. Однако повышение производительности требует более сложных подходов к управлению. Можно провести аналогию с аэрокосмической промышленностью. Первыми самолетами было легко управлять, но они обладали малой эффективностью, а современные летательные аппараты, такие как космические челноки, невозможно поднять в воздух без сложного программного обеспечения [3].

В данной работе рассматривается робот, приводимый в движение двигателем постоянного тока с редуктором. В этом случае модель двигателя является линейной и хорошо изученной, однако наличие системы передач в редукторе создает дополнительное трение и люфты.

В дальнейшем будем рассматривать независимое совместное управление манипулятором. В этом случае каждое сочленение манипулятора будет представлять собой SISO-систему. Любые эффекты, создаваемые движением других сочленений манипулятора, рассматриваются как внешнее возмущение. Основная структура SISO-системы с обратной связью представлена на рис. 3.

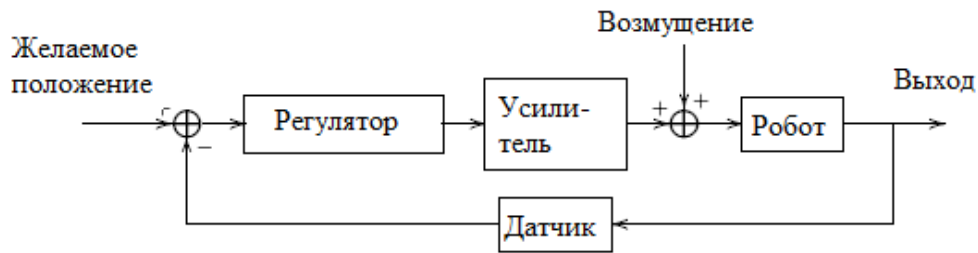


Рис. 3. Структура системы управления с обратной связью

Цель синтеза системы управления заключается в выборе регулятора таким образом, чтобы выходной сигнал совпадал с задаваемым. В качестве управляющего сигнала действует не только входной. Возмущения системы также влияют на поведение выходной переменной. Таким образом, разрабатываемый регулятор должен также компенсировать и отклонения, создаваемые возмущением. Слежение за выходным сигналом и компенсация возмущения являются основой для синтеза регулятора.

Выше было показано, что уравнение движения манипулятора можно записать в виде:

$$d(q(t))\ddot{q}(t) + h(q(t), \dot{q}(t))\dot{q}(t) + c(q(t)) = \tau(t). \quad (4)$$

Важно понимать физический смысл, заключенный в этой формуле. Данное управление представляет динамику связанных между собой идеальных жестких звеньев, на сочленения которых действуют обобщенные силы. Они создаются приводом и могут иметь электрическую, гидравлическую или механическую природу. Хотя рассматриваемое уравнение и является сложным, тем не менее оно является идеализацией и не учитывает некоторых эффектов, например трения между сочленениями. Кроме того, звенья робота не являются абсолютно жесткими.

В данной работе рассмотрим динамику двигателя постоянного тока. Привода данного типа состоят из неподвижного статора и вращающегося ротора, помещенного внутрь статора. Статор создает магнитный поток ϕ . По ротору протекает ток i , в результате чего он начинает вращаться. При этом создается момент:

$$\tau_m = K_1 \phi i_a, \quad (5)$$

где τ_m – момент, создаваемый приводом, ϕ – магнитный поток, i_a – протекающий по обмотке ротора ток, K_1 – константа.

При движении ротора в магнитном поле возникает противо-ЭДС V_b :

$$V_b = K_2 \phi \omega_m, \quad (6)$$

где ω_m – угловая скорость вращения ротора, K_2 – константа.

Схема, иллюстрирующая работу привода, представлена на рис. 4.

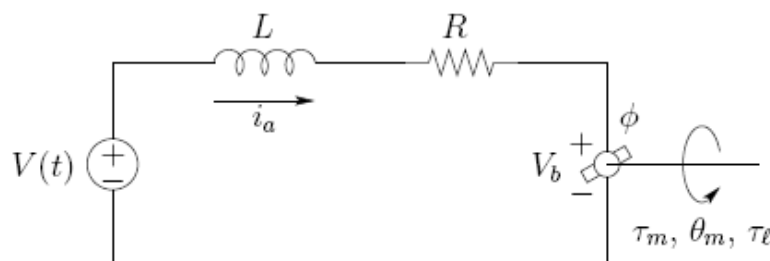


Рис. 4. Электрическая схема двигателя постоянного тока: V – напряжение, подаваемое на двигатель, L, R – индуктивность и сопротивление обмотки ротора

Дифференциальное уравнение, описывающее ток в обмотке ротора, можно представить в виде:

$$L \frac{di}{dt} + Ri = V - V_b. \quad (7)$$

С учетом того, что поток в установившемся режиме постоянный,

$$\tau_m = K_m i, \quad (8)$$

где K_m – постоянная момента.

$$V_b = K_b \frac{d\theta_m}{dt}. \quad (9)$$

Зависимость момента от скорости вращения при постоянном напряжении, подаваемом на двигатель, изображена на рис. 5.

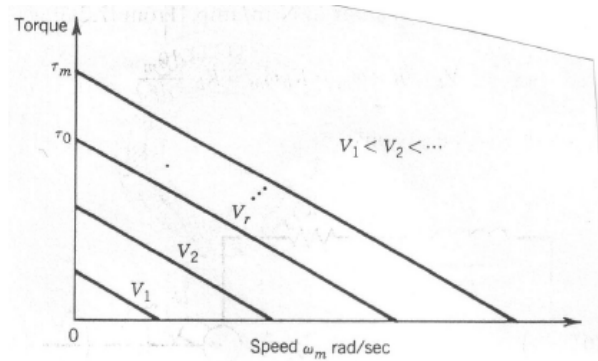


Рис. 5. Механическая характеристика двигателя постоянного тока

Пусть при остановке двигателя на валу создается момент τ_0 . Тогда из уравнения (8) получим:

$$K_m = \frac{R\tau_0}{V_b}, \quad V_b = Ri_a = \frac{R\tau_0}{K_m}. \quad (10)$$

Как правило, для увеличения приложенного к звену момента сил и снижения скорости вращения звена к приводу подключается редуктор (рис. 6).

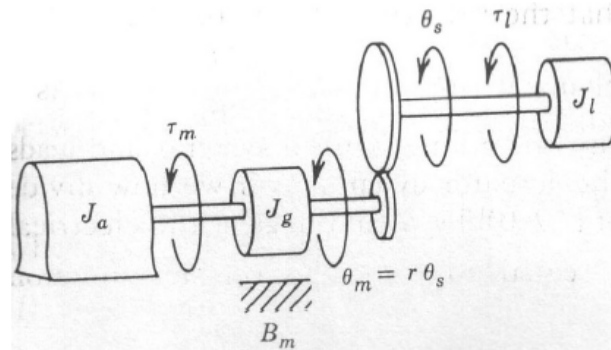


Рис. 6. Модель привода с редуктором

Обозначим через r передаточное число редуктора, $I_m = I_a + I_g$ – сумму моментов инерции привода и редуктора. Тогда получим:

$$J_m \frac{d^2\theta_m}{dt^2} + B_m \frac{d\theta_m}{dt} = \tau_m - \frac{\tau_l}{r} = K_m i_a - \frac{\tau_l}{r}. \quad (11)$$

Применим преобразование Лапласа к уравнениям (8), (10), (11):

$$(Ls + R)I_a(s) = V(s) - K_b s\theta_m(s), \quad (12)$$

$$(J_m s^2 + B_m s)\theta_m(s) = K_i I_a(s) - \tau_l(s)/r. \quad (13)$$

Отсюда найдем передаточные функции двигателя по координате и по силе, пренебрегая значениями малых величин.

$$\frac{\theta_m(s)}{V(s)} = \frac{K_m/R}{s(J_m s + B_m + K_b K_m/R)}, \quad (14)$$

$$\frac{\theta_m(s)}{V(s)} = - \frac{1}{s(J_m s + B_m + K_b K_m/R)}. \quad (15)$$

Объединяя уравнения (14) и (15), получим дифференциальное уравнение второго порядка, описывающее привод постоянного тока.

$$J_m \ddot{\theta}_m(t) + \left(B_m + \frac{K_b K_m}{R} \right) \dot{\theta}_m = \frac{K_m}{R} V(t) - \frac{\tau_l(t)}{r}. \quad (16)$$

Свяжем обобщенные координаты и моменты с координатами и моментами на выходе редуктора.

$$\theta_{mk} = r_k q_k; k = 1, \dots, n, \quad (17)$$

$$\tau_{lk} = \tau_k; k = 1, \dots, n. \quad (18)$$

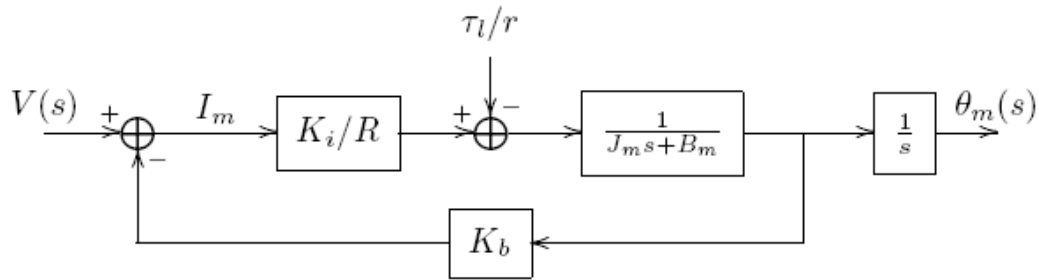


Рис. 7. Схема двигателя постоянного тока

Также был произведен расчет системы управления в задаче слежения с использованием пропорционально-дифференциального (ПД) и пропорционально-дифференциально-интегрального (ПИД) регуляторов. Данный тип управления применим при использовании манипуляторов с не очень большими скоростями движения и большим передаточным числом редуктора, что соответствует рассматриваемой задаче.

В дальнейшем для простоты положим, что:

$$\theta_{mk} = r_k q_k; k = 1, \dots, n, \quad (19)$$

$$\tau_{lk} = \tau_k; k = 1, \dots, n. \quad (20)$$

Тогда уравнение движения каждого звена манипулятора можно представить в виде:

$$\sum_{j=1}^n d_{jk}(q) \ddot{q}_j + \sum_{i,j=1}^n c_{ijk}(q) \dot{q}_i \dot{q}_j + g_k(q) = \tau_k, \quad (21)$$

$$J_{mk} \ddot{\theta}_{mk} + \left(B_{mk} + \frac{K_{bk} K_{mk}}{R_k} \right) \dot{\theta}_{mk} = \frac{K_{mk}}{R_k} V_k - \tau_k / r_k. \quad (22)$$

Подставим уравнение (21) в (22)

$$\left(J_{mk} + \frac{1}{r_k^2} d_{kk}(q) \right) \ddot{\theta}_{mk} + \left(B_{mk} + \frac{K_{bk} K_{mk}}{R_k} \right) \dot{\theta}_{mk} = (K_{mk} / R_k) V_k - d_k. \quad (23)$$

Выполнив замену переменных, получим:

$$d_k = \frac{1}{r_k} \sum_{j \neq k} \ddot{q}_j + \sum_{i,j} c_{ijk} \dot{q}_i \dot{q}_j + g_k. \quad (24)$$

Заметим, что коэффициент при $\ddot{\theta}_{mk}$ является нелинейной функцией, зависящей от конфигурации манипулятора. Тем не менее редуктор с большим передаточным числом снижает влияние этой величины, в результате чего его заменяют средней величиной, приближенной к выражению $J_{mk} + \frac{1}{r_k^2} d_{kk}(q)$. Введем также замену переменных:

$$B_{effk} = B_{mk} + \frac{K_{bk} K_{mk}}{R_k}, \quad u_k = (K_{mk} / R_k) V_k. \quad (25)$$

Преимущество полученной модели заключается в линейности динамики привода. Применим преобразование Лапласа к (23) и построим схему.

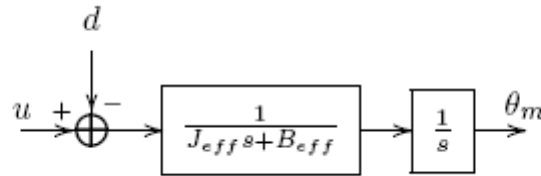


Рис. 8. Схема привода в сочленении

Задача управления сводится к достижению θ_m некоторого заданного значения θ^d с учетом внешних возмущений d . Структурная схема замкнутой системы с ПД-регулятором изображена на рис. 9.

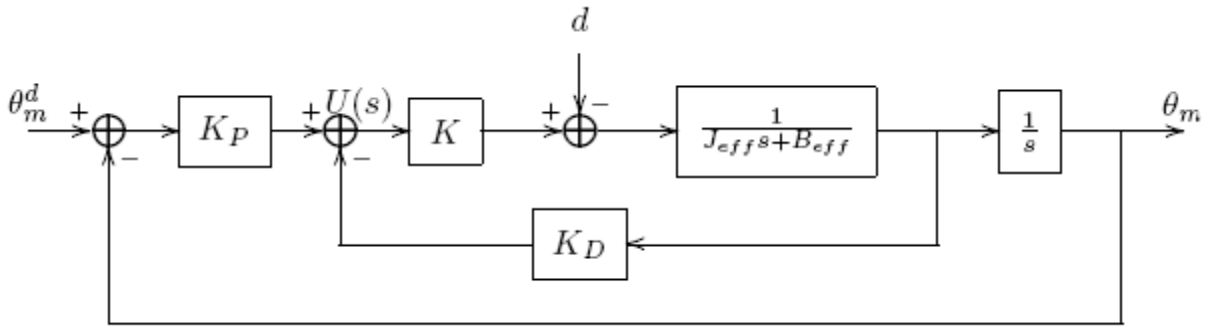


Рис. 9. Замкнутая система с ПД-регулятором.

Управляющее приводом воздействие задается в виде:

$$U(s) = K_p(\theta^d(s) - \theta(s)) - K_D s\theta(s), \quad (26)$$

где K_p и K_D – пропорциональный и дифференциальный коэффициенты усиления соответственно.

Используя преобразование Лапласа от уравнения (25) и уравнение (26) получим:

$$\theta_m(s) = \frac{KK_p}{\Omega(s)}\theta^d(s) - \frac{1}{\Omega(s)}D(s), \quad (27)$$

где $\Omega(s)$ – характеристический полином замкнутой системы.

$$\Omega(s) = J_{eff}s^2 + (B_{eff} + KK_D)s + KK_p. \quad (28)$$

Установившуюся ошибку полученной системы можно рассчитать по формуле:

$$E(s) = \Omega^d(s) - \theta_m(s) = \frac{J_{eff}s^2 + (B_{eff} + KK_D)s}{\Omega(s)}\theta^d(s) + \frac{1}{\Omega(s)}D(s). \quad (29)$$

Откуда для постоянного возмущения и постоянного задающего воздействия получим:

$$e_{ss} = \lim_{s \rightarrow 0} sE(s) = -\frac{D}{KK_p}. \quad (30)$$

Воздействие, вносимое возмущением, обратно пропорционально передаточному числу редуктора и ошибка, создаваемая возмущающим воздействием, может быть уменьшена увеличением пропорционального коэффициента регулятора.

Пусть передаточная функция привода будет задаваться уравнением:

$$W(s) = \frac{2}{3s^2 + 5s}.$$

Для удобства оператора биологического манипулятора системе необходимо обеспечить минимальное перерегулирование, что соответствует биномиальному распределению корней характеристического полинома (28), откуда получим необходимые значения коэффициентов регулятора (Табл. 1)

Таблица 1
Зависимость времени переходного процесса от параметров ПД-регулятора

K_P	K_D	$t_{п}, c$
16	7	1.383
64	15	0.607
144	23	0.387

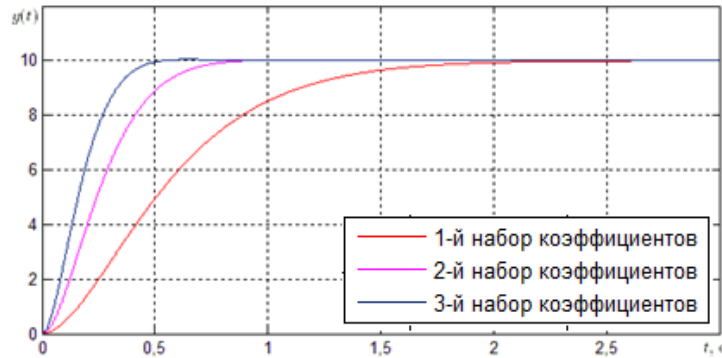


Рис. 10. Переходные процессы системы с ПД-регулятором ($u=10, D=0$)

Характеристики переходных процессов в системе с возмущениями приведены в Табл. 2 и на рис. 11.

Таблица 2
Переходные процессы в системе с возмущениями

K_P	K_D	$t_{п}, c$	e
16	7	1,222	0,310
64	15	0,608	0,080
144	23	0,388	0,035

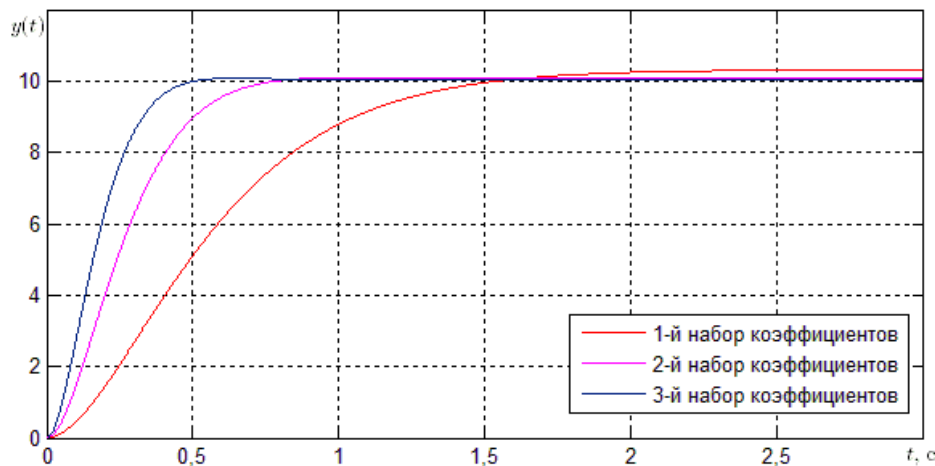


Рис. 11. Переходные процессы в системе с ПД-регулятором и возмущениями ($u=10, D=10$)

Из результатов моделирования можно сделать вывод: чем больше параметры регулятора, тем меньше время переходного процесса и больше энергетические затраты на управление. Однако рост параметров может вызвать в системе колебательные процессы.

Возмущения создают в системе установившуюся ошибку, величина которой обратно пропорциональна K_P .

Из уравнений, полученных выше, следует, что для достижения маленькой установившейся ошибки при постоянном возмущающем воздействии требуются большие

коэффициенты пропорциональности регулятора. С использованием интегрального регулятора можно достигнуть нулевой установившейся ошибки при небольших коэффициентах пропорциональности. Использование ПИД-регулятора позволяет обрабатывать постоянные задающие сигналы с компенсацией возмущений при условии сохранения устойчивости замкнутой системы (рис. 12).

При использовании ПИД-регулятора задающее воздействие принимает вид:

$$C(s) = K_p + K_D s + \frac{K_I}{s}. \quad (31)$$

В этом случае замкнутую систему третьего порядка можно описать уравнением:

$$\theta_m(s) = \frac{(K_D s^2 + K_p s + K_I)}{\Omega_2(s)} \theta^d(s) - \frac{rs}{\Omega_2(s)} D(s), \quad (32)$$

$$\Omega_2 = J_{eff} s^3 + (B_{eff} + K K_D) s^2 + K K_p s + K K_I. \quad (33)$$

В соответствии с алгебраическим критерием устойчивости Гурвица

$$K_I < \frac{(B_{eff} + K K_D) K_p}{J_{eff}}. \quad (34)$$

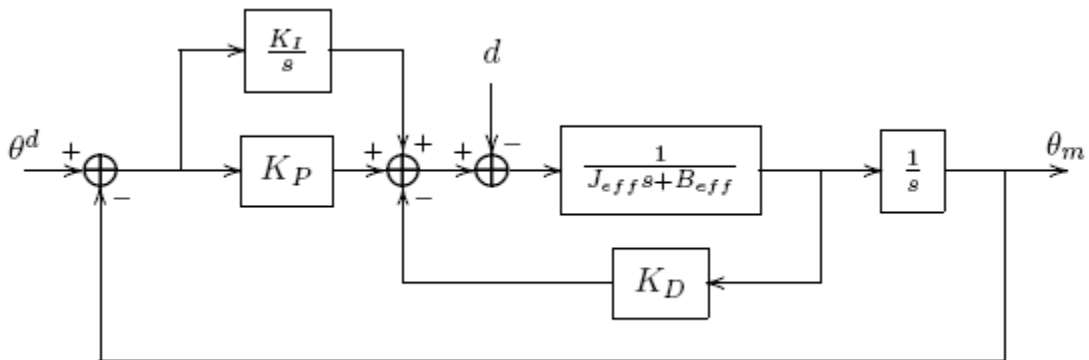


Рис. 12. Замкнутая система с ПИД-регулятором

Для обеспечения нулевого перерегулирования в системе с ПИД-регулятором выберем биномиальное распределение корней. В этом случае $K_p=15,033$, $K_I=1,910$, $K_D=17,818$.

Как видно из результатов моделирования (рис. 13), ПИД-регулятор позволяет достичь нулевой установившейся ошибки при наличии возмущений, но при этом в системе возникает перерегуливание.

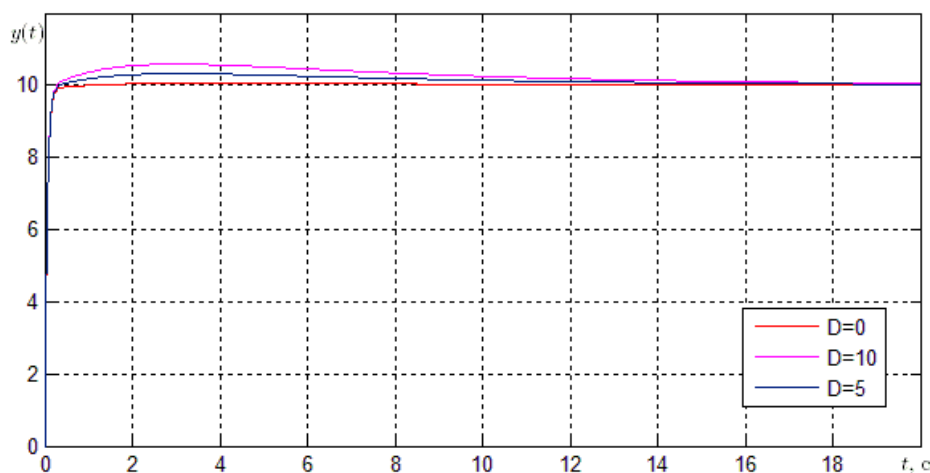


Рис. 13. Переходные процессы в системе с ПИД-регулятором и возмущениями

Заключение

В ходе данной работы был проведен поиск и анализ существующих моделей биологических манипуляторов, представляющих по своей кинематической схеме верхнюю конечность человека. Были решены прямая и обратная задачи кинематики для различных конфигураций манипулятора с использованием представления Денавита-Хартенберга. Проверка конфигурации манипулятора при изменении обобщенных координат подтвердила правильность проведенных расчетов.

Было получено уравнение динамики манипулятора с использованием метода Эйлера-Лагранжа. Применяя его, можно рассчитать динамику любого биологического манипулятора.

Разработан алгоритм независимого управления сочленениями робота и проведено сравнение ПД- и ПИД-регуляторов аналитически и на основании данных, полученных в результате моделирования.

Результаты данной работы могут применяться для расчета системы управления биологическими манипуляторами и использоваться в устройствах управления экзоскелетами верхних конечностей и рук человекоподобных роботов.

Публикации

1. Zimenko K.A., Borgul A.S., Margun A.A., Kremlev A.S., Krasnov A.Y. Intuitive control for robotic rehabilitation devices by human-machine interface with EMG and EEG Signals // 17th international conference on Methods and Models in Automation and Robotics (MMAR – 2012). Miedzyzdroje, Poland, 2013. P. 308–311.
2. Боргуль А.С., Зименко К.А., Маргун А.А., Кремлев А.С. Многофункциональный активный протез руки // Известия высших учебных заведений. Приборостроение. 2013. № 4, С. 86–89.
3. Боргуль А.С., Зименко К.А., Маргун А.А. Биоинформационные технологии в управлении техническими системами // Сб. тезисов докладов XIX Всероссийской научно-методической конференции «Телематика – 2012». СПб., 2012. С. 147.
4. Боргуль А.С., Зименко К.А., Маргун А.А. Биотехническая система управления устройством реабилитации // Сб. тезисов докладов I Всероссийского конгресса молодых ученых. СПб., 2012. С. 197.

Литература

1. Spong M.W., Hutchinson S., Vidyasagar M. Robot Modeling and Control. New York, 2005. – 407 p.
2. Шаньгин Е.С. Управление роботами и робототехническими системами. Уфа, 2005. – 188 с.
3. Юревич Е.И. Управление роботами и робототехническими системами. СПб., 2001. – 171 с.



Мелихов Иван Федорович
Факультет естественнонаучный
Кафедра высшей математики
E-mail:ivan.melikhov@gmail.com

ПРИМЕНЕНИЕ МЕТОДА ХАРТРИ-ФОКА В ЗАДАЧЕ ХРАНЕНИЯ ЧАСТИЦ

И.Ф. Мелихов

Научный руководитель – профессор И.Ю. Попов

Введение

В современной физике важную роль играет вопрос о существовании связанных состояний квантовой системы. Для иллюстрации значимости данного вопроса можно привести проблему хранения квантовых частиц в нанослоистых структурах. Исследования в данной области открывают новые возможности для безопасного хранения водорода в аккумуляторах, составленных из множества нанослоев. Другое применение задача о хранении частиц может найти в реализации некоторых операций в квантовом компьютере, требующих локализации нескольких электронов в одной области пространства.

Целью данной работы является исследование проблемы локализации и хранения квантовых частиц в деформированном волноводе. Известно, что в искривленном волноводе можно задержать частицу. С точки зрения математики наличие подобного связанного состояния можно интерпретировать как существование квадратично интегрируемого решения уравнения Шредингера для данной физической системы. При отсутствии деформации система не имеет подобных решений и дискретный спектр соответствующего гамильтониана пуст. Но возмущение границ волновода ведет к появлению дискретного спектра, причем чем больше величина возмущения, тем больше собственных чисел имеет гамильтониан [1–7]. Во многих приложениях требуется хранить больше одной частицы, и, следовательно, приходится иметь дело с многочастичной проблемой [8, 9].

Задача данной работы – нахождение минимального значения деформации нанослоя, при которой в нем можно локализовать заданное количество частиц. В построенной модели рассматривается система из нескольких электронов с заданными спинами, взаимодействующих посредством дельта-потенциала, помещенных в двумерный деформированный волновод.

Исследуются различные виды деформации волновода: волновод с локальным возмущением на одной из границ, два волновода, связанные друг с другом через «окно», и изогнутый волновод. Для упрощения уравнения Шредингера для системы частиц используется метод Хартри-Фока, далее полученные уравнения решаются с помощью метода конечных элементов.

В ходе численного эксперимента была получена зависимость порогового значения геометрического параметра, сохраняющего при заданной силе дельта-взаимодействия связанное состояние системы, для различного числа частиц и их суммарного спина. Полученные результаты позволяют лучше понять механизм хранения частиц в деформированных нанослоях и возможность оптимальной физической реализации подобной системы.

Результаты данной работы могут найти применение при проектировании безопасных хранилищ водорода и водородных аккумуляторов, основанных на нанослоистых структурах. Подобная технология могла бы решить проблему транспортировки и использования топлива для водородных двигателей.

Теория

Рассмотрим несколько взаимодействующих частиц, помещенных в деформированный нанослой или волновод. Если пренебречь спин-орбитальным взаимодействием, гамильтониан такой системы будет иметь вид:

$$\hat{H} = \sum_k \left(-\frac{\hbar^2}{2m} \nabla_k^2 + U_k \right) + \frac{1}{2} \sum_{\substack{j,k \\ j \neq k}} V_{jk} = \sum_k \hat{H}_k + \frac{1}{2} \sum_{\substack{j,k \\ j \neq k}} \hat{H}_{jk},$$

где суммирование производится по всем частицам, m – масса частицы, \hbar – постоянная Планка, U_k – потенциал внешнего поля, V_{ik} – потенциал взаимодействия частиц.

С ростом числа частиц размерность задачи возрастает очень быстро, и решать ее в таком виде практически невозможно, поэтому уравнение нужно упростить.

В статье [10] было описано применение метода Хартри для решения двухчастичной задачи. Этот метод довольно прост в реализации, но имеет один существенный недостаток. Дело в том, что волновая функция для фермионов, которыми являются, в частности, электроны и протоны, должна быть антисимметричной относительно перестановки координат двух частиц, что не выполняется в приближении Хартри. Этот недостаток был исправлен советским физиком В.А. Фоком. Идея его метода заключается в том, чтобы представить волновую функцию многочастичной системы $\psi(x_1, \dots, x_n)$ в виде детерминанта Слейтера:

$$\psi(x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(x_1) & \psi_2(x_1) & \dots & \psi_n(x_1) \\ \psi_1(x_2) & \psi_2(x_2) & \dots & \psi_n(x_2) \\ \dots & \dots & \dots & \dots \\ \psi_1(x_n) & \psi_2(x_n) & \dots & \psi_n(x_n) \end{vmatrix}.$$

Здесь n – число частиц, ψ_i – одночастичная функция, x_i представляет собой набор из четырех координат: трех пространственных и одной спиновой. Применяя вариационный принцип [11], можно получить уравнения на одночастичные функции $\psi_k(x)$ и соответствующие им уровни энергии E_k , называемые уравнениями Хартри-Фока:

$$\left[\hat{H}_k + \sum_{\substack{j \\ j \neq k}} \int \psi_j^*(x_j) \hat{H}_{jk} \psi_j(x_j) dx_j \right] \psi_k(x_k) - \\ - \sum_{\substack{j \\ j \neq k}} \left[\int \psi_j^*(x_j) \hat{H}_{jk} \psi_k(x_j) dx_j \right] \psi_j(x_k) = E_k \psi_k(x_k) \quad (k = \overline{1, n}).$$

При нахождении дискретного спектра системы необходимо учитывать, что соответствующие уровни энергии $E_{1..n}$ должны быть ниже порогового значения – границы непрерывного спектра.

Численная модель

В работе строится численная модель поведения частиц в волноводе. Предполагается отсутствие внешнего поля и дельта-потенциал взаимодействия частиц. При этих допущениях уравнения Хартри-Фока упрощаются:

$$\left[-\frac{\hbar^2}{2m} \nabla_k^2 + \sum_{\substack{j \\ j \neq k}} U_0 (1 - \delta_{s_j, s_k}) |\psi_j(r_k)|^2 \right] \psi_k(r_k) = E_k \psi_k(r_k) \quad (k = \overline{1, n}),$$

где $\delta_{i,j}$ – дельта-символ Кронекера, s_i, s_j – спины частиц, U_0 – константа, характеризующая взаимодействие между частицами.

В данной работе решались двумерные задачи трех различных геометрических конфигураций:

- волновод с деформированной границей (рис. 1),

- два нанослоя, связанных через «окно» (рис. 2),
- изогнутый волновод (рис. 3).

На границах волновода задаются условия Дирихле, а на концах – Неймана. В первой конфигурации возмущение задается аналитической функцией. Параметр деформации – H/L (ширина деформации b постоянна); параметр второй конфигурации – a/L , последней – L/R (длина искривленной дуги постоянна). Во всех случаях большему значению геометрического параметра соответствует большая величина деформации.

Для нахождения решения уравнений была написана программа, использующая метод конечных элементов (применяется пакет FreeFEM++). Алгоритм решения аналогичен описанному в [10].

На первом шаге все волновые функции полагаются равными нулю. Затем решается первое уравнение, и его результаты подставляются во второе, которое решается относительно второй волновой функции. Процесс продолжается итеративно до тех пор, пока разница между двумя последними аппроксимациями не станет достаточно мала.

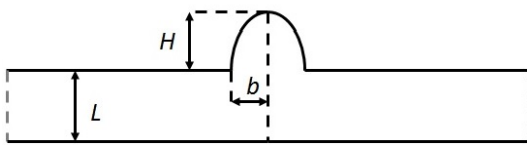


Рис. 1. Деформированный волновод

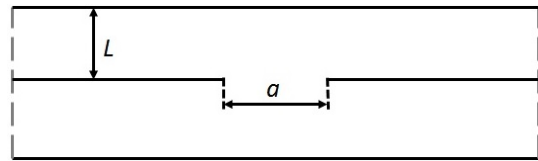


Рис. 2. Связанные волноводы

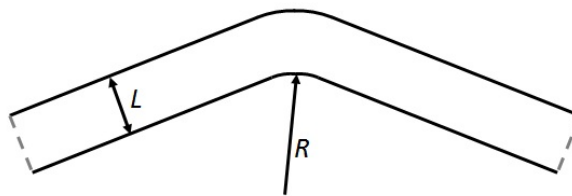


Рис. 3. Изогнутый волновод

Результаты

В ходе работы были проведены вычисления для систем с различным числом частиц и суммарным значением спина. В ходе численного эксперимента фиксировалась сила взаимодействия частиц U_0 и измерялось критическое значение геометрического параметра – минимальное значение, при котором сохраняется связанное состояние выбранной системы. Результаты приведены на рис. 4–6.

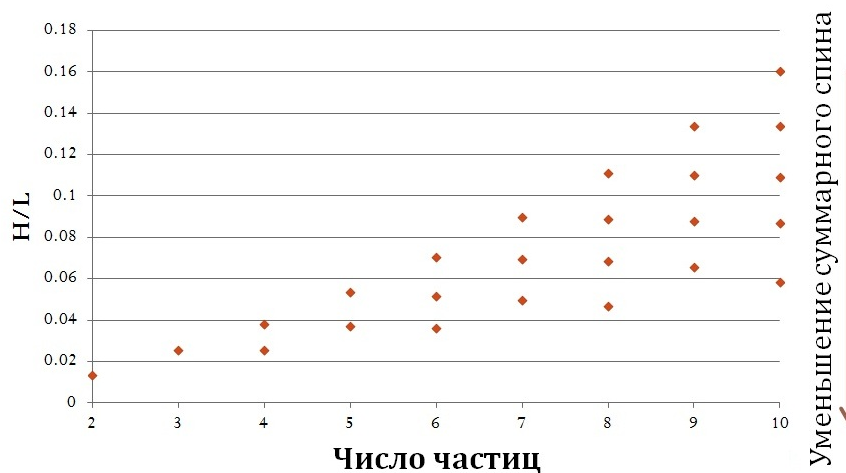


Рис. 4. Деформированный волновод: критическое значение геометрического параметра

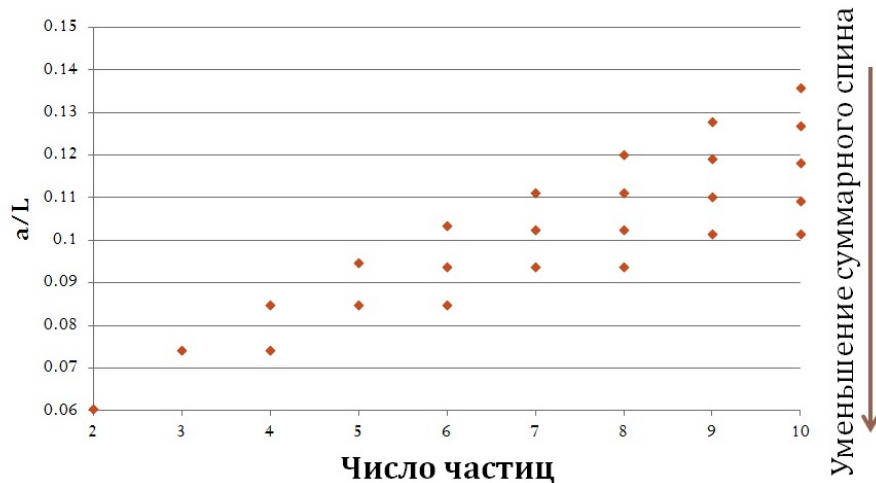


Рис. 5. Связанные волноводы: критическое значение геометрического параметра

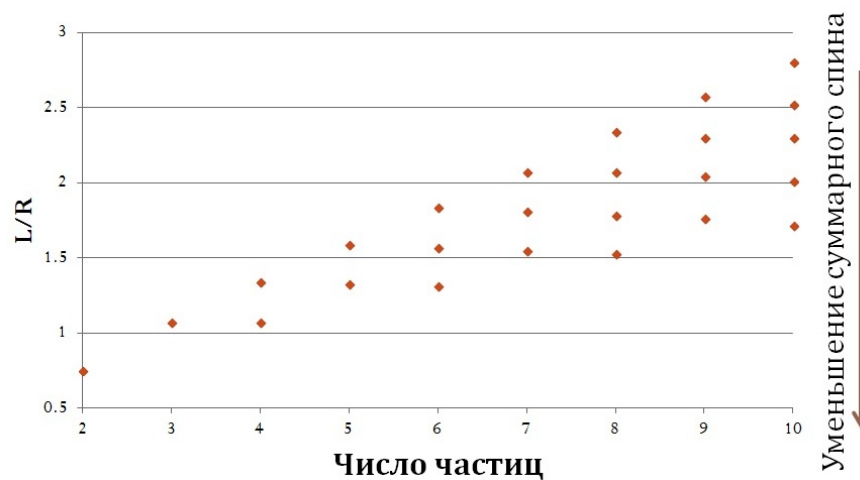


Рис. 6. Изогнутый волновод: критическое значение геометрического параметра

Как и ожидалось, чем больше количество частиц и больше суммарный спин системы, тем значительно требуется деформация, приводящая к связанному состоянию. Можно сказать, что хранить больше частиц труднее. Но кроме этого, на графиках заметен интересный эффект, проявляющийся во всех типах геометрии. Видно, что хранить четное число частиц с нулевым спином более выгодно, чем хранить на одну частицу меньше. Сравнение наклонов графиков на рис. 4–6 показывает, что два связанных волновода – наиболее эффективная конфигурация из рассмотренных. Случай изогнутого волновода, напротив, самый низкоэффективный.

Заключение

В работе строится и изучается модель хранения частиц в нанослоистых структурах и исследуется система взаимодействующих частиц, помещенных в деформированный волновод. Были получены следующие результаты:

- многоэлектронное уравнение Шредингера с помощью метода Хартри-Фока сводится к системе более простых уравнений – уравнений Хартри-Фока;
- для решения полученной системы уравнений была написана программа для пакета расчетов с помощью метода конечных элементов;
- исследованы три различных вида деформации нанослоя и системы с разным количеством электронов и их суммарным спином;
- получены зависимости минимальной деформации нанослоя, сохраняющей связанные состояния системы, от количества частиц и их спина;
- на основе полученных зависимостей сделано заключение о наибольшей эффективности деформации типа «два волновода, связанные через окно» и системы, состоящей из четного количества частиц с нулевым суммарным спином.

В дальнейшем работу можно развивать в нескольких направлениях, например: рассмотреть другие виды взаимодействия частиц кулоновского типа или изучать трехмерные системы. Все эти шаги ведут к описанию реальных систем, и результаты могут быть применены на практике.

Публикации

1. *Melikhov I.F., Popov I.Yu.* Hartree-Fock approximation for the problem of particle storage in deformed nanolayer // Наносистемы: Физика, химия, математика. 2013. Т. 4. № 4. С. 559–563.
2. *Melikhov I.F.* Particle storage in a nanolayer: Hartree-Fock approximation // Days on Diffraction – 2013. International conference (Saint Petersburg, May 27–31, 2013). Abstracts. Saint Petersburg, 2013. P. 63.
3. *Мелихов И.Ф.* Метод Хартри-Фока для нахождения связанных состояний системы взаимодействующих квантовых частиц в деформированном нанослое // Сб. тезисов докладов конгресса молодых ученых. Вып. 2. СПб., 2013. С. 216.

Литература

1. *Duclos P., Exner P., Stovicek P.* Curvature induced resonances in a two-dimensional Dirichlet tube // Annales de l'institut Henri Poincaré. (A) Physique théorique. 1995. Vol. 62. No 1. P. 81–101.
2. *Duclos P., Exner P., Krejcirik D.* Locally curved quantum layers // Ukrainian Journal of Physics. 2000. Vol. 45. P. 595–601.
3. *Exner P., Vugalter S.A.* Bound states in a locally deformed waveguide: The critical case // Letters in Mathematical Physics. 1997. Vol. 39. P. 59–68.
4. *Exner P., Vugalter S.A.* Asymptotic estimates for bound states in quantum waveguides coupled laterally through a narrow window // Annales de l'institut Henri Poincaré. (A) Physique théorique. 1996. Vol. 65. No 1. P. 109–123.
5. *Exner P., Seba P., Tater M., Vanek D.* Bound states and scattering in quantum waveguides coupled laterally through a boundary window // Journal of Mathematical Physics. 1996. Vol. 37. P. 4867–4887.
6. *Popov I.Yu.* Asymptotics of bound state for laterally coupled waveguides // Reports on Mathematical Physics. 1999. Vol. 43. No 3. P. 427–437.
7. *Popov I.Yu.* Asymptotics of resonances and bound states for laterally coupled curved quantum waveguides // Physics Letters A. 2000. Vol. 269. P. 148–153.
8. *Linde H.* Geometrically induced two-particle binding in a waveguide // Journal of Physics A: Mathematical and General. 2006. Vol. 39. No.18. P. 5105.
9. *Exner P., Vugalter S.A.* On the number of particles that a curved quantum waveguide can bind // Journal of Mathematical Physics. 1999. Vol. 40. P. 4630–4638.
10. *Popov S.I., Gavrilov M.I., Popov I.Yu.* Two interacting particles in deformed nanolayer: Discrete spectrum and particle storage // Physica Scripta. 2012. Vol. 86. No 3. P. 17–23.
11. *Messiah A.M.L.* Quantum mechanics. Vol. 2. Russian translation. Moscow, 1979. – 583 p.



Мелихова (Аникевич) Алина Семеновна
Факультет естественнонаучный
Кафедра высшей математики
E-mail:alina.s.melikhova@gmail.com

СПЕКТРАЛЬНАЯ ЗАДАЧА ДЛЯ Y-РАЗВЕТВЛЕННОЙ ЦЕПОЧКИ СЛАБО СВЯЗАННЫХ ШАРООБРАЗНЫХ РЕЗОНАТОРОВ

А.С. Мелихова (Аникевич)

Научный руководитель – профессор И.Ю. Попов

Введение

В последнее время одним из приоритетных научных направлений является изучение наноструктур. Применение подобных технологий находит свое отражение в молекулярной биологии, электротехнике и других областях. Уже сейчас можно сказать, что в ближайшем будущем нанотехнологии повлияют практически на все сферы жизни человека: от производства обычных тканей до создания ракетного топлива.

Квантовая механика позволяет описывать физические явления на уровне атомов, молекул, электронов и других элементарных частиц. Сейчас большое внимание уделяется изучению квантовых [1–3] и декорированных графов [4]. Систему, рассмотренную в данной работе, можно отнести именно к этому типу структур.

Статья посвящена решению спектральной задачи для модели Y-разветвленной цепочки слабо связанных шарообразных резонаторов и является естественным продолжением работы «Спектральная задача для цепочек слабо связанных шарообразных резонаторов» [5]. Основная цель – изучение отрицательных собственных значений для системы с Y-разветвлением. В работе рассмотрено два типа соединений резонаторов: δ - и δ' -соединения. Решение задачи проведено на основе теории расширений [6], а также при использовании метода матриц монодромии [7].

Итак, Y-разветвленная система цепочек шарообразных резонаторов может быть описана как шар (называемый центральным) единичного радиуса и три полубесконечные цепочки, составленные из идентичных центральному шаров.

При решении данной задачи считается, что система находится вне действия внешних полей и используется система единиц, в которой $\hbar=2m=1$, где m – масса частицы, «живущей» внутри шаров, \hbar – редуцированная постоянная Планка. Для демонстрации метода решения, не умаляя общности, рассмотрена упрощенная конфигурация Y-разветвления:

- предполагаем, что у каждой ветви рассматриваемой системы есть ось, проходящая через центры шаров, формирующих эту ветвь, и что все три оси лежат в одной плоскости;
- считаем, что шары пронумерованы в соответствии с рис. 1;
- центры шаров под номерами (1,1); (2,1) и (3,1) являются вершинами равностороннего треугольника.

Система при условии наличия указанных выше предположений далее называется «базовым Y-разветвлением цепочки».

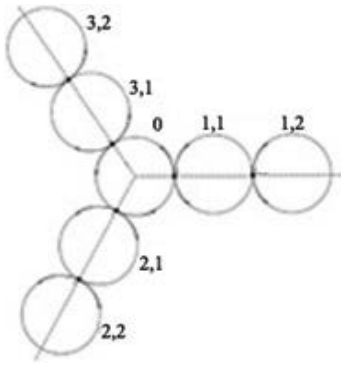


Рис. 1. Базовое Y-разветвление цепочки

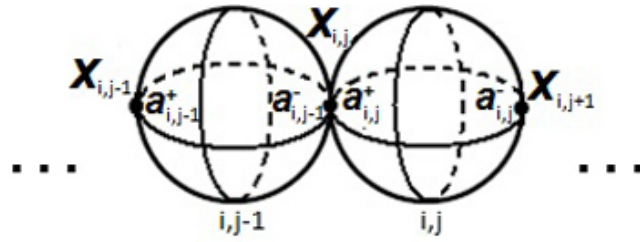


Рис. 2. Схема прямого участка цепи: определение коэффициентов

Резонаторы в рассматриваемой системе являются слабо связанными, то есть предполагается, что все взаимодействия между элементарными ячейками цепи осуществляются через точечный прокол [8–10] в точке соединения резонаторов.

Было исследовано стационарное состояние бесспиновой нерелятивистской частицы, волновая функция $\psi(x)$ которого удовлетворяет стационарному уравнению Шредингера:

$$\mathcal{H}(x)\psi(x) = \lambda\psi(x),$$

где $\mathcal{H}(x) = -\Delta$ – Гамильтониан системы, действующий, как отрицательный Лапласиан, а λ – энергия частицы при движении, описываемом волновой функцией $\psi(x)$.

Предполагалось также, что на границе шара выполнено условие Неймана: $\left. \frac{\partial \psi}{\partial n} \right|_{\partial \Omega} = 0$.

Для решения данной задачи была привлечена теория расширений: мы воспользовались процедурой «сужение-расширение», то есть сначала построили симметрическое сужение исходного самосопряженного оператора, затем рассмотрели его самосопряженное расширение, отличающееся от исходного. Итак, сузим оператор $-\Delta$ на множество всех функций из $D(-\Delta)$, которые обращаются в нуль в точках соединения $\mathbf{x}_{i,j}$. Таким образом, получили симметрический оператор $-\Delta_0$, дефектные элементы которого есть функции Грина $G(\mathbf{x}, \mathbf{x}_j, \lambda)$ [11]:

$$G(r, \theta, \varphi, r_j, \theta_j, \varphi_j, \lambda) = \sum_{k=1}^{\infty} \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{\Psi_{klm}(r, \theta, \varphi) \overline{\Psi_{klm}(r_j, \theta_j, \varphi_j)}}{\lambda_{lk} - \lambda},$$

где $\lambda_{lk} = x_{lk}^2$ – собственные числа, $\Psi_{klm}(r, \theta, \varphi) = N_{lk} j_l(x_{lk} r) Y_m^l(\theta, \varphi)$ – собственные векторы задачи Неймана для оператора Лапласа в шаре, $Y_m^l(\theta, \varphi)$ – сферические гармоники, $j_l(x)$ – сферические функции Бесселя, x_{lk} – k -ый корень уравнения $j_l'(x) = 0$, N_{lk} – нормировочный коэффициент.

Как уже упоминалось ранее, оператор $-\Delta_0$ является симметрическим, но не самосопряженным. Учитывая то, что все самосопряженные расширения оператора $-\Delta_0$ являются сужениями сопряженного оператора, был построен оператор $-\Delta_0^*$.

Любая функция U такая, что $U \in D(-\Delta_0^*)$, имеет вид: $U = U_0 + \frac{a}{|\mathbf{x} - \mathbf{x}_0|} + b$,

где $U_0 \in D(-\Delta_0^*)$, $U_0(\mathbf{x}_0) = 0$; a , b – некоторые коэффициенты. Здесь и далее были использованы следующие обозначения: $a_{i,j}^+$, $a_{i,j}^-$ – коэффициенты, относящиеся к точке соединения резонаторов под номером (i, j) в соответствии с рис. 2 (коэффициенты $b_{i,j}^+$, $b_{i,j}^-$ определяются по аналогичным соображениям).

При данных обозначениях самосопряженное расширение может быть описано следующим соотношением между коэффициентами:

$$\begin{cases} a_{i,j}^+ = -a_{i,j-1}^-, \\ b_{i,j}^+ = b_{i,j-1}^-. \end{cases} \quad (1)$$

Предположим, что в точках соединения резонаторов присутствует точечный потенциал, и рассмотрим два типа соединения: δ - и δ' -соединение.

Случай δ -соединения резонаторов

Наличие в точках соединения резонаторов так называемого δ -потенциала с параметром α накладывает следующие требования на коэффициенты $a_{i,j}^\pm, b_{i,j}^\pm$:

$$\begin{cases} a_{i,j}^+ = -a_{i,j-1}^-, \\ b_{i,j}^+ - b_{i,j-1}^- = \alpha a_{i,j-1}^-, \end{cases} \quad (2)$$

где коэффициенты $b_{i,j}^+, b_{i,j-1}^-$ имеют следующий вид:

$$b_{i,j}^+ = a_{i,j}^+ \lim_{\mathbf{x} \rightarrow \mathbf{x}_{i,j}} (G(\mathbf{x}, \mathbf{x}_{i,j}, \lambda) - G(\mathbf{x}, \mathbf{x}_{i,j}, \lambda_0)) + a_{i,j}^- G(\mathbf{x}_{i,j+1}, \mathbf{x}_{i,j}, \lambda), \quad (3)$$

$$b_{i,j-1}^- = a_{i,j-1}^- \lim_{\mathbf{x} \rightarrow \mathbf{x}_{i,j}} (G(\mathbf{x}, \mathbf{x}_{i,j}, \lambda) - G(\mathbf{x}, \mathbf{x}_{i,j}, \lambda_0)) + a_{i,j-1}^+ G(\mathbf{x}_{i,j-1}, \mathbf{x}_{i,j}, \lambda), \quad (4)$$

$$b_{i,0}^- = a_{i,0}^- \lim_{\mathbf{x} \rightarrow \mathbf{x}_{i,1}} (G(\mathbf{x}, \mathbf{x}_{i,1}, \lambda) - G(\mathbf{x}, \mathbf{x}_{i,1}, \lambda_0)) + \sum_{n \neq i} a_{in,0}^+ G(\mathbf{x}_{n,1}, \mathbf{x}_{i,1}, \lambda), \quad i, n = 1, 2, 3. \quad (5)$$

Сопоставление уравнений (2) и (3, 4) для разных точек $\mathbf{x}_{i,j}$ позволило получить уравнение на дискретный спектр.

Связь между коэффициентами $a_{i,j}^\pm$ и $a_{i,j-1}^\pm \quad \forall j \geq 2, i = 1, 2, 3$ была получена путем подстановки выражений (3, 4) и первого уравнения системы (2) во второе уравнение системы:

$$\begin{pmatrix} a_{i,j}^+ \\ a_{i,j}^- \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ \frac{G(\mathbf{x}_{i,j-1}, \mathbf{x}_{i,j}, \lambda)}{G(\mathbf{x}_{i,j+1}, \mathbf{x}_{i,j}, \lambda)} & \frac{\alpha + 2(G(\mathbf{x}, \mathbf{x}_{i,j}, \lambda) - G(\mathbf{x}, \mathbf{x}_{i,j}, \lambda_0))|_{\mathbf{x}=\mathbf{x}_{i,j}}}{G(\mathbf{x}_{i,j+1}, \mathbf{x}_{i,j}, \lambda)} \end{pmatrix} \begin{pmatrix} a_{i,j-1}^+ \\ a_{i,j-1}^- \end{pmatrix} = \mathbf{M}_j \begin{pmatrix} a_{i,j-1}^+ \\ a_{i,j-1}^- \end{pmatrix},$$

где \mathbf{M}_j – матрица монодромии. Нетрудно найти ее собственные числа μ_j^\pm и соответствующие им собственные векторы ν_j^\pm . Положив первую компоненту вектора единицей, был получен следующий вид для собственного вектора:

$$\nu_j^\pm = \begin{pmatrix} 1 \\ -\mu_j^\pm \end{pmatrix}.$$

Связь между коэффициентами $a_{i,1}^\pm$ и $a_{i,0}^\pm$, где $i = 1, 2, 3$, была получена с помощью подстановки выражений (3, 5) в систему (2).

Предположив теперь равенство (с точностью до мультипликативной константы) одного из векторов (для определенности, вектора $(a_{1,1}^+, a_{1,1}^-)^T$) собственному вектору ν , а также воспользовавшись линейной зависимостью, получаем следующую систему:

$$\begin{cases} (\mu_2^\pm - X)c + Aa_{2,0}^- + Ba_{3,0}^- = 0 \\ Ac + (\mu_2^\pm - X)a_{2,0}^- - Ca_{3,0}^- = 0, \\ Bc - Ca_{2,0}^- (\mu_2^\pm - X)a_{3,0}^- = 0 \end{cases} \quad (6)$$

где c – константа и где для краткости записи были использованы следующие обозначения:

$$X = \alpha + 2 \lim_{x \rightarrow x_{i,1}} \left(G(\mathbf{x}, \mathbf{x}_{i,1}, \lambda) - G(\mathbf{x}, \mathbf{x}_{i,1}, \lambda_0) \right), \quad A = G(\mathbf{x}_{2,1}, \mathbf{x}_{1,1}, \lambda), \quad B = G(\mathbf{x}_{3,1}, \mathbf{x}_{1,1}, \lambda), \\ C = G(\mathbf{x}_{3,1}, \mathbf{x}_{2,1}, \lambda).$$

Очевидно, что система (6) имеет нетривиальное решение относительно c , $a_{2,0}^-$ и $a_{3,0}^-$ тогда, и только тогда, когда ее детерминант равен нулю:

$$(X - \mu_2^\pm G)^3 - (X - \mu_2^\pm G)(A^2 + B^2 + C^2) + 2ABC = 0. \quad (7)$$

Уравнение (7) и являлось основным предметом дальнейших исследований. В связи с тем, что мы пока ограничились только базовым Y -разветвлением цепочки (а значит, $A=B=C$), можно переписать это уравнение в более простой форме:

$$(X - \mu_2^\pm G)^3 - 3A^2(X - \mu_2^\pm G) + 2A^3 = 0. \quad (8)$$

Следует отметить, что уравнения (7, 8) являются уравнениями на дискретный спектр только в том случае, если соответствующее собственное число μ удовлетворяет условию:

$$|\mu_2^-| < 1 \Leftrightarrow \frac{X}{2G} > 1, \quad (9)$$

$$|\mu_2^+| < 1 \Leftrightarrow \frac{X}{2G} < -1. \quad (10)$$

Исследовав решения уравнения (8) при выполнении условий (9, 10), можно сформулировать некоторую теорему относительно количества этих решений, которая будет рассмотрена далее.

Случай δ' -соединения резонаторов

В данном разделе описывается решение рассмотренной выше задачи, но при наличии в точках соединения резонаторов δ' -соединения. В этом случае предполагается выполнение следующего (вместо (1)) условия на коэффициенты:

$$\begin{cases} a_{i,j}^+ + a_{i,j-1}^- = \alpha b, \\ b_{i,j}^+ = b_{i,j-1}^- = b. \end{cases} \quad (11)$$

Аналогично был получен явный вид матрицы монодромии M' для этого случая:

$$M' = \begin{pmatrix} \alpha G & -X' \\ X' & \frac{1-X'^2}{\alpha G} \end{pmatrix},$$

где $X' = 1 - \alpha \lim_{x \rightarrow x_{i,1}} \left(G(\mathbf{x}, \mathbf{x}_{i,1}, \lambda) - G(\mathbf{x}, \mathbf{x}_{i,1}, \lambda_0) \right) = 1 - \alpha g$, а также ее собственные числа μ^{\pm}

и собственные векторы ν^{\pm} . Условия на модуль собственного числа примет следующий вид:

$$|\mu^{+}| < 1 \Leftrightarrow \alpha G + \frac{1-X'^2}{\alpha G} < -2, \quad (12)$$

$$|\mu^{-}| < 1 \Leftrightarrow \alpha G + \frac{1-X'^2}{\alpha G} > 2. \quad (13)$$

Подставив в условие δ' -соединения (11) выражения для $b_{i,j}^\pm$, получаем связь между коэффициентами $a_{i,1}^\pm$ и $a_{i,0}^\pm$, а затем, используя линейную зависимость, и основное уравнение на дискретный спектр для данного типа соединений:

$$\left(g - \frac{G - \mu'G}{\alpha(G - \mu'g) + \mu'}\right)^3 - \left(g - \frac{G - \mu'G}{\alpha(G - \mu'g) + \mu'}\right)(A^2 + B^2 + C^2) + 2ABC = 0 \quad (14)$$

С учетом приведенных ранее соображений представим уравнение (14) в более простой форме:

$$\left(g - \frac{G - \mu'G}{\alpha(G - \mu'g) + \mu'}\right)^3 - 3A^2 \left(g - \frac{G - \mu'G}{\alpha(G - \mu'g) + \mu'}\right) + 2A^3 = 0, \quad (15)$$

а выводы о количестве его решений, удовлетворяющих некоторым требованиям, представлены ниже.

Основные результаты

Итак, в данном разделе приведены основные результаты, полученные для каждого типа соединения, и методы, использованные для их доказательства.

В случае наличия условия δ -соединения резонаторов можно сформулировать следующую теорему.

Теорема. Система с базовым Y-разветвлением цепочек слабосвязанных шарообразных резонаторов имеет не более одного отрицательного собственного значения в случае наличия условий δ -соединения с константой связи α в точках соединения резонаторов (где

$$\alpha < -\frac{9}{2} \sum_{k=1, l=0}^{\infty} \frac{N_{lk}}{x_{lk}^2} - \frac{3}{2\pi\lambda_0}).$$

Доказательство теоремы проведено с помощью оценки сверху рядов, фигурирующих в записи функций X, G, A слева от нуля. При этом уравнение (8) становится однородным относительно $\frac{X}{G}$, и можно найти три его решения. Однако условиям на модуль собственного числа удовлетворяет только одно из них. Следовательно, данная система действительно имеет не более одного отрицательного собственного значения при определенных значениях параметра α .

Для случая δ' -соединения резонаторов можно сформулировать аналогичное предыдущей теореме, но более слабое утверждение.

Утверждение. Система с базовым Y-разветвлением цепочек слабосвязанных шарообразных резонаторов имеет не более восьми отрицательных собственных значений в случае наличия условий δ' -соединения с константой связи α в точках соединения резонаторов.

Доказательство утверждения также было проведено на основании того факта, что

уравнение (15) является однородным относительно $\frac{g - \frac{G - \mu'G}{\alpha(G - \mu'g) + \mu'}}{A}$, и его решения были

найжены. Далее, зафиксировав значение свободного параметра λ_0 , получаем два различных уравнения: одно из них – уравнение четвертой степени, а второе – квадратное, но в двух экземплярах за счет кратности корня уравнения (15). Следовательно, уравнение имеет восемь решений. Использование некоторых оценок сверху дает возможность утверждать, что и система имеет не более восьми решений. А значит, при ее численном моделировании, имея полученные выше решения уравнений, остается только удовлетворить соответствующим условиям (12, 13). В работе показано, что они могут быть переписаны в следующем виде:

$$\left[\begin{array}{l} \left\{ \begin{array}{l} F_4(\lambda, \alpha, R) \geq 0 \\ F_2(\lambda, \alpha, R) \geq 0 \end{array} \right. \\ \left\{ \begin{array}{l} F_4(\lambda, \alpha, R) \leq 0 \\ F_2(\lambda, \alpha, R) \leq 0 \end{array} \right. \\ \left\{ \begin{array}{l} F_4(\lambda, \alpha, R) \leq 0 \\ F_2(\lambda, \alpha, R) \geq 0 \end{array} \right. \\ \left\{ \begin{array}{l} F_4(\lambda, \alpha, R) \geq 0 \\ F_2(\lambda, \alpha, R) \leq 0 \end{array} \right. \end{array} \right. ,$$

где $F_4(\lambda, \alpha, R)$ и $F_2(\lambda, \alpha, R)$ – функции четвертой и второй степеней соответственно относительно λ , а $R = \sum_{k=1, l=0}^N \frac{N_{lk}}{x_{lk}^2}$ при фиксированном N . Данная система накладывает дополнительные ограничения на значения параметра α , что в дальнейшем позволит усилить сформулированное утверждение, уменьшив количество отрицательных собственных значений.

Заключение

В работе описано решение спектральной задачи на основе теории расширений для Y-разветвленной цепочки шарообразных резонаторов. Рассмотрено два типа соединения: δ - и δ' -соединение. Для обоих случаев получены основное уравнение и условия для нахождения дискретного спектра.

Доказано, что в случае δ -соединения система имеет не более одного отрицательного собственного значения при определенных значениях параметра α .

В случае δ' -соединения доказано слабое утверждение о количестве связанных состояний. Оно усилено за счет наложения дополнительных требований на параметр α .

Следует отметить, что полученные уравнения (7) и (14) описывают зависимость спектральных характеристик от угловых параметров системы, что позволяет описывать все более и более сложные наноструктуры. Таким образом, данная работа может быть использована при расчете наносистем.

Публикации

1. Аникевич А.С. Спектральная задача для цепочек слабо связанных шарообразных резонаторов // Аннотированный сб. научно-исследовательских выпускных квалификационных работ бакалавров УНИВЕРСИТЕТ ИТМО. СПб., 2012. С. 13–15.
2. Аникевич А.С. Спектральная задача для цепочек слабо связанных резонаторов // Сб. тезисов докладов I Всероссийского конгресса молодых ученых. Вып. 2. СПб., 2012. С. 249.
3. Аникевич А.С. Спектральная задача для цепочек слабо связанных шарообразных резонаторов // Наносистемы: Физика, химия, математика. 2012. Т. 3. № 3. С. 23–30.
4. Аникевич А.С. Спектральная задача для цепочек слабо связанных шарообразных резонаторов // 17-я Санкт-Петербургская ассамблея молодых ученых и специалистов: Сб. тезисов. СПб., 2012. С. 36.
5. Anikevich A. Negative eigenvalues of Laplacian for the Y-bent chain of weakly coupled conglobated resonators // Days on Diffraction – 2013. International conference (Saint Petersburg, May 27–31, 2013). Abstracts. Saint Petersburg, 2013. P. 17.
6. Anikevich A. Negative eigenvalues of the y-type chain of weakly coupled ball resonators // Nanosystems: Physics, chemistry, mathematics. 2013. Vol. 4. No 4. P. 545–549.
7. Аникевич А.С. Отрицательные собственные значения оператора Лапласа для цепочки слабо связанных шарообразных резонаторов с у-разветвлением // Сб. тезисов докладов конгресса молодых ученых. Вып. 2. Математическое моделирование и математическая физика. СПб., 2013. С. 209.

Литература

1. Duclos P., Exner P., Turek O. On the spectrum of a bent chain graph // *Journal of Physics A: Mathematical and Theoretical*. 2008. Vol. 41. No 41. P. 415206–415224.
2. Gerasimenko N.I., Pavlov B.S. Scattering problems on noncompact graphs // *Theoretical and Mathematical Physics*. 1988. Vol. 74. No 3. P. 230–240.
3. Popov I.Yu., Smirnov P.I. Spectral problem for branching chain quantum graph // *Physics Letters A*. 2013. Vol. 377. No 6. P. 439–442.
4. Korotyaev E., Lobanov I. Schrödinger operators on zigzag graphs // *Annales de l'institut Henri Poincaré*. 2007. Vol. 8. No 6. P. 1151–1176.
5. Аникевич А.С. Спектральная задача для цепочек слабо связанных шарообразных резонаторов // *Наносистемы: Физика, химия, математика*. 2012. Т. 3. № 3. С. 23–30.
6. Павлов Б.С. Теория расширений и явнорешаемые модели // *Успехи математических наук*. 1987. Т. 42. Вып. 6. С. 99–131.
7. Bagraev N., Martin G., Pavlov B.S., Yafyasov A. Landau-Zener effect for a quasi-2D periodic sandwich // *Nanosystems: Physics, chemistry, mathematics*. 2011. Vol. 2. No 4. P. 32–50.
8. Brown R., Hislop P.D., Martinez A. Eigenvalues and resonances for domains with tubes: Neumann boundary conditions // *Journal of Differential Equations*. 1995. Vol. 115. No 2. P. 458–476.
9. Popov I.Yu. The resonator with narrow slit and the model based on the operator extensions theory // *Journal of Mathematical Physics*. 1992. Vol. 33. No 11. P. 3794–3801.
10. Brüning J., Martin G., Pavlov B. Calculation of the Kirchhoff coefficients for the Helmholtz resonator // *Russian Journal of Mathematical Physics*. 2009. Vol. 16. No 2. P. 188–207.
11. Ахиезер Н.И., Глазман И.М. Теория линейных операторов в гильбертовом пространстве. М., 1966. – 543 с.



Мельник Максим Владимирович
Факультет фотоники и оптоинформатики
Кафедра фотоники и оптоинформатики
E-mail: Maxim.Melnick@gmail.com

РАСЧЕТ ВРЕМЕНИ КОГЕРЕНТНОСТИ ФЕМТОСЕКУНДНОГО СПЕКТРАЛЬНОГО СУПЕРКОНТИНУУМА В КВАРЦЕВОМ ОПТИЧЕСКОМ ВОЛОКНЕ И ВЫЯВЛЕНИЕ ОПТИМАЛЬНЫХ УСЛОВИЙ ЕГО ГЕНЕРАЦИИ

М.В. Мельник

Научный руководитель – инженер-исследователь А.Н. Цыпкин

Введение

Развитие лазерной техники привело к созданию в последнее десятилетие лазерных систем, генерирующих импульсы длительностью всего в несколько периодов колебаний светового поля. Характерной особенностью этих систем является столь сильная фазовая модуляция излучения в нелинейной среде, что ширина его спектра $\Delta\omega$ становится соизмеримой с центральной частотой:

$$\langle \omega \rangle = \int_0^{\infty} \omega |G|^2 d\omega / \int_0^{\infty} |G|^2 d\omega,$$

где G – спектральная плотность импульса. Такое суживание спектра интенсивных световых импульсов в настоящее время принято называть явлением генерации спектрального суперконтинуума.

Генерация суперконтинуума в волокнах хорошо изученный процесс. Она может включать различные нелинейные эффекты, такие как само- и кросс-модуляции, солитонные эффекты, комбинационное рассеяние, модуляционная неустойчивость и четырехволновые смешения. Большинство спектральных суперконтинуумов генерируются в области аномальной дисперсии групповой скорости, в данном случае в процессе уширения спектра доминируют солитонные эффекты [1–4]. Они приводят к тому, что спектр становится чувствительным к импульсу накачки и изменяется от импульса к импульсу, что способствует низкой согласованности между соседствующими импульсами, а следовательно, и низкой когерентности. Избавиться от этого можно двумя способами: либо тщательно выбирать входные параметры, либо генерировать спектральный суперконтинуум в области с нормальной дисперсией групповых скоростей, в котором солитоны формироваться не могут, а генерация спектрального суперконтинуума происходит в основном из-за фазовой самомодуляции.

Моделирование распространения высококогерентного фемтосекундного спектрального суперконтинуума в оптическом волокне в области нормальной, нулевой и аномальной групповой дисперсий

С целью изучения распространения предельно коротких импульсов (ПКИ) и генерации суперконтинуума, используя уравнение [5, 6]

$$\frac{\partial \vec{E}}{\partial z} - a \frac{\partial^3 \vec{E}}{\partial \tau^3} + b \int_{-\infty}^{\tau} \vec{E} d\tau' + g \vec{E}^2 \frac{\partial \vec{E}}{\partial \tau} = 0, \quad (1)$$

где E – электрическое поле светового импульса; z – направление распространения импульса;

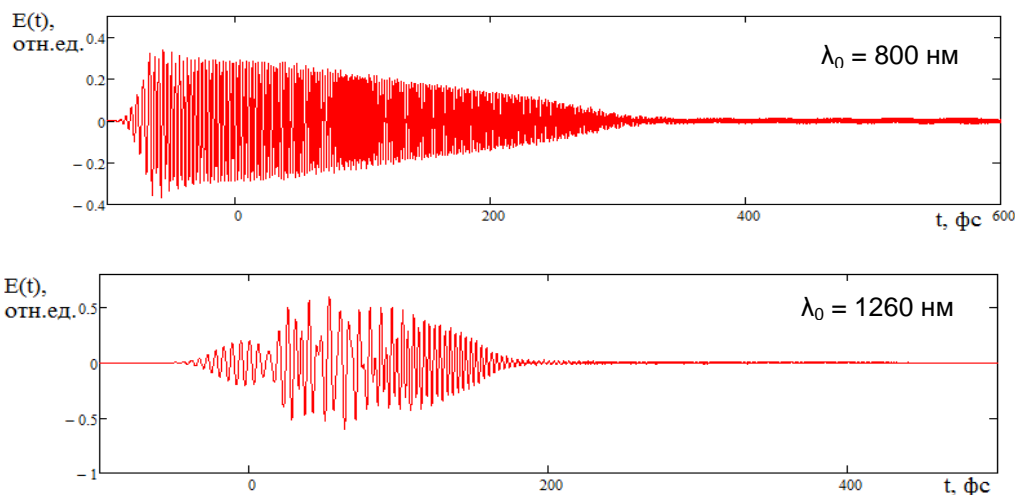
τ – вспомогательная переменная; $g = \frac{\gamma_6}{\omega_{e1}^2 \omega_{e2}^2}$ – константа при нелинейном члене волнового

уравнения; γ – характеризует величину отклика на вынуждающее воздействие; ω_{e1} , ω_{e2} – феноменологические параметры среды, характеризующие дисперсию линейного и нелинейного поляризационного отклика электронной и электронно-колебательной природы;

$\tau = t - \frac{N_0}{c} z$ – время в движущейся с импульсом системе координат; параметры N_0 , a и b – эмпирические константы дисперсии среды и описывают линейный показатель преломления среды и его дисперсию:

$$n(\omega) = N_0 + a c \omega^2 - c \frac{b}{\omega^2},$$

мною были численно смоделированы ПКИ со спектром в различных областях групповой дисперсии. Результаты моделирования приведены на рис. 1, 2.



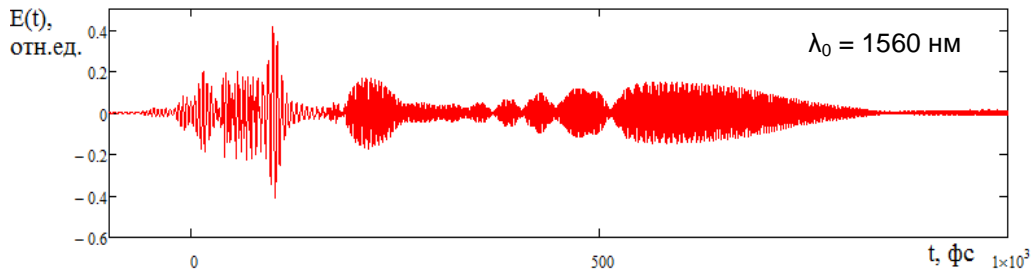


Рис. 1. Электрическое поле фемтосекундного импульса с пиковой интенсивностью $I = 10^{13} \text{ Вт/см}^2$, длительностью 20 фс и различными длинами волн: $\lambda_0 = 800 \text{ нм}$ (нормальная дисперсия); $\lambda_0 = 1260 \text{ нм}$ (нулевая дисперсия); $\lambda_0 = 1560 \text{ нм}$ (аномальная дисперсия)

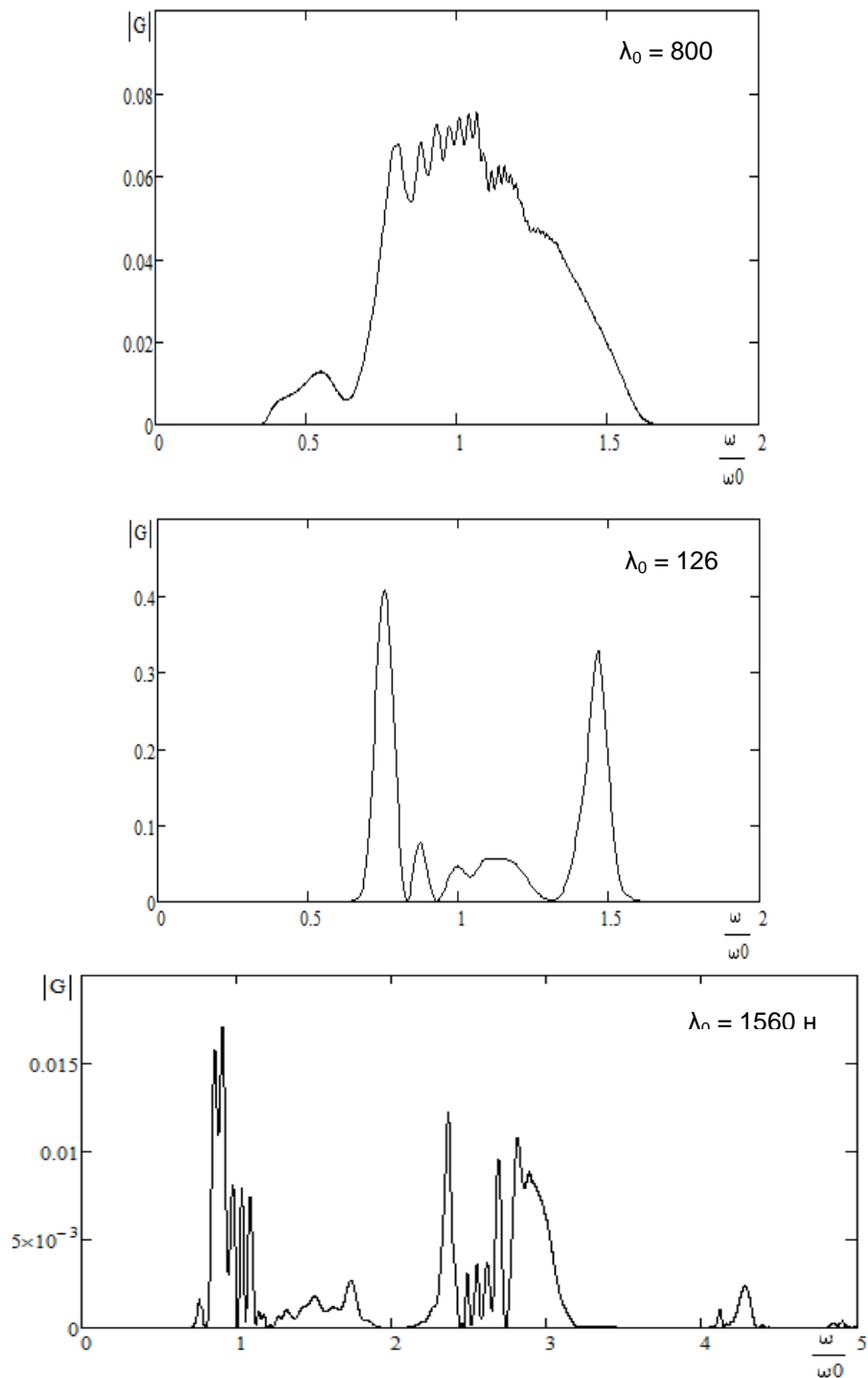


Рис. 2. Модуль спектра фемтосекундного импульса с пиковой интенсивностью $I = 10^{13} \text{ Вт/см}^2$, длительностью 20 фс и различными длинами волн

На рис. 3, 4 представлены электрическое поле и модуль спектра фемтосекундных импульсов с одинаковой длительностью 20 фс и интенсивностями $I = 10^{12}$ Вт/см² и $I = 10^{13}$ Вт/см². Как видно из рисунков, при увеличении интенсивности исходного импульса длительность импульса на выходе из среды увеличивается. Это также сопровождается уширением спектра на выходе из среды.

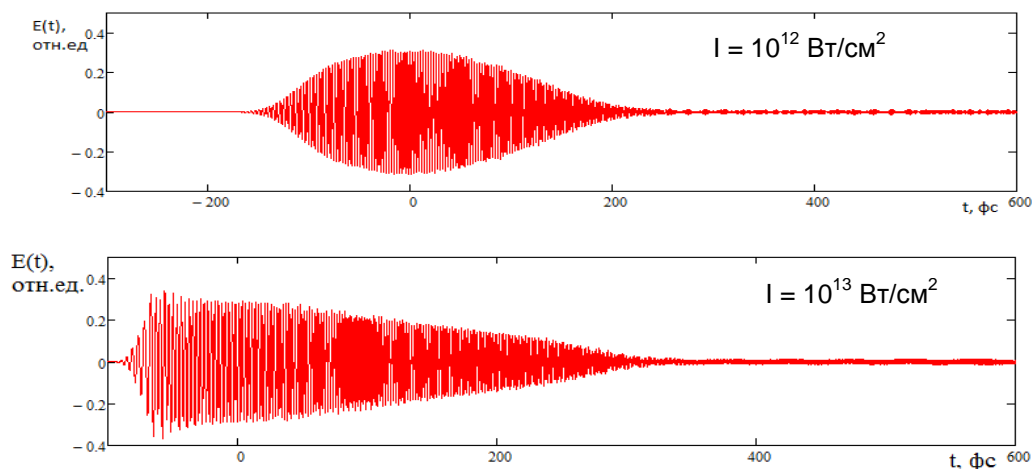


Рис. 3. Электрическое поле фемтосекундных импульсов длительностью 20 фс и центральной длиной волны $\lambda_0 = 800$ нм (нормальная дисперсия)

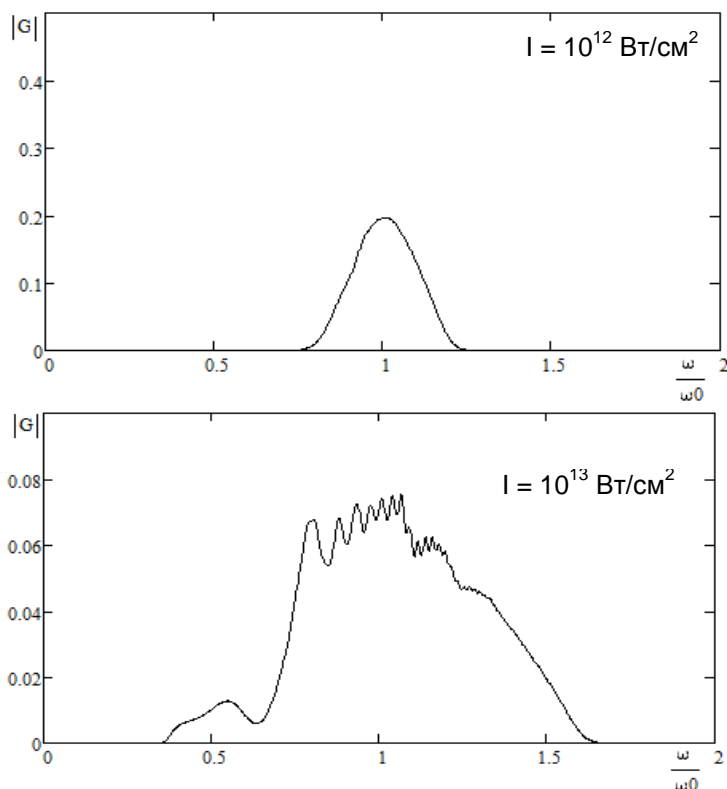


Рис. 4. Модуль спектра фемтосекундных импульсов с длительностью 20 фс и центральной длиной волны $\lambda_0 = 800$ нм (нормальная дисперсия)

На рис. 5, 6 представлены электрическое поле и модуль спектра фемтосекундных импульсов с пиковой интенсивностью $I = 10^{13}$ Вт/см² и длительностями 20 и 40 фс. Из рисунков можно сделать вывод, что при увеличении длительности исходного импульса длительность импульса после прохождения среды возрастает, а спектр уширяется.

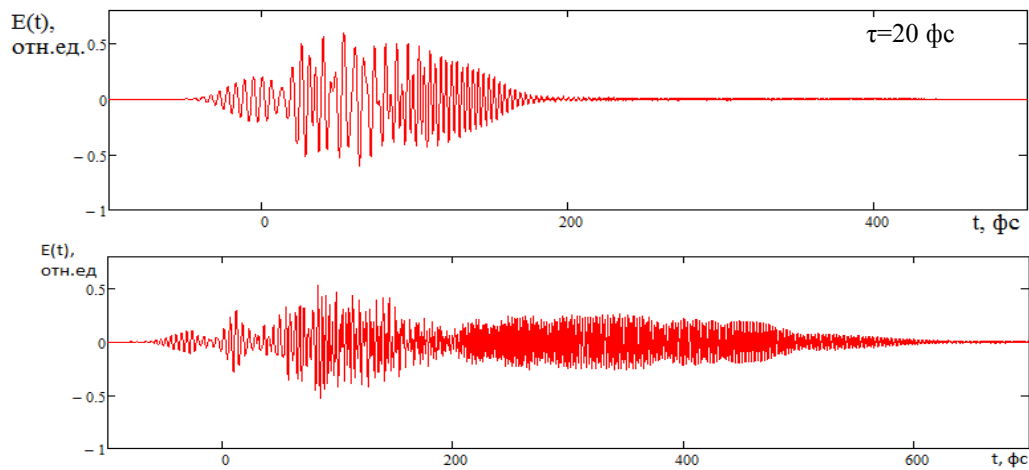


Рис. 5. Электрическое поле фемтосекундных импульсов с центральной длиной волны $\lambda_0 = 1260$ нм (нулевая дисперсия)

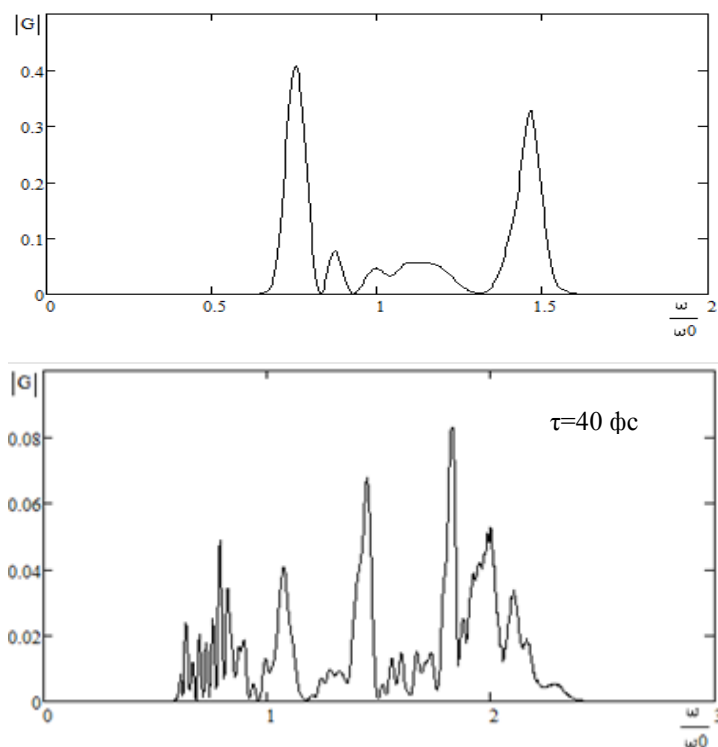


Рис. 6 Модуль спектра фемтосекундных импульсов с центральной длиной волны $\lambda_0 = 1260$ нм (нулевая дисперсия)

Расчет времени когерентности фемтосекундного спектрального суперконтинуума в кварцевом оптическом волокне

Для определения времени когерентности рассмотрим взаимодействие светового фемтосекундного импульса с самим собой, сдвинутым на временную задержку. Зная временную структуру поля на выходе из оптической среды при генерации спектрального суперконтинуума можно определить интерференционный сигнал по следующей формуле [7]:

$$I(\Delta\tau) = \int_{-\infty}^{+\infty} (E(t) + E(t + \Delta\tau))^2 dt, \quad (2)$$

где $I(\Delta\tau)$ – зависимость интенсивности интерференционного поля от временной задержки; $E(t)$, $E(t+\Delta\tau)$ – временная структура поля на выходе из оптической среды и то же поле, сдвинутое во времени; $\Delta\tau$ – временной сдвиг между интерферирующими импульсами.

На рис. 7 представлены результаты расчета интерференционного сигнала для временной структуры на выходе из кварцевого оптического волокна длиной 10 мм при генерации спектрального суперконтинуума для исходного импульса длительностью $\tau_p = 40$ фс, пиковый

интенсивностью $I = 10^{13}$ Вт/см² и с центральными длинами волн излучения: $\lambda_0 = 800$ нм (область нормальной групповой дисперсии кварцевого оптического волокна), $\lambda_0 = 1260$ нм (область нулевой групповой дисперсии кварцевого оптического волокна), $\lambda_0 = 1560$ нм (область аномальной групповой дисперсии кварцевого оптического волокна), определенными по формуле (2).

Из этого рисунка видно, что при смещении центральной длины волны в область аномальной групповой дисперсии оптической среды время когерентности импульса со сверхшироким спектром значительно уменьшается.

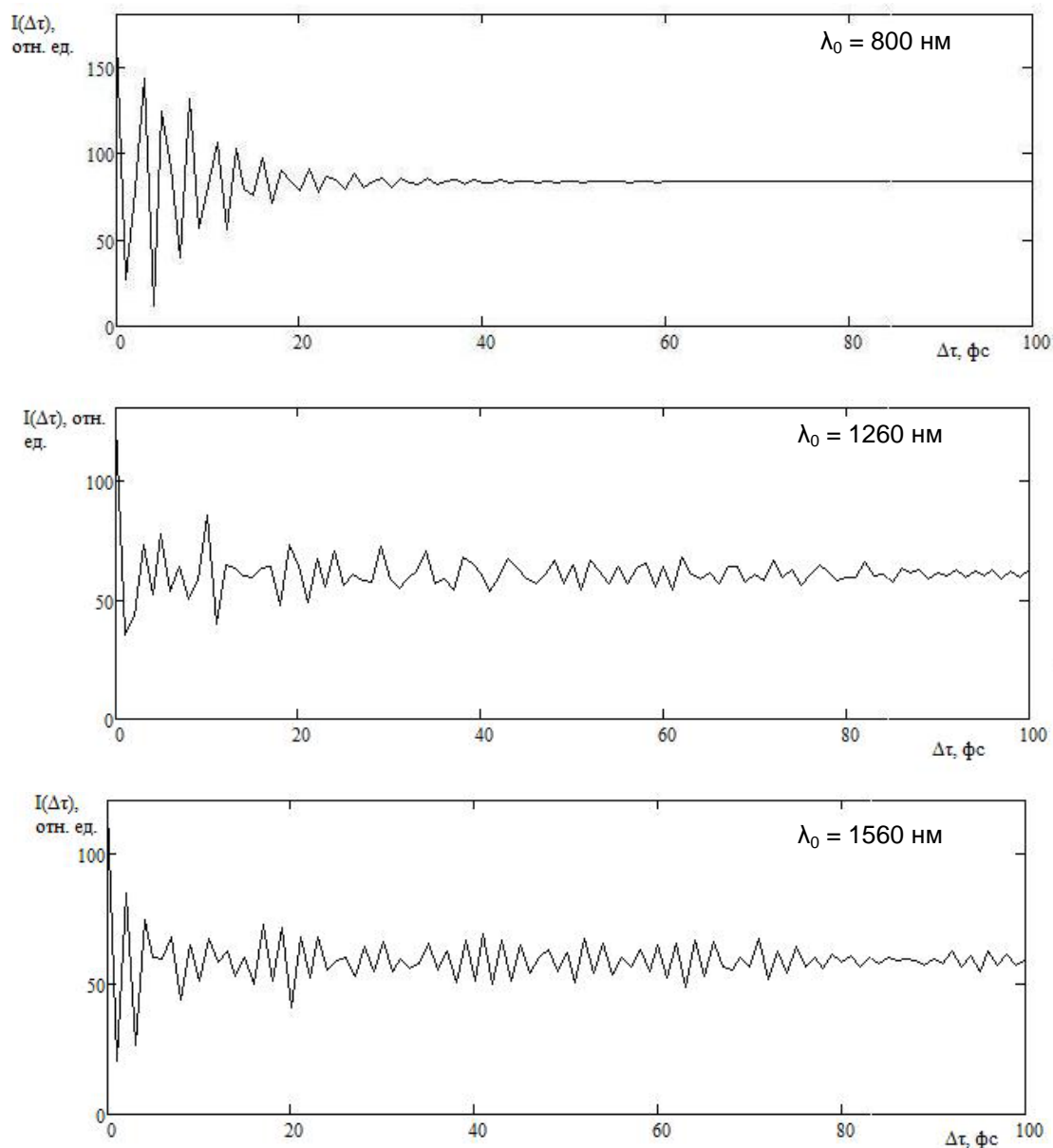


Рис. 7. Интерференционный сигнал для временной структуры на выходе из кварцевого оптического волокна

На рис. 8 приведен пример графического определения времени когерентности фемтосекундного спектрального суперконтинуума. В качестве огибающей $U(\Delta\tau)$ используется экспоненциальная функция (кривая 2), накладываемая на функцию интерференционного сигнала (кривая 1), для временной структуры на выходе из кварцевого оптического волокна, которую можно описать следующим выражением:

$$U(\Delta\tau) = e^{-\left(\frac{\Delta\tau}{\tau_{\text{ког}}}\right)^2},$$

где $\tau_{\text{ког}}$ – время когерентности фемтосекундного спектрального суперконтинуума.

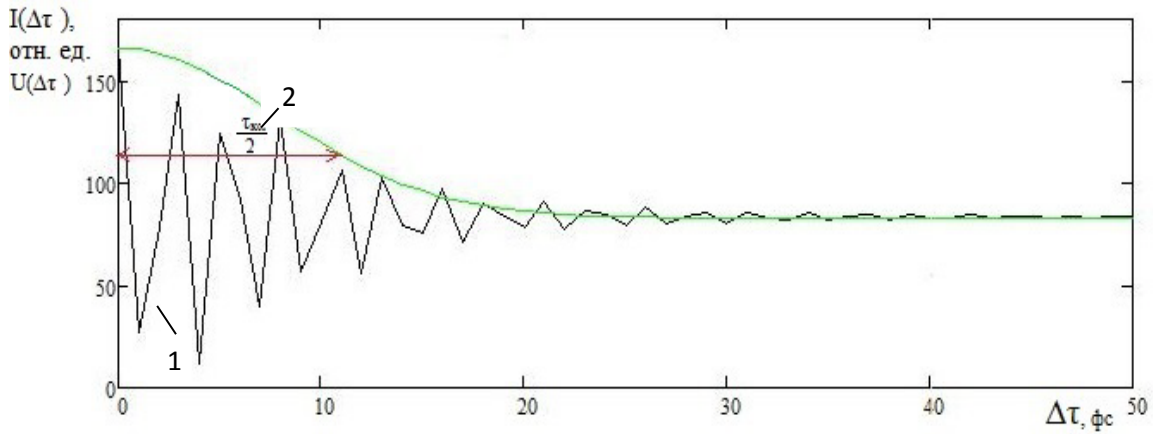
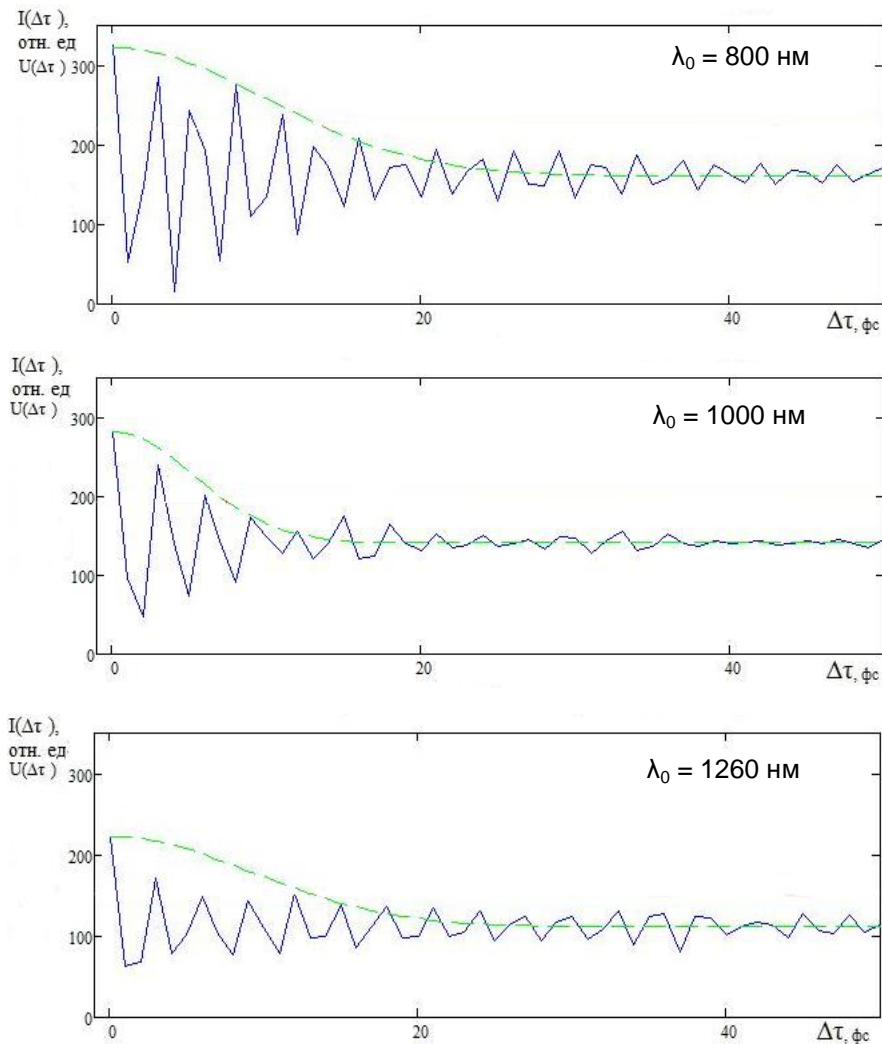


Рис. 8. Расчет времени когерентности импульса со сверхшироким спектром: кривая 1 – интерференционный сигнал; кривая 2 – экспоненциальная огибающая.

На рис. 9 представлены результаты расчета интерференционного сигнала и определения времени когерентности фемтосекундного спектрального суперконтинуума для временной структуры на выходе из кварцевого оптического волокна длиной 10 мм, при генерации спектрального суперконтинуума для исходного импульса длительностью $\tau_p = 80$ фс, пиковой интенсивностью $I = 10^{13}$ Вт/см² и с центральными длинами волн излучения: $\lambda_0 = 800$ нм, $\lambda_0 = 1000$ нм, $\lambda_0 = 1260$ нм, $\lambda_0 = 1560$ нм, определенными по формуле (2).

Из этого рисунка видно, что при смещении центральной длины волны в область аномальной групповой дисперсии оптической среды время когерентности импульса со сверхшироким спектром значительно уменьшается.



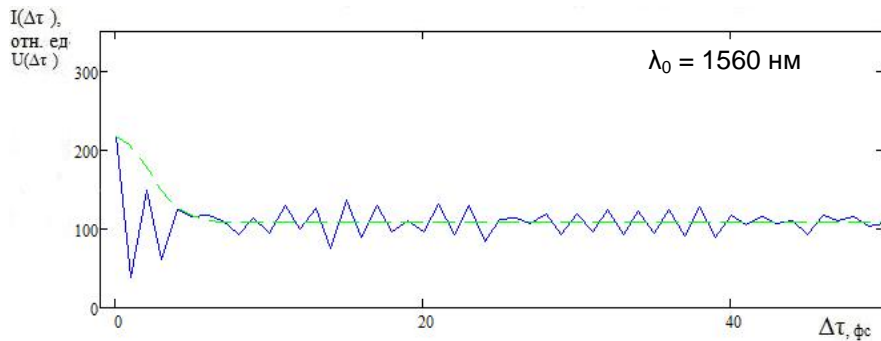


Рис. 9. Расчет времени когерентности импульса со сверхшироким спектром на выходе из кварцевого оптического волокна

Зависимость времени когерентности от центральной длины фемтосекундного лазерного импульса на входе была получена при определении времени когерентности описанным выше методом для различных длин волн (рис. 10).

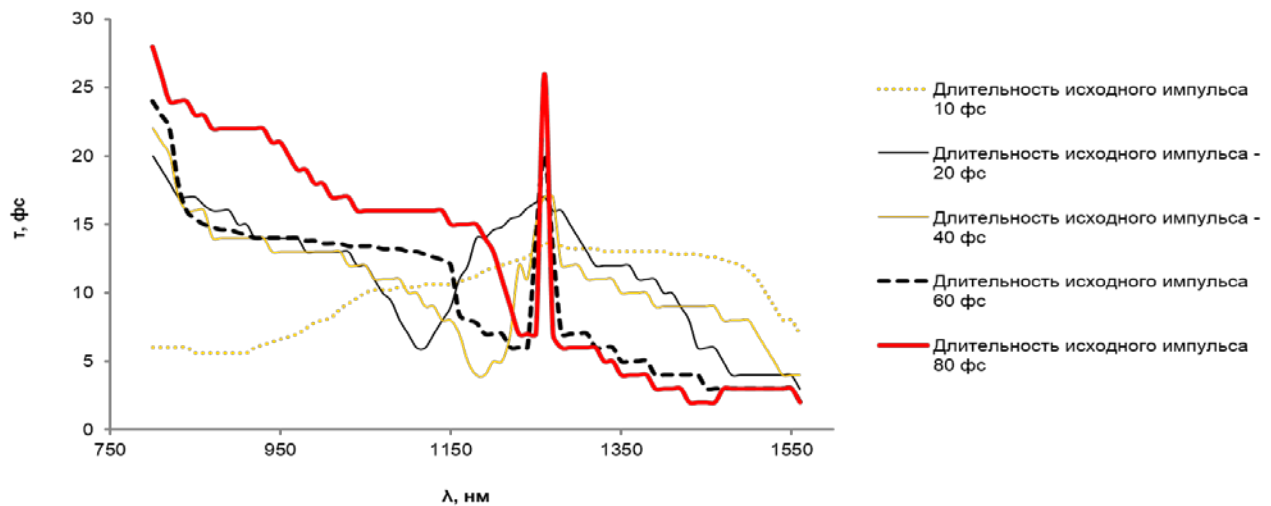


Рис. 10. Зависимость времени когерентности спектрального суперконтинуума от центральной длины фемтосекундного лазерного импульса на входе при пиковой интенсивности $I = 10^{13}$ Вт/см² и различных длительностях исходного импульса

Из полученной зависимости видно, что при увеличении центральной длины волны время когерентности значительно уменьшается, однако в области нулевой групповой дисперсии оптической среды существует «скачок» его увеличения. Например, для длительности исходного импульса 40 фс и пиковой интенсивности $I = 10^{13}$ Вт/см² на длине волны 800 нм время когерентности равно 22 фс, при увеличении длины волны время когерентности падает до 4 фс на длине волны 1180 нм. В области нулевой групповой дисперсии время когерентности резко увеличивается до 20 фс, после чего опять уменьшается, достигая минимума в 4 фс на длине волны 1560 нм.

Для сравнения с полученными результатами были проведены численное моделирование распространения ПКИ и расчет времени когерентности для исходного импульса с интенсивностью $I = 10^{12}$ Вт/см². Результаты представлены на рис. 11.

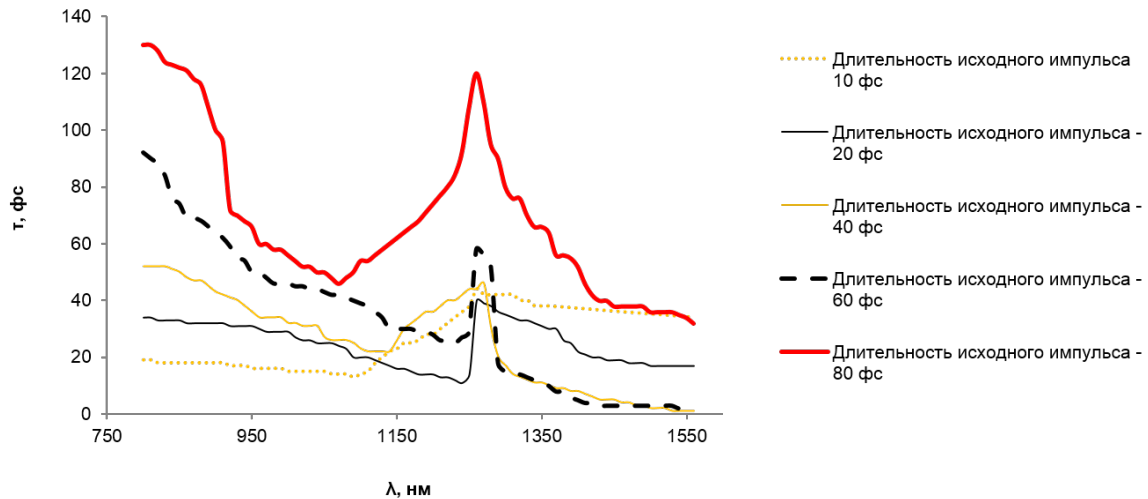


Рис. 11. Зависимость времени когерентности спектрального суперконтинуума от центральной длины фемтосекундного лазерного импульса на входе при пиковой интенсивности $I = 10^{12}$ Вт/см² и различных длительностях исходного импульса

Из полученной зависимости следует, что для различных исходных длительностей импульса и пиковой интенсивности $I = 10^{12}$ Вт/см² в области нулевой групповой дисперсии оптической среды также существует «скачок» увеличения времени когерентности, как и для пиковой интенсивности $I = 10^{13}$ Вт/см². Значение времени когерентности в области нулевой групповой дисперсии оптической среды соизмеримо со значением времени когерентности в области нормальной групповой дисперсии. Например, для длительности исходного импульса 80 фс и пиковой интенсивности $I = 10^{12}$ Вт/см² на длине волны исходного импульса 800 нм время когерентности равно 130 фс, а на длине волны 1260 нм – 120 фс.

Сравнивая зависимости времени когерентности от длины волны для исходных импульсов с одинаковой интенсивностью и разными длительностями, можно сделать вывод, что при увеличении длительности исходного импульса ширина «скачка» времени когерентности в области нулевой групповой дисперсии кварцевого волокна уменьшается. Так для зависимости, изображенной на рис. 12, для исходных импульсов с пиковой интенсивностью $I = 10^{13}$ Вт/см² и длительностях исходного импульса 40 и 80 фс ширина «скачка» времени когерентности равна 50 и 20 нм соответственно.

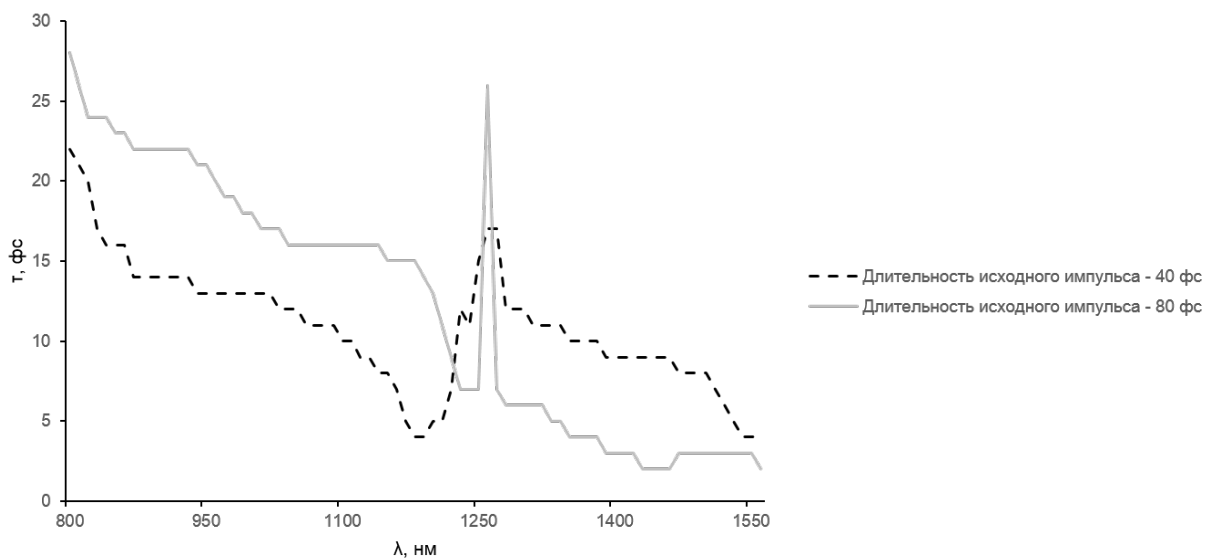


Рис. 12. Зависимость времени когерентности спектрального суперконтинуума от центральной длины фемтосекундного лазерного импульса на входе при пиковой интенсивности $I = 10^{13}$ Вт/см² и длительностях исходного импульса 40 и 80 фс

При дальнейшем сравнении полученных зависимостей можно заметить, что при увеличении пиковой интенсивности исходного импульса для одной и той же длительности значение времени когерентности уменьшается в два раза. На рис. 13 приведена зависимость времени когерентности от длины волны для исходного импульса, длительностью 40 фс и пиковыми интенсивностями $I = 10^{12}$ Вт/см² и $I = 10^{13}$ Вт/см². Значения времени когерентности на длине волны 1260 нм равны 44 и 20 фс соответственно.

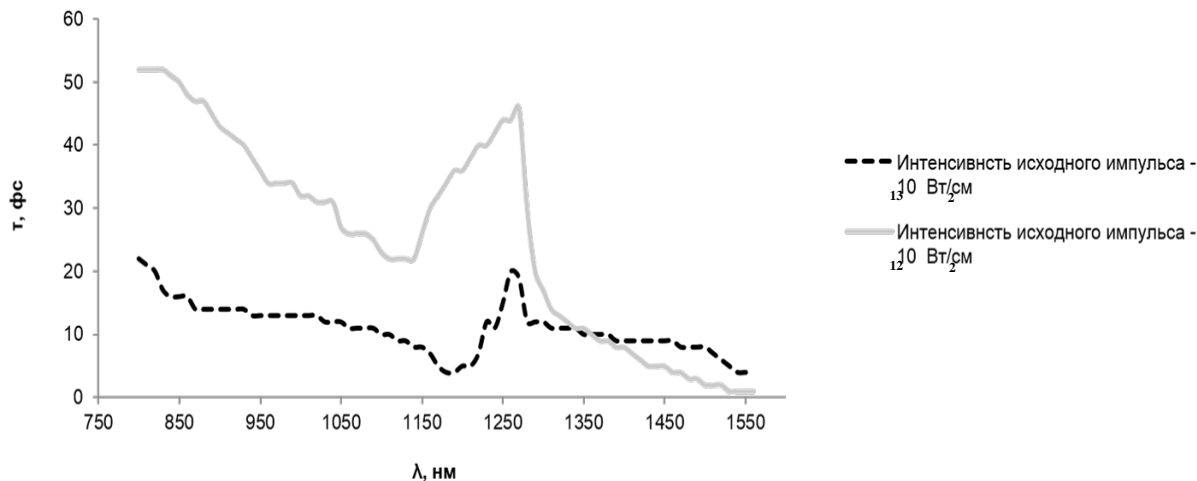


Рис. 13. Зависимость времени когерентности спектрального суперконтинуума от центральной длины фемтосекундного лазерного импульса на входе при длительности исходного импульса 40 фс. И пиковых интенсивностях $I = 10^{13}$ Вт/см² и $I = 10^{12}$ Вт/см²

Заключение

Проведены анализ методов распространения высокоинтенсивных фемтосекундных импульсов в оптических средах и анализ методов определения параметров когерентности для спектрального суперконтинуума. Последний заключался в определении времени когерентности для различного генерированного спектрального суперконтинуума. Было рассчитано время когерентности фемтосекундного спектрального суперконтинуума в оптическом волокне для различных интенсивностей и длительностей исходного излучения, проведен сравнительный анализ полученных зависимостей и теоретически апробированы возможности использования спектральных суперконтинуумов в различных приложениях.

Результаты исследования:

1. При увеличении центральной длины волны время когерентности значительно уменьшается, однако в области нулевой групповой дисперсии оптической среды существует «скачок» увеличения времени когерентности.

2. Для различных исходных длительностей импульса и пиковой интенсивности $I = 10^{12}$ Вт/см² в области нулевой групповой дисперсии оптической среды также существует «скачок» увеличения времени когерентности, как и для пиковой интенсивности $I = 10^{13}$ Вт/см². Значение времени когерентности в области нулевой групповой дисперсии оптической среды соизмеримо со значением времени когерентности в области нормальной групповой дисперсии.

3. При увеличении длительности исходного импульса ширина пика «скачка» времени когерентности в области нулевой групповой дисперсии кварцевого волокна уменьшается.

4. При увеличении пиковой интенсивности исходного импульса для одной и той же длительности на один порядок значение времени когерентности уменьшается в два раза.

Из полученных зависимостей можно сделать вывод, что оптимальными условиями генерации высококогерентного фемтосекундного спектрального суперконтинуума в оптическом волокне являются: пиковая интенсивность – $I = 10^{12}$ Вт/см², длительность исходного импульса – $\tau = 80$ фс. Длина волны исходного импульса должна находиться в области нормальной или нулевой групповой дисперсии (800 нм и 1260 нм соответственно). Использование излучения спектрального суперконтинуума, генерируемого в данных условиях, является целесообразным, например в системах передачи информации по причине высокого значения времени когерентности.

Области возможного использования

Использование излучения спектрального суперконтинуума, генерируемого в области нормальной групповой дисперсии, целесообразно в системах передачи информации [8, 9] по причине высокого значения времени когерентности и квазилинейной частотной модуляции временной структуры. Знание о времени когерентности спектрального суперконтинуума также может быть использовано, например для расчета максимального разрешения оптической когерентной томографии [10–12].

Публикации

1. Мельник М.В., Цыпкин А.Н. Выявление оптимальных условий генерации высококогерентного спектрального суперконтинуума в областях нормальной, нулевой и аномальной дисперсии групповых скоростей // Сб. трудов VII Международной конференции «Фундаментальные проблемы оптики – 2012». Санкт-Петербург, 2012. С. 459–460.
2. Мельник М.В. Выявление оптимальных условий генерации высококогерентного фемтосекундного спектрального суперконтинуума в оптическом волокне // Сб. тезисов II Всероссийского конгресса молодых ученых «КМУ – 2013». Вып. 3. Санкт-Петербург, 2013. С. 196–198.
3. Мельник М.В., Цыпкин А.Н. Расчет времени когерентности фемтосекундного спектрального суперконтинуума в кварцевом оптическом волокне // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 3 (85). С. 42–46.

Литература

1. Dudley J.M., Coen S. Supercontinuum generation in photonic crystal fiber // *Reviews of Modern Physics*. 2006. Vol. 78. No 4. P. 1135–1184.
2. Islam M.N., Sucha G., Bar-Joseph I. et al. Femtosecond distributed soliton spectrum in fibers // *Journal of the Optical Society of America B*. 1989. Vol. 6. No 6. P. 1149–1158.
3. Herrmann J., Griebner U., Zhavoronkov N. et al. Experimental evidence for supercontinuum generation by fission of higher-order solitons in photonic fibers // *Physical Review Letters*. 2002. Vol. 88. No 17. P. 173901 (4 pages).
4. Corwin K.L., Newbury N.R., Dudley J.M. et al. Fundamental amplitude noise limitations to supercontinuum spectra generated in a microstructured fiber // *Applied Physics B: Lasers and Optics*. 2003. Vol. 77. No 2–3. P. 269–277.
5. Козлов С.А., Сазонов С.В. Нелинейное распространение импульсов длительностью в несколько колебаний светового поля в диэлектрических средах // *Журнал экспериментальной и теоретической физики*. 1997. Т. 111. № 2. С. 404–418.
6. Шполянский Ю.А. Сценарии развития фемтосекундного спектрального суперконтинуума // *Проблемы когерентной и нелинейной оптики*. СПб., 2000. С. 136–153.
7. Беспалов В.Г., Козлов С.А., Крылов В.Н., Путилин С.Э. Фемтосекундная оптика и фемтотехнологии. СПб., 2010. – 234 с.
8. Bakhtin M.A., Kozlov S.A. Generation of the discrete spectral supercontinuum in two intensive ultrashort pulses interaction // *Optical Memory and Neural Network*. 2006. Vol. 15. No 1. P. 1–10.
9. Козлов С.А., Дроздов А.А., Цыпкин А.Н. Устройство для формирования последовательности фемтосекундных световых импульсов (2009). Патент 87058.
10. Povazay B., Bizheva K., Unterhuber A. et al. Submicrometer axial resolution optical coherence tomography // *Optics Letters*. 2002. Vol. 27. No 20. P. 1800–1802.
11. Humbert G., Wadsworth W., Leon-Saval S. et al. Supercontinuum generation system for optical coherence tomography based on tapered photonic crystal fiber // *Optics Express*. 2006. Vol. 14. No 4. P. 1596–1603.
12. Hartl I., Li X.D., Chudoba C. et al. Ultrahigh-resolution optical coherence tomography using continuum generation in an air-silica microstructure optical fiber // *Optics Letters*. 2001. Vol. 26. No 9. P. 608–610.



Николаева Татьяна Юрьевна
Факультет фотоники и оптоинформатики
Кафедра фотоники и оптоинформатики
E-mail: paltanya@mail.ru

РАЗРАБОТКА СХЕМНЫХ РЕШЕНИЙ ДЛЯ РЕГИСТРАЦИИ ВЫСОКИХ ПРОСТРАНСТВЕННЫХ ЧАСТОТ ДЛЯ БЕЗЛИНЗОВЫХ МЕТОДОВ ФОРМИРОВАНИЯ ИЗОБРАЖЕНИЙ

Т.Ю. Николаева

Научный руководитель – старший научный сотрудник Н.В. Петров

Введение

Обычно изображение, зарегистрированное в дальней зоне около оптической оси, не содержит высоких пространственных частот, так как их интенсивность на удалении от оптической оси мала по сравнению с низкочастотными компонентами. В силу ограниченности динамического диапазона регистрирующих сигнал матричных фотодетекторов величиной 8, 10 или 12 бит малые возмущения волнового поля, соответствующие высоким пространственным частотам, расходящимся под большими углами от оптической оси, не удается зафиксировать одновременно с высокоинтенсивными низкочастотными компонентами. Таким образом, теряется информация о мелких деталях в изображении исследуемого объекта.

Известные теневые методы измерений (метод темного поля или темнопольная микроскопия) позволяют преодолеть указанный недостаток и увеличить контраст изображения только за счет регистрации света, рассеянного изучаемым образцом. Они основаны на подавлении главного максимума путем размещения непрозрачного экрана прямо на оптической оси. В результате могут быть зарегистрированы низкоинтенсивные высокие пространственные частоты. Основным ограничивающим фактором данного метода является то, что изображение формирует только малая часть падающего света, и приходится применять достаточно мощные источники, что иногда приводит к повреждениям образца [1].

Однако в связи с определяющей ролью использования человечеством ЭВМ, мощный стимул к развитию получили вычислительные методы формирования изображений. Они основаны на решении дифракционных интегралов и использовании различных математических методов оптики. В этом случае, как правило, удается либо снизить требования к качеству оптической схемы прибора, либо увеличить его ключевые характеристики. При решении некоторых задач, обходятся вовсе без оптической системы (особенно при использовании когерентного излучения). В этом случае для преобразования полученных на матричном фотоприемнике сигналов изображения применяются уравнения распространения волн в свободном пространстве. Такие вычислительные методы называются безлинзовыми.

Для улучшения разрешающей способности безлинзовых методов формирования изображений (включающих в себя итерационные методы восстановления фазы) предложено рассмотреть некоторые схемы регистрации высоких пространственных частот, альтернативные осевым, и оценить целесообразность их использования. Безлинзовые (или дифракционные) методы формирования изображений комбинируют использование теоретических методов и компьютерных алгоритмов [2].

Разработка программного обеспечения для моделирования схем регистрации высоких пространственных частот

Предложены две схемы регистрации высоких пространственных частот. Для первой схемы в качестве тестового объекта использовалась протяженная в объеме среда с расположенными в ней частицами. Предполагается, что формирование изображения будет

происходить с использованием когерентного лазерного излучения. Данная задача является типовой, например при изучении планктона [3], динамики частиц в сверхтемпературных газовых потоках [4] и пр.

Для исключения дифракционных эффектов на краях апертуры рассматривался волновой фронт, имеющий в поперечном сечении супергауссов профиль. Для этого была написана программа расчета функции супергаусса. Объем среды моделировался в специально разработанном программном обеспечении. Он разбивался на продольные сегменты, каждый из которых содержал определенное количество частиц, распределенных случайным образом. Далее через эту объемную среду и систему расположенных друг за другом элементов (кольцевых диафрагм и линз) до плоскости регистрации моделировалось распространение волнового поля.

В качестве второй предложенной схемы регистрации высоких пространственных частот предлагается использовать оптический клин.

В обоих случаях процесс моделирования производился с использованием метода распространения углового спектра плоских волн, реализованного через два быстрых преобразования Фурье. Это способ численного расчета распространения волнового поля u из плоскости с координатами (x', y') в плоскость с координатами (x, y) действует, начиная с расстояний $l \gg \lambda$ (λ – длина волны излучения) [5]:

$$U(x, y) = \hat{F}^{-1} \left\{ H(f_x, f_y, l) \cdot \hat{F} [u(x', y')] \right\}, \quad (1)$$

где \hat{F} и \hat{F}^{-1} – операторы прямого и обратного преобразований Фурье соответственно, а $H(f_x, f_y, l)$ – передаточная функция, l – расстояние от плоскости объекта до плоскости регистрации.

Использование уравнения (1), по сравнению с классическим преобразованием Френеля, является более предпочтительным. Для параметров установки, работающей в терагерцовом диапазоне частот, уравнение удовлетворяет ограничениям на расчет в ближней зоне, а расчет в дальней зоне может быть выполнен с помощью процедуры заполнения нулями – пикселями нулевой яркости. Максимальное расстояние, на котором может быть рассчитано распределение волнового поля методом распространения углового спектра без искажений при равных размерах пикселя в плоскостях объекта и регистрации, определяется как:

$$l_{\max} = \frac{\Delta x D}{\lambda}, \quad (2)$$

где l – расстояние от плоскости объекта до плоскости регистрации, Δx – размер пикселя в плоскости объекта, равный размеру пикселя в плоскости регистрации, D – линейный размер апертуры.

Согласно модели распространения углового спектра плоских волн, угловой спектр $U_0(f_x, f_y)$ определяет амплитуду каждой из компонент Фурье. Таким образом, функцию $u(x, y, 0)$, описывающую распределение поля сразу за исследуемым объектом, можно разложить на совокупность бесконечных плоских пространственных волн [6]:

$$u(x, y, 0) = \iint \exp[i2\pi(f_x x + f_y y)] U_0(f_x, f_y) df_x df_y. \quad (3)$$

Эти плоские волны $\exp[i2\pi(f_x x + f_y y)]$ имеют период $(f_x^2 + f_y^2)^{-1/2}$, а направление их распространения задается направляющими косинусами:

$$\begin{aligned} \cos \alpha &= \lambda f_x, \\ \cos \beta &= \lambda f_y, \end{aligned} \quad (4)$$

$$\cos \gamma = \lambda \sqrt{\frac{1}{\lambda^2} - (f_x^2 + f_y^2)}.$$

В случае использования произвольного полупрозрачного транспаранта каждую его точку можно рассматривать как источник излучения набора плоских волн с различными пространственными частотами, каждая из которых распространяется в своем определенном направлении. Наиболее высокочастотные плоские волны распространяются под большими углами относительно оптической оси. Для самых удаленных точек объекта эти волны пересекаются в ближней зоне на расстоянии l_0 :

$$l_0 = \frac{\Delta x D}{\lambda}, \quad (5)$$

где λ – длина волны, Δx – размер пикселя в плоскости регистрации (благодаря использованию при расчете быстрого преобразования Фурье шаг дискретизации в плоскости объекта также равен Δx), D – линейный размер апертуры.

Ближняя зона ($l \leq l_0$) имеет большое значение при решении задачи восстановления волнового фронта из распределений интенсивности. Во-первых, зарегистрированные в этой области пространства распределения интенсивности содержат наиболее высокие пространственные частоты, которые переносят информацию о мельчайших деталях восстанавливаемого изображения объекта. Во-вторых, именно в ней происходит «перемешивание» пространственных частот от всех точек объекта. Регистрация входных данных в этой области является одним из условий восстановления изображения объекта с высоким качеством [7].

На рис. 1 представлена интерфейсная панель разработанного программного обеспечения, с помощью которого производится моделирование объемной среды, с расположенными в ней в случайном порядке частицами, и распространение волнового поля через нее. В центре находится изображение распределения амплитуды после прохождения излучения через объем среды. Программа также позволяет наблюдать процесс распространения волнового поля в реальном времени, то есть видеть распределение амплитуды после прохождения излучения через каждый продольный сегмент среды.

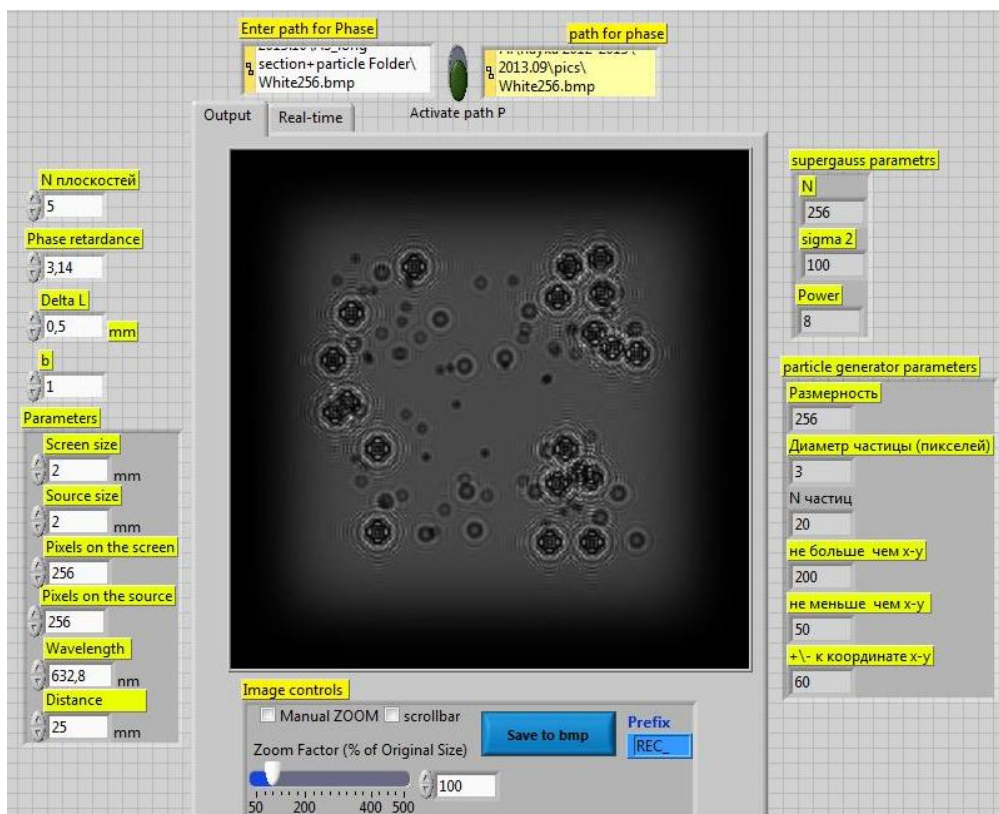


Рис. 1. Интерфейсная панель программного обеспечения, моделирующего процесс дифракции на частицах в объеме среды

На рис. 2 представлена блок-диаграмма программы. На ней выделены три области, на которые разработаны отдельные подпрограммы, оформленные как подприборы, что позволяет упростить реализацию главной программы:

- область 1 соответствует подпрограмме, реализующей моделирование объемной среды и распределение частиц в ней;
- область 2 – подпрограмме, реализующей расчет функции супергаусса;
- область 3 – главный прибор, который рассчитывает распространение волнового поля методом углового спектра плоских волн.

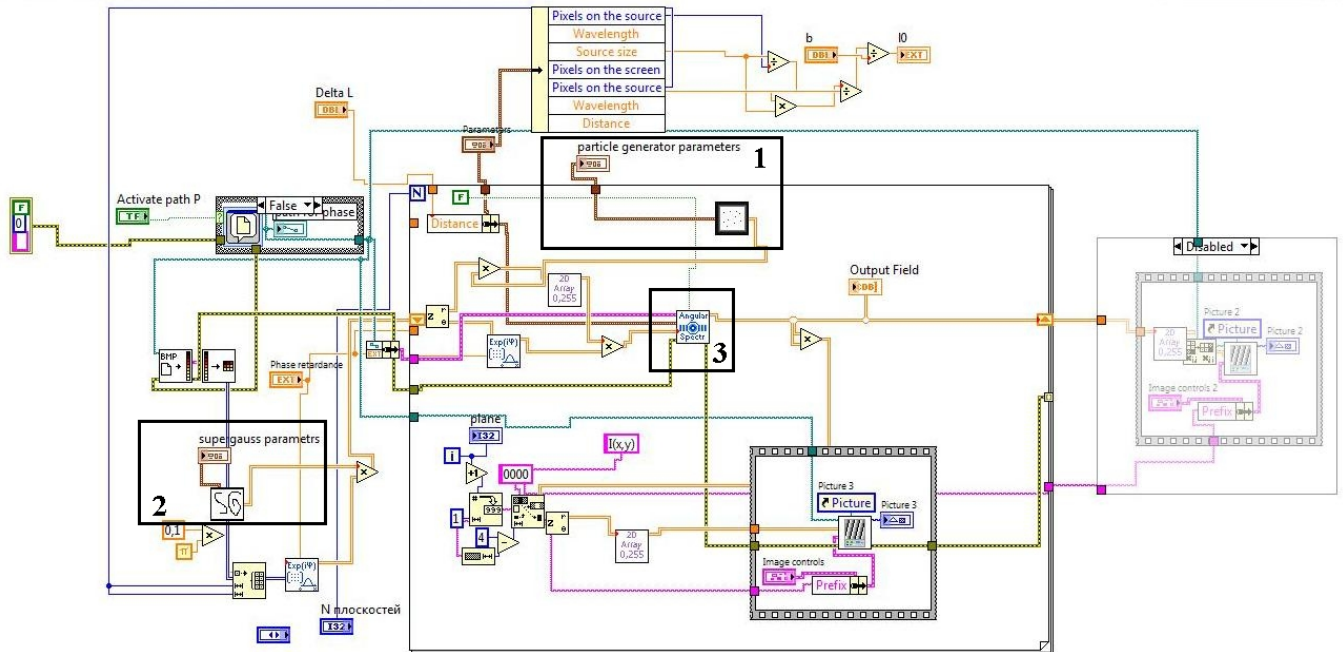


Рис. 2. Блок-диаграмма программного обеспечения, моделирующего процесс дифракции на частицах в объеме среды

Моделирование процесса дифракции на объектах, соответствующих предложенным схемам регистрации высокочастотных компонент

Моделирование процесса дифракции в случае использования схемы с внеосевой регистрацией высоких пространственных частот проводилось для простой круглой диафрагмы, а в случае использования схемы с кольцевыми диафрагмами – для объема среды, в котором в случайном порядке распределены частицы.

Рассмотрим первую предложенную схему, которая предполагает использование оптического клина (рис. 3) [8].

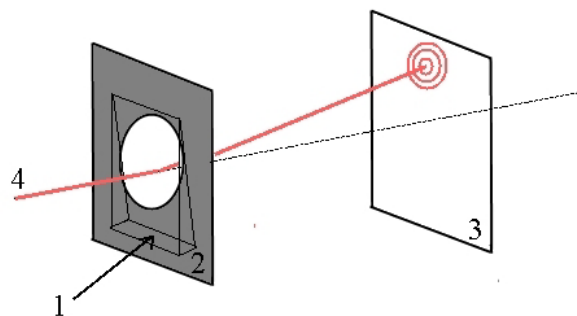


Рис. 3. Схема регистрации высоких пространственных частот, использующая оптический клин: 1 – наклонный фазовый экран; 2 – тестовый объект; 3 – плоскость регистрации волнового поля; 4 – падающее на тестовый объект излучение

На рис. 4 представлены тестовый объект и фазовое распределение интенсивности наклонного фазового экрана – оптического клина. Здесь необходимым условием расчета распространения волнового поля методом углового спектра плоских волн являлось

применение процедуры заполнения нулевыми элементами области памяти амплитудного и фазового распределений интенсивности тестового объекта.



Рис. 4. а – тестовый объект; б – фазовое распределение наклонного фазового экрана

На рис. 5 изображены смоделированные поперечные и фазовые распределения интенсивности.

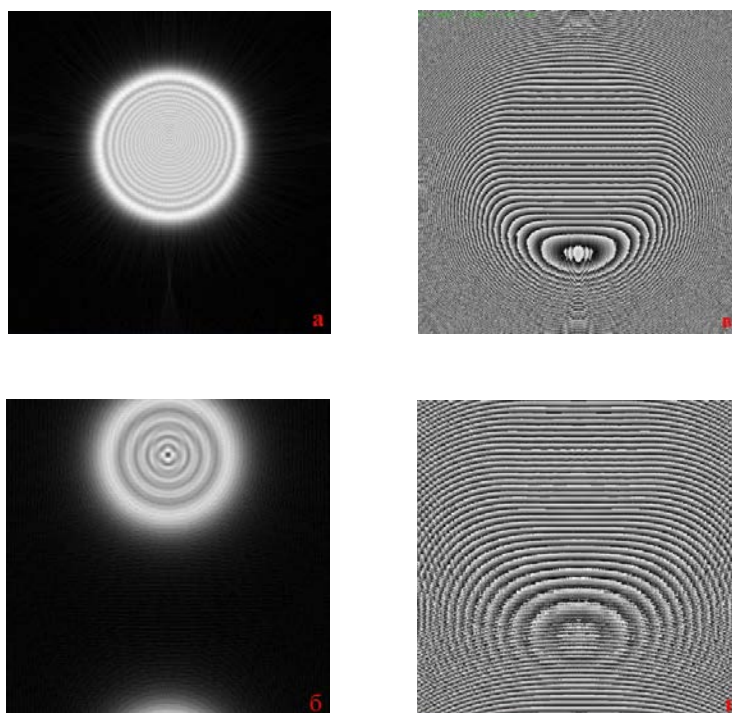


Рис. 5. а, б – распределения интенсивности, перпендикулярные оптической оси на 7 и 24 итерациях; в, г – фазовые распределения на 7 и 24 итерациях

На рис. 6 представлено полученное продольное распределение интенсивности за круглой диафрагмой и наклонным фазовым экраном. При удалении от плоскости, где расположен тестовый объект, благодаря использованию оптического клина, наблюдаем отклонение лучей от оптической оси.

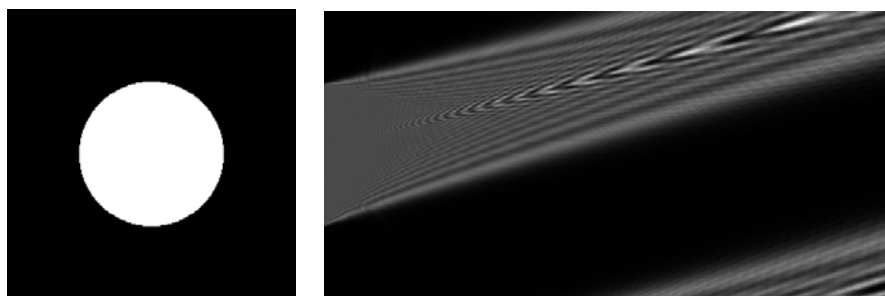


Рис. 6. Продольное распределение интенсивности за круглой диафрагмой и оптическим клином

Далее рассмотрим вторую схему регистрации высоких пространственных частот, использующую кольцевые диафрагмы. При наблюдении каких-либо объемных скоплений частиц в когерентном излучении его рассеяние на них порождает спеклы и не дает возможности отличить изображение одной из частиц от другой. Метод, основанный на использовании двух кольцевых диафрагм, позволяет выделять только узкую область из ансамбля частиц по продольной оси в объеме среды. Жертвуя светосилой, мы выигрываем в отношении сигнал/шум.

В специально разработанном программном обеспечении, описанном выше, была смоделирована объемная среда и распределение частиц в ней. Электромагнитное излучение, проходя через объем среды с распределенными в ней случайным образом частицами, дифрагирует на каждом из них в каждом из продольных сегментов объема. Дифракционные картины при этом накладываются одна на другую. Волновое поле распространяется через оптическую схему, состоящую из двух кольцевых диафрагм и линзы. Эти кольцевые диафрагмы нужны, для того чтобы выделить определенные высокие пространственные частоты, которые переносят информацию о необходимой узкой области объемной среды с частицами. На рис. 7 представлена схема регистрации высоких пространственных частот, использующая предложенный метод. Когерентное излучение падает на тестовый объем среды и дифрагирует на частицах 1. Далее с помощью кольцевой диафрагмы 2 выделяется определенная область дифрагировавшего излучения и фокусируется с помощью линзы 3 и кольцевой диафрагмы 4 (идентичной 2) в плоскости регистрации фотоприемного матричного устройства.

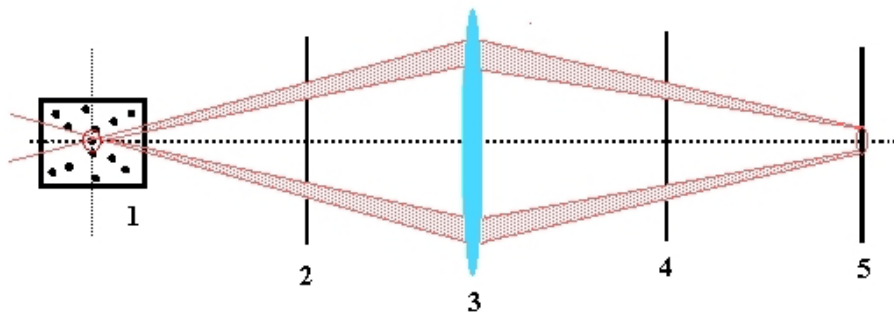


Рис. 7. Схема регистрации высоких пространственных частот с использованием кольцевых диафрагм

Определение наиболее оптимального схемного решения для регистрации высоких пространственных частот

Далее было необходимо определить наиболее оптимальную из предложенных внеосевых схем регистрации высоких пространственных частот. У каждой из описанных в работе схем есть ряд достоинств и недостатков.

Схема регистрации с использованием оптического клина обладает таким преимуществом, как простота реализации и отсутствие дополнительных оптических элементов. Также следует отметить, что данный метод хорошо работает в терагерцовом диапазоне при использовании широкополосного терагерцового излучения. Однако применение этой схемы регистрации ограничивается рядом оптических задач. Из недостатков стоит отметить необходимость выполнения процедуры заполнения нулевыми элементами пикселей памяти в процессе моделирования схемы.

Схема регистрации высоких пространственных частот, использующая кольцевые диафрагмы несомненно более сложна в реализации, однако применима для ряда важных оптических и физических задач, таких как обнаружение частиц, в том числе и прозрачных, в объеме. Недостаток данной схемы заключается в том, что, выигрывая в соотношении сигнал/шум, мы теряем в светосиле. Также остается пока невозможным выделение с ее помощью отдельных конкретных частиц, так как в плоскости регистрации будет фокусироваться область их ансамбля.

Заключение

В ходе выполнения работ по проекту были предложены две внеосевые схемы регистрации высоких пространственных частот. Первая из них предполагает использование оптического клина для обеспечения фазового наклонного фронта. Вторая использует две кольцевые диафрагмы, с помощью которых происходит выделение ансамбля частиц в объеме среды. В программной среде LabVIEW был разработан ряд программ, моделирующих предложенные схемы, произведен расчет распространения в них волнового поля. Проведено сравнение предложенных схем регистрации высоких пространственных частот, перечислены их достоинства и недостатки.

Области возможного использования результатов проекта

Полученные результаты исследования помогут повысить разрешающую способность безлинзовых методов формирования изображения, в том числе и итерационных методов восстановления волновых фронтов. Изучение различных схем регистрации высоких пространственных частот полезно при разработке новых подходов к регистрации изображений для последующего восстановления в целях повышения качества восстановления волновых полей. Также рассмотренная задача выделения узкой области объемной среды является типовой, например при изучении прозрачных живых объектов и исследовании динамики частиц в сверхтемпературных газовых потоках.

Публикации

1. Petrov N.V., Galiaskarov A.N., Nikolaeva T.Yu., Bepalov V.G. The features of optimization of a phase retrieval technique in THz frequency range // Proceedings of SPIE. 2012. Vol. 8413. Speckle 2012: V International Conference on Speckle Metrology, 84131T (September 11, 2012); doi: 10.1117/12.978688.
2. Николаева Т.Ю., Петров Н.В. Исследование продольных дифракционных картин для видимого и терагерцевого излучения // Сб. трудов VII Международной конференции «Фундаментальные проблемы оптики – 2012». Санкт-Петербург, 2012. С. 529–531.
3. Николаева Т.Ю. Исследование свойств пространственных спектров квазимонохроматических пучков излучения в ближней и дальней зонах дифракции с использованием методов компьютерной оптики // Сб. тезисов докладов II Всероссийского конгресса молодых ученых «КМУ – 2013». Вып. 3. СПб., 2013. С. 199–201.

Литература

1. Villiger M., Pache C. Dark-field optical coherence microscopy // Optics Letters. 2010. Vol. 35. No 20. P. 3489–3491.
2. Science of Microscopy / P. Hawkes, J.C.H. Spence (eds.). New York, 2006. – 1196 p.
3. Демин В.В., Макаров А.В., Половцев И.Г. Регистрация планктона с помощью имитатора погружаемой голографической камеры // Оптика атмосферы и океана. 2006. Т. 19. № 4. С. 312–318.
4. Воронецкий А.В., Михайлов В.Н., Петров Н.В., Стаселько Д.И. Измерение пространственно-временных параметров движения самосветящихся частиц в сверхзвуковом высокотемпературном потоке // Оптический журнал. 2012. Т. 79. № 1. С. 18–24.
5. Гудмен Дж. Введение в Фурье-оптику. Пер. с англ. М., 1970. – 364 с.
6. Петров Н., Беспалов В., Волков М. Восстановление волнового фронта при безопорной цифровой ПЗС-регистрации мультиспектральных спекл-картин // Наносистемы: Физика, химия, математика. 2011. Т. 2. № 1. С. 82–90.
7. Petrov N.V., Galiaskarov A.N., Nikolaeva T.Yu., Bepalov V.G. The features of optimization of a phase retrieval technique in THz frequency range // Speckle – 2012: V International Conference on Speckle Metrology. Proceedings of SPIE. 2012. Vol. 8413. P. 84131.
8. Kanghee L., Jaewook A. Single-pixel coherent imaging // Applied Physics Letters. 2010. Vol. 97. P. 241101.



Петухова Дарья Борисовна
Факультет оптико-информационных систем и технологий
Кафедра оптико-электронных приборов и систем
E-mail: petuxovadarja@yandex.ru

ИССЛЕДОВАНИЕ И РАЗРАБОТКА ПРИНЦИПОВ ОБРАБОТКИ ДАННЫХ В ВИДЕОИНФОРМАЦИОННЫХ СИСТЕМАХ ЦВЕТОВОГО АНАЛИЗА МИНЕРАЛОВ

Д.Б. Петухова

Научный руководитель – доцент А.Н. Чертов

Цель и задачи работы

Целью работы является исследование принципов цветового анализа применительно к решению задачи селекции минеральных объектов с использованием видеоинформационных средств наблюдения, а также разработка алгоритма функционирования подобных систем с учетом необходимости определения параметров разделения объектов.

В данной работе поставленная цель достигается путем решения следующих задач:

- теоретический обзор принципов цветового анализа в видеоинформационных системах;
- экспериментальное исследование существующих алгоритмов цветовой интерполяции с целью определения оптимального из них;
- анализ методов выделения объектов на фоне и цветовой сегментации для последующей реализации одного из них в программном обеспечении видеоинформационной системы цветового анализа минеральных объектов;
- разработка алгоритма функционирования системы цветового анализа, а также формирование основного тела ее программного обеспечения;
- проведение ряда экспериментальных исследований разработанной видеоинформационной системы.

Методы исследования

В работе применяются:

- элементы классической теории цвета и его измерения, в том числе модели представления и методы преобразования цвета;
- методы цветовой интерполяции применительно к обработке изображений минеральных объектов;
- основы методов выделения объектов на фоне;
- элементы базового способа интерактивной цветовой сегментации Magic Wand;
- методы объектно-ориентированного программирования на языке C++.

Основные результаты

Теоретический обзор основных принципов цветового анализа показал, что для получения окончательного алгоритма цветового анализа минеральных объектов необходимо последовательно реализовать следующие функции:

1. Получение данных с матрицы и построение полноцветного изображения. Метод интерполяции должен обеспечивать необходимый уровень качества, при этом его быстродействие не должно ощутимо замедлять работу установки.

2. Выделение объекта на относительно равномерном фоне. Алгоритм должен правильно отделять от объекта наличие тени, что имеет большое значение для получения правильного результата цветового анализа.

3. Цветовая сегментация на выделенном объекте. Система должна быть достаточно чувствительной для различения близких оттенков.

4. Автоматическая классификация объекта на основе выявленных критериев разделения объектов.

При этом следует отдельное внимание уделить программной оболочке разработанного алгоритма цветового анализа, которая должна обеспечивать интерактивность и быть удобной в использовании.

Следующим шагом стала разработка алгоритма на основе представленных этапов. Качество полученного изображения напрямую зависит от применяемого метода интерполяции. Зная это, для решения поставленной задачи цветового анализа минеральных объектов необходимо тщательно выбрать оптимальный способ, который обеспечивал бы корректное восстановление мелких вкраплений, блесток и других характерных для минералов включений.

На сегодняшний день существует множество методов цветовой интерполяции. Для сравнения был выбран двадцать один метод различного уровня сложности. Восстанавливая «сырой» формат данных тестового изображения [1] до «полноцветного» различными методами интерполяции и вычисляя значение *PSNR* (отношение пикового уровня сигнала к шуму), можно оценить качество каждого из них. Далее была сформирована Табл. 1, где алгоритмы расположены в порядке увеличения качества проинтерполированного изображения.

Таблица 1
Результаты сравнения методов цветовой интерполяции

№	Название алгоритма	PSNR
1	Nearest	17,56183
2	Chang-Tan	19,43616
3	Malvar He Cutler	19,96013
4	Laroche-Prescott	20,03506
5	Adaptive	20,18636
6	Freeman	20,28008
7	Wang-Lin-Xue	20,33354
8	VNG	20,34372
9	Hamilton-Adams	20,43820
10	Hamilton	20,43820
11	Bilinear	20,53772
12	David cok	20,53772
13	Chuan-kai Lin	20,56421
14	PMI	20,57906
15	Adams	20,63675
16	Ron Kimmel	20,64211
17	Lukin-Kubasov	20,64211
18	Bicubic	20,70012
19	Wemmiao Lu	20,75577
20	PRI	20,78996
21	AHD	20,85858

На основе визуального сравнения и анализа полученных значений *PSNR* был сделан выбор в пользу метода *Wemmiao Lu*, который показал наилучшую способность к восстановлению мелких деталей.

Для оконтуривания необходимой зоны изображения были испробованы несколько распространенных на сегодняшний день способов: дифференциальные методы, оператор Робертса, Лапласиан, оператор Собела, выделение контуров через оператор Кени, с помощью межкадровой разности в совокупности с оператором Кени [2].

Первые четыре метода на реальных объектах не обеспечивали необходимого качества результата. Объяснялось это недостаточностью контраста анализируемых изображений. При

тех же условиях применение оператора Кени эффективно решало задачу. Неудобство заключалось в том, что отличающиеся по яркости объекты требовали задания индивидуальных параметров (например, градиентные пороги). Это приводило к дополнительным трудностям при работе алгоритма в автоматическом режиме. Метод межкадровой разности обычно используется при динамическом эксперименте, когда один кадр отличается от предыдущего лишь участком с движущимся объектом. В случае «статике», когда объект неподвижно лежит перед системой регистрации, за первый кадр можно принять фон до момента помещения туда исследуемого объекта.

По результатам серии экспериментов установлено, что алгоритм с использованием межкадровой разности в совокупности с оператором Кени (рис. 1) обладает необходимым для решения поставленной задачи быстродействием, при этом гибок и почти не нуждается в индивидуальных параметрах при изменении объекта, что и определило выбор.



Рис. 1. Выделение контуров с помощью межкадровой разности в совокупности с оператором Кени

Для решения задачи выделения необходимого цвета использовалась интерактивная цветовая сегментация. Проведя сравнительный анализ таких модификаций известного метода Magic Wand, как пороговое выделение, выделение с использованием «чистых цветов», пороговое выделение с использованием радиуса в цветовом пространстве, было решено реализовать выделение цвета следующим образом (рис. 2). Оператор ставит метку на цветовой области, которую необходимо выделить. Определив координаты метки, алгоритм вычисляет среднее значение по всем трем каналам в области 5x5 вокруг данной метки. Далее каждый пиксель анализируемого изображения проверяется на выполнение следующего условия:

$$\sqrt{(r_{cp} - r)^2 + (g_{cp} - g)^2 + (b_{cp} - b)^2} > R^2,$$

где R – радиус окружности вокруг метки в цветовой системе RGB , а x_{cp} – цветовые координаты центра этой окружности.

В случае не удовлетворения условию пиксель обнуляется. Таким образом, на изображении остаются лишь пиксели, которые попадают в окружности с радиусом R в кубе цветовой системы RGB . Радиус, по умолчанию, равен 20, но оператор в случае необходимости может изменить это значение.

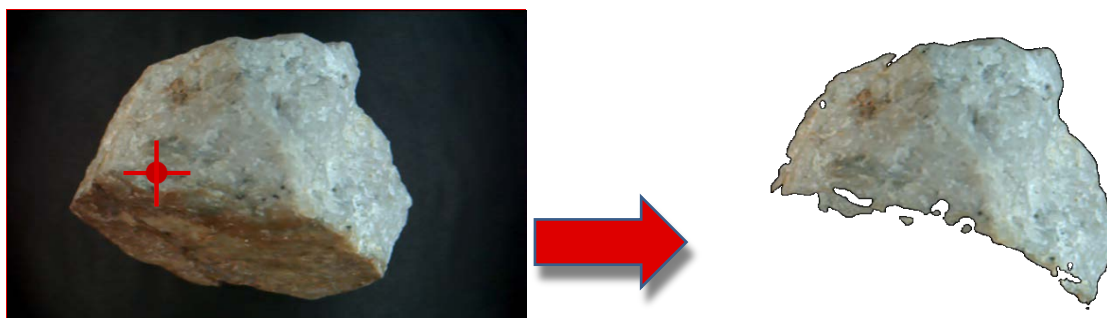


Рис. 2. Цветовая сегментация

Далее на основе описанных исследований основных этапов цветового анализа в видеоинформационных системах был разработан алгоритм и сформирована его программная оболочка.

Для проведения исследований процесса цветового анализа минеральных объектов была собрана экспериментальная установка (рис. 4). Узел освещения включает в себя два протяженных источника оптического излучения. В качестве приемника оптического излучения используется многоэлементная цветная КМОП-матрица, которая через USB-интерфейс связана с узлом управления.

Узел управления представляет собой персональный компьютер. Также в систему входит разработанный программный комплекс, реализующий алгоритм цветового анализа и управления настройками КМОП-матрицы.

Установка предназначена для анализа минеральных объектов в статике, то есть неподвижных объектов. Поэтому узлы освещения и регистрации закрепляются на вертикальной стойке с кронштейном (рис. 4), позволяющим менять их положение по высоте, тем самым реализуется возможность изменения дистанции от узла регистрации до объекта. Сам объект неподвижно лежит в зоне анализа – на площадке.

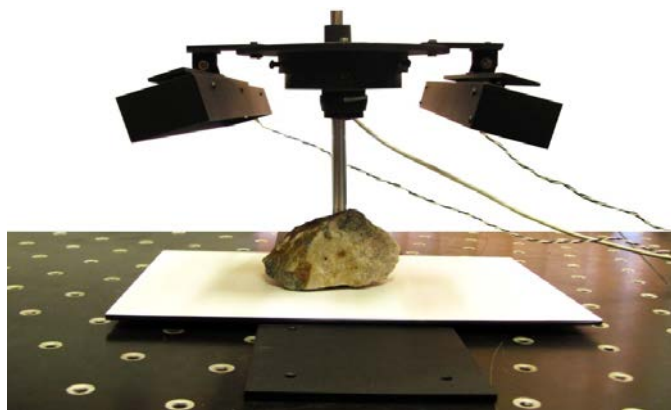


Рис. 4. Экспериментальная установка – вертикальная модификация

Исследование образцов руды происходит партиями, поступающими с одного месторождения. Для проведения первого лабораторного эксперимента использовались минеральные объекты золотосодержащей руды, представленные на рис. 5. Известно, что о содержании золота в образцах можно судить по наличию белых (кварц) и светло-серых (березит) оттенков. Пустая порода характеризуется наличием темно-серых оттенков (гранодиорит).



Рис. 5. Объекты золотосодержащей руды

Объекты были отобраны таким образом, чтобы каждая группа – концентрат, промежуточный продукт и хвосты – была представлена одинаковым числом минеральных

образцов. Визуально легко заметить, что процентное соотношение полезной части к неполезной уменьшается от группы к группе.

Анализ объектов проводится в двух режимах. Первый, «Режим диагностики», включает в себя процедуры, связанные с получением необходимых для сепарации параметров: цветовой диапазон, который необходимо выделить на объектах, допустимое процентное соотношение полезной части объекта к неполезной, и др. Получив от узла регистрации изображение объекта, оператору необходимо отнести каждый к одной из трех групп: «Концентрат», «Хвосты» и «Промежуточный продукт» (рис. 6). Последняя группа представляет собой граничные объекты, по которым можно судить о минимально возможном процентном соотношении полезной части к неполезной.

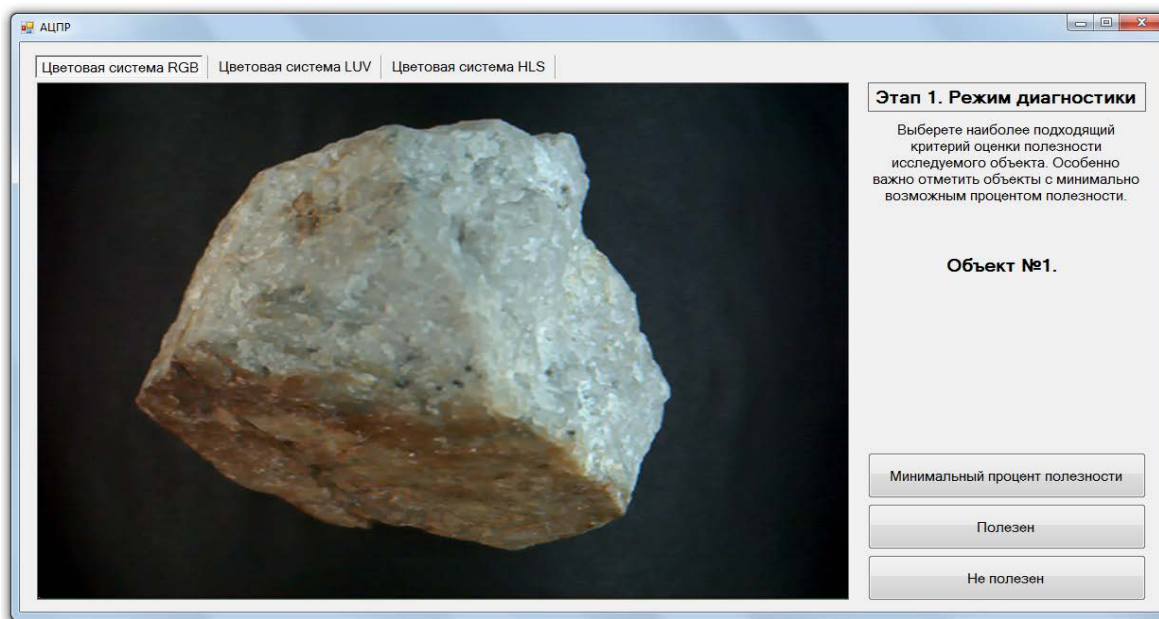


Рис. 6. Пользовательский интерфейс разработанного ПО

Следующий этап заключается в выделении оператором цветового диапазона полезной части объекта. Курсором мышки оператор ставит метку на необходимом участке на изображении объекта (рис. 7). Для более точного выделения интерфейс предоставляет также группу ползунков, передвижением которых можно подкорректировать выделенный участок. Для каждой из цветовых систем подбирается положение ползунков, при котором в выделенную область попадают все полезные оттенки.

Алгоритм отделяет анализируемый объект от фона и вычисляет процентное соотношение полезной части ко всей площади объекта. Далее усредняются полученные с помощью оператора параметры для каждой цветовой системы: выделенный цветовой диапазон, процентное соотношение полезной части к общей площади объекта и др.

Далее запускается режим проверки, в ходе которого алгоритм уже автоматически принимает решение о наличии/отсутствии полезной части на каждом объекте, основываясь на вычисленных ранее параметрах. По окончании работы алгоритма, пронаблюдав процент ложной классификации в каждой из цветовых систем, можно сделать вывод о возможности или невозможности сепарации минеральных объектов данного месторождения представленным алгоритмом. о окончании исследования одного месторождения программное обеспечение сохраняет информацию по объекту в структурированную базу данных (Табл. 2).

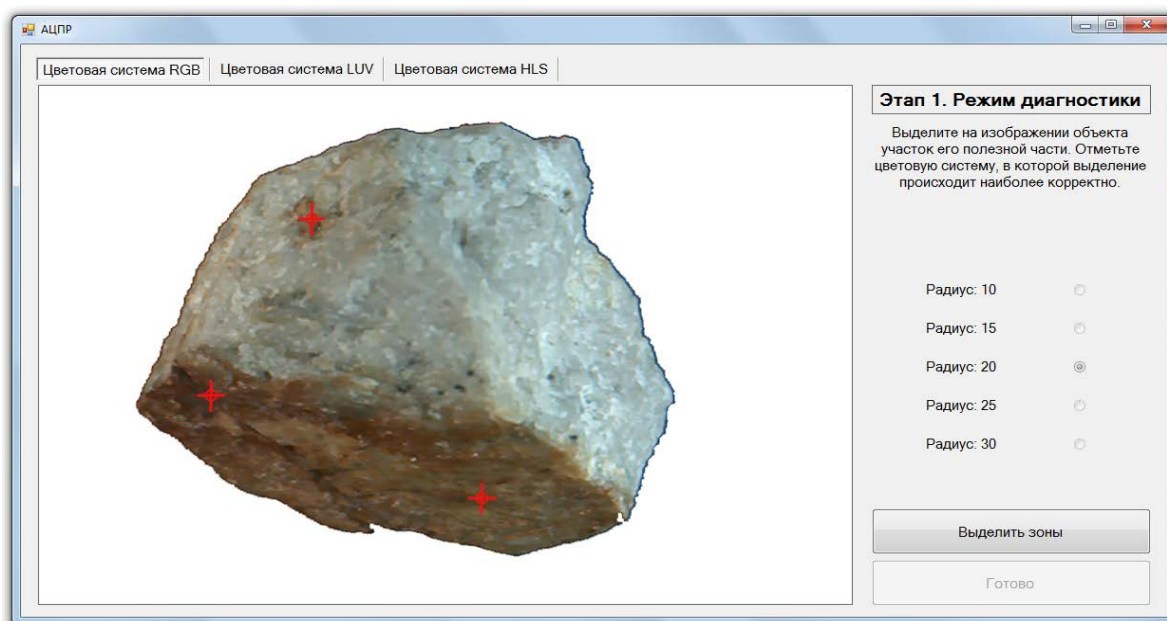


Рис. 7. Пример выделения оператором необходимого цвета

Таблица 2

Отчет об исследовании золотосодержащей руды

Месторождение	№ эксперимента	Полезность	Проценты	Режим
Зарница (Кварц, 15)	1	Граничный	39,5711652652619	Диагностика
Зарница (Кварц, 15)	2	Не полезен	10,8231643098872	Диагностика
Зарница (Кварц, 15)	3	Не полезен	0,8734566867879	Диагностика
Зарница (Кварц, 15)	4	нет	8,81656850649099	Проверка
Зарница (Кварц, 15)	5	Полезен	62,5280663602283	Проверка
Зарница (Кварц, 15)	6	Полезен	57,2689741513362	Проверка
Зарница (Кварц, 15)	7	нет	38,0052168768656	Проверка
Зарница (Кварц, 15)	8	нет	32,4490573884114	Проверка
Зарница (Кварц, 15)	9	Полезен	78,5718058649013	Проверка
Зарница (Кварц, 15)	10	Полезен	56,5442700621854	Проверка
Зарница (Кварц, 15)	11	нет	3,3219941632478	Проверка
Зарница (Кварц, 15)	12	нет	12,098226763144	Проверка
Зарница (Кварц, 15)	13	нет	16,5121879756741	Проверка
Зарница (Кварц, 15)	14	нет	22,2418063654505	Проверка
Зарница (Кварц, 15)	15	Полезен	40,9671130938434	Проверка

Анализ объектов в режиме проверки показал, что алгоритм справился с поставленной задачей, так как все пятнадцать минеральных объектов были отнесены в правильную группу. При этом порог полезности составил тридцать девять процентов. Следует отметить повторяемость результатов, получаемых при анализе разработанным программным обеспечением, что доказывается проведением данного эксперимента несколько раз.

Также была проанализирована партия медно-никелевой руды, представленная пятьюдесятью объектами и прошедшая предварительные испытания на реальном оптическом сепараторе (Табл. 3, где *RGB/HLS/LUV* – цветовые системы). Примеры объектов данного месторождения можно посмотреть на рис. 8.

Порог полезности рассчитывался по следующей формуле:

$$\text{Пор. полезности} = (G_{cp} - H_{cp})/2 - H_{cp}$$

где G_{cp} – среднее значение процентного отношения полезной части к неполезной граничных объектов, H_{cp} – среднее значение полезной части к неполезной неполезных объектов.

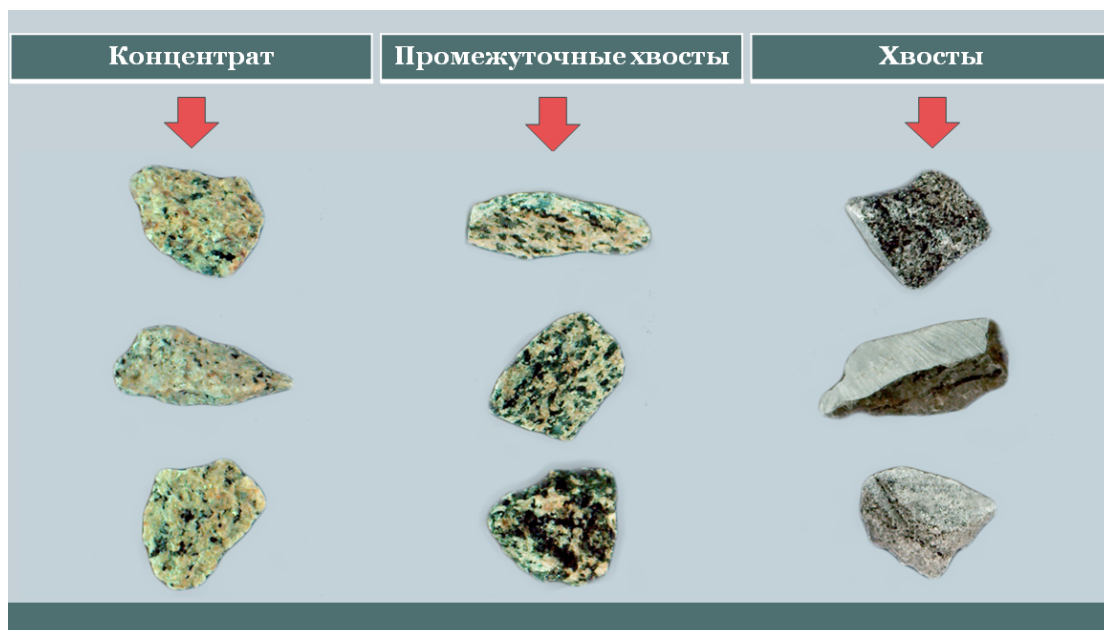


Рис. 8. Объекты медно-никелевой руды

Таблица 3
Отчет об исследовании медно-никелевой руды

Месторождение	№ эксп.	Полезность	RGB	HLS	LUV	Режим
Зарница	1	Полезен	77,70996248	83,0885625	72,58566978	Диагностика
Зарница	2	Граничный	43,62464183	48,6328284	42,89807613	Диагностика
Зарница	3	Полезен	69,26643032	68,6419416	61,65803108	Диагностика
Зарница	4	Полезен	73,48073942	74,1670356	63,99712198	Проверка
Зарница	5	Полезен	71,32527969	76,2378005	63,44025232	Проверка
Зарница	6	Нет	2,422749852	4,86654879	4,390839437	Проверка
Зарница	7	Граничный	37,55648891	40,1775745	37,31015294	Проверка
Зарница	8	Граничный	27,55341785	35,7544995	29,93616455	Проверка
Зарница	9	Граничный	32,99950747	39,8867181	32,84353964	Проверка
Зарница	10	Полезен	69,88572708	72,5520950	67,09985809	Проверка
Зарница	11	Полезен	76,38794084	82,8725824	71,46757679	Проверка
Зарница	12	Граничный	27,75485918	27,4057913	26,18802062	Проверка
Зарница	13	Нет	12,49980478	15,6294410	15,62631767	Проверка
Зарница	14	Граничный	27,94979079	30,8442037	35,72729510	Проверка
Зарница	15	Нет	5,126696092	5,51904528	5,751185221	Проверка
Зарница	16	Полезен	66,35072073	64,7163745	64,67489370	Проверка
Зарница	17	Полезен	61,78856973	70,3583061	59,96446550	Проверка
Зарница	18	Полезен	73,64823876	82,4200266	68,19670805	Проверка
Зарница	19	Нет	3,029069294	5,81793013	3,796514941	Проверка
Зарница	20	Нет	3,205729388	6,58197629	2,745332082	Проверка
Зарница	21	Граничный	31,28444652	33,2087809	32,08780943	Проверка
Зарница	22	Граничный	27,30771152	30,2563846	27,65755410	Проверка
Зарница	23	Нет	2,652932036	6,32445186	2,382459572	Проверка
Зарница	24	Граничный	28,28976034	31,8191721	29,70315904	Проверка

Месторождение	№ эксп.	Полезность	RGB	HLS	LUV	Режим
Зарница	25	Граничный	40,38833776	43,1224292	40,05968223	Проверка
Зарница	26	Граничный	23,51247600	29,8464491	22,82012613	Проверка
Зарница	27	Полезен	76,86542305	82,7115256	70,84166111	Проверка
Зарница	28	Нет	22,97228900	21,2893837	24,92724281	Проверка
Зарница	29	Нет	18,54314818	15,2066115	18,57005573	Проверка
Зарница	30	Нет	7,152583099	10,2923508	7,120544653	Проверка
Зарница	31	Граничный	17,72315653	20,2529826	18,30769967	Проверка
Зарница	32	Граничный	20,67469070	24,1632300	21,86391100	Проверка
Зарница	33	Полезен	59,29762097	70,9095322	60,12299724	Проверка
Зарница	34	Граничный	14,69597241	13,3717285	16,36107193	Проверка
Зарница	35	Граничный	36,36068764	42,9559952	35,86333657	Проверка
Зарница	36	Полезен	74,50280767	84,0255030	67,85212915	Проверка
Зарница	37	Полезен	75,27364076	83,6981876	70,45576888	Проверка
Зарница	38	Полезен	75,61165384	79,1475749	69,11040027	Проверка
Зарница	39	Полезен	69,33083096	71,0199326	66,29562516	Проверка
Зарница	40	Полезен	66,56336755	68,4356640	62,01820646	Проверка
Зарница	41	Полезен	71,92789766	75,7009949	67,55394753	Проверка
Зарница	42	Полезен	76,54520404	85,3817283	72,27421730	Проверка
Зарница	43	Полезен	46,36728421	53,0863143	45,77819516	Проверка
Зарница	44	Полезен	76,60583565	80,2247654	72,71883699	Проверка
Зарница	45	Полезен	59,14029736	69,3004193	57,51048417	Проверка
Зарница	46	Граничный	34,30433779	35,5301107	36,89720812	Проверка
Зарница	47	Граничный	45,23001740	48,6772021	44,29658797	Проверка
Зарница	48	Полезен	40,70914014	34,4862821	51,36667858	Проверка
Зарница	49	Полезен	31,68015848	33,9686443	34,52881101	Проверка
Зарница	50	Полезен	38,11994958	44,6067165	43,83942471	Проверка

Анализируя представленный отчет, можно увидеть, что три объекта классифицированы в неправильную группу, следовательно, погрешность классификации не превышает 1,5 %.

Таким образом, проведенные экспериментальные исследования показали, что разработанное программное обеспечение эффективно справляется с поставленной задачей разделения минеральных объектов на группы полезности. Важно отметить повторяемость результатов, получаемых при анализе, что доказывается при нескольких повторениях экспериментов.

Исследование медно-никелевой руды показало, что анализ слабоконтрастных минералов реализуется с некоторой погрешностью. В конкретном эксперименте погрешность классификации объектов не превышает 1,5 %, что укладывается в рамки допустимых отклонений для обогащательного оборудования.

Области возможного использования

Основная область применения результатов проекта – оптико-электронные приборы и комплексы цветового анализа промышленного назначения, в том числе автоматизированные, например, для:

- горнодобывающей промышленности – создание высоко-производительных комплексов сортировки полезных ископаемых, селективируемых на основе анализа совокупности физических параметров;
- перерабатывающей промышленности – сортировка отходов;

– пищевой и фармацевтической отраслей промышленности – контроль качества продукции.

Публикации

1. *Петухова Д.Б.* Алгоритм принятия решения системы технического зрения, основанный на цветовом анализе // Прикладная оптика – 2012: Сб. докладов. СПб., 2012. С. 267–269.
2. *Горбунова Е.В., Петухова Д.Б.* Особенности алгоритма цветового анализа минералов // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 5 (81). С. 156.
3. *Горбунова Е.В., Петухова Д.Б., Чертов А.Н.* Некоторые аспекты автоматической оценки применимости оптического метода для обогащения руд твердых полезных ископаемых // Научно-технический журнал «Приборостроение» (в печати).
4. *Петухова Д.Б.* Особенности идентификации и селекции объектов на изображении на основании анализа их цветовых характеристик // Сб. тезисов докладов I Всероссийского конгресса молодых ученых «КМУ – 2012». Вып. 2: Труды молодых ученых. СПб., 2012. С. 108.
5. *Петухова Д.Б.* Формирование и предварительная обработка цветного изображения в системах технического зрения специального назначения // Актуальные теоретические и практические вопросы современного оптико-электронного приборостроения: Сб. трудов молодых ученых. СПб., 2012. С. 49–50.
6. *Alekhin A.A., Chertov A.N., Gorbunova E.V., Petuhova D.B.* Optical-electronic system for express analysis of mineral raw materials dressability by color sorting method // Proceedings of SPIE. 2013. Vol. 8791. Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection, 87911N (May 24, 2013); doi: 10.1117/12.2020460.
7. *Алехин А.А., Горбунова Е.В., Коротаев В.В., Ольховский А.М., Петухова Д.Б., Чертов А.Н.* Возможности применения оптико-электронной системы для экспресс-анализа руд твердых полезных ископаемых оптическим методом // Научно-технический журнал «Приборостроение» (в печати).

Подана заявка на регистрацию разработанной программы «Цветовой анализ минеральных объектов».

Литература

1. *Alekhin A.A., Chertov A.N., Gorbunova E.V., Petuhova D.B.* Optical-electronic system for express analysis of mineral raw materials dressability by color sorting method // Proceedings of SPIE. 2013. Vol. 8791. Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection, 87911N (May 24, 2013); doi: 10.1117/12.2020460.
2. *Петухова Д.Б.* Алгоритм принятия решения системы технического зрения, основанный на цветовом анализе // Прикладная оптика – 2012: Сб. докладов. СПб., 2012. С. 267–269.



Подкидышева Валерия Евгеньевна
Факультет компьютерных технологий и управления
Кафедра безопасных информационных технологий
E-mail: v.scherzo@gmail.com

РАЗРАБОТКА ПРОЕКТА ГЛОССАРИЯ ПО ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ДЛЯ ПАРЛАМЕНТСКОЙ АССАМБЛЕИ ОРГАНИЗАЦИИ ДОГОВОРА О КОЛЛЕКТИВНОЙ БЕЗОПАСНОСТИ

В.Е. Подкидышева

Научный руководитель – доцент М.А. Вус

Введение

Научно-технический прогресс, достижения в сфере информатики, развитие средств информационно-вычислительной техники привели к росту роли и значения информационных ресурсов, повышению актуальности и усложнению задач защиты информации. Глобализация информационных процессов повлекла за собой возникновение целого спектра проблем информационной безопасности, а само это понятие превратилось из узкотехнического в общеупотребительное. Исследование проблем обеспечения информационной безопасности не только на национальном, но и на региональном и международном уровнях, сейчас очень актуально.

Одним из важных приоритетных направлений сотрудничества государств в рамках международных организаций является скоординированная информационная политика. Обеспечение совместных усилий по созданию системы коллективной безопасности позволяет более эффективно оказывать противодействие современным вызовам и угрозам. В контексте глобальных тенденций развития экономических отношений и социальных процессов сотрудничество государств сопровождается расширением соответствующей терминологии. При этом терминологическое многообразие и слабая определенность используемого в различных документах понятийного аппарата становится все более актуальной проблемой.

Впервые проблема международной информационной безопасности была поднята Российской Федерацией в рамках Генеральной Ассамблеи ООН в 1999 г., при этом были обозначены три основных взаимосвязанных вида угроз военно-политического, криминального и террористического характера. Вопросы терминологии для сферы международной информационной безопасности на глобальном уровне впервые стали темой научной проработки для Международного исследовательского консорциума по информационной безопасности, учрежденного в Гармиш-Паркенкирхен в 2009 г. Выработка общепризнанной терминологии должна позволить разговаривать «на одном языке». Подписание правительствами государств-участников Шанхайской организации сотрудничества (Екатеринбург, 2009 г.) соглашения стало первым практическим шагом на пути формирования региональных систем обеспечения международной информационной безопасности. В том же году было заключено российско-бразильское межправительственное соглашение о сотрудничестве в сфере международной информационной безопасности. В этих документах нашли отражение и вопросы понятийного аппарата.

После распада СССР в постсоветском пространстве возникло Содружество Независимых Государств (СНГ). В качестве одного из основных направлений сотрудничества страны содружества рассматривают сближение и гармонизацию ключевых положений законодательных актов, в том числе в информационной сфере.

В 2012 г. Межпарламентская ассамблея государств-участников СНГ приняла Рекомендации по совершенствованию и гармонизации национального законодательства государств-участников СНГ в сфере обеспечения информационной безопасности. В этом документе гармонизацию законодательства предложено рассматривать на трех уровнях:

- концептуально-научном (в рамках выработки концепций, рекомендаций, доктрин информационной безопасности);
- международном и региональном (разработка модельного законодательства, подготовка и заключение межгосударственных договоров и соглашений);
- национальном (развитие и совершенствование национального законодательства с учетом общих концептуальных подходов, приведение его в соответствие с модельным законодательством Межпарламентской ассамблеи (МПА) СНГ, ратификация достигнутых соглашений).

Предложенный разработчиками рекомендаций подход к решению проблем правового регулирования в области обеспечения информационной безопасности будет также способствовать развитию сотрудничества государств-участников СНГ по противодействию другим угрозам [1].

Исключительно важным становится сегодня вопрос проработки и однозначного толкования правовых дефиниций. Законодатель зачастую вместо создания целостной структуры понятийного аппарата выхватывает его отдельные элементы (понятия), которым дает определения в разных законах чаще с пометкой «для целей настоящего закона», что по логике исключает использование такой дефиниции относительно других правовых актов. Отдельные понятия нередко употребляются в прагматическом аспекте, легального определения в модельных актах им не дается. Как следствие, в настоящее время в модельном законодательстве, насчитывающем более 300 законодательных актов, сфера информационной безопасности представлена довольно фрагментарно, включая и принятый в ноябре 2012 г. Модельный информационный кодекс для государств-участников СНГ. В свете сказанного актуальной задачей является не просто систематизация выборки понятийного аппарата международного законодательства, но и его дополнение, согласование и гармонизация. Разработчиками названных рекомендаций было предложено рассматривать унификацию понятийно-категориального аппарата как одно из приоритетных направлений совершенствования законодательства и правоприменения [2].

В 2012 г. был выпущен словарь-справочник терминов и определений понятий модельного законодательства государств-участников СНГ, содержащий порядка 3 тыс. дефиниций [3]. Работа над ним наглядно показала необходимость в тщательной проработке понятийного аппарата.

Организация договора о коллективной безопасности (ОДКБ) – международная региональная организация военно-технического сотрудничества. Ее целями являются укрепление мира, международной и региональной безопасности и стабильности, защита на коллективной основе независимости, территориальной целостности и суверенитета государств-членов, приоритет в достижении которых отдается политическим средствам.

Для реализации уставных целей государства-члены ОДКБ принимают совместные меры к формированию в рамках ОДКБ действенной системы коллективной безопасности. В соответствии с уставом ОДКБ они координируют и объединяют свои усилия в борьбе с угрозами безопасности, принимают меры по развитию договорно-правовой базы, гармонизации национального законодательства по вопросам обороны, военного строительства и безопасности. Сегодня, как заявил Генеральный секретарь ОДКБ, в свете организованного в киберпространстве негативного воздействия на наши государства и на сознание людей «информационная составляющая становится все более важной областью деятельности ОДКБ» [4].

Вопросы информационной безопасности для ОДКБ как для организации военно-технического сотрудничества имеют свою специфику. Процессы глобализации в экономике и политике, расширяющееся повсеместное использование информационно-коммуникационных технологий, развитие культурного диалога отражается на регулирующих инструментах. При этом в области юрисдикции государств-участников ОДКБ как во внутригосударственной сфере, так и в международном пространстве большую роль играет правильный выбор конкретных значений применяемых терминов и понятий, их точное толкование и применение в различных договорах, соглашениях и иных актах, прежде всего, в актах имеющих

политическое, военное, экономическое и юридическое значение. Вышеизложенное свидетельствует в пользу актуальности исследования вопросов понятийного аппарата для сферы информационной безопасности в рамках ОДКБ.

Проект является этапом большого объема работ по систематизации понятийного аппарата международного законодательства в области информационной безопасности в интересах ОДКБ. Его цель – разработка макета глоссария по информационной безопасности для Парламентской ассамблеи ОДКБ (ПА ОДКБ).

Исследование понятийного аппарата правового поля информационной безопасности

Проект направлен на систематизацию, упорядочение и гармонизацию понятийного аппарата в сфере информационной безопасности, его легализацию и легитимизацию в рамках ОДКБ. Динамика информационных процессов, развитие научных знаний требовали как уточнения перечня понятий, так и отражения новых признаков в их содержании. В связи с этим возникла необходимость системного сравнительно-сопоставительного исследования наиболее значимых терминов и понятий. Языковой материал для такого исследования составили лексические единицы из национальных и международных нормативных документов, регулирующих сферу информационной безопасности в рамках СНГ, ОДКБ, Таможенного союза, ЕврАзЭС и др. В силу того обстоятельства, что в сфере информатики и информационной безопасности (прежде всего, вследствие технологического лидерства западных стран) широкоупотребительной является англоязычная терминология, необходимо провести также сравнение с английской лексикой родственной тематической группы.

В процессе выполнения работы был проанализирован ряд законодательств международных организаций (СНГ, ОДКБ, ШОС) и сформирована выборка из понятийного аппарата основополагающих документов в области информационной безопасности национальных законодательств государств-участников этих организаций, как то: федеральный закон Российской Федерации от 27.07.2006 №149-ФЗ «Об информации, информатизации и защите информации», закон Республики Беларусь от 10.11.2008 № 455-З «Об информации, информатизации и защите информации», закон Республики Казахстан от 11.01.2007 № 217-III «Об информатизации» и др.

Анализ полученной выборки показал, что на основе существующей легальной терминологической базы каждого из государств довольно затруднительно составить единый для всех государств-участников ОДКБ понятийный аппарат. Главной причиной являются расхождения в толкованиях основополагающих понятий области, а иногда и в различной формулировке самих понятий. Например, понятие «документ» («документированная информация») или понятие «информационная система» встречаются в большинстве национальных законодательств приблизительно в одинаковой формулировке, а такие понятия, как «конфиденциальная информация» или «информационные ресурсы», либо не определены вообще, либо имеют существенные расхождения в определениях. Составление адекватной выборки в немалой степени затруднено отсутствием официальных юридически значимых переводов некоторых национальных законодательных актов на русский язык (рабочий язык ПА ОДКБ).

Также сформирована более узкая выборка из понятийного аппарата российского законодательства, в которую вошли термины не только из федеральных законов, но также из подзаконных актов, государственных стандартов и руководящих документов. В ее состав включено более трехсот понятий, имеющих англоязычные аналоги. Большая часть терминов определена в государственных стандартах, но не в федеральных законах. Кроме того, сравнительный анализ данной выборки и Словаря-справочника терминов и определений понятий модельного законодательства государств-участников СНГ показал, что подавляющая часть терминов из исследуемой выборки не определена в модельном законодательстве МПА СНГ, хотя при осуществлении деятельности по обеспечению информационной безопасности они используются достаточно часто.

На основе вышеизложенного был сделан вывод, что формирование согласованного унифицированного легального понятийного аппарата на основе существующей терминологической базы национальных законодательств весьма затруднительно.

Разработка и анализ макета глоссария

Формирование понятийного аппарата на основе системного подхода

Сложность разработки глоссария заключается в необходимости создать фактически универсальную лексическую систему, включающую при всем многообразии лексики тематической группы «информационная безопасность» совокупность самых необходимых законодателю правовых терминов. Универсальность глоссария должна сочетаться с минимизированностью экспликации языкового материала, чтобы сделать этот документ удобным для практического использования парламентариями. В качестве методологической основы использовался системный подход к формированию понятийного аппарата [5]. Согласно этому, основные понятия и термины информационной безопасности могут быть объединены в три группы:

- термины, определяющие концептуальную научную основу информационной безопасности;
- термины, определяющие предметную область информационной безопасности;
- термины, определяющие основы деятельности по обеспечению информационной безопасности.

Термины внутри каждой группы должны быть упорядочены не в алфавитном порядке, а согласно принципу, в соответствии с которым определение предыдущего термина должно способствовать пониманию каждого последующего. Полученный список терминов и определений образует версию глоссария, представленную в Таблице.

В первую группу вошли устойчивые, семантически унифицированные, стилистически нейтральные понятия: «информация», «коммуникация», «угроза», «опасность» и пр. Определения данного уровня являются основополагающими и впоследствии могут быть использованы для формирования понятийного аппарата в любой области знаний. Однозначное определение таких терминов позволит избежать дальнейших разночтений в их значениях. Например, в модельном законодательстве МПА СНГ имеется несколько терминов, отражающих понятие «безопасность» («экологическая безопасность», «технологическая безопасность», «безопасность туризма» и пр.), и толкуются они по-разному: безопасность как процесс, безопасность как состояние защищенности, безопасность как комплекс мер. Подобные разногласия обусловлены как раз тем, что понятие безопасности не имеет однозначного толкования в правовом поле модельного законодательства.

Во вторую группу вошли термины, обозначающие понятия и их структурные связи с другими понятиями исследуемой специфической области знаний: «информационная система», «информационный поток», «информационный процесс» и пр. По сути, термины данной группы являются проекцией основополагающих терминов, выделенных ранее, в области информационной безопасности.

Третью группу составил широкий круг понятий различного уровня, связанных непосредственно с деятельностью по обеспечению информационной безопасности: «информационная защищенность», «система защиты информации», «средства защиты информации» и пр. При рассмотрении понятия информационной безопасности, кроме технического аспекта, необходимо учитывать также социокультурный и другие, особенно если речь идет о международной информационной безопасности [6]. На данном этапе работы рассматриваются преимущественно понятия, определяющие деятельность по защите информации, так как эта сфера является наиболее изученной. Впоследствии данную группу терминов планируется дополнить.

Сформированная выборка терминов сопоставлена с полученными на предыдущем этапе работы выборками и с модельным законодательством МПА СНГ. Как оказалось, по большей части эти понятия в нем не определены. Даже Модельный информационный кодекс для государств-участников СНГ от 23.11.2012, призванный систематизировать правовые нормы

и используемый в них понятийный аппарат, не позволил продвинуться в этой области (этот кодекс практически не содержит определений по тематике информационной безопасности).

Анализ рабочей версии глоссария

С целью структуризации выделенного понятийного аппарата были использованы методы системного анализа. Для установления связей между понятиями использовался алгоритм построения сети знаний профессора В.Я. Розенберга для составления семантической сети [7]. В качестве элемента системы знаний рассматривается понятие, состоящее из названия, идентификатора и описания его содержания. В нашем случае понятия идентифицируются в соответствии с Таблицей. Каждое из понятий системы раскрывается с помощью других понятий. Суть алгоритма заключается в выявлении связей между ними и построении иерархической сети. На первом (нижнем) уровне располагаются понятия, в определении которых не используются другие понятия группы, на следующем – понятия, в определении которых задействованы понятия первого уровня, и так далее. На верхнем уровне располагаются понятия, которые не применяются в определении других понятий.

В целях упрощения процедуры проведения анализа данный алгоритм был применен не для всей совокупности определений, а для каждой из трех групп в отдельности. Следует отметить, что связи выявлялись строго в соответствии с алгоритмом, то есть только в том случае, если понятие более низкого уровня использовалось в определении понятия более высокого уровня. Например, понятие «объект обеспечения информационной безопасности» не считался связанным с понятием «информационная безопасность», так как в определении (описании содержания понятия) данный термин не используется. Это значит, что если мы исключим из системы понятие «информационная безопасность», понятие «объект обеспечения информационной безопасности» останется определенным.

В первую группу входят основополагающие понятия, и связь здесь почти линейная, как видно из рис. 1. Понятия располагаются на пяти уровнях.

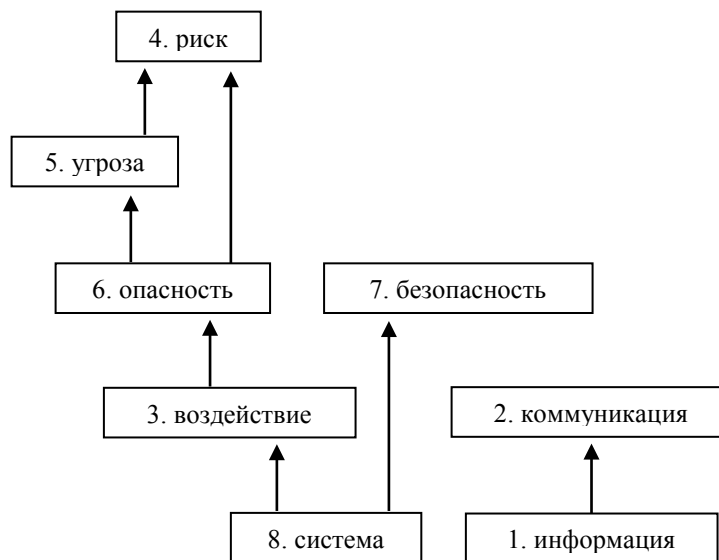


Рис. 1. Связи понятий первой группы



Рис. 2. Не связанные между собой понятия второй группы

Связи понятий второй группы уже более сложные. На рис. 2 отображены понятия, не относящиеся напрямую к другим понятиям группы, только к понятию «информация» из первой группы. На рис. 3 отображены остальные, связанные между собой (и также с понятием «информация») понятия на пяти уровнях.

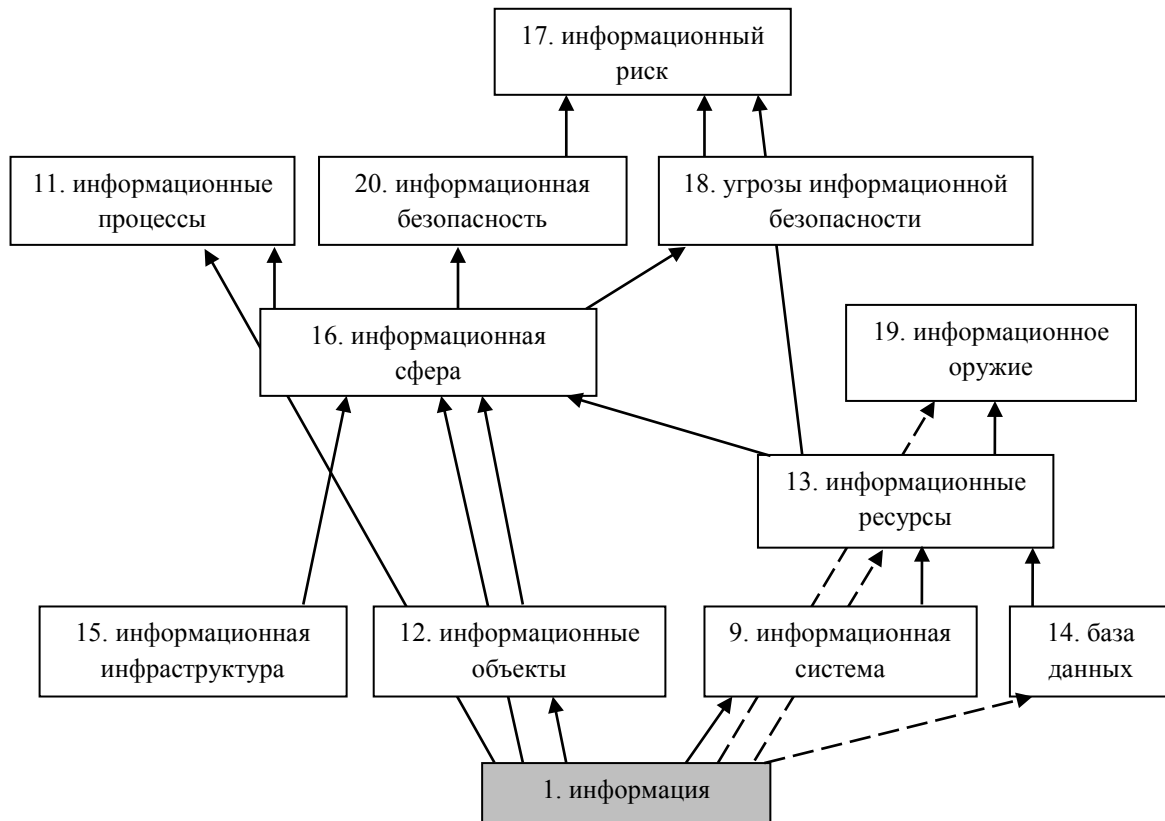


Рис. 3. Связи понятий второй группы

В процессе структурирования выяснилось, что зачастую в понятие высшего уровня входит не собственно термин, определяющий понятие, а его содержание (определение). Например, есть такое определение: «информационный поток – генерируемая совокупность сведений (сообщений, данных) независимо от формы их представления, циркулирующих в системе, а также между системой и внешней средой, предназначенный для реализации управляющих функций». При этом часть определения «генерируемая совокупность сведений (сообщений, данных) независимо от формы их представления» есть содержание понятия «информация», то есть связь между понятиями присутствует. Но сам термин «информация» в определении не используется. К тому же в нашей таблице обозначено более подробное определение информации, значит, строго говоря, связи здесь нет. Таких нестрогих связей, при которых в определение понятия транслируется содержание (частично или полностью) других понятий, достаточно много. Часть из них отображена на рис. 2 и 3 в виде пунктирных линий.

В третьей группе получилось всего три уровня, как видно из рис. 4а, 4б. Также эти понятия связаны с предыдущими группами (блоки серого цвета на рисунках).

Здесь также имеются нестрогие связи, при которых содержание понятия низшего уровня частично транслируется в определение понятия высшего уровня. В некоторых из таких случаев сложно определить, связаны ли рассматриваемые понятия вообще. Например, «объект защиты информации – носитель информации или информационный процесс, в отношении которых необходимо обеспечить защиту в соответствии с поставленной целью». Требуется уточнение употребленного здесь понятие «защита»: подразумевается «защита информации» или иной вид защиты. Другой пример: «безопасность информации – состояние защищенности информации, в том числе обрабатываемой средствами вычислительной техники, от угроз деструктивного воздействия». Здесь используется понятие «защищенность информации», то есть состояние сведений (данных), а среди наших определений есть «информационная защищенность» как состояние элемента информационной системы, то есть элемента совокупности сведений.

Вопрос, можно ли в данном случае заменить «защищенность информации» на «информационную защищенность» или это разные понятия, требует уточнения. Подобные связи также частично отмечены на рисунках пунктирной линией.

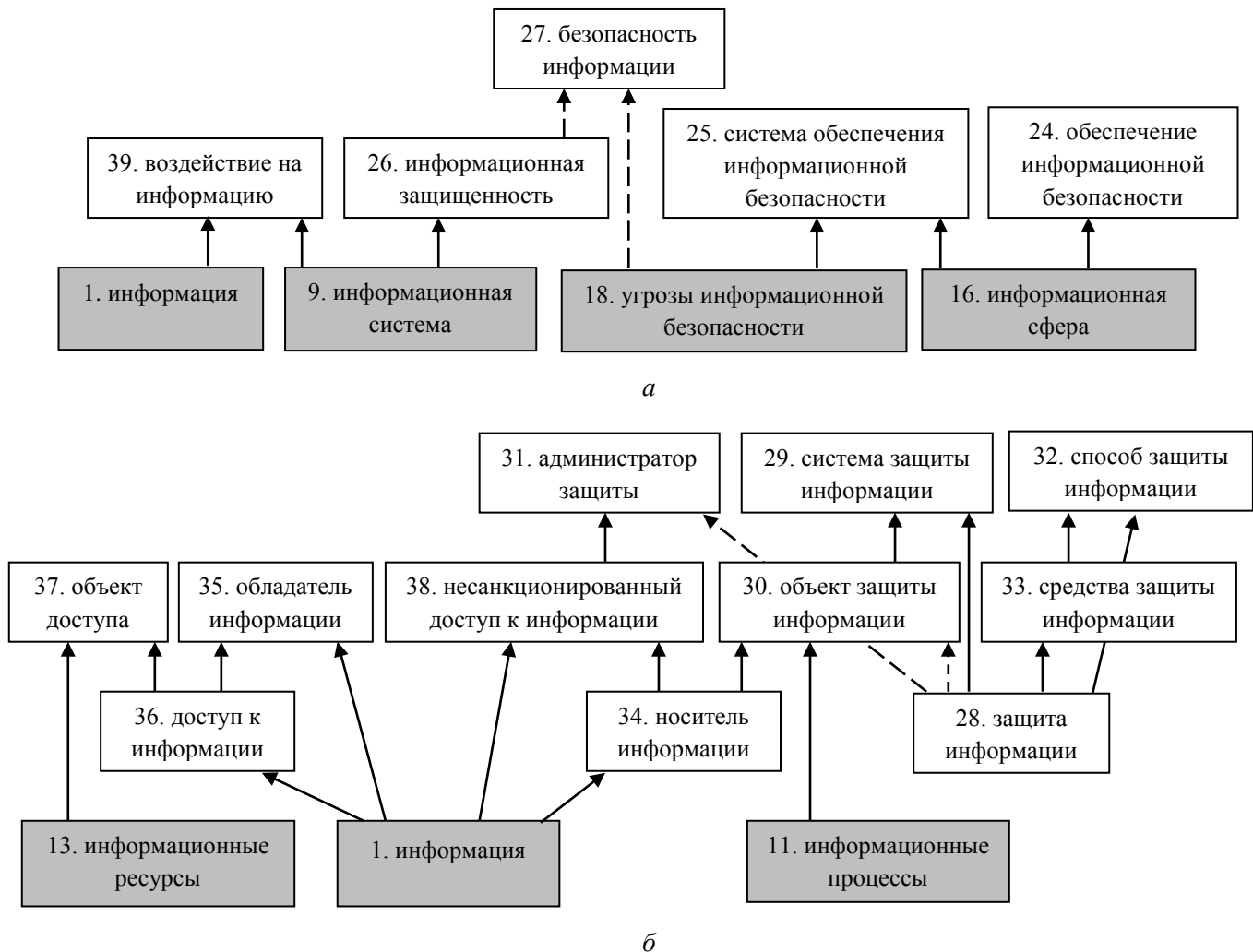


Рис. 4. Связи понятий третьей группы

Возможные направления совершенствования глоссария

В целом сформированная система терминов является достаточно прозрачной, термины не имеют циклических ссылок друг на друга. Слабую связанность некоторых понятий можно усилить путем уточнения их определений.

Проведенная систематизация показала необходимость расширения и дополнения первоначального набора понятий. Представляется целесообразным добавить такие понятия, как «защищенность», «ресурс», «информационная среда», «информационные технологии» и др. Необходимо также учитывать, что сфера информационной безопасности динамично развивается, и в правовом поле регулярно появляются новые нормативно-правовые акты, на которые следует ориентироваться при уточнении и дополнении понятий. Также следует обращать внимание на документы министерств и ведомств, посвященные проблематике информационной безопасности. Например, только в Российской Федерации в 2011 г. появился документ Министерства обороны РФ «Концептуальные взгляды на деятельность вооруженных сил Российской Федерации в информационном пространстве», в 2012 г. опубликован документ Совета безопасности РФ «Основные направления государственной политики в области обеспечения безопасности автоматизированных систем управления производственными и технологическими процессами критически важных объектов инфраструктуры Российской Федерации», а в 2013 г. – проект документа «Основы государственной политики Российской Федерации в области международной информационной безопасности на период до 2020 года».

Заключение

В результате проведенного исследования:

- сформированы выборки легального и легитимного понятийного аппарата из международных и национальных законодательных актов. По результатам их анализа сделан вывод, что этот набор понятий не вполне адекватен задачам формирования глоссария по информационной безопасности для ПА ОДКБ;

- с использованием системного подхода и с опорой на доступные нормативные материалы и литературные источники была сформирована первая рабочая версия макета Глоссария по информационной безопасности, которая будет положена в основу дальнейшей работы по совершенствованию разрабатываемой понятийной базы;

- результаты анализа и систематизации полученной лексической базы позволили выработать предложения по уточнению отдельных формулировок определяемых понятий в целях расширения и совершенствования лексической базы;

- полученные материалы будут использованы в качестве основы для аналитической записки, готовящейся к рассмотрению на очередном заседании экспертно-консультативного совета при ПА ОДКБ.

Имеющиеся на данный момент словарные материалы и документы, в которых собраны понятия и определения в сфере информационной безопасности (включая документы ОДКБ), по большей части представляют собой слабо систематизированную выборку понятий, не имеют четкой структуры, а потому крайне неудобны в использовании. Слабая доступность таких материалов, ограниченный тираж печатных версий, отсутствие электронного представления затрудняют работу с ними. Подготовленная на данном этапе работы версия глоссария (1.0) отличается системностью и структурированностью, а также представлена в электронной версии. Это позволит дополнять и совершенствовать имеющуюся систему терминов. Требуют дополнительной проработки вопросы сопоставления полученных понятий с их англоязычными версиями.

Основное использование результатов проекта предполагается в законодательной, правотворческой и правоприменительной деятельности МПА СНГ, ПА ОДКБ и других международных организаций. Единые лексические формы, согласованные в рамках ПА ОДКБ понятия и их определения в переводе на национальные языки послужат целям гармонизации межгосударственных отношений. Выработанный подход к формированию понятийного аппарата может быть использован в работе по совершенствованию как международного, так и национального законодательства в различных предметных областях. База легальных понятий и определений для области информационной безопасности может быть использована также в научно-исследовательской деятельности специалистов, а также для обучения студентов.

Таблица

Версия 1.0 макета глоссария по информационной безопасности для ОДКБ

1. Термины, определяющие концептуальную научную основу информационной безопасности

№п/п	Понятие (термин)	Содержание понятия (определение)
1	Информация	результат (следы) взаимодействия объектов (предметов, средств) и явлений (процессов) материального мира, воспринимаемых человеком или техническими устройствами в виде сведений (данных) в различной форме представления и служащие ему источником получения и передачи знаний
2	Коммуникация	информационный обмен, заключающийся в процессе передачи информации между социальным субъектом, называемым коммуникатором, и социальным объектом – коммуникантом, или группами этих социальных объектов
3	Воздействие	возбуждение (или торможение) в управляемой системе таких процессов, которые стимулируют желательный для управляющей стороны выбор

№п/п	Понятие (термин)	Содержание понятия (определение)
4	Риск	угроза или опасность потенциально возможной (вероятностной) потери части ресурсов или деятельность, связанная с преодолением неопределенности в ситуации выбора варианта достижения цели
5	Угроза	потенциальная возможность возникновения совокупности факторов и условий, создающих опасность нанесения ущерба жизненно важным интересам человека, общества и государства
6	Опасность	следствие действия деструктивных (вредных и опасных) факторов и условий на определенный объект (предмет, средство) воздействия
7	Безопасность	состояние (положение) объекта (предмета, средства, процесса) при котором его характеристики (свойства, способности) и условия деятельности (протекания, проявления) сводят к минимуму эффективность деструктивного воздействия за счет наличия и успешного использования системы мероприятий, обеспечивающих защищенность от известных средств и методов подобного воздействия
8	Система	нечто целое, представляющее собой единство закономерно расположенных и находящихся во взаимной связи частей (элементов, предметов, явлений), объединение которых придает системе свойства, не присущие каждой из частей в отдельности

II. Термины, определяющие предметную область информационной безопасности

№п/п	Понятие (термин)	Содержание понятия (определение)
9	Информационная система	естественная (биологическая) или целевая (организационно-техническая) совокупность закономерно расположенных и находящихся во взаимной связи совокупность сведений об объекте (предмете, средстве, явлении, процессе), его деятельности (проявлении, протекании), а также используемые информационные технологии
10	Информационный поток	генерируемая совокупность сведений (сообщений, данных) независимо от формы их представления, циркулирующих в системе, а также между системой и внешней средой, предназначенная для реализации управляющих функций
11	Информационные процессы	генерирование (производство), добывание, переработка и передача информации субъектом и функционирование естественных и технических средств в информационной сфере
12	Информационные объекты	объекты материального мира, реализующие функции генерации (производства, воспроизводства), добывания, переработки и передачи информации
13	Информационные ресурсы	сведения сами по себе и сведения в базах данных информационных систем
14	База данных	объективная форма представления и организации совокупности данных, систематизированных таким образом, чтобы эти данные могли быть найдены и обработаны с помощью ЭВМ
15	Информационная инфраструктура	совокупность систем, средств и технологий генерации, добывания, переработки и передачи информации, а также систем и средств, обеспечивающих защищенность информационных систем и информационного ресурса от угроз деструктивного воздействия

№п/п	Понятие (термин)	Содержание понятия (определение)
16	Информационная сфера	– (как пространство) совокупность информационного ресурса, информационной инфраструктуры, а также информационных объектов и субъектов, использующих информацию – (как деятельность) связана с функционированием естественных и технических объектов (средств, предметов) генерации, распространения и потребления информации
17	Информационный риск	– угроза информационной безопасности или информационная опасность потенциально возможной (вероятностной) потери части информационного ресурса – деятельность, связанная с преодолением неопределенности в обеспечении информационной безопасности
18	Угрозы информационной безопасности (информационные опасности)	возможность возникновения такой совокупности условий и факторов (ситуации, явления, события), следствием которой может стать нанесение ущерба жизненно важным интересам человека, общества и государства в информационной сфере
19	Информационное оружие	специальные устройства и средства, конструктивно предназначенные для воздействия на информационный ресурс, информационную среду, носители информации, информационные процессы, а также на субъекты, использующие информацию в своей деятельности
20	Информационная безопасность	состояние защищенности человека, общества и государства, их жизненно важных интересов от угроз деструктивного воздействия в информационной сфере
21	Объекты обеспечения информационной безопасности	объекты материального мира, реализующие в себе функции генерации, добычания, переработки, передачи информации и обладающие программами реализации этих функций
22	Информационный потенциал	возможность своевременно генерировать, добывать, перерабатывать и передавать заданное (необходимое) количество информации
23	Информационное превосходство	способность более эффективно добывать, перерабатывать и распределять непрерывный поток информации о складывающейся в реальном масштабе времени обстановке, препятствуя противоборствующей стороне делать то же самое

III. Термины, определяющие основы деятельности по обеспечению информационной безопасности

№п/п	Понятие (термин)	Содержание понятия (определение)
24	Обеспечение информационной безопасности	процесс, содержание которого составляет проведение единой государственной политики в виде системы мер организационно-технического, правового, социального и иного характера, адекватных угрозам жизненно важным интересам человека, общества и государства в информационной сфере
25	Система обеспечения информационной безопасности	целенаправленная совокупность взаимосвязанных организационно-технических, правовых и социальных норм и механизмов, направленных на противодействие угрозам информационной безопасности с целью сведения до минимума возможного ущерба в информационной сфере
26	Информационная защищенность	состояние элемента информационной системы, при котором обеспечены необходимые и достаточные условия его надежного, успешного и эффективного функционирования

№п/п	Понятие (термин)	Содержание понятия (определение)
27	Безопасность информации	состояние защищенности информации, в том числе обрабатываемой средствами вычислительной техники, от угроз деструктивного воздействия
28	Защита информации	процесс обеспечения сохранности, целостности и надежности сведений (данных)
29	Система защиты информации	целенаправленная совокупность органов и (или) исполнителей, используемой ими техники защиты информации, а также объектов защиты
30	Объект защиты информации	носитель информации или информационный процесс, в отношении которых необходимо обеспечить защиту в соответствии с поставленной целью
31	Администратор защиты	субъект, ответственный за защиту автоматизированной системы от несанкционированного доступа к информации
32	Способ защиты информации	обоснованная совокупность мероприятий и действий по защите информации, составляющих сущность защиты информации, а также определенный порядок применения средств защиты информации при решении конкретных задач
33	Средства защиты информации	технические, программные средства, а также системы управления комплексами средств, предназначенные и используемые для защиты информации
34	Носитель информации	любой материальный объект, в том числе физическое поле, в котором информация отображается в виде символов, образов, явлений или, например технических решений и процессов
35	Обладатель информации	физическое или юридическое лицо, общественная организация или государство, самостоятельно создавшие информацию, либо получившие на основании закона или договора право разрешать или ограничивать доступ к информации
36	Доступ к информации	возможность получения информации, в том числе и с целью ее дальнейшего использования
37	Объект доступа	единица информационного ресурса, доступ к которому регламентируется правилами разграничения доступа
38	Несанкционированный доступ к информации	процесс, нарушающий принятые или установленные кем-либо правила разграничения доступа, следствием которого могут быть утечка, искажение, разрушение, уничтожение, блокирование информации, а также утрата, уничтожение или сбой функционирования носителя информации
39	Воздействие на информацию	система мероприятий, оказывающих влияние на количественное или качественное (смысловое) содержание информации, циркулирующей в информационных системах, проводимых с целью преднамеренного искажения, задержки распространения, разрушения или уничтожения в процессе ее генерации, переработки или передачи

Публикации

1. *Сергеев А.И., Александрова О.А., Анцев Г.В., Бондуровский В.В., Вус М.А., Зикратов И.А., Коржук В.М., Кучерявый М.М., Подкидышева В.Е., Юсупов Р.М.* Словарь-справочник терминов и определений понятий модельного законодательства государств-участников СНГ. СПб., 2012. – 360 с.
2. *Вус М.А., Кучерявый М.М., Коржук В.М., Подкидышева В.Е., Шуклина В.А.* Информационное пространство СНГ: О правовом базисе и понятийном аппарате сотрудничества // Международная научно-практическая конференция «Право на доступ

к информации: Возможности и ограничения в электронной среде» (Президентская библиотека им. Б.Н. Ельцина, Санкт-Петербург, 13.04.2012). СПб., 2012.

3. Подкидышева В.Е. Медиа- и информационная грамотность // Юбилейная XIII Санкт-Петербургская международная конференция «Региональная информатика (РИ – 2012)» (Санкт-Петербург, 24–26 октября 2012 г.): Материалы конференции. СПб., 2012. С. 143.

4. Вус М.А., Кучерявый М.М., Подкидышева В.Е. Понятийный аппарат модельного законодательства Содружества Независимых Государств // Экономика, социология и право. 2012. № 12. С. 205–207.

5. Вус М.А., Подкидышева В.Е. В интересах укрепления национальной и международной информационной безопасности // Сб. трудов I Международной научно-практической конференции «Информационная безопасность в свете Стратегии Казахстан-2050» (12 сентября 2013 г., г. Астана). Астана, 2013. С. 200–206.

Литература

1. Бачило И.Л., Вус М.А., Кучерявый М.М., Макаров О.С. К вопросу о концептуальных направлениях организационно-правового обеспечения информационной безопасности государств-участников СНГ // Информатизация и связь. 2012. № 4. С. 4–11.

2. Бачило И.Л., Бондуrowsкий В.В., Вус М.А. и др. О совершенствовании и гармонизации национального законодательства государств-участников СНГ в сфере обеспечения информационной безопасности // Информационное право. 2013. № 1 (32). С. 24–27.

3. Сергеев А.И., Александрова О.А., Анцев Г.В. и др. Словарь-справочник терминов и определений понятий модельного законодательства государств-участников СНГ. СПб., 2012. – 360 с.

4. Бордюжа Н. События на Ближнем Востоке могут дать толчок для «конфликтной эпидемии» // Информационный бюллетень Секретариата ОДКБ. «Новости – Казахстан» (9 июля 2013 г.): www.odkb-csto.org.

5. Вус М.А., Кучерявый М.М., Шакин Д.Н., Юсупов Р.М. Эскиз системного подхода к формированию понятийного аппарата информационной безопасности // Информатика и связь. 2012. № 9. С. 7–15.

6. Сб. трудов I Международной научно-практической конференции «Информационная безопасность в свете Стратегии Казахстан-2050» (12 сентября 2013 г., г. Астана). Астана, 2013. – 639 с.

7. Федоров Д.Ю. Применение структуризации знаний в информационной безопасности // Материалы Международной научно-практической конференции «Фундаментальные и прикладные исследования в современном мире» (13–15 марта 2013 г.). СПб., 2013. С. 40–43.



Посмитная Яна Станиславовна
Факультет точной механики и технологий
Кафедра нанотехнологий и материаловедения
E-mail: posmitnaya@rambler.ru

РАЗРАБОТКА ГИБРИДНОГО МИКРОФЛЮИДНОГО ЧИПА ДЛЯ ПОЛИМЕРАЗНОЙ ЦЕПНОЙ РЕАКЦИИ

Я.С. Посмитная

Научный руководитель – доктор технических наук А.А. Евстрапов
(Институт аналитического приборостроения РАН)

Введение

По данным аналитического отчета Yole Development [1], более 200 компаний по всему миру задействовано в производстве микрофлюидных устройств (МФУ).

МФУ являются основой систем «лаборатория на чипе», ориентированных на применение в медицине, биотехнологии, криминалистике, космических исследованиях и военных технологиях.

Суммарная стоимость материалов, элементов устройства и технологий формирования функциональных структур определяет стоимость устройства в целом. Выбор материала предполагает и определение подходящего технологического процесса с учетом требуемых точностных характеристик получаемых структур и экономических затрат при организации необходимого объема производства устройств. В последнее время при создании МФУ наблюдается значительное увеличение доли используемых полимерных материалов.

Планарным вариантом МФУ является микрофлюидный чип (МФЧ). Прототипы таких чипов в настоящее время создаются на базе Института аналитического приборостроения РАН [2] для анализа биологических проб, в частности для методов молекулярной диагностики, основанных на полимеразной цепной реакции (ПЦР). Для того чтобы реализовать данный метод, требуется проведение комплексных исследований и работ, связанных с:

- выбором материала микрочипа;
- определением метода изготовления функциональных структур в микрочипе;
- выбором способов обработки рабочей поверхности;
- особенностями загрузки реакционной смеси;
- изучением влияния свойств материала чипа на ПЦР;
- изучением способов и методов эффективной герметизации микрочипа и т. д.

Целью данного проекта являлась разработка и изготовление прототипа гибридного (стеклянно-полимерного) МФЧ для ПЦР в жидкой среде. Была предложена топология и конструкция чипа, разработана и изготовлена мастер-форма для получения структур в полимере методом «мягкой» литографии, исследованы методы обработки поверхностей материалов и способы их герметизации, изготовлены и апробированы прототипы МФЧ на модельных объектах.

При изготовлении чипов применялись материалы: стекло марки Крон 8 (ГОСТ 3514-76, Россия), полидиметилсилоксан марки Sylgard® 184 (Dow Corning Corporation, США).

Литературный обзор

Процесс создания аналитического МФЧ включает следующие стадии: определение метода анализа, реализуемого на чипе; разработку топологии и конструкции микрочипа; выбор материалов; определение технологий формирования функциональных структур микрочипа; изготовление микро- и наноразмерных структур и элементов на подложке и их контроль;

обработку поверхности; создание интерфейсных устройств; герметизацию полученных структур и элементов; контроль и испытание микрочипа, как готового изделия; его упаковку.

Методы и приборы для полимеразной цепной реакции на основе микрочиповых технологий

ПЦР является одним из лидирующих методов анализа ДНК при генетических исследованиях, обнаружении особо опасных инфекций и патогенов, а также в медицинской диагностике, криминалистике и т. п. [3]. Метод основан на многократном копировании (амплификации) определенных фрагментов нуклеиновых кислот при помощи ферментов и реакционной среды. Процесс проходит в искусственных условиях при повторяющихся температурных циклах. Существует ряд трудностей проведения ПЦР: проблема контаминации, снижение эффективности реакции амплификации, ингибирование из-за содержания в пробе специфичных агентов, сложность очистки ДНК от примесей и ряд других.

К тенденциям развития методов ПЦР можно отнести повышение общей надежности метода и увеличение эффективности реакции, существенное сокращение времени исследования, а также широкое внедрение количественной ПЦР в медицинскую диагностику [4]. Обусловлено это и тем, что одновременное проведение качественного анализа (амплификации) и количественного измерения амплифицируемых молекул ДНК (с применением флуоресцентных специфичных меток), позволяет достичь более высокой чувствительности, воспроизводимости и специфичности. Таким образом, проведение ПЦР в режиме реального времени упрощает и ускоряет процедуру традиционного анализа за счет увеличения полученной от образцов информации и уменьшения времени анализа [5].

Более точным количественным методом анализа является цифровая ПЦР (цПЦР), применяемая при предельно малых концентрациях определяемых молекул в исходном образце и тем самым позволяющая проводить медицинскую диагностику заболеваний на ранних стадиях. Методы цПЦР по сравнению с ПЦР в реальном времени (ПЦР-РВ): не требуют построения градуировочных зависимостей и использования дорогостоящих реагентов; дают возможность количественно определять концентрацию молекул ДНК в сверхмалых объемах исходной пробы; обладают одномолекулярной чувствительностью [6].

Одним из направлений совершенствования приборных методов для молекулярной диагностики является адаптация метода ПЦР для микрофлюидных платформ (микрочипов). Такое решение позволяет не только значительно уменьшить расход пробы и дорогостоящих реагентов, но и существенно увеличить скорость нагрева (охлаждения) реакционной камеры и сократить время анализа, а также одновременно проводить множество реакций в одинаковых условиях.

Первыми исследованиями по разработке приборов для ПЦР на основе микрочипов являлись работы Л. Криска [7, 8].

Вариант проточных МФЧ был предложен компанией Microfluidic ChipShop GmbH [9]. Однако подобная конструкция чипов имеет ряд недостатков: образование пузырьков воздуха в каналах, снижающих эффективность реакции; размытие образца; сложность контроля скорости перемещения и создания стабильности реакционной смеси [2].

Возможной альтернативой проточных МФЧ является использование ПЦР-чипов, содержащих миниатюрные стационарные реакционные камеры. Такие чипы можно разделить на однокамерные, мультикамерные и чипы с «виртуальными» камерами. В последней конструкции применяется гидрофобная-олеофобная поверхность, на которой размещается капля масла, выступающая в качестве своеобразной «виртуальной» реакционной камеры, содержащей внутри себя каплю ПЦР-смеси. Такой подход был предложен исследователями из Institute of Bioengineering and Nanotechnology (Singapore) [10].

К микрочипам со стационарными реакционными камерами можно отнести и некоторые устройства для цПЦР. Первоначально цПЦР была реализована на стандартных микротитрационных планшетах. Для детектирования результатов измерений применялись высокочувствительные системы фирмы Agilent Technology [11], позволяющие осуществлять анализ экспрессии генов, проводить метилирование ДНК, обнаружение патогенов и т. д. Но

существенными недостатками такого подхода являлись чрезвычайно высокая трудоемкость измерений и необходимость большого количества исследуемой пробы. Устранить их позволило появление новых более прогрессивных систем на базе устройств с реакционными камерами гораздо меньших размеров: «цифровыми» панелями фирм Fluidigm [12] и Applied Biosystems [13], микрофлюидными устройствами (например, устройство с техникой SlipChip [14]) и др.

Невысокая стоимость устройства, возможность быстрого прототипирования в условиях исследовательских лабораторий, малый расход пробы и реагентов, возможность использования стандартного оборудования как для проведения термоциклирования, так и для детектирования результатов ПЦР, малые затраты на специальное оборудование позволили создать прототип МФЧ для количественной ПЦР в жидкой среде.

Особенности выбора материалов и изготовления микроразмерных структур

Непосредственное влияние на выбор материалов МФЧ оказывают методы формирования микроструктур. Традиционными материалами для микрочипового производства являются кремний, кварц и стекло, так как для них хорошо отработаны технологии изготовления микроразмерных структур, а именно: методы фото- или лазерной литографии и кислотного или ионно-реактивного травления. Несмотря на то что метод травления структур является доступным для серийного производства, он имеет ряд недостатков: высокую стоимость, трудоемкость и длительное время изготовления, применение агрессивных сред, влияние на свойства материала процессов обработки (что может приводить к ингибированию и снижению эффективности ПЦР), а также ограничение на глубину получаемых структур (не более 200–250 мкм) и на аспектное соотношение (соотношение глубины к ширине получаемой реакционной камеры) – не более одного. Тем не менее кремний и стеклянные материалы остаются востребованными при создании микроаналитических систем, в которых предъявляются повышенные требования к метрологическим характеристикам.

В последнее время наметилась устойчивая тенденция использования полимерных материалов, таких как полидиметилсилоксан (ПДМС), поликарбонат (ПК), полиметилметакрилат (ПММА), полиимид (ПИ) и пр. [15, 16] благодаря их низкой себестоимости и относительно низкой стоимости изделий при массовом производстве. Это позволяет создавать одноразовые чипы и открывает перспективу применения полимерных материалов в конструкциях аналитических МФУ, особенно микрочипов для ПЦР.

По сравнению с массовым производством иностранными фирмами (например, Fluidigm Corporation [17]) устройств для медицинской диагностики (в том числе для проведения ПЦР), предназначенных для использования в крупных диагностических центрах, в условиях исследовательских лабораторий высока потребность в менее производительных, но более гибких и недорогих системах. Так экономически не выгодными являются прецизионные мастер-формы для методов горячего тиснения. Они имеют высокую стоимость и быстро изнашиваются (к тому же существует вероятность появления разнообразных источников погрешности: например, при грубой отделке мастер-форм все дефекты формы переходят на формируемую деталь). Поэтому наиболее привлекательными являлись бы доступные и оперативные технологии изготовления МФЧ, позволяющие снизить стоимость готового изделия. Благодаря развитию в последнее время метода «мягкой» литографии (soft lithography) стало возможным переносить отпечаток с мастер-формы, получаемой методами литографии и жидкостного травления, на мягкую подложку, при этом топология формы определяет конечную структуру подложки.

Методы изготовления микроразмерных структур зависят от применяемого материала и в настоящее время совершенствуются. Они связаны с активным воздействием на поверхность материала и приводят к изменению его поверхностных свойств, что следует учитывать при создании МФЧ. Применяются следующие методы изготовления микроструктур на подложке:

- литография [18–20];
- лазерная механическая обработка [21–23];

- получение отпечатков с помощью мастер-форм (горячее тиснение, литье под давлением, «мягкая» литография (soft lithography)) [24, 25];
- LIGA – технологии (литография, гальваника, импринтинг) [26] и пр.

Полимеры подразделяются в соответствии с молекулярной структурой и их свойствами на группы: термопласты, эластомеры и термореактивные полимеры.

Термопластичные полимеры (термопласты) – линейные или разветвленные полимеры, которые могут быть расплавлены при тепловом воздействии. Они не являются ковалентно сшитыми и способны переходить в эластичное состояние при нагреве. С точки зрения серийного производства устройств «лаборатория на чипе», основное преимущество термопластов заключается в том, что они могут формоваться при температуре, близкой к температуре плавления (размягчения) и имеют высокую жесткость, что обеспечивает защиту микрофлюидных каналов. Существуют коммерческие варианты МФУ из термопластов, предназначенные для применения в медицинской диагностике. Наиболее часто в микрофлюидике применяются такие материалы как: ПММА, ПК и циклический олефиновый сополимер.

Термореактивные полимеры (бакелит, полиимид) представляют собой ковалентно сшитые полимеры, которые не плавятся и не растворяются, обладают жесткостью и хрупкостью. Их механическая обработка сложна. Термореактивные полимеры формируются в процессе полимеризации. Основными их преимуществами для систем «лаборатория на чипе» является геометрическая стабильность и устойчивость ко многим растворителям. Обычно для макетирования конструкций устройств используются: фоторезист SU-8 (MicroChem, США) и фотоотверждаемые полимеры (например, оптический клей NOA81 (Норленд Products Inc., США) [27].

Изделия из термопластичных полимеров обычно изготавливаются методом горячего тиснения, а из термореактивных – методом литья.

Эластомеры – поперечно слабосвязанные полимеры, эластичные и упругие, восстанавливающие свою форму после снятия напряжения. Они могут быть термопластичными или термореактивными. Среди термореактивных эластомеров ПДМС является доминирующим материалом для создания прототипов МФУ. Впервые использование ПДМС в качестве материала для микрофлюидики было рассмотрено в работах G. Whitesides [28] и S. Quake [29].

Материал обладает высокой оптической прозрачностью, биосовместимостью и высокой адгезией к кремнию и стеклу, устойчив к высоким температурам, однако существенным недостатком ПДМС являются трудности, связанные с модификацией поверхности, обусловленные высокой подвижностью полимерных цепей. Кроме того, полимерная структура в виде сетки обладает способностью поглощать небольшие молекулы [30], наблюдается вымывание молекул неотвержденных мономеров [31], а в растворителях происходит набухание полимера [32]. Все это является мотивацией исследований, направленных на изменение и улучшение свойств ПДМС.

Особенности физической и химической обработок материалов, способы герметизации

ПДМС (рис. 1) в исходном состоянии обладает гидрофобными свойствами, которые обусловлены наличием метильных групп в структуре материала.

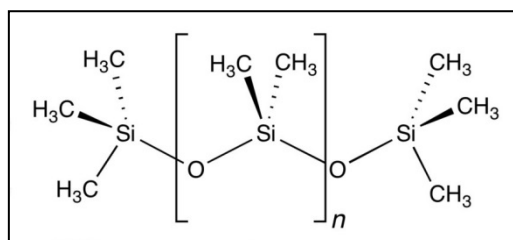


Рис. 1. Структурная формула полидиметилсилоксана

Одним из наиболее распространенных методов обработки поверхности ПДМС является плазменное воздействие в среде кислорода, при котором происходит замена метильных групп ($=\text{Si}-\text{CH}_3$) на поверхности на гидроксильные ($=\text{Si}-\text{OH}$), которые взаимодействуют с гидроксильными группами на поверхности стекла, что приводит к формированию сильной связи ($=\text{Si}-\text{O}-\text{Si}=\text{O}$) и образованию молекул воды.

Модификация поверхностных свойств может быть достигнута различными способами: прививкой полимерных цепей на поверхности [33], химическим присоединением монослоев молекул [34] или физической адсорбцией молекул на подложке [35].

В случае ПДМС можно использовать систему из двух слоев: один слой с избытком основания (где преимущественно – винильные группы), а другой – с избытком отвердителя ($\text{Si}-\text{H}$ группы). Они могут быть ковалентно связаны друг с другом без помощи плазменной обработки [36]. В работе [37] представлен метод соединения ПММА и ПДМС, позволяющий осуществлять герметизацию микрофлюидных чипов. Для формирования гидроксильных групп на поверхности ПММА предложено использовать силаны. Гидроксильные группы повышают прочность сцепления пластин ПДМС – ПММА без засорения функциональных микроструктур (каналов) или видимой их деформации. Средняя прочность сцепления при применении силана составила 2,5 МПа.

Среди методов герметизации можно выделить: методы термического и анодного связывания (используются для кремниевых и стеклянных пластин), оптического контакта, герметизации скотчами и полимерными пленками, методы склеивания полимерными материалами (в том числе термо- и фотоотверждаемыми композитами). Эти методы зависят от материалов соединяемых изделий, параметров микро- и наноструктур, наличия тонкослойных элементов (токопроводящие дорожки, пленки и т. п.).

Цель и задачи проекта

Цель проекта – разработка и изготовление прототипа гибридного (стеклянно-полимерного) МФЧ для ПЦР в жидкой среде.

Основные задачи:

1. Анализ топологий и конструкций МФЧ для ПЦР.
2. Анализ литературных источников по методам, технологиям и процессам изготовления МФЧ для ПЦР.
3. Разработка и расчет топологии и конструкции МФЧ.
4. Разработка и изготовление мастер-формы с заданной топологией методом лазерной литографии и кислотного травления для получения микроструктур в ПДМС.
5. Формирование структур в полимере методом «мягкой» литографии (soft lithography).
6. Исследование физико-химических методов и выбор условий обработки материалов для эффективной модификации рабочих поверхностей чипа, а также с целью создания герметичного соединения стеклянной подложки с ПДМС отпечатком.
7. Изготовление прототипов МФЧ для ПЦР в жидкой среде.
8. Тестирование и апробация полученных прототипов МФЧ на модельных объектах.
9. Анализ полученных результатов, формирование рекомендаций для дальнейших исследований, подготовка материалов для публикации и участия в конференции.

Методы исследования, оборудование и материалы

Изготовление мастер-формы осуществлялось с помощью методов лазерной литографии и кислотного травления. Напыление на стекло слоя хрома проходило на установке термического/электронного напыления Auto 500 Coater (Edwards, Англия). При нанесении слоя фоторезиста использовалась центрифуга Delta 6 RC (SUSS MicroTec, Германия). Экспонирование фоторезиста осуществлялось на установке лазерной литографии DWL 66FS (Heidelberg Instruments Mikrotechnik GmbH, Германия). Для проявления резиста использовался проявитель AZ 756 MIF (MicroChemicals GmbH, Германия). Высота полученных структур контролировалась на профилометре Ambios Technology (США).

Для формирования структур в ПДМС применялся метод «мягкой» литографии. Поверхность исследуемых материалов подвергалась физической и химической обработке, результаты воздействия контролировались методами микроскопии ближнего поля (зондовая нанолaborатория NTEGRA Solaris, (НТ-МДТ, Россия)), лежащей капли (экспериментальная установка на базе оптического микроскопа PrimaExpert (ЛОМО, Россия)) и оптической спектроскопии (спектрофотометр Hitachi UV 3210 (Hitachi High-Technologies Corporation, Япония)).

Плазменная обработка поверхности материалов проводилась на вакуумной напылительной установке SPI Module Sputter/Carbon Coater (Structure Probe, Inc., США) в аргонной среде.

Ультрафиолетовое воздействие на образцы осуществлялось при помощи УФ ртутной лампы 50 Вт (OSRAM GmbH, Германия) со спектральным диапазоном 220–450 нм.

Термоциклирование полученных чипов проводилось на устройстве Swift MaxPro Thermal Cyclers (ESCO Healthcare MicroPte Ltd., Сингапур).

Контроль герметичности полученных прототипов до и после термоциклирования проводился методом оптической микроскопии высокого разрешения (конфокальный лазерный сканирующий микроскоп Leica TCS SL (Leica Microsystems CMS GmbH, Германия)).

Объектами исследования служили пластины 24×24 мм из: стекла марки Крон 8 (К8) (ГОСТ 3514-76, Россия), ПДМС марки Sylgard® 184 (Dow Corning Corporation, США), а также чипы, изготовленные в рамках данного проекта из стекла и ПДМС с двумя реакционными камерами для проведения ПЦР в жидкой среде (рис. 2в).

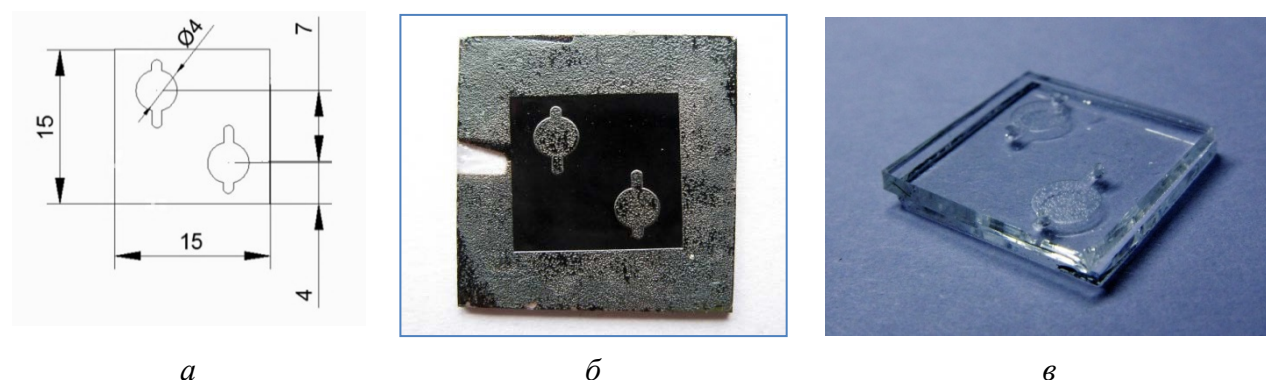


Рис. 2. Объекты исследования: а – эскиз топологии гибридного МФЧ с двумя реакционными камерами (размеры в мм); б – изображение мастер-формы из стекла К8; в – изображение гибридного чипа (глубина 50 мкм, диаметр реакционной камеры 4 мм)

Разработка и расчет топологии и конструкции микрофлюидного чипа

Сочетание свойств стеклянного (прозрачность, биосовместимость, высокий теплообмен) и полимерного (прозрачность, биосовместимость, доступная технология формирования микроструктур в ПДМС) материалов делает оправданным их объединение в единую конструкцию.

ПДМС имеет высокую адгезию к стеклу и кремнию, поэтому обратимая герметизация микроструктур в этом полимере может осуществляться при помощи ван-дер-Ваальсовых сил путем прикрепления к подложке. Однако такой способ связывания не рассчитан на воздействие высоких температур, применяемых при проведении ПЦР. Он может быть использован для создания МФУ, в каналах которого не требуется высокое давление. Необратимая герметизация может быть достигнута при обработке в среде кислородной плазмы или после химической модификации поверхности при связывании ПДМС со стеклянными пластинами, кремнием и полистиролом. При этом прочность соединения выдерживает давление до 200 кПа. Для многократного использования стеклянной подложки чипа его погружают в раствор щелочи на несколько часов, затем пленка ПДМС отделяется от подложки, которая может применяться для герметизации другого чипа. Принимая во внимание особенности проведения полимеразной цепной реакции, нами была выбрана двухкамерная

топология: в одну реакционную камеру загружается исследуемая проба, вторая требуется для осуществления отрицательного контроля.

С учетом приборных особенностей (размера оправы для МФЧ, площади зоны детектирования) и требуемого объема анализируемого образца были выбраны соответствующие топология и габаритные размеры, которые приведены на рис. 2а. Конструкция полученного прототипа МФЧ представила собой герметично соединенную стеклянную пластину толщиной 1 мм с пленкой ПДМС толщиной 1 мм. Глубина реакционных камер и подводящих каналов составила 50 мкм, диаметр отверстий для ввода пробы – 0,8 мм. Объем камеры – 0,6 мкл, объем подводящих каналов 0,3 мкл. Таким образом, общий объем загружаемой в чип пробы составляет ~1 мкл.

Разработка и изготовление мастер-формы для получения микроструктур в полидиметилсилоксане методом «мягкой» литографии

На основании эскиза топологии МФЧ был выполнен чертеж, по которому изготавливалась мастер-форма из стекла марки К8 методами лазерной литографии и жидкостного травления в буферном растворе плавиковой кислоты в соответствии с последовательностью технологических операций:

- поверхность стекла в напылительной установке покрывается слоем хрома толщиной около 100 нм;
- слой фоторезиста наносится на центрифуге при частоте вращения ротора 4000 об/мин (пленка резиста отжигается при температуре 100 °С в течение 50 с);
- экспонирование фоторезиста осуществляется на установке лазерной литографии (проявление резиста проводилось в проявителе в течение 60 с);
- выполняется снятие слоя хрома, а затем травление и промывание заготовки поочередно в буферном растворе для травления (HF:NH₄F:H₂O, соотношение 1:1:1) и дистиллированной воде в течение 10 мин. Высота полученных структур составила 50 мкм;
- проводится очистка поверхности от остатков хрома и резиста, покрытие структуры слоем хрома толщиной около 70 нм, который препятствует прилипанию отпечатка из ПДМС к мастер-форме.

Полученная мастер-форма последовательно промывается в ацетоне, этиловом спирте, дистиллированной воде и высушивается при комнатной температуре. Фотография готовой мастер-формы приведена на рис. 2б.

После этого методом «мягкой литографии» осуществляется получение отпечатков из ПДМС. Процесс проходит в несколько этапов:

1. В чашку Петри с закрепленными маячками высотой 1 мм, обеспечивающими изготовление пленки фиксированной толщины, помещается раствор полимера (соотношение основа-отвердитель 10:1 (по массе)). Мастер-форма, ориентированная структурами вниз, устанавливается в чашку поверх маячков.
2. Чашка Петри с жидким полимером помещается в эксикатор на 30 мин для удаления растворенного воздуха.
3. Затем происходит отверждение полимера в течение 2 час при температуре 80 °С.
4. Отвержденный полимер отделяется от формы, в подложке формируются отверстия для загрузки пробы.

Исследование физико-химических методов обработки материалов и способов их герметизации

В качестве физических методов обработки поверхности были выбраны: плазменная обработка материалов, которая проводилась на вакуумной напылительной установке SPI Module Sputter/Carbon Coater (Structure Probe, Inc., США) в аргоновой среде; воздействие ультрафиолетовым излучением на образцы, которое осуществлялось при помощи УФ ртутной лампы 50 Вт (OSRAM GmbH, Германия) со спектральным диапазоном 220–450 нм.

Результаты воздействия различных методов обработки на поверхность образцов контролировались методом лежащей капли при измерении контактного угла с помощью

экспериментальной установки на базе оптического микроскопа PrimaExpert (ЛОМО, Россия). Проводилось по 6–8 измерений контактного угла для каждого образца, объем капли 20 мкл.

Обработка плазмой в течение 30 с поверхности стекла К8 (исходное значение контактного угла θ составляет $47\pm 2^\circ$) приводит к увеличению угла смачивания ($72\pm 2^\circ$). В случае с ПДМС, напротив, поверхность приобретает гидрофильные свойства: значение угла меняется со $107\pm 1^\circ$ на $\sim 42^\circ$. В последнем случае удалось получить приблизительную оценку, так как капля при нанесении на поверхность имеет неправильную форму. После обработки ПДМС в течение двух минут поверхность приобретает супергидрофильные свойства ($\theta \rightarrow 0^\circ$).

Воздействие ультрафиолетовым излучением в течение полутора часов на поверхность ПДМС практически не приводит к изменению свойств смачивания (значение угла после обработки составляет $104\pm 2^\circ$).

Был проведен ряд экспериментов по изучению возможности изменения свойств поверхности ПДМС различными методами химической обработки: выдерживанием в растворах соляной кислоты, растворе «пиранья» ($\text{H}_2\text{SO}_4:\text{H}_2\text{O}_2$) с различными соотношениями элементов и щелочи NaOH. Кроме того, изучалось влияние соотношения компонентов ПДМС (основа:отвердитель 10:1, 3:1) на свойства поверхности при кипячении в воде, а также исследовалась модификация полимера путем введения в состав исходной смеси избыточного количества одного из компонентов отвердителя. Наиболее эффективным способом обработки согласно Табл. 1 с точки зрения наибольшего изменения свойств и оптимального времени воздействия является выдерживание образца полимера в растворе «пиранья» с соотношением 3:1 в течение одной минуты. Экспериментально было установлено, что обработка в указанном растворе более длительное время приводит к тому, что поверхность полимера становится матовой.

Таблица 1

Изменение контактного угла ПДМС (соотношение основа:отвердитель 10:1) в зависимости от способов и времени обработки

Метод обработки	Время обработки	Контактный угол, град
Исходный образец	-	107 ± 1
Раствор «пиранья» H_2SO_4 (конц): H_2O_2 (30 %), (3:2)	1 час	91 ± 3
Раствор «пиранья» H_2SO_4 (конц): H_2O_2 (30 %), (3:2)	18 час	88 ± 2
Раствор «пиранья» H_2SO_4 (конц): H_2O_2 (30 %), (3:1)	1 мин	93 ± 1
Раствор $\text{H}_2\text{O}:\text{H}_2\text{O}_2:\text{HCl}$ (3:3:1)	10 мин	103 ± 2
	40 мин	100 ± 1
1M NaOH	18 час	108 ± 2

С учетом ранее полученных результатов были проведены исследования по изучению влияния времени обработки раствором «пиранья» H_2SO_4 (конц): H_2O_2 (30 %) (3:2) в течение 10, 20, 30 и 40 мин на величину контактного угла. Также была проведена серия экспериментов, включающих выдерживание образцов в растворе «пиранья» с последующей обработкой 1M раствором щелочи KOH в течение равных промежутков времени. Полученные результаты свидетельствуют о том, что существенное изменение свойств поверхности происходит при кратковременной (в течение 10 мин) обработке полимера раствором «пиранья». Последующее выдерживание в растворе не приводит к значительному изменению свойств поверхности, хотя контактный угол уменьшается на 7° при обработке в течение 30 мин (при этом общее время обработки составило один час).

В работе [38] предложен способ увеличения гидрофильности поверхности отвержденного ПДМС путем его обработки кипящей водой продолжительностью от 30 мин до двух часов. Авторы считают, что в реакции участвуют избыточные Si-H группы, присутствующие на поверхности полимера. Изменение свойств смачивания при этом рассматривалось для образцов с соотношением основа:отвердитель 10:1. Однако при попытке повторить данный эксперимент были получены результаты (Табл. 2), не соответствующие литературным данным. Поэтому было решено провести дополнительные исследования по изучению изменения

контактного угла после кипячения образцов с соотношением основа:отвердитель 3:1. Результаты последних экспериментов являются оригинальными (получены впервые).

Таблица 2

Изменение контактного угла поверхности ПДМС с различными соотношениями основы к отвердителю в зависимости от времени кипячения в воде

Соотношение основа:отвердитель	Контактный угол, град.				
	Исходный образец	Время кипячения, час			
		0,5	1	1,5	2
10:1	107±1	104±2	106±1	103±2	105±2
3:1	102±1	96±1	92±1	93±1	94±1

Согласно полученным результатам в случае стандартного соотношения основа:отвердитель (10:1) для ПДМС не наблюдается существенного изменения свойств поверхности. Возможно, причиной этого является неполная полимеризация материала. Поэтому были проведены исследования по изучению влияния кипячения на величину контактного угла для соотношения основа:отвердитель 3:1. После одного часа кипячения наблюдалось снижение угла смачивания на 10°.

Для увеличения концентрации водород-содержащих (гидридных) групп на поверхности отвержденного ПДМС было предложено ввести в состав исходной смеси избыточное количество одного из компонентов отвердителя, декларированного производителем и содержащего 50 % (моль) гидридных групп – гидросилоксана (ГС) и провести реакцию гидролиза в кипящей воде. Для экспериментов были изготовлены образцы пленок ПДМС с добавками 5 % (масс.) и 10 % (масс.) ГС по отношению к массе основы. Соотношение основы и отвердителя составляло 10:1 (по массе). Известно, что в случае мономерных соединений для возникновения реакции достаточно присутствия следов щелочи. Поэтому в качестве катализатора при кипячении в воду было добавлено небольшое количество щелочи (50 мкл КОН на 250 мл воды). Результаты измерений представлены в Табл. 3.

Таблица 3

Изменение контактного угла ПДМС и его модификаций в зависимости от способа обработки

Состав образца	Способ обработки	Контактный угол, град.
исходный состав	-	103±1
исх. сост. + 5 % ГС	-	101±2
исх. сост. + 5 % ГС	кипячение, 1 час	102±1
исх. сост. + 10 % ГС	-	96±1
исх. сост. +10 % ГС	кипячение, 1 час	104±2

В данном случае изучались пленки ПДМС другой партии, поэтому значение исходного контактного угла отличается от приведенного ранее. Из полученных данных следует, что добавление 10 % гидросилоксана в состав ПДМС приводит к уменьшению контактного угла на 7°, при этом стадию кипячения следует исключить, так как поверхность приобретает исходные свойства.

С целью получения неразъемного герметичного соединения необходимо также обрабатывать поверхность стекла. Влияние способа обработки на шероховатость и неоднородность поверхности стекла марки К8 изучалось при измерении необработанной и обработанной поверхностей методом микроскопии поперечных сил (зондовая нанолaborатория NTEGRA Solaris, (НТ-МДТ, Россия)). После обработки образца раствором «пиранья» H₂SO₄(конц):H₂O₂(30 %) (3:1) в течение 10 мин наблюдается увеличение средней шероховатости с Ra=0,5 нм (для необработанного образца) до Ra=2,6 нм. В этом случае происходит травление поверхности стекла (кислота взаимодействует с оксидами металлов, входящими в его состав, что приводит к образованию солей). После данной обработки поверхность приобретает супергидрофильные свойства ($\theta \rightarrow 0^\circ$). При выдерживании образца в растворе щелочи 0,5M NaOH в течение 12 час также происходит изменение рельефа

поверхности ($R_a=1,0$ нм) за счет выщелачивания стекла. Исследования контактного угла смачивания после обработки образцов стекла растворами щелочей были проведены в работе [39].

Для создания герметичного соединения между стеклянной подложкой и ПДМС отпечатком были выбраны химические и физические методы. Контроль герметичности прототипов до и после термоциклирования проводился на конфокальном лазерном сканирующем микроскопе Leica TCS SL (Leica Microsystems CMS GmbH, Германия).

Обработка в растворе «пирания» H_2SO_4 (конц.): H_2O_2 (30 %) в соотношении 3:1

Выдерживание стекла в этом растворе в течение 10 мин приводит к тому, что поверхность материала становится супергидрофильной. Для ПДМС время обработки было выбрано экспериментально и составило 1 мин, так как при дальнейшем выдерживании поверхность становилась матовой. После обработки пластины обоих материалов промывались в воде, высушивались и соединялись, далее выдерживались 12 час при комнатной температуре. Затем проводилось заполнение реакционных камер раствором флуоресцеина $C=10^{-4}$ М и выполнялся контроль герметичности до и после термоциклирования (циклический нагрев и охлаждение МФЧ от 60 до 97 °С в течении 15 циклов). Термоциклирование проводилось на устройстве Swift MaxPro Thermal Cyclers (ESCO Healthcare MicroPte Ltd., Сингапур). Полученные изображения приведены на рис. 3.

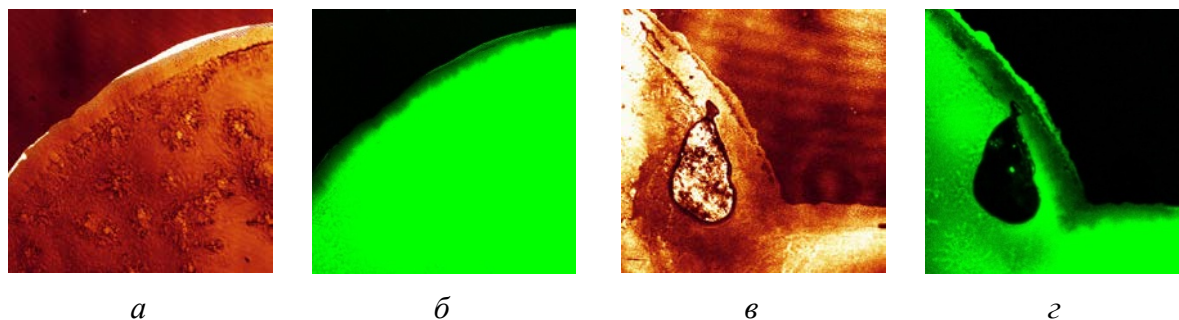


Рис. 3. Изображения границы герметизированной реакционной камеры путем обработки раствором «пирания» (а, в – режим отраженного света; б, г – флуоресценции): а, б – до процесса термоциклирования; в, г – после. Размеры кадров $1,2 \times 1,2$ мм

Плазменная обработка в течение 2 мин в среде аргона

При данной модификации ПДМС поверхность приобретала супергидрофильные свойства. Выполнялись аналогичные действия, описанные выше для контроля герметичности. Полученные изображения приведены на рис. 4.

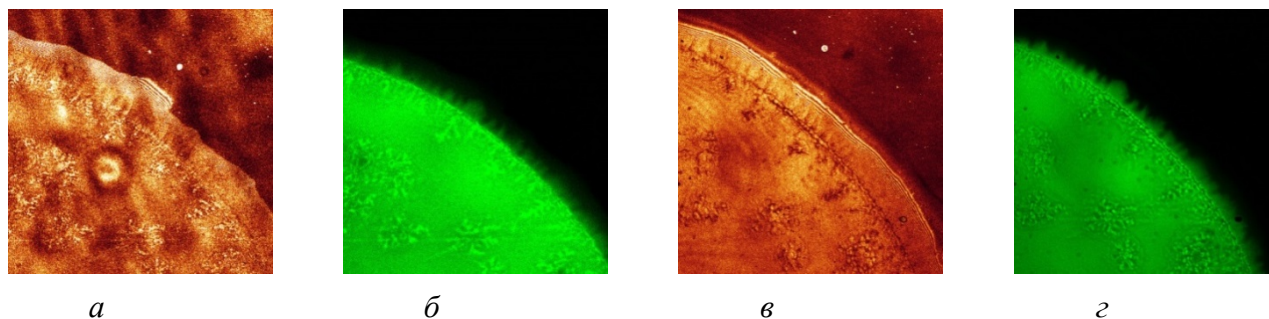


Рис. 4. Изображения границы герметизированной реакционной камеры путем плазменной обработки (а, в – режим отраженного света; б, г – флуоресценции): а, б – до процесса термоциклирования; в, г – после. Размеры кадров $1,2 \times 1,2$ мм

Данные результаты позволяют сделать вывод о том, что предложенные методы обработки дают возможность получить герметичное соединение, которое сохраняется после воздействия циклически меняющихся температур, что подтверждается изображениями флуоресценции, наблюдаемой только в области заполненной реакционной камеры.

Исследование влияния процессов обработки и модификации ПДМС на спектральные зависимости светопропускания в ближней ИК-области

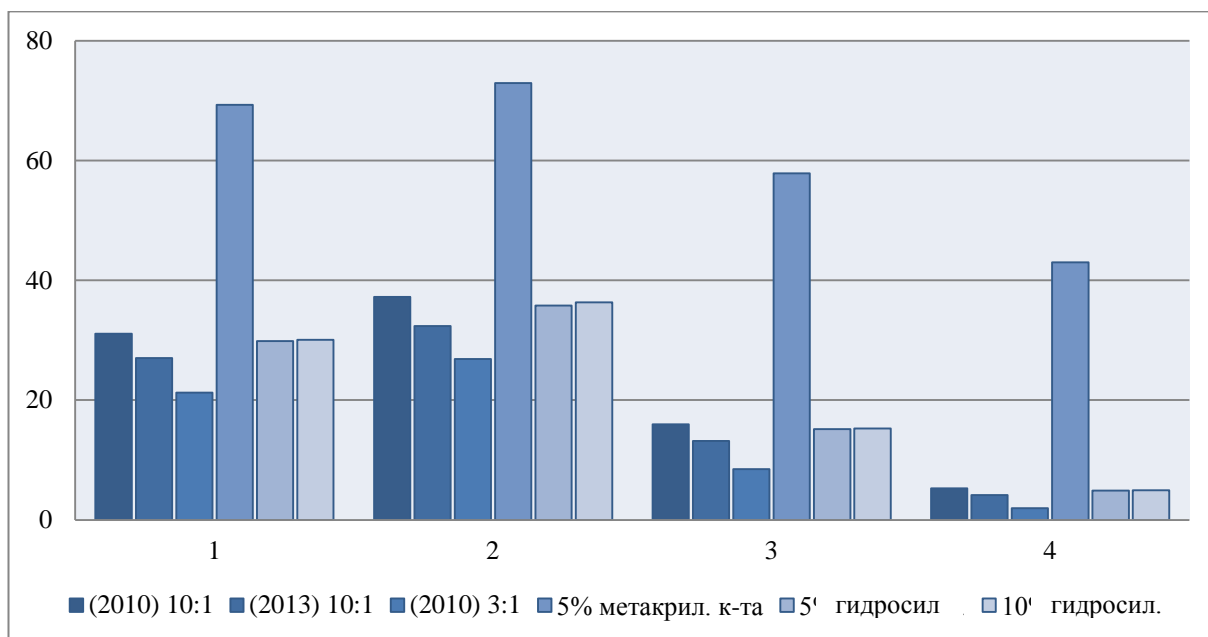
ПДМС, как уже упоминалось, обладает высокой прозрачностью в видимом спектральном диапазоне, показатель преломления 1,4035 (при 25 °С на 589 нм). Известно, что для ПДМС характерно наличие частот поглощения, обусловленных валентными колебаниями групп CH_3 - (2964 см^{-1} или 3374 нм), деформационные колебания связей $\text{CH}_3\text{-Si}$ на частоте 1261 см^{-1} (7930 нм), слабый уширенный сигнал колебаний OH -групп в области $3400\text{--}3475 \text{ см}^{-1}$ (2940–2875 нм) и т. д. [38, 40]. ПДМС имеет не только сильные полосы поглощения в средней ИК-области, но и определенные полосы обертонов в ближней ИК-области (БИК). В работе [41] показано, что поглощение ПДМС в БИК диапазоне в основном вызвано колебательными обертонами групп CH_3 полимера, а также приведены данные по длинам волн поглощения для диапазона от 620 до 1850 нм.

Задачами спектрофотометрических измерений образцов ПДМС в БИК-области после их обработки являлось определение возможности контроля групп OH и выявление спектральных диапазонов, где наблюдаются наибольшие изменения светопоглощения.

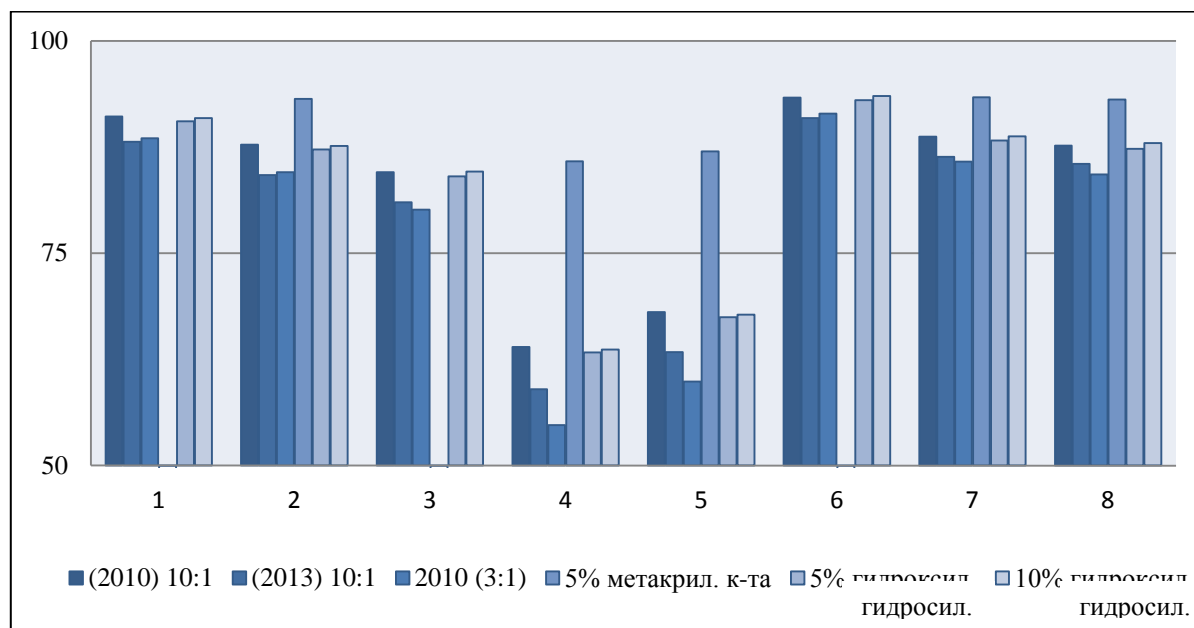
На основании полученных измерений спектральных зависимостей светопропускания образцов были определены минимумы пропускания (максимумы поглощения), построены соответствующие таблицы, а затем, для наглядности, – гистограммы. На рис. 5 приведены гистограммы светопропускания (в двух спектральных диапазонах) для образцов ПДМС разных партий, разного соотношения основы и отвердителя, при добавлении метакриловой кислоты и гидросилоксана.

Из Табл. 4, на основании которой построены гистограммы, и данных работы [41], следует, что наблюдаемые полосы поглощения обертонов метильных групп (CH_3) отражают состояние их колебательных уровней ($\nu_1 = 2905 \text{ см}^{-1}$, $\nu_2 = 2963 \text{ см}^{-1}$, $\nu_3 = 1263 \text{ см}^{-1}$, $\nu_4 = 1412 \text{ см}^{-1}$). Обертон комбинированных колебаний групп OH ($2\nu+\delta$) совпадает с обертоном метильной группы ($3 \nu_1$), что делает невозможным оценку групп OH . Кроме того, нет явно выраженного пика, присущего обертонам валентных колебаний OH (в области 1450 нм). Если предположить, что этому обертону соответствует пик пропускания, зарегистрированный на 1480 нм (подобный сдвиг частот вполне возможен в сложных молекулах), то исходя из данных, полученных при спектральных измерениях после разных видов обработки полимера (Табл. 5), можно считать, что метод БИК в данном случае не подходит для оценки групп OH .

Полученные данные являются мотивацией для проведения дальнейших исследований по изучению процессов модификации ПДМС спектрофотометрическими методами и сопоставлению результатов спектральных исследований с изменением свойств поверхности.



а



б

Рис. 5. Изменение светопропускания образцов ПДМС: а – на длинах волн (1 – 2474 нм, 2 – 2405 нм, 3 – 2374 нм, 4 – 2296 нм); б – на длинах волн (1 – 1976 нм, 2 – 1850 нм, 3 – 1784 нм, 4 – 1746 нм, 5 – 1700 нм, 6 – 1529 нм, 7 – 1401 нм, 8 – 1186 нм)

Таблица 4

Минимумы светопропускание образцов ПДМС различных составов

Длина волны, нм	Соответств. поглощ. обертонов метильных групп [43]	ПДМС, 2010 10:1	ПДМС, 2013 10:1	ПДМС, 2010 3:1	ПДМС +5 % метакр. кислоты	ПДМС +5 % гидросилоксан	ПДМС +10 % гидросилоксан
2471		31,06	27,04	21,27	69,27	29,87	30,06
2405		37,21	32,34	26,86	72,91	35,8	36,34
2374		15,98	13,2	8,45	57,85	15,13	15,25
2296		5,24	4,13	1,95	42,99	4,89	4,92
1976		91,08	88,12	88,52	-	90,52	90,9
1850	$v_1 + 2 v_3$	87,79	84,21	84,52	93,17	87,23	87,64
1784	$v_2 + v_3 + v_4$	84,54	81	80,12	-	84,06	84,6
1746	$v_1 + v_2$	63,95	59	54,75	85,82	63,31	63,67
1700	$2 v_2$	68,08	63,36	59,88	87	67,47	67,76
1529	-	93,31	90,89	91,43	-	93	93,5
1401	$2v_1 + v_4$	88,71	86,35	85,77	93,34	88,27	88,77
1186	$3 v_1$	87,68	85,52	84,27	93,09	87,3	87,98

Таблица 5

Светопропускание образцов ПДМС в интервале длин волн 1475–1480 нм после обработки поверхности

ПДМС 2010, 10:1		ПДМС 2010, 3:1		ПДМС 2010 + 5 % гидросилоксана		ПДМС 2010 + 10 % гидросилоксана	
исходное	после обработки «пираньей»	исходное	после кипячения	исходное	после кипячения	исходное	после кипячения
94,86	94,27	93,42	94,19	94,54	94,98	95,04	95,12

Заключение

Показана перспективность применения ПДМС Sylgard 184 (Dow Corning Corporation, США) для решения задач проекта, в том числе:

- создания реакционных камер методами быстрого прототипирования;
- обработки рабочей поверхности материала для снижения гидрофобных свойств;
- создания герметичного соединения стеклянной подложки с полимерным отпечатком;
- изготовления прототипов МФЧ для ПЦР в жидкой среде.

Проведены экспериментальные исследования по изучению возможности гидрофилизации поверхности ПДМС Sylgard 184 различными методами физической (УФ-излучение) и химической обработки, в том числе:

- в растворе «пиранья» H_2SO_4 (конц.): H_2O_2 (30 %) с различными соотношениями;
- кипячения в воде;
- в растворе щелочи;
- в растворе $\text{H}_2\text{O}:\text{H}_2\text{O}_2:\text{HCl}$ (3:3:1).

Кроме того, подтверждено, что введение в состав исходной смеси избыточного количества одного из компонентов отвердителя – гидросилоксана (5 % и 10 %) позволяет изменить величину контактного угла. Наиболее существенное его изменение наблюдается при обработке ПДМС в растворе «пиранья».

Исследования по изучению влияния соотношения основа-отвердитель на изменение контактного угла после кипячения образца ПДМС показали, что наиболее существенны они для образца ПДМС (соотношение основа:отвердитель 3:1) при кипячении в течение 1 час.

При анализе спектральных зависимостей светопропускания образцов ПДМС было выявлено, что:

- контроль групп ОН в ближней области спектра затруднен в связи наличием обертонов метильных групп, проявляющихся на этих частотах;
- после обработки ПДМС наиболее существенные изменения спектра наблюдаются в диапазоне 2500–2200 нм, что может быть использовано для контроля качества модифицированной поверхности.

Публикации

1. *Посмитная Я.С., Кухтевич И.В.* Исследование свойств поверхности материалов микрофлюидных чипов методами лежащей капли и атомно-силовой микроскопии // Сб. тезисов докладов I Всероссийского конгресса молодых ученых. Вып. 2. Труды молодых ученых. СПб., 2012. С. 358.
2. *Посмитная Я.С., Кухтевич И.В., Евстапов А.А., Тупик А.Н.* Целевая обработка поверхности материалов микрофлюидных чипов // Научное приборостроение. 2012. Т. 22. № 4. С. 38–45.
3. *Евстапов А.А., Буляница А.Л., Рудницкая Г.Е., Посмитная Я.С.* Оценка результатов полимеразной цепной реакции в реальном времени кДНК онкомаркеров СК-19 методом электрофореза на микрофлюидном чипе // Научное приборостроение. 2012. Т. 22. № 4. С. 77–85.
4. *Посмитная Я.С., Тупик А.Н., Рудницкая Г.Е., Лукашенко Т.А., Цымбалов А.И., Евстапов А.А.* Полимерные микрочиповые устройства для полимеразной цепной реакции // Сб. статей IV Международной научно-практической конференции «Высокие технологии, фундаментальные и прикладные исследования в физиологии и медицине» (15–16.11.2012, Санкт-Петербург). Т. 1. СПб., 2012. С. 252–254.
5. *Посмитная Я.С.* Исследование методов целевой обработки поверхности стеклянных и полимерных материалов в микрофлюидных устройствах // XVII Санкт-Петербургская ассамблея молодых ученых и специалистов: Сб. тезисов. СПб., 2012. С. 90.
6. *Посмитная Я.С.* Исследование методов контроля и диагностики поверхности полимерных микрочипов после целевой обработки для полимеразной цепной реакции // Сб. тезисов докладов II Всероссийского конгресса молодых ученых. Труды молодых ученых. СПб., 2013 (в печати).

Литература

1. Microfluidic substrates, market and processing trends. Market forecast update of January 2012 (Technology, application, industry and market report): http://www.i-micronews.com/upload/Rapports/Yole_Microfluidic_Substrates%20_July_2011_Report_sample-2012_update.pdf.
2. Рудницкая Г.Е., Евстратов А.А. Микрочиповые устройства для полимеразной цепной реакции. Ч. 1. Основные принципы ПЦР; конструкция и материалы микрочипов (Обзор) // Научное приборостроение. 2008. Т. 18. № 3. С. 3–20.
3. Kaltenboeck B., Wang C. Advances in real-time PCR: Application to clinical laboratory diagnostics // *Advances in Clinical Chemistry*. 2005. Vol. 40. P. 220–259.
4. Ребриков Д.В. и др. ПЦР в реальном времени. М., 2011. – 223 с.
5. Maskay I.M. Real-time PCR in microbiology: From diagnosis to characterization. Norfolk, UK, 2007. – 454 p.
6. Антонова О.С., Рудницкая Г.Е., Буляница А.Л. и др. Полимеразная цепная реакция: Приборная и методическая реализация. Обзор аналитических характеристик // Научное приборостроение. 2011. Т. 21. № 4. С. 5–21.
7. Shoffner M.A., Cheng J., Kricka L.J. et al. Chip PCR. I. Surface passivation of microfabricated silicon-glass chips for PCR // *Nucleic Acids Research*. 1996. Vol. 24. No 2. P. 375–379.
8. Cheng J., Shoffner M.A., Kricka L.J. et al. Chip PCR. II. Investigation of different PCR amplification systems in microfabricated silicon-glass chips // *Nucleic Acids Research*. 1996. Vol. 24. No 2. P. 380–385.
9. URL: <http://www.microfluidic-chipshop.com>.
10. Neuzil P., Pippner J., Hsieh T.M. Disposable real-time microPCR device: Lab-on-a-chip at a low cost // *Molecular BioSystems*. 2006. Vol. 2. No 6–7. P. 292–298.
11. URL: <http://www.genomics.agilent.com>.
12. URL: <http://www.fluidigm.com/biomark-system.html>.
13. URL: <http://appliedbiosystems.com>.
14. Shen F., Du W., Kreutz J.E. et al. Digital PCR on a SlipChip // *Lab Chip*. 2010. Vol. 10. No 20. P. 2666–2672.
15. Song S., Lee K.Y. Polymers for microfluidic chips // *Macromolecular Research*. 2006. Vol. 14. No 2. P. 121–128.
16. Chen Y., Zhang L., Chen G. Fabrication, modification, and application of poly(methyl methacrylate) microfluidic chips // *Electrophoresis*. 2008. Vol. 29. No 9. P. 1801–1814.
17. URL: www.fluidigm.com.
18. Thienot E., Domingo F., Cambriel E., Gosse C. Reactive ion etching of glass for biochip applications: Composition effects and surface damages // *Microelectronic Engineering*. 2006. Vol. 83. No 4–9. P. 1155–1158.
19. Akashi T., Yoshimura Y. Deep reactive ion etching of borosilicate glass using an anodically bonded silicon wafer as an etching mask // *Journal of Micromechanics and Microengineering*. 2006. Vol. 16. No 5. P. 1051–1056.
20. Altissimo M. E-beam lithography for micro- / nanofabrication // *Biomicrofluidics*. 2010. Vol. 4. No 2. P. 026503 (6 p).
21. Heng Qi, Xiansong Wang, Tao Chen et al. Fabrication and characterization of a polymethyl methacrylate continuous-flow PCR microfluidic chip using CO₂ laser ablation // *Microsystem Technologies*. 2009. Vol. 15. No 7. P. 1027–1030.
22. Евстратов А.А., Лукашенко Т.А., Горный С.Г., Юдин К.В. Микрофлюидные чипы из полиметилметакрилата: Метод лазерной абляции и термического связывания // Научное приборостроение. 2005. Т. 15. № 2. С. 72–81.
23. Евстратов А.А., Поздняков А.О., Горный С.Г., Юдин К.В. Формирование края микроразмерного канала в полиимидах методом лазерной абляции // Письма в Журнал технической физики. 2005. Т. 31. Вып. 13. С. 10–17.
24. Abdelgawad M., Watson M.W.L., Young E.W.K. et al. Soft lithography: Masters on demand // *Lab Chip*. 2008. Vol. 8. No 8. P. 1379–1385.

25. Kim P., Kwon K.W., Park M.C. *et al.* Soft lithography for microfluidics: A review // *Biochip Journal*. 2008. Vol. 2. No 1. P. 1–11.
26. Zhu X., Liu G., Xiong Y. *et al.* Fabrication of PMMA microchip of capillary electrophoresis by optimized UV-LIGA process // *Journal of Physics: Conference Series*. 2006. Vol. 34. P. 875–879.
27. Bartolo D., Degre G., Nghe P., Studer V. Microfluidic stickers // *Lab Chip*. 2008. Vol. 8. No 2. P. 274–279.
28. Duffy D.C., McDonald J.C., Schueller O.J.A., Whitesides G.M. Rapid prototyping of microfluidic systems in poly(dimethylsiloxane) // *Analytical Chemistry*. 1998. Vol. 70. P. 4974–4984.
29. Quake S.R., Scherer A. From micro- to nanofabrication with soft materials // *Science*. 2000. Vol. 290. P. 1536–1540.
30. Toepke M.W., Beebe D.J. PDMS absorption of small molecules and consequences in microfluidic applications // *Lab Chip*. 2006. Vol. 6. No 12. P. 1484–1486.
31. Yunus S., de Crombrughe de Loringhe C., Poleunis C., Delcorte A. Diffusion of oligomers from polydimethylsiloxane stamps in microcontact printing: Surface analysis and possible application // *Surface and Interface Analysis*. 2007. Vol. 39. No 12–13. P. 922–925.
32. Lee J.N., Park C., Whitesides G.M. Solvent compatibility of poly(dimethylsiloxane)-based microfluidic devices // *Analytical Chemistry*. 2003. Vol. 75. P. 6544–6554.
33. Wu D., Zhao B., Dai Z. *et al.* Grafting epoxy-modified hydrophilic polymers onto poly(dimethylsiloxane) microfluidic chip to resist nonspecific protein adsorption // *Lab Chip*. 2006. Vol. 6. No 7. P. 942–947.
34. Grainger D.W., Greef C.H., Gong P., Lochhead M.J. Current microarray surface chemistries // *Methods in Molecular Biology*. 2007. Vol. 381. P. 35–57.
35. Carlborg C.F., van der Wijngaart W. Sustained superhydrophobic friction reduction at high liquid pressures and large flows // *Langmuir*. 2011. Vol. 27. No 1. P. 487–493.
36. Unger M.A., Chou H.-P., Thorsen T. *et al.* Monolithic microfabricated valves and pumps by multilayer soft lithography // *Science*. 2000. Vol. 288. P. 113–116.
37. Kim K., Park S.W., Yang S.S. The optimization of PDMS-PMMA bonding process using silane primer // *BioChip Journal*. 2010. Vol. 4. No 2. P. 148–154.
38. Park J.Y., Ahn D., Choi Y.Y. Surface chemistry modification of PDMS elastomers with boiling water improves cellular adhesion // *Sensors and Actuators B: Chemical*. 2012. Vol. 173. P. 765–771.
39. Посмитная Я.С., Кухтевич И.В., Евстратов А.А., Тупик А.Н. Целевая обработка поверхности материалов микрофлюидных чипов // *Научное приборостроение*. 2012. Т. 22. № 4. С. 38–45.
40. Ихден З., Мекки А., Беккар Джеллул Сайах З. и др. Изучение удерживания ФОС и летучих органических соединений на полиметилдисилоксане методом обращенной ГХ // *Методы и объекты химического анализа*. 2013. Т. 8. № 2. С. 83–90.
41. Cai D.K., Neyer A., Kuckuk R., Heise H.M. Optical absorption in transparent PDMS materials applied for multimode waveguides fabrication // *Optical Materials*. 2008. Vol. 30. No 7. P. 1157–1161.



Рудакова Мария Сергеевна
Факультет инженерно-физический
Кафедра лазерной техники и биомедицинской оптики
E-mail: RudakovaMS@bk.ru

МИКРОПРОЕКТОРНЫЕ СИСТЕМЫ СМЕШАННОЙ И ДОПОЛНЕННОЙ РЕАЛЬНОСТИ НА ПЛОСКИХ ВОЛНОВОДНЫХ ЭКРАНАХ

М.С. Рудакова

Научный руководитель – доцент С.А. Смирнов

Введение

Объектом исследования являются микропроекторные системы смешанной и дополненной реальности на плоских волноводных экранах, предназначенные для вывода дополнительной служебной информации в поле зрения наблюдателя на фоне окружающего пространства.

Настоящий проект посвящен исследованию комбинированного метода построения плоского экрана (комбинера) на основе геометрических методов транспортировки плоскополяризованного излучения через селективно отражающие/пропускающие слои (ТР-слои).

Задача создания такого рода систем, как комбинеры, не является типовой задачей теории оптического изображения и не решена с точки зрения обеспечения сквозного расчета от плоскости предметов (с учетом характера освещающего поля) до плоскости изображения (сетчатка глаза). Поэтому актуальность данной работы связана с решением указанных проблем и созданием цветных систем индикации приборной информации. Область применения такого устройства простирается в широких пределах, дальнейшая разработка и реализация его открывает большие возможности.

Конечной целью исследования была разработка методов расчета и проектирования компактных оптических систем современных микропроекторов, предназначенных для вывода дополнительной служебной информации в поле зрения наблюдателя на фоне окружающего пространства. Были рассмотрены отдельные блоки систем индикации: блок осветителя совместно с генератором изображения и объективом и блок комбинера. Произведены теоретические расчеты, обосновывающие критерии выбора характеристик составных блоков системы индикации. Разработка архитектуры системы проведена в следующих направлениях:

- оптимизация характеристик блока формирования изображения (осветительный узел, генератор изображений, коллиматорный объектив).
- разработка моделей и алгоритмов расчета комбинера совместно с глазом.

Основными задачами проекта являлись моделирование устройства (комбинера), при котором будут выполнены все необходимые требования, предъявляемые к системам такого типа и анализ полученных распределений в плоскости изображения по ряду параметров: равномерности освещенности, яркости изображения, потерь излучения в тракте, отсутствию слепых зон в изображении. Поскольку комбинер может строиться на различных физических принципах, в работе были рассмотрены и возможные варианты его построения. Кроме того, был выполнен анализ существующих зарубежных и отечественных разработок в этой области.

Моделирование всех вариантов проводилось в программном пакете TracePro с использованием встроенной системы глаза для анализа распределений на сетчатке.

Теоретическая часть

Теоретическое обоснование характеристик блока формирования изображения

Если рассматривать авиационную область применения микропроекторных систем смешанной и дополненной реальности, то индикатор на лобовом стекле представляет собой устройство, позволяющее создавать изображение шкал, меток и другой вспомогательной

информации в поле зрения пилота. Экран этого устройства размещается на линии визирования между пилотом и лобовым стеклом кабины. Он практически прозрачен и поэтому информация видна на фоне окружающей обстановки. Изображение создается системой на бесконечности с помощью коллиматорного объектива.

Примерные требования к объективу системы индикации на лобовом стекле будут выглядеть следующим образом:

- угловое поле $30 \times 20^\circ$;
- размер выходного зрачка должен обеспечивать обзор всего поля зрения даже при смещениях глаз наблюдателя из расчетного положения;
- дисторсия объектива должна быть небольшой для обеспечения точности воспроизведения информации.

Проблему создания системы индикатора на лобовом стекле, в конструкцию которого входит рассматриваемый комбинер, можно разделить на две.

1. Создание коллиматорного объектива с характеристиками: угловое поле $2\omega \approx 30^\circ$, небольшая величина дисторсии; фокусное расстояние должно обеспечивать при использовании конкретного генератора символов необходимую величину поля. Объектив должен обладать необходимым разрешением – не меньшим чем используемый генератор изображения. Габариты объектива должны быть как можно меньше.

2. Создание виртуального дисплея (комбинер), обеспечивающего обзор внешнего пространства с коэффициентом пропускания не менее 70 % и отражающего генерируемое изображение так, чтобы обеспечить необходимую яркость символов.

Согласно патентному поиску и ряду проведенных расчетов были выявлены основные требования, которые необходимо соблюдать при построении схем и габаритном расчете.

Для общего блока коллиматорного объектива:

- необходимо обеспечивать полную засветку рабочей площадки отражательного микродисплея световым потоком от источника света. Из этого условия следует, что должны быть согласованы передняя числовая апертура конденсора и размер микродисплея;
- целесообразно, чтобы числовая апертура коллиматорного объектива была равна (или немного меньше) углового размера источника после конденсора, так как микродисплей отражает световой поток от источника зеркально;
- из условий работы комбинера следует, что диаметр выходного зрачка объектива необходимо по возможности увеличивать;
- между конденсором и микродисплеем необходимо установить призму-куб с диагональным поляризационным покрытием.

Учет перечисленных требований приводит к тому, что источник света должен располагаться в передней фокальной плоскости конденсора, микродисплей – в задней фокальной плоскости конденсора и в передней фокальной плоскости объектива, выходной зрачок должен находиться в задней фокальной плоскости объектива.

Таким образом, предварительный расчет показывает, что оптическая система коллиматорного объектива должна удовлетворять следующим основным характеристикам:

- фокусное расстояние $f'_{об} = 32,38$ мм;
- относительное отверстие $1/8$, диаметр выходного зрачка $D' = 4$ мм, числовая апертура в пространстве предметов $A = 0,0625$;
- линейное поле объектива в пространстве предметов не менее $2y = 11,13$ мм, а угловое в пространстве изображений $2\omega' = 38^\circ$;
- телецентрический ход главного луча в пространстве предметов, вынос выходного зрачка не менее 10 мм.

Основные оптические характеристики конденсора:

- фокусное расстояние $f'_к = 16$ мм;
- числовая апертура в пространстве предметов $A = 0,643$, размер освещаемой площадки в пространстве изображений $17,35 \times 13,95$ мм;
- линейное поле конденсора в пространстве предметов 2×2 мм.

Что же касается оптической системы ввода излучения в комбинер, то вариантов может быть несколько. Конденсор в узле осветителя отражательного микродисплея должен соответствовать следующим основным характеристикам:

- фокусное расстояние $f_k = 16$ мм;
- числовая апертура в пространстве предметов $A = 0,643$, размер освещаемой площадки в пространстве изображений $17,35 \times 13,95$ мм;
- линейное поле конденсора в пространстве предметов 2×2 мм.

Вследствие установки освещаемого дисплея в заднюю фокальную плоскость конденсора и особенностей отражения от поверхности дисплея дополнительным требованием является телецентрический ход главного луча в пространстве предметов.

Конденсор может быть решен как двухлинзовая система, как линза Френеля или возможны трех-, четырехлинзовые системы. Но основное требование к качеству работы конденсора – равномерность распределения светового потока от источника излучения по всей площадке дисплея. При угле охвата более 80° конденсор обычно делают трех- или четырехлинзовым, так как в таких системах достигается хорошая собираемость светового потока. Расчет двухлинзового конденсора необходим с точки зрения уменьшения массы оптических элементов. Линза Френеля в качестве конденсора имеет предпочтение по массогабаритным характеристикам перед линзовыми системами. Но с ее применением сложнее добиться высокой степени равномерности засветки поверхности микродисплея.

Основные варианты построения плоских волноводных экранов

Рассмотрим способы построения плоских волноводных экранов на основе анализа существующих зарубежных и отечественных разработок.

Возможным способом обеспечения избирательной отражающей способности (ИЛС) является применение отражающих голограмм. Такая голограмма записывается в фоточувствительной среде двумя когерентными лазерными пучками. Запись создает синусоидальное изменение коэффициента отражения по толщине слоев материала с периодом, равным половине длины волны света, который нужно отражать. Записывающий материал очень хрупок, поэтому его необходимо защищать от внешних механических воздействий, из-за этого такой комбинер получается толстым и тяжелым. К тому же голограмма является высокоэффективным отражателем только для света, падающего под определенным углом. При отклонении угла падения света от оптимального, отражающая способность резко падает.

Предпринимаются попытки создания ИЛС с синтезированной голограммой. Такая голограмма представляет собой ряд слоев прозрачного диэлектрика с разными коэффициентами отражения. Когда коэффициент отражения изменяется от слоя к слою по синусоидальному закону, то покрытие ведет себя как голограмма, отражая только свет выбранной длины волны. Диэлектрики являются прочными материалами, поэтому они не требуют дополнительной защиты и могут быть нанесены на внешнюю поверхность субстрата. Это позволяет уменьшить массу комбинера. Синтезированная голограмма имеет более высокий коэффициент отражения, что позволяет получить эффективное отражение в более широком диапазоне углов падения света.

К системам призматического типа можно отнести устройства, которые представляют собой поверхность типа линзы Френеля. Профиль ступеней подбирается таким образом, чтобы свет, идущий от генератора изображений, отражался от профиля по направлению к наблюдателю, а свет, идущий от внешнего пространства, проходил практически без изменений. Указывается [1], что в таком случае можно достичь необходимого соотношения: от внешнего пространства проходит 70 % света, 30 % отражается от профиля. Профиль штрихов несимметричной линзы, рассчитанный соответствующим образом, позволяет не только наблюдать изображение и внешнюю среду одновременно, но и не дает паразитных и двойных изображений.

Также существует устройство, которое представляет собой световод, на некоторой части которого нанесен профиль из зеркал (решетка), позволяющий вывести часть света по направлению к наблюдателю [2]. На профиль нанесено полупрозрачное покрытие, так что оно

частично пропускает свет, который распространяется далее по световоду за счет полного внутреннего отражения, а частично отражает свет в сторону наблюдателя. В такой схеме также можно подобрать положение глаз наблюдателя таким образом, чтобы освещенность в пределах поля менялась не более чем на 30 %, что обеспечивает комфортное наблюдение.

Существует еще много способов реализации экранов комбинера и здесь представлена лишь часть из них. С учетом проведенного патентно-информационного обзора и предварительных оценок можно сделать следующие выводы.

1. Комбинер может быть выполнен как с применением голографических элементов, так и без них. Голографические элементы имеют некоторые существенные недостатки, в частности их отражательная способность сильно зависит от угла падения и длины волны света, что препятствует созданию цветных широкоугольных систем.

2. Комбинер в виде линзы Френеля подходит для многих применений, однако поскольку линза Френеля обладает оптической силой, то она может исказить восприятие внешнего пространства, что нежелательно или даже недопустимо при применении в гражданской авиации.

3. В системах с комбинером, который представляет собой набор элементов с селективно отражающими покрытиями, предложено выводить излучение за счет отражения от покрытий, нанесенных на грани призматических элементов составного экрана комбинера.

4. Комбинер, выполненный на основе волноводной подложки с широким полем зрения, а также с возможностью использования искривленного субстрата, обеспечивает высокую прозрачность (до 90 %) [3].

Была разработана расчетная модель основного элемента системы – плоского оптического световодного устройства (комбинера). Оно состоит из двух элементов: мультипликатора горизонтального поля (МГП) и вертикального поля (МВП). Мультипликаторы представляют собой плоскопараллельные пластины толщиной d с n избирательно отражающими поверхностями. ТР-слои расположены по высоте H с шагом T (период). Угол наклона всех слоев относительно наружных поверхностей равен α . На рис. 1 соответственно представлены возможные варианты реализации мультипликатора в зависимости от выбора угла наклона α ТР-слоя (или угла в основании входной прямоугольной призмы β):

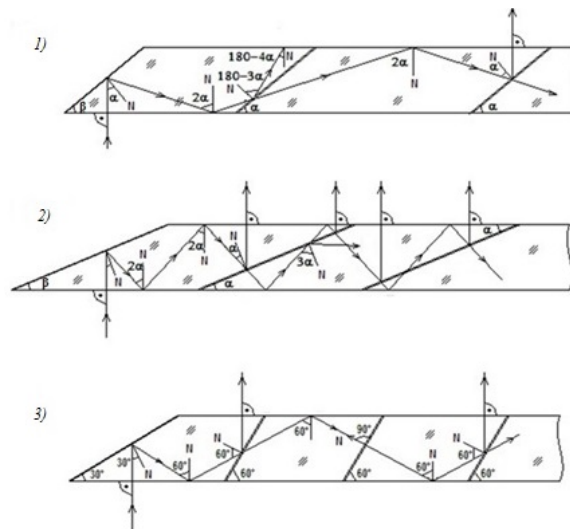


Рис. 1 Мультиплицирование потока излучения на входных гранях призматических элементов

- вариант 1 – наклон ТР-слоя под углом $\alpha = \beta > 30^\circ$;
- вариант 2 – наклон ТР-слоя под углом $\alpha = \beta < 30^\circ$;
- вариант 3 – наклон в основании прямоугольной призмы под углом $\beta = 30^\circ$.

Мультипликаторы осуществляют тиражирование лучей в горизонтальной и вертикальной плоскостях соответственно.

Результаты моделирования

Результаты, полученные при первоначальном моделировании в программном пакете TracePro, оказались неудовлетворительными. Из предложенных вариантов только два подходили по требованиям, предъявляемым к данным системам, но при этом равномерность освещения экрана в обоих случаях сильно отличалась от желаемой. Согласно проведенному анализу хода лучей было выявлено, что слои мультипликаторов не выполняют равномерного деления светового потока. При этом, помимо потерь, связанных с поглощением излучения материалом, существует еще два рода потерь:

- Flux Threshold (пороговые потери). Они сильно сказываются именно на МВП. Их максимальная величина для первого блока составила 35 Вт (для варианта 23,2°/23,2° – 271 Вт), чем и может быть вызвано такое резкое падение коэффициента пропускания на блоке;

- Escaped Model. Это потери, вызванные выходом лучей за грани объекта, в нашем случае за счет несоблюдения закона полного внутреннего отражения. Для МГП они составляют сотые доли, а для МВП эта величина достигла 54 Вт (начальная мощность 271 Вт). При увеличении количества лучей до 202021 (начальная мощность 1 Вт), величина этих потерь составила 0,19 Вт.

Для МВП прослеживалась одинаковая тенденция по сохранению величины пропускающей способности слоев. Данные можно улучшить, если подобрать оптимальные пороговые потери и разобраться, почему происходит выход излучения за пределы объекта. Ведь именно на втором блоке МВП ухудшается пропускающая способность, что ведет к неравномерности освещенности.

Для МГП таких проблем с потерями нет, так как там преобладают потери на поглощение, и других нет или они составляют очень малые величины. Интерес представляет причина, по которой для двух вариантов реализации мультипликаторов пропускающая способность различна. Вероятно, это связано с неоднократным падением излучения на слои для варианта 23,2°/23,2°. Но тогда неравномерное распределение для варианта 30°/23,2° может объясняться только несоответствием углов падения с профилирующим покрытием.

Таким образом, с целью устранить потери, приводящие к столь неравномерному распределению освещенности и падению пропускающей способности слоев, было проведено повторное моделирование.

Моделирование микропроекторных систем индикаторов

Построение экрана комбинера, как и прежде, было выполнено в программном пакете TracePro. Источник излучения имел круглое сечение диаметром 3 мм. Начальное количество лучей составило 119401 общей мощностью 1 Вт. Длина волны 0,589 мкм. Показатель преломления материала 1,65104 (стекло ТК-21). Поле по горизонту составляет $\pm 11,3^\circ$, по вертикали $\pm 8,5^\circ$. Для удобства моделирования ТР-слои, которые должны отражать все падающее на них излучение, выполнены в виде зеркал. На них было нанесено профилирующее покрытие, обеспечивающее необходимый коэффициент отражения в заданном диапазоне углов падения.

Анализ распределения плотности мощности осуществлялся по яркости изображения (к дисплеям с диагональю 20 см предъявляется требование по яркости > 150 нит), равномерности освещения изображения, наличию слепых зон и потерям в оптическом тракте.

Предполагаемая яркость экрана микродисплея $B = 6\ 000$ нит при величине излучающей площадки $S = 7\ \text{мм}^2$ и полном апертурном угле $14,4^\circ$ (в стерadiansах $\Omega = 0,0632\ \text{ср}$). Примем величину зрачка глаза наблюдателя равной $d = 5\ \text{мм}$ ($S' = 19,625\ \text{мм}^2$) и угловое поле $22,6 \times 17^\circ$ (в стерadiansах $\Omega' = 0,117\ \text{ср}$). Коэффициент пропускания зрачка τ рассчитывается как отношение прошедшего количества лучей к начальному. Во всех случаях зрачковая область располагалась на том же расстоянии от устройства (мультипликатора), что и экран.

Яркость изображения B' рассчитывалась по выражению:

$$B' = \tau B S \Omega / S' \Omega'. \quad (1)$$

Результаты расчета модели комбинера с параметрами: $\alpha = 37^\circ/23$; $d = 4/4$; $n = 4/3$. Здесь и далее в числителе приводятся данные для мультипликатора МГП, а в знаменателе – для МВП.

На рис. 2а представлено распределение плотности мощности, которое получено на экране, расположенном на расстоянии 26,8 мм от устройства. В данном случае количество лучей после прохождения устройства увеличилось в 1,7 раза. Мощность уменьшилась в 2,36 раз. Смещение центра зрачка относительно центра экрана составило: $x = -3$ мм, $y = -5,5$ мм. Мощность в зрачковой области 0,024 Вт (рис. 2б). Коэффициент пропускания для зрачка: $\tau = 11034/119401 = 0,092$ (или 9,2 %). Яркость изображения для наблюдателя составит $B' = 106,35$ нит.

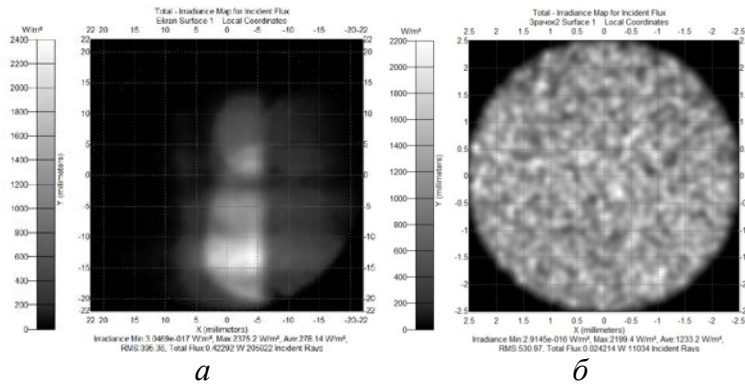


Рис. 2. Распределение плотности мощности для модели $\alpha = 37^\circ/23$; $d = 4/4$; $n = 4/3$:
а – на экране; б – в зрачковой области

По яркости данный вариант не совсем плох, нет явных слепых зон, но при этом нет и равномерности освещенности изображения. Однако в данном случае мы имеем слабое увеличение количества лучей после прохождения комбинера, что указывает на неполное их тражирование.

Результаты расчета модели комбинера с параметрами: $\alpha = 23,2^\circ/23,2^\circ$; $d = 4/4$, $n = 3/3$. На рис. 3а представлено распределение плотности мощности, которое получено на экране, расположенном на расстоянии 24,8 мм от устройства. Количество лучей, достигшее экрана, увеличилось в 6,2 раза. Это произошло в результате мультиплицирования их на TP-слоях. Мощность уменьшилась в 1,42 раза. В этом варианте смещение центра зрачка относительно центра экрана составило: $x = -10,3$ мм, $y = 10,7$ мм. Мощность в зрачковой области 0,0034 Вт (рис. 3б). Коэффициент пропускания для зрачковой области: $\tau = 7452/119401 = 0,0625$ (или 6,25 %).

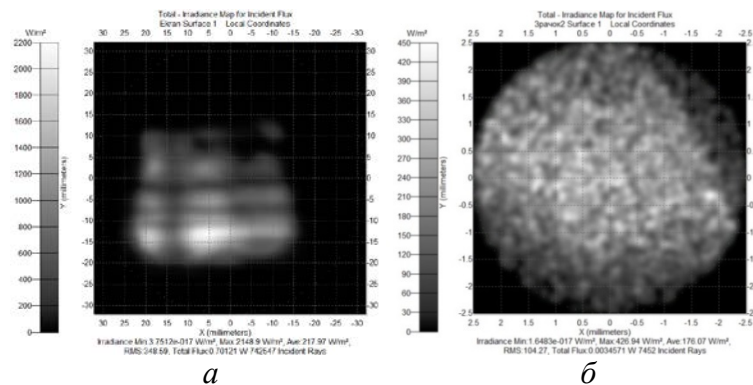


Рис. 3. Распределение плотности мощности для модели $\alpha = 23,2^\circ/23,2^\circ$; $d = 4/4$, $n = 3/3$:
а – на экране; б – в зрачковой области

Яркость изображения для наблюдателя составляет: $B' = 72,36$ нит. Она упала в 1,47 раз по сравнению с вариантом $\alpha=37^\circ/23,2^\circ$. Кроме того имеется слепая зона, но само по себе распределение достаточно равномерное. По сравнению с предыдущими данными, здесь явно уменьшились потери, исчезли слепые зоны.

Результаты расчета модели комбинера с параметрами: $\alpha = 30^\circ - 60^\circ/30^\circ - 60^\circ$, $d = 4/6$, $n = 6/5$. Здесь угол $\alpha = 30^\circ$ лежит в основании призмы (β), а угол $\alpha = 60^\circ$ у ТР-слоев.

На рис. 4а представлено распределение плотности мощности, которое получено на экране, расположенном на расстоянии 13,8 мм от устройства. Количество лучей, достигшее экрана, уменьшилось в 1,15 раза. Мощность уменьшилась в 1,42 раза. Смещение центра зрачка относительно центра экрана: $x = -3$ мм, $y = 2,9$ мм. Мощность в зрачковой области равна 0,14 Вт (рис. 4б). Коэффициент пропускания для зрачковой области: $\tau = 20988/119401 = 0,175$ (или 17,5 %).

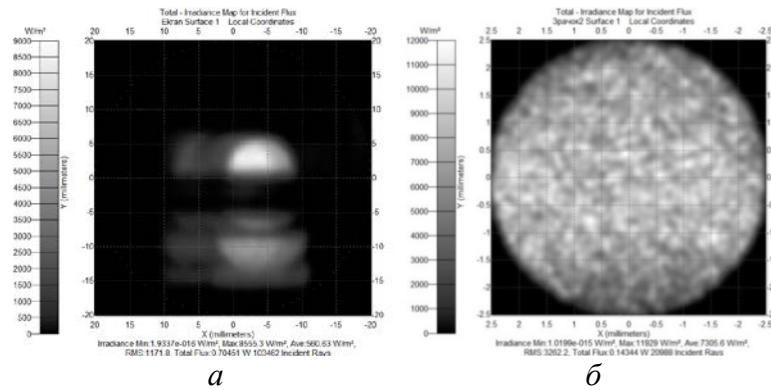


Рис. 4. Распределение плотности мощности для модели $\alpha = 30^\circ - 60^\circ/30^\circ - 60^\circ$, $d = 4/6$, $n=6/5$:
 а – на экране; б – в зрачковой области

Яркость изображения для наблюдателя составит $V' = 202,3$ нит. Здесь мы имеем наилучший результат по яркости, но не происходит тиражирования лучей (количество лучей после прохождения уменьшилось).

Результаты расчета модели комбинера с параметрами: $\alpha = 30^\circ/23,2^\circ$, $d = 4/4$, $n = 3/3$. На рис. 5а представлено распределение плотности мощности, которое получено на экране, расположенном на расстоянии 26,8 мм от устройства. Количество лучей, достигшее его, увеличилось в 6,85 раза. Мощность уменьшилась в 1,3 раза. Смещение центра зрачка относительно центра экрана составляет: $x = -2,4$ мм, $y = -14,5$ мм. Мощность в зрачковой области 0,016 Вт (рис. 5б). Коэффициент пропускания для зрачковой области: $\tau = 23968/119401 = 0,2$ (или 20 %).

Яркость изображения $V' = 231,2$ нит. Результат является самым приемлемым, слепые зоны отсутствуют, но опять наблюдается небольшая неравномерность заполнения экрана.

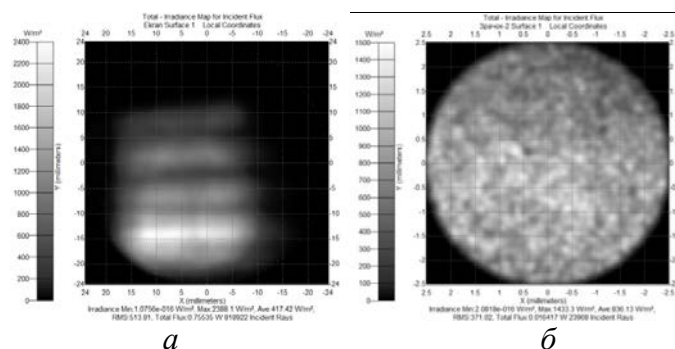


Рис. 5. Распределение плотности мощности для модели $\alpha = 30^\circ/23,2^\circ$, $d = 4/4$, $n = 3/3$:
 а – на экране; б – в зрачковой области

Таким образом, в большинстве случаев в результате мультиплицирования лучей на ТР-слоях количество их на выходе из устройства увеличилось, а для варианта $30^\circ-60^\circ/30^\circ-60^\circ$ – уменьшилось. Это произошло в силу того, что ТР-слои в этом случае не работают надлежащим образом, то есть не происходит расщепления луча на прошедший и падающий. Значит, этот вариант не может подходить для реализации, даже при том условии, что имеет неплохой коэффициент пропускания и яркость. Стоит отметить, что находились лучи, выходящие за пределы устройства (вследствие несоблюдения закона полного внутреннего отражения).

Для варианта $37^\circ/23,2^\circ$ имеется неплохой результат, но при более детальном рассмотрении видно, что для МГП не происходит тиражирования всех лучей, что приводит к ухудшению качества изображения, а значит, мультипликатор не работает в полной мере. Кроме того, наблюдаются лучи, выходящие за пределы мультипликатора.

Для двух оставшихся случаев тиражирование лучей происходит нормально и результат также вполне удовлетворителен, но для случая $30^\circ/23,2^\circ$ имеется более равномерное заполнение экрана, как по освещенности, так и по отсутствию слепых зон.

Как уже было замечено, на сегодняшний день проведение расчетов с существующими специализированными программными продуктами затруднено большими массивами лучей и поверхностей, которые необходимо рассмотреть при моделировании. В программе TracePro также невозможно проведение большого количества лучей, и не все они там отображаются, поэтому трудно чисто визуально найти плоскость для расположения зрачковой области. Выбиралась она исходя из наибольшей освещенности полученного на экране изображения. Но результат по заполнению экрана был достигнут не совсем удовлетворительный, так как равномерности распределения плотности мощности не наблюдается, то есть не соблюдается равномерное деление излучения ТР-слоями.

Анализ хода лучей

Для выявления причин неравномерного деления потока ТР-слоями был проведен анализ хода лучей для варианта $23,2^\circ/23,2^\circ$. По исходным данным, первый слой должен пропускать 67 %, второй – 50 % падающего излучения. Третий слой отражает все излучение.

Пусть мощность излучения каждого луча 1 Вт, тогда при полном количестве лучей 271, имеем на входе 271 Вт. Данные по анализу представлены в Таблице.

Таблица

Мощность излучения на различных ТР-слоях (блоках) на выходе их мультипликаторов

МГП	Мощность, Вт		МВП	Мощность, Вт	
	$23,2^\circ/23,2^\circ$	$30^\circ/23,2^\circ$		$23,2^\circ/23,2^\circ$	$30^\circ/23,2^\circ$
Призма входная	290,5	265,9	Призма входная	300,5	301,5
Блок 1	376,9	280,6	Блок 1	391,1	385,6
Блок 2	318,7	182,1	Блок 2	313,2	309,2
Блок 3	155,4	95,8	Блок 3	150,7	153,6

Таким образом, для МГП пропускание первого слоя составило 84,6 % (вместо 67 %), второго слоя – 48,7 % (вместо 50 %). Для МВП пропускание первого слоя 80,1 % (вместо 67 %), второго – 48,1 % (вместо 50 %). Стоит отметить, что для МГП количество лучей, пришедших на входную призму больше, чем заявленных изначально. Это связано с тем, что крайние лучи успели отразиться от призмы и граней мультипликатора и повторно «упасть» на входную призму. Та же самая картина наблюдается и для первого блока мультипликатора. Пропускание блоков хоть и отличается от заявленных, но, тем не менее, результат по коэффициенту пропускания для второго слоя близок к желаемому. Для первого же слоя коэффициент пропускания выше. Вследствие этого на экране мы получаем яркие пятна, которым и соответствует большой коэффициент пропускания. Такое положение может объясняться неправильным ходом лучей в мультипликаторе.

Также был рассмотрен случай с 119401 лучом. Полученная картина распределения плотности мощности практически ничем не отличается от предыдущего варианта, то есть пропускание слоев сохраняется.

Для того чтобы строить какие-либо предположения или делать выводы, необходимо провести анализ хода лучей для варианта $30^\circ/23,2^\circ$. Пусть мощность излучения каждого луча 1 Вт, тогда при количестве лучей 271, на входе имеем 271 Вт. Данные по анализу представлены в Таблице.

Итак, для МГП пропускание первого слоя составило 64,1 % (вместо 67 %), второго – 52,6 % (вместо 50 %). Для МВП пропускание первого слоя 80,1 % (вместо 67 %), второго – 49,6 % (вместо 50 %).

В данном случае количество лучей на входной призме МГП не увеличивается. Это можно объяснить другим углом наклона призмы, что соответственно не приводит к двойному падению излучения на слой. Для второго слоя мультипликаторов прослеживается та же картина с пропускающей способностью, а для первых слоев мы замечаем различный коэффициент пропускания. Здесь, в отличие от варианта $23,2^\circ/23,2^\circ$, пропускание в МГП близко к идеальному.

При рассмотрении случая с 119401 лучом мощностью 1 Вт полученная картина также не меняется. Но в данном случае уменьшились потери:

- Flux Threshold (пороговые потери). Максимальная величина их для первого блока МВП составила 15 Вт (для варианта $23,2^\circ/23,2^\circ$ – 271 Вт – 271 луч), но для случая 119401 луча эта величина является несущественной, а значит, уже не может так сильно влиять на результат. Потери снизились в 2,3 раза по сравнению с первоначальными результатами;

- Escaped Model. Для МГП потери такого рода составляют тысячные доли (для 271 луча), а для МВП она равна 0,59 Вт (мощность 271 Вт). При увеличении количества лучей до 119401 (мощность 1 Вт), величина этих потерь мала и они не могут существенно влиять на потерю мощности и неравномерность освещения.

Итак, вопрос по поводу потерь можно считать решенным, так как их величины не могут влиять на общую картину распределения плотности мощности. Но вопрос о равномерности распределения все равно остается открытым. Конкретно по мультипликаторам:

- для МВП прослеживается тенденция по сохранению величины пропускающей способности первого слоя (около 80 %). Однако данные полученные при анализе не согласуются с заданными (67 %). Скорее всего, именно по этой причине распределение плотности мощности неравномерное;

- для МГП особых проблем с потерями нет. Для варианта $30^\circ/23,2^\circ$ величина коэффициента пропускания является удовлетворительной, а вот для случая $23,2^\circ/23,2^\circ$ все-таки далека от желаемой.

Так как из всех вариантов наибольший интерес представляет $30^\circ/23,2^\circ$, то система была промоделирована еще раз, но теперь на комбинер падал параллельный пучок лучей (рис. 6а). Количество лучей после прохождения устройства увеличилось в девять раз, что и следовало ожидать. Так как количество слоев в каждом мультипликаторе составляет три, то их произведение, равное девяти, должно в случае параллельного падения пучка лучей, давать в девять раз большее количество лучей на выходе. Это говорит о том, что тиражирование происходит правильно. Но, здесь имеются более яркие пятна, что говорит о неравномерности деления светового потока. Кроме того, ожидалось увидеть картину из девяти пятен, но из-за многократного падения лучей, а также за счет переложения пятен друг на друга, картина выглядит немного другим образом. Если моделировать составные части комбинера, то можно проследить какой именно элемент приводит к неравномерности освещения (рис. 6б, 6в). Для МГП имеется равномерное распределение освещенности, но из-за неоднократного отражения наблюдается перекрытие пятен. Для МВП также наблюдается переложение пятен и для первого слоя – более яркое пятно.

При построении всех вариантов с параллельным пучком лучей получают картины из отдельных пятен, архитектура построения которых визуально похожа на картины, полученные с расходящимся пучком, но имеются те же слепые зоны (рис. 6), однако количество лучей больше. Для варианта $23,2^\circ/23,2^\circ$ количество лучей после прохождения устройства увеличилось в пять раз, что говорит о том, что четыре слоя не являются рабочими. Так как в данном случае были изменены лишь покрытия слоев, то следует, что такое распределение получается из-за неправильного хода лучей, который ведет к изменению распространения лучей в дальнейшем направлении. Для того чтобы получить иной ход лучей, необходимо изменить период расположения ГР-слоев или изменить расположение источника излучения относительно входной призмы МГП. Варьируя эти величины можно получать различные варианты расположения пятен плотности мощности на экране.

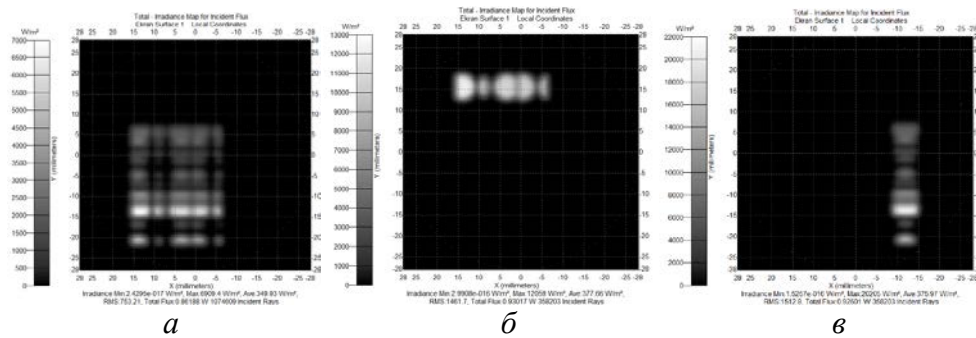


Рис. 6. Распределение плотности мощности для модели $\alpha = 30^\circ/23,2^\circ$, $d = 4/4$, $n = 3/3$:
 а – на экране; б – для МГП; в – для МВП

Расчет комбинера совместно с глазом

Изображение, которое получено на экране, не дает возможности в полной мере судить о том, что в итоге будет видеть человек. Анализируемое изображение надо рассматривать на сетчатке. Поэтому в систему и, соответственно, программу TracePro была встроена модель глаза, для которой уже были определены все его необходимые параметры. Моделирование мультипликаторов проводилось с прежними параметрами и для тех же четырех вариантов, несмотря на то, что два из них нельзя считать удовлетворительными. В статье представлено распределение плотности мощности на сетчатке лишь для двух вариантов. Система глаза располагалась на том же месте, что и зрачковая область.

Как известно, если плоскопараллельный пучок излучения падает на сетчатку глаза, то можно увидеть уменьшенное изображение. Моделирование системы глаза показало, что если на сетчатку направлять расходящийся пучок излучения, то на ней формируется увеличенное изображение. Таким образом, ожидаемый результат на сетчатке глаза – это увеличенное изображение с микродисплея.

Результаты расчета модели комбинера с параметрами $\alpha = 23,2^\circ/23,2^\circ$, $d = 4/4$, $n = 3/3$.

На рис. 7а представлено распределение плотности мощности, полученное на сетчатке глаза, удаленной от последней поверхности комбинера на расстоянии 45,1 мм. Количество лучей, достигшее сетчатки, уменьшилось в 4,4 раза. Мощность на сетчатке составила 0,0146 Вт. В данном случае наблюдается расплывчатое изображение, смещенное от центра сетчатки. Это единственный случай, при котором на сетчатку глаза пришло уменьшенное количество лучей.

Результаты расчета модели комбинера с параметрами $\alpha = 30^\circ/23,2^\circ$, $d = 4/4$, $n = 3/3$.

На рис. 7б представлено распределение плотности мощности, полученное на сетчатке глаза, расположенной на расстоянии 47,1 мм от последней поверхности комбинера. Количество лучей, достигшее ее, увеличилось в 1,47 раза. Мощность на сетчатке составила 0,139 Вт.

Таким образом, если рассматривать два варианта $23,2^\circ/23,2^\circ$ и $30^\circ/23,2^\circ$, то для первого имеем неудовлетворительный результат: малая мощность и уменьшенное количество лучей, а второй удовлетворяет по всем параметрам. Естественно, при добавлении модели глаза необходимо учитывать и различные его динамические свойства, в том числе и психофизические, которые в полной мере описывают то, что видит человек. В моделируемом варианте система глаза статична, но, тем не менее, это дает возможность проанализировать распределения плотности мощности на сетчатке глаза и тем самым выявить необходимые закономерности в построении изображения и сравнить его с тем, которое получается на микродисплее.

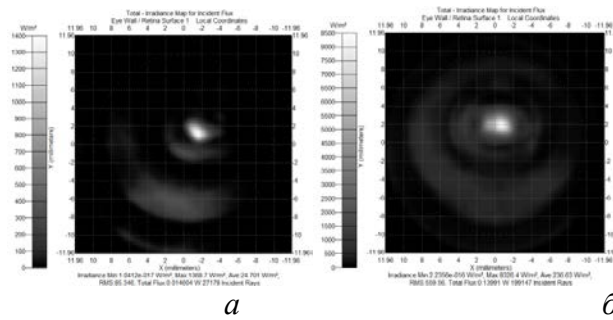


Рис. 7. Распределение плотности мощности на сетчатке для модели $\alpha = 30^\circ/23,2^\circ$, $d = 4/4$, $n = 3/3$: а – $\alpha = 23,2^\circ/23,2^\circ$; б – $\alpha = 30^\circ/23,2^\circ$

Заключение

В результате проделанной работы были выявлены основные требования и характеристики, налагаемые на составные элементы устройства, определена его архитектура, разработаны оптические схемы составных элементов: конденсора, коллиматорного объектива и комбинера; выполнены необходимые габаритно-абберационные расчеты и оценка качества оптических систем.

Согласно требуемым характеристикам были разработаны несколько вариантов построения системы. В первом варианте используется двухлинзовый конденсор с параметрами: $f = 18$ мм, $F = 8$; коллиматорный объектив: $f = 32,38$ мм, $F = 8$, $D = 4$ мм; комбинер из полистирола толщиной $d = 5$ мм. Во втором варианте – четырехлинзовый светосильный конденсор: $f = 10$ мм, $F = 0,5$; коллиматорный ортоскопический объектив: $f = 32,12$ мм, $F = 5$, $D = 6,4$ мм; комбинер из оптического стекла марки К8 или его аналога с меньшими потерями N-BK7 толщиной $d = 8$ мм.

Был проведен анализ качества изображения системы и оценено соответствие ее характеристик габаритным и абберационным требованиям. Таким образом, рассмотренный объектив удовлетворяет всем требованиям и обладает новизной. Предложенная схема объектива дает потенциальную возможность по дальнейшему увеличению углового поля и относительного отверстия (до 40° и F4). Сам объектив имеет простую пятилинзовую схему и технологичен, как по выбору марок стекол, так и по допускам на его изготовления, он легкий и компактен.

Основные выводы по расчету конденсора: полученные значения поперечных аббераций для оптической системы позволяют сделать вывод, что конденсор удовлетворяет всем требованиям. К его недостатку можно отнести наличие асферической поверхности. Возможно, более оптимальным был бы конденсор на основе линзы Френеля, что могло бы позволить сократить габариты и вес изделия.

На основе предложенного варианта построения системы был сделан расчет влияния параметров и допусков объектива, что позволило оформить на него оптический выпуск, схему оптическую принципиальную канала объектива и разработать чертежи оптических элементов.

Что касается непосредственно экрана комбинера, то отсутствие в нем голографических элементов является преимуществом, так как они не позволяют создавать цветные системы. Голографические оптические элементы могут представлять собой решетки поверхностного рельефа. Они хорошо подходят для масштабного производства. Напротив, решение с частично отражающими зеркалами, используемыми в рассмотренной конструкции комбинера, обычно предполагает более сложные технологические процессы. Чем их больше, тем этот процесс сложнее и дороже. Число зеркал возрастает с уменьшением толщины световода. Это означает, что последнее решение не конкурентоспособно, если требуется создать световод толщиной, например менее 3 мм. С другой стороны, для световодов большей толщины (более 5 мм) во избежание неоднородной яркости поля обзора может возникнуть необходимость изменять коэффициент отражения покрытия не только между соседними зеркалами, но и на отдельно взятом зеркале [4].

Таким образом, если обобщить все полученные данные моделирования, то можно сделать вывод, что из четырех вариантов, наиболее оптимальным являются два: $23,2^\circ/23,2^\circ$ и $30^\circ/23,2^\circ$. Но в первом имеется слепая зона в распределении плотности мощности, что затрудняет определение наиболее освещенной зоны и неблагоприятно сказывается на данных моделирования. Кроме того, при рассмотрении этих систем с учетом модели глаза, видно, что на сетчатке наиболее оптимальное (равномерное и симметричное) распределение соответствует варианту $30^\circ/23,2^\circ$.

Проблемы, связанные с потерями, были выявлены и решены после анализа хода лучей в мультипликаторах. Неразрешенной осталась проблема неравномерного распределения освещенности. Ее причина – неправильный ход лучей, в результате которого часть из них отклоняется от желаемого направления. Поэтому следующий шаг в исследовании – модификация уже существующих параметров комбинера: периода следования слоев T , толщины устройства d , расположения источника излучения. Особенно в данном случае интересуют МВП, именно для этого элемента необходимо пересмотреть рабочие зоны в профилирующих покрытиях.

Аналогичных расчетных схем систем комбинера такого рода нет, поэтому данная может использоваться при моделировании похожих. Результаты проделанной работы предполагается использовать в учебном процессе в качестве дополнения и модернизации учебных курсов по дисциплинам «Теория и методы проектирования оптических систем», «Прикладная физика» и «Формирование и управление параметрами лазерного излучения».

Был выполнен патентный поиск по способам построения плоских волноводных экранов, выявлены их преимущества и недостатки; произведены расчеты и обоснование характеристик для блока формирования изображения. Полностью промоделирован комбинер в различных вариациях своего построения, совместно с глазом и при параллельном ходе лучей. Были сделаны соответствующие выводы, согласно которым блок формирования изображения полностью удовлетворяет заданным требованиям. Световодный элемент нуждается в некоторой доработке, хотя полученные данные во много раз лучше предыдущих. Были решены вопросы с потерями; выявлены основные проблемы, связанные с неравномерностью распределения плотности мощности и предложены варианты по их устранению. Возможно, будет целесообразно рассчитать и промоделировать и другие варианты построения экрана, например с использованием волноводной подложки.

Публикации

1. Рудакова М.С. Изображающие свойства двухкоординатного комбинера на плоских световодных элементах призмного типа // Сб. тезисов докладов конгресса молодых ученых. Вып. 2. СПб., 2013. С. 92–93.
2. Гримм В.А., Рудакова М.С., Смирнов С.А. Микропроекторные системы смешанной и дополненной реальности на плоских волноводных экранах // Известия вузов. Приборостроение. 2013. Т. 56. Вып. 9. С. 84–89.
3. Бахолдин А.В., Васильев В.Н., Гримм В.А., Гришина Н.Ю., Романова Г.Э., Рудакова М.С., Смирнов С.А. Отчет по НИР № 411392 «Виртуальные дисплеи микропроекторных систем на плоских оптических световодных элементах».

Литература

1. Патент EP 2372407 A2 (Lisa Masuda, Shigeru Shimakawa, Naotada Okada, Kazuo Horiuchi, Haruhiko Okumura, Takashi Sasaki, Hiromi Suzuki).
2. Патент US 6825987 B2 (Piermario Repetto, Elena Borello, Stefano Bernard).
3. Патент US 2010/0157400 A1 (Fedor Dimov, Redondo Beach; Tin Aye, Mission Viejo; Kevin Yu, Temple City; Svetlana Soboleva, Torrance; Khin Swe Yin, Alhambra; Myo Kyaw, Rosemead; Dmitry Volochenko, Torrance). Holographic substrate-guided wave-based see-through display.
4. Бахолдин А.В., Васильев В.Н., Воронцов Д.Н. и др. Отчет по НИР № 211153 «Разработка составных частей комплекса унифицированных базовых элементов бортового оборудования открытой архитектуры на основе интегрированной модульной авионики с применением

отечественного коммутатора шины ARINC – 664 в 2011 г.: вычислительные и интерфейсные средства, средства индикации и управления, система информационного обеспечения экипажа».



Сергеев Максим Михайлович
Факультет инженерно-физический
Кафедра лазерных технологий и экологического приборостроения
E-mail: maks-sv-32@yandex.ru

ЛАЗЕРНО-ИНДУЦИРОВАННАЯ ЛОКАЛЬНАЯ МОДИФИКАЦИЯ ПОРИСТЫХ СТЕКОЛ

М.М. Сергеев

Научный руководитель – профессор Е.Б. Яковлев

Введение

Одним из основных направлений в развитии современных лазерных технологий является создание модифицированных областей (МО) в стекле и стеклокерамическом материале с отличными от исходного материала оптическими свойствами как на поверхности, так и в объеме с размерами 1–500 мкм. Такие МО способны локализовать, отражать, преломлять, а также фокусировать или рассеивать падающее излучение [1–5].

Создание МО находит широкое применение. Это технологии по разработке микрооптических элементов (МОЭ), лаборатории на чипах, микрохимические реакторы, фоточувствительные оптические элементы, микроэлементы оптоэлектронных приборов и сенсоров, МОЭ для волоконной оптики, фотоники и оптоинформатики [6]. Технологии создания всех вышеперечисленных микроэлементов широко развиваются в настоящее время и используются при разработке передовых информационных, оптоэлектронных и оптических устройств. В связи с этим перспективным направлением является исследование процессов локальной лазерно-индуцированной модификации в объеме и на поверхности стеклообразных материалов, а также разработка новых методов формирования МОЭ, основанных на полученных результатах.

Для создания в объеме стеклообразных материалов сложных трехмерных структур необходимо обеспечить проникновение лазерного излучения на заданную глубину. Для этого в настоящее время в основном используют источники мощного лазерного излучения с ультракороткой длительностью импульса (10^{-9} – 10^{-15} с), с высокой частотой их следования (10–100 МГц) и длиной волны излучения ~ 800 нм, для которой стекло оптически прозрачно [6].

Новые методы обработки оптически прозрачных материалов, появившиеся в последнее время, основаны на уменьшении мощности лазерного излучения и увеличении длительности импульса до 10^{-6} – 10^{-9} с. Традиционно интересными для формирования МО на основе методов лазерной обработки остаются пористые стекла (ПС) – продукты сквозного химического травления одной из фаз (менее стойкой) щелочноборосиликатных стекол [7, 8]. Применение пластин ПС в качестве матриц для формирования на их поверхности и по всему объему МО различного размера и формы с измененными оптическими характеристиками представляется перспективным [9–12]. В этом случае требуется мощность лазерного излучения на несколько порядков меньшая, чем у лазеров с фемтосекундной длительностью импульса.

Пропитка пластин ПС веществом с высокой степенью поляризации, предшествующая локальной лазерно-индуцированной модификации в объеме, использовалась для подтверждения предположения [10, 11] о массопереносе вещества, частично заполняющего

каналы ПС под действием лазерного излучения, сфокусированного в ее объеме. В качестве такого вещества был выбран широкодоступный глицерин, физико-химические и оптические характеристики которого детально изучены [13].

Целью настоящего проекта являлось исследование процессов локальной модификации оптических свойств на поверхности и в объеме композитных материалов на основе нанопористых силикатных матриц, пропитанных веществом с высокой степенью поляризации, под действием лазерного излучения с длиной волны, частично поглощаемой материалом.

Основной задачей работы стало исследование влияния физических явлений, возникающих при взаимодействии лазерного излучения с материалом, а также динамики процессов формирования МО на их размеры, формы, свойства и характеристики.

В отчете представлены экспериментальные результаты локальной модификации в объеме пластин ПС, пропитанных глицерином, под действием лазерного излучения с длиной волны 1,07 мкм, слабо поглощаемого пластиной. При последующем спекании пластины ПС в печи до состояния кварцоида анализировалось проявление и сохранение сформированных МО. Осуществлено исследование процессов формирования МО с измененными оптическими свойствами в области фокусировки лазерного пучка.

Методы исследования

В исследовании процессов локальной лазерно-индуцированной модификации использовались пластины ПС, представляющие собой матрицу кремнеземного каркаса с пористостью $\sigma = 0,26 \text{ см}^3/\text{см}^3$ и средним радиусом пор $\sim 2 \text{ нм}$, следующего состава: $94,73\text{SiO}_2 - 4,97 \text{ В}_2\text{O}_3 - 0,30 \text{ Na}_2\text{O}$ с возможным присутствием следов ($\leq 0,1 \text{ мас}\%$) Al_2O_3 . В эксперименте использовались плоскопараллельные пластины размером $15 \times 15 \times 1,5 \text{ мм}$. Пластины ПС были изготовлены в Лаборатории физической химии стекла Института химии силикатов им. И.В. Гребенщикова Российской Академии наук.

Пластины ПС пропитывались чистым глицерином в течение 72 час при комнатной температуре и нормальном атмосферном давлении. До проведения эксперимента они хранились в глицерине. За 30 мин до начала эксперимента пластины ПС вынимались из глицерина, просушивались с помощью фильтровальной бумаги и закреплялись на координатном столе для последующего облучения (рис. 1).

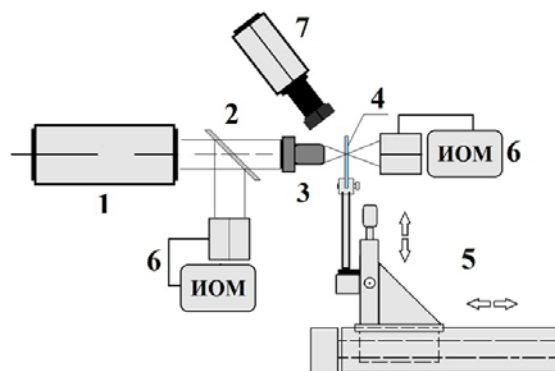


Рис. 1. Схема экспериментальной установки: 1 – волоконный лазер; 2 – делительная пластина; 3 – микробиоъектив; 4 – пластина ПС, пропитанная глицерином; 5 – координатный стол; 6 – измеритель оптической мощности; 7 – тепловизионная камера

При извлечении пластины из глицерина наблюдалось заметное ее помутнение, которое, на наш взгляд, связано с попаданием незначительного количества воздуха в каналы пластины. Непосредственно перед пропиткой пластины ПС не подвергались сушке при повышенной температуре, а при сушке в условиях комнатной температуры в каналах кремнеземного каркаса присутствовала вода (около 7 – 10 %).

На пластины ПС, пропитанные глицерином, воздействовали излучением непрерывного иттербиевого волоконного лазера с длиной волны $\lambda = 1,07 \text{ мкм}$, шириной линии излучения $\Delta\lambda = 0,003 \text{ мкм}$ и нестабильностью выходной мощности 1 %. Излучение было сфокусировано с помощью микробиоъектива (10^{\times} , 0,25) в объем до размера области перетяжки пучка

$r_0 = 25$ мкм. Мощность лазерного излучения, прошедшего через микрообъектив и падающего на пластину ПС, в зависимости от условий эксперимента составляла одну из величин: 9,6; 13,1 или 16,5 Вт. Длительность облучения во всех экспериментах находилась в пределах от 270 до 300 с.

Мощность лазерного излучения, падающего ($P_{об}$) на пластину ПС, пропитанную глицерином, и прошедшего через нее ($P_{пр}$), фиксировалась с помощью измерителя оптической мощности Gentec Solo PE-2M с точностью 1 % от измеряемой величины. Измерения $P_{об}$ и $P_{пр}$ позволяли оценивать поглотительную способность ПС в процессе формирования МО. Температура в области облучения, соответствующей форме перетяжки лазерного пучка в объеме пластины ПС, фиксировалась с помощью тепловизионной камеры «Flip Titanium» 520 M с разрешением по размеру измеряемой области 30x30 мкм и точностью измерения 25 мК для интервала температур 20–300 °С. После лазерной обработки пластины ПС МО исследовались с помощью оптического микроскопа Axio Imager Carl Zeiss в проходящем неполяризованном и в линейно-поляризованном свете при скрещенных поляризаторе и анализаторе.

На заключительном этапе обработки пластины ПС вместе с МО спекались в печи при температуре 850–870 °С в течение 10–15 мин до уплотнения кремнеземного каркаса и формирования кварцоидного стекла. Глицерин в ходе спекания испарялся при температуре $T_u \sim 300$ °С. Для предотвращения разрушения пластин ПС в ходе спекания, когда давление в каналах ПС увеличивается, становясь максимальным при температурах испарения глицерина, пластины ПС выдерживались при $T \sim 250$ –300 °С в течение 20–30 мин. МО в объеме пластин кварцоида, сохраненные после спекания, вновь исследовались с помощью микроскопа. Оптические характеристики кварцоидного стекла близки к плавленому кварцу, а физико-химические свойства намного стабильнее, чем у ПС. Сформированные таким образом МО находятся в монолите кварцоида, что позволяет сохранять их целостность и функциональность при длительной эксплуатации.

Основные достигнутые результаты

Создание модифицированных областей в объеме пластин пористого стекла, пропитанных глицерином

Как было установлено, в ходе эксперимента под действием лазерного излучения пластины ПС нагревались в области облучения до температур 70 ± 10 °С при минимальной мощности падающего излучения ($P_{об} = 9,6$ Вт) и до 100 ± 10 °С при максимальной мощности ($P_{об} = 16,5$ Вт) (рис. 2).

Известно [13], что при давлении $\sim 0,93 \cdot 10^5$ Па температура кипения чистого глицерина составляет 290 °С. При смешивании его с водой в соотношении 9:1 температура кипения изменяется не более чем на 2,5 °С. Кипение глицерина при максимальной температуре ~ 100 °С, фиксируемой в эксперименте, маловероятно, но возможно испарение воды, приводящее к обогащению раствора глицерином, и соответственно, к повышению его температуры кипения. Из анализа рис. 2 следует, что в первые 120–150 с, в зависимости от мощности падающего излучения, происходит постепенное повышение температуры, при 120–150 с скорость роста температуры заметно снижается, а при дальнейшем облучении, вплоть до конца, значение температуры стабилизируется (рис. 2). Длительный прогрев пластины ПС в процессе облучения при расположении перетяжки лазерного пучка в объеме может быть связан с низкой температуропроводностью материала, относительно невысокой мощностью падающего излучения и конвекцией тепла с широких поверхностей пластины.

Известно [8], что заметное изменение в размерах пластин ПС при спекании в печи начинается при температурах, превышающих 750 °С. Этот экспериментальный факт указывает на то, что термоуплотнение кремнеземного каркаса ПС при температуре ~ 100 °С маловероятно и формирование МО на основе термических процессов является не убедительным. С другой стороны, давление, создаваемое парами глицерина при температуре в 150 °С не превышает 131,6 Па или 1 мм. р. ст. [13], и поэтому также не может служить причиной формирования МО в проводимых экспериментах.

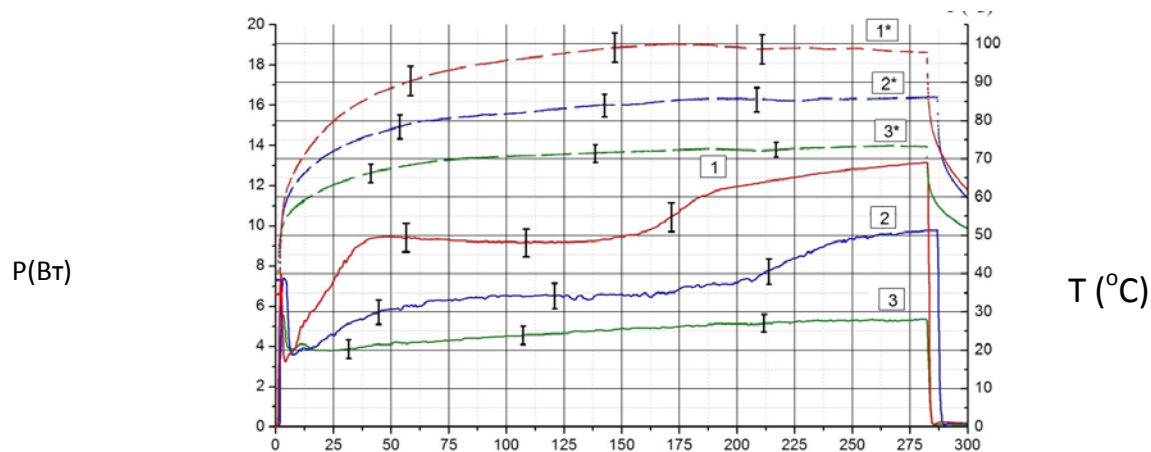


Рис. 2. Временные зависимости мощности прошедшего излучения $P_{пр}$ (1, 2 и 3) и температуры в области облучения T (1^* , 2^* и 3^*) от времени воздействия при мощности падающего излучения (Вт): 9,6 (1); 13,1 (2) и 16,5 (3)

Мощность прошедшего сквозь пластину ПС лазерного излучения варьировалась в зависимости от длительности облучения, и это указывало на изменение оптических характеристик ПС в процессе воздействия, в частности поглощательной способности и показателя преломления в области фокусировки лазерного пучка (рис. 2). Характерный вид зависимости мощности прошедшего излучения от длительности воздействия регистрировался в результате неоднократных экспериментов (~ 10 раз).

Постоянство мощности прошедшего излучения в интервале 50–200 с, скорее всего, указывало на то, что формирование центральной части МО в основном завершалось. Начиная с 200 с и до конца процесса облучения наблюдалось повышение мощности прошедшего излучения, что свидетельствовало об окончательном заполнении пространства пор центральной части МО и о более значительном заполнении пространства пор шарового слоя, примыкающего к центральной части в пределах области фокусировки лазерного пучка.

Зная экспериментальные значения падающей $P_{об}$ и прошедшей $P_{пр}$ мощности, в каждом из измерений определялся коэффициент пропускания. Погрешность определения коэффициента пропускания (доверительный интервал при доверительной вероятности 0,95) не превышала $\pm 7\%$ для $P_{об} = 9,6$ Вт и $\pm 4\%$ для $P_{об} = 13,1$ и 16,5 Вт. На погрешность могли влиять различные факторы, проявляющиеся при проведении экспериментов. Например, неравновесная пропитка пластин ПС по всей толщине образца приводила к различной концентрации глицерина по глубине его проникновения в объем ПС.

Сразу же после проведения экспериментов пластины с МО исследовались под микроскопом в проходящем (рис. 3а, 3в) и поляризованном свете (рис. 3б, 3г). Было установлено, что процесс формирования очертаний и размера МО либо происходил непосредственно во время воздействия лазерным излучением, либо заканчивался в течение нескольких секунд после окончания облучения. МО располагались преимущественно на глубине 500 мкм от поверхности пластины ПС, а их размер составлял около 100 мкм. Четкого очертания границ частей МО, свидетельствующего о перераспределении вещества в пределах МО и о значительном различии показателей преломления ее частей, не наблюдалось. Граница МО с окружающей пластиной была хорошо заметна при наблюдении под микроскопом и это, скорее всего, указывало на разницу в показателях преломления МО и пластины ПС.

Облучение длительностью 360 с в диапазоне мощностей падающего излучения 17–20 Вт приводило к увеличению различия в показателях преломления МО и пластины ПС и тем самым способствовало формированию более четкой границы МО – ПС (рис. 4). Дальнейшее повышение длительности облучения, а также мощности падающего излучения приводило к стремительному структурно-фазовому преобразованию и, как следствие, разрушению образца.

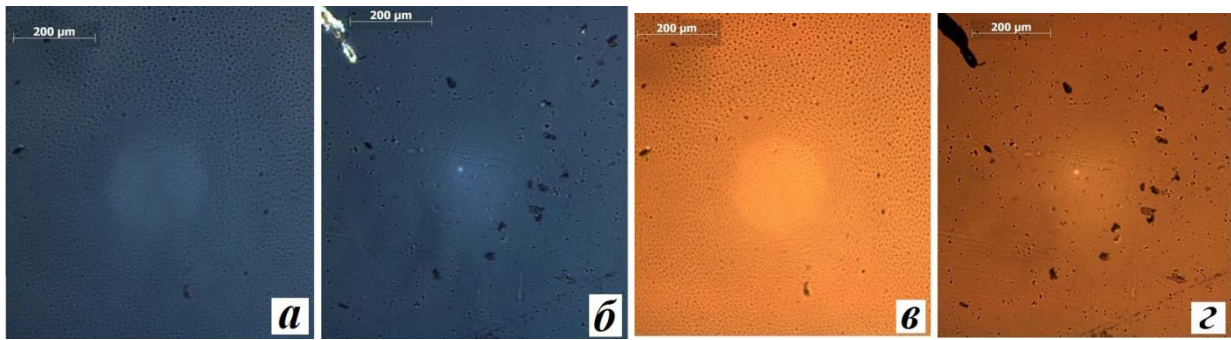


Рис. 3. Фотографии: а, в – нижней; б, з – верхней поверхностей пластины ПС, пропитанной глицерином, вместе с МО, сформированной в объеме ПС в: а, б – поляризованном; в, з – проходящем свете (масштаб 200 мкм)

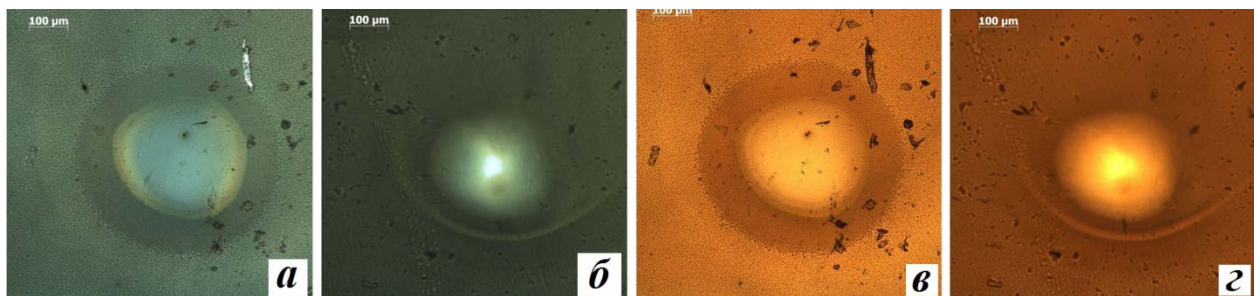


Рис. 4. Фотографии: а, в – нижней; б, з – верхней поверхностей пластины ПС, пропитанной глицерином, вместе с МО, сформированной в объеме ПС в: а, б – поляризованном; в, з – проходящем свете при длительности облучения 360 с и мощности 17–20 Вт (масштаб 100 мкм)

На заключительном этапе пластины ПС вместе с МО подвергались спеканию в печи. После спекания удалось обнаружить только те МО, у которых разность в значениях показателя преломления на границе МО – пластина кварцеидного стекла оказалась достаточной для обнаружения при наблюдении в проходящем свете (МО, сформированные при облучении с мощностью 13,1–16,5 Вт). Что касается МО, сформированных при мощности излучения 9,6 Вт, то их исчезновение в процессе спекания, по-видимому, было связано с недостаточной интенсивностью процесса массопереноса вещества в пределах области перетяжки от края МО к ее центральной части. Это привело к неполному заполнению пространства пор центральной части и незначительному освобождению пространства пор на границе МО – пластина ПС, о чем и свидетельствует фактическое отсутствие снижения прошедшего излучения в интервале 5–15 с (рис. 2). Таким образом, при дальнейшем спекании ПС в печи усадка кремнеземного каркаса в пределах МО не отличалась от его усадки в области всей пластины, поэтому различить МО с помощью оптического микроскопа не представлялось возможным.

Ряд экспериментов по формированию МО в объеме пластин ПС, пропитанных глицерином, проводился с использованием импульсного волоконного лазера с длительностью импульса $\tau = 100$ нс, частотой следования импульсов $f = 50$ кГц и длиной волны излучения $\lambda = 1,07$ мкм (такой же, как и у непрерывного волоконного лазера, применяемого во всех экспериментах, описанных ранее). Изменение температуры в зоне облучения в зависимости от длительности воздействия было аналогично представленному на рис. 2. Максимальное значение температуры при переходе в стационарный режим составляло 110–115 °С (рис. 5). Процесс пропитки пластин ПС в глицерине осуществлялся способом, описанным ранее, в течение 35 час 30 мин. Использование импульсного лазерного излучения позволило уменьшить диаметр МО до 60 мкм. Формирование таких МО происходило при средней мощности излучения ~25 Вт и длительности общего облучения ~90 с (рис. 5).

Исследование МО, сформированных в объеме пластины ПС, под микроскопом в проходящем и поляризованном свете, выявило возможность МО фокусировать проходящее сквозь нее излучение от осветительной системы микроскопа. МО изменяла фокальное

расстояние оптической системы микроскопа таким образом, что нижняя поверхность пластины ПС сквозь МО была четко видна, в то время как изображение поверхности вне МО оказывалось размытым. Данный факт свидетельствовал о способности МО уменьшать фокальное расстояние оптической системы микроскопа, то есть она работала, как положительная оптическая линза, а значит, показатель преломления в пределах МО был больше, чем в пластине ПС.

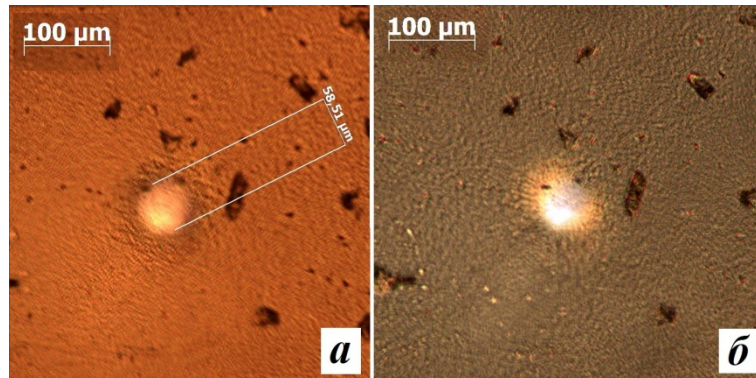


Рис. 5. Фотографии МО, сформированной в объеме пропитанной пластины ПС с помощью излучения импульсного лазера: а – проходящем; б – поляризованном свете

Нагревание ПС в области перетяжки лазерного пучка

Характер зависимости температур в центре зоны облучения от длительности облучения во всех случаях и на всем временном интервале облучения хорошо согласовывался с процессами нагревания вещества лазерным излучением с длиной волны, слабо поглощаемой материалом, которые описывались выражением [14]:

$$T(t) = \frac{AP_{об}}{4\pi\delta k} \cdot \ln \left(19.4 \cdot \frac{kt}{\rho c r_0^2} \right) + T_0, \quad t \gg \frac{r_0^2}{k} \rho c, \quad \delta \gg \sqrt{\frac{k}{\rho c} t}, \quad (1)$$

где t – время, за которое устанавливалась данная температура после начала облучения; r_0 – размер области перетяжки пучка; δ – глубина проникновения света в вещество ($\delta = 2 \cdot 10^{-3}$ м); ρ – плотность, k – теплопроводность, которая для «сухого» ПС приблизительно равна ~ 0.8 Вт/м·К, а c – теплоемкость пластин ПС; $P_{об}$ – значение мощности падающего излучения, A – поглощательная способность пластины ПС. Плотность и теплоемкость принимались равными: $\rho = 1644$ кг/м³, $c = 1233$ Дж/кг·К и $\rho = 2292$ кг/м³, $c = 1513$ Дж/кг·К соответственно, для пластин «сухого» ПС и пропитанного глицерином. Значение поглощательной способности пластин ПС, пропитанных глицерином, определялось по результатам экспериментальных данных с использованием методики из работы [10] и во время воздействия составляло в среднем $A = 0,06$ – $0,11$. Для сравнения, поглощательная способность такой же пластины «сухого» ПС, определенная экспериментально с использованием методики [10], принимала значение $A \approx 0,004$. Теплоемкость пропитанных пластин ПС также была найдена экспериментально. Метод ее определения будет приведен ниже.

Определение значения поглощенного излучения в процессе облучения представляло собой сложную задачу, практически невыполнимую с достаточной точностью из-за сложности интерпретации регистрируемого излучения, проходящего через пластину в процессе формирования МО. На это указывала зависимость мощности прошедшего излучения от длительности облучения, которая изменялась при постоянной мощности излучения. Тем не менее, из зависимости температуры в зоне облучения от длительности облучения видно, что для интервала 50–200 с вид температурной кривой мог быть описан выражением (1), так как в этом интервале формирование МО в основном завершилось. Об этом свидетельствует постоянство мощности прошедшего излучения (рис. 2).

Так как ранее было высказано предположение о нетермическом характере формирования МО, то оценивалось не только значение плотности поглощенной мощности q_a ,

но и значения плотности мощности падающего q_0 и прошедшего q_τ излучений в области перетяжки пучка с размером $r_0 \sim 25$ мкм по выражениям [14]

$$q_0 = \frac{P_{об}}{\pi r_0^2}, \quad q_\tau = \tau \cdot q_0, \quad q_\alpha = A \cdot q_0. \quad (2)$$

Для диапазона мощностей непрерывного излучения 9,6–16,5 Вт плотность мощности падающего излучения оказалась в пределах значений $q_0 \approx (4,9-8,4) \cdot 10^5$ Вт/см², плотность мощности прошедшего излучения для пропускания $\tau \sim 0,85$ – $q_\tau \approx (4,2-7,2) \cdot 10^5$ Вт/см², а плотность мощности поглощенного излучения при максимальном значении поглощательной способности $A \sim 0,11$ – $q_\alpha \approx (5,4-9,2) \cdot 10^4$ Вт/см². Потери излучения при рассеянии были незначительны и намного меньше, чем потери при поглощении, поэтому при расчетах они не учитывались.

В процессе эксперимента по формированию МО в объеме ПС осуществлялась регистрация температуры в зоне облучения с помощью тепловизионной камеры (рис. 6). Приведенные выше теплофизические характеристики пластин «сухого» ПС были взяты из работы [9], где определялись экспериментально. Для ПС, пропитанных глицерином, плотность, теплоемкость и теплопроводность были получены при анализе теплофизических характеристик «сухого» ПС, а также глицерина.

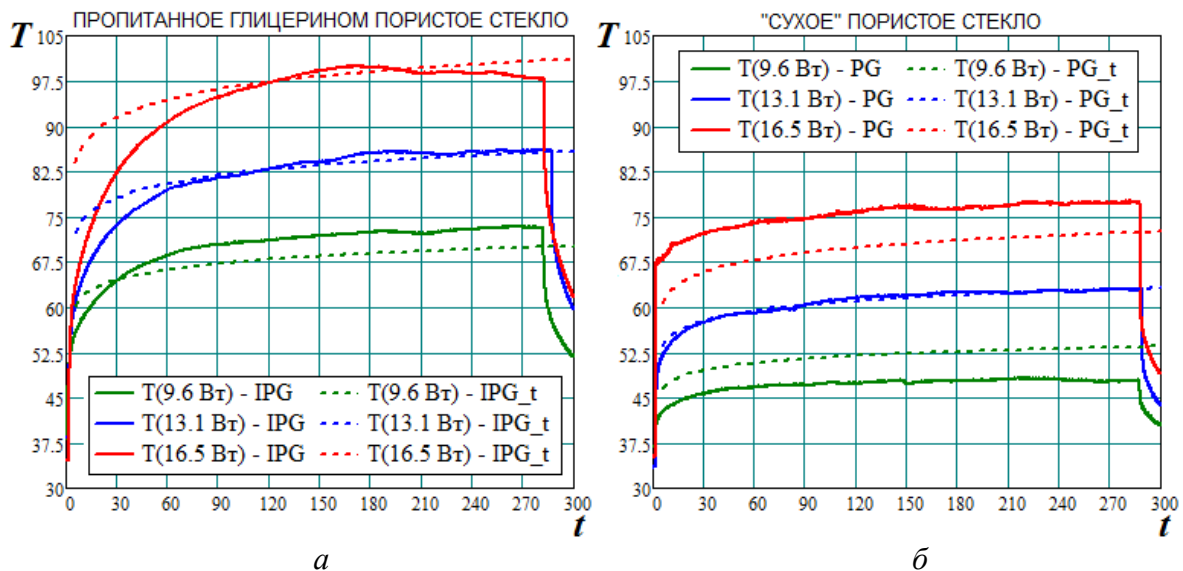


Рис. 6. Зависимости температуры T от длительности облучения в области перетяжки лазерного пучка, расположенной: а – в объеме пропитанного ПС; б – «сухого» ПС.

Непрерывной линией указаны экспериментальные значения температуры;
пунктирной – рассчитанные по выражению (1)

По выражению (1) были построены зависимости температуры в зоне облучения от длительности воздействия. При этом значения плотности, теплоемкости и теплопроводности подбирались таким образом, чтобы отклонение экспериментальных значений температуры от значений температуры, полученных из выражения (1), были минимальны. В итоге значение теплопроводности пропитанных ПС оказалось $k \sim 2,36$ Вт/м·К.

С помощью тепловизионной камеры также было зафиксировано пространственное распределение температуры в плоскости пластины ПС, пропитанной глицерином, в зоне облучения и прилегающих к ней областях, а также в плоскости пластины «сухого» ПС при аналогичных мощностях падающего излучения (рис. 7). По данным зависимостям температуры пластин ПС от расстояния до центра зоны облучения можно оценить скорость понижения температуры в зависимости от расстояния от точечного источника тепла – области, где располагалась перетяжка лазерного пучка.

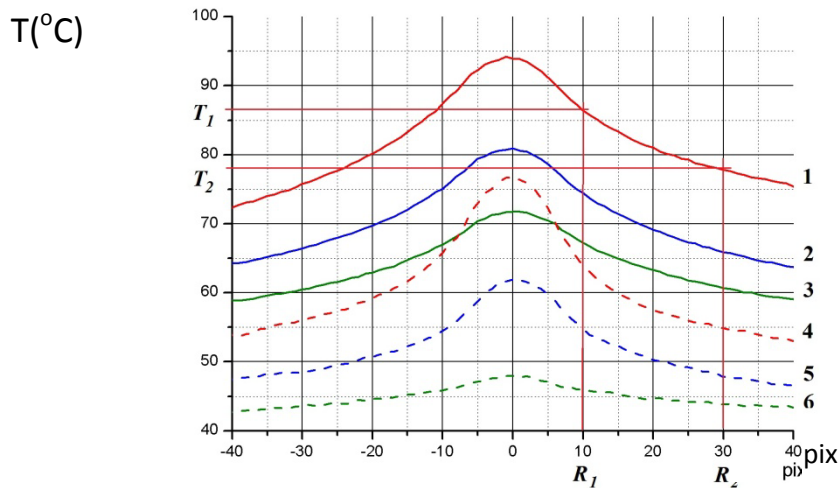


Рис. 7. Зависимость температуры от расстояния до центра области облучения в плоскости пластин ПС, пропитанных глицерином (1, 2 и 3), и «сухого» ПС (4, 5 и 6). Мощность падающего излучения (Вт): 9,6 (1, 4), 13,1 (2, 5) и 16,5 (3, 6)

Допустим, что распределение температуры в зоне облучения в зависимости от расстояния r от центра облученной области в стационарном режиме нагревания соответствовало распределению Гаусса:

$$T(r, t_s) = T_{\max}(t_s) \exp\left(-\frac{r^2}{r_0^2}\right), \quad (3)$$

причем максимальное значение температуры в центре облученной области T_{\max} находилось из выражения (1) спустя $t_s = 160$ с после начала облучения, когда процесс нагревания достигал стационарного режима. Изменение температуры с изменением расстояния r от центра области облучения в предельном значении определяется, как

$$-\frac{\partial T(r, t_s)}{\partial r} = \frac{AP_{об}r}{2\pi\rho c\delta ar_0^2} \cdot \ln\left(19.4 \cdot \frac{at_s}{r_0^2}\right) \cdot \exp\left(-\frac{r^2}{r_0^2}\right), \quad (4)$$

где a – температуропроводность, связанная с теплопроводностью выражением:

$$a = \frac{k}{\rho c}. \quad (5)$$

Выражение (4), учитывая распределение температуры в пластине ПС в зависимости от расстояния от центра области облучения (рис. 7), было преобразовано для оценки значений температуропроводности – a «сухого» ПС и ПС, пропитанного глицерином к следующему виду:

$$a\rho c - C(t_s) \cdot D(T, R) = D(T, R) \cdot \ln(a), \quad (6)$$

$$C(t_s) = -\ln\left(19.4 \frac{t_s}{r_0^2}\right), \quad D(T, R) = \frac{AP_{об}}{2\pi r_0^2 \delta} \cdot \frac{(R_2 - R_1)^2}{T_2 - T_1} \cdot \exp\left(-\frac{(R_2 - R_1)^2}{r_0^2}\right), \quad (7)$$

где T_1 и T_2 – значения температуры на расстоянии R_1 и R_2 от центра области облучения; $C(t)$ и $D(T, R)$ – переменные, зависящие от времени облучения, температуры и расстояния соответственно. Уравнение (6) может быть решено графически нахождением области пересечения двух функций:

$$F_1(a) = a\rho c - C(t_s) \cdot D(T, R), \quad F_2(a) = D(T, R) \cdot \ln(a). \quad (8)$$

Оценочные расчеты, проведенные по выражениям (6, 7), позволили определить значение a для пластины «сухого» ПС ($a \sim (1 \pm 0,7) \cdot 10^{-6} \text{ м}^2/\text{с}$), а также для пластины ПС, пропитанной глицерином ($a \sim (2 \pm 0,5) \cdot 10^{-5} \text{ м}^2/\text{с}$). Температуропроводность пластин «сухого» ПС, определенная экспериментально для образцов – дисков толщиной 5 мм и диаметром 15 мм в работе [9], при комнатной температуре 300 К соответствовала величине $a \approx (0,39 \pm 0,04) \cdot 10^{-6} \text{ м}^2/\text{с}$, что с учетом погрешности удовлетворительно совпадало со

значением, полученным из выражения (8). Значения a , оцененные с учетом экспериментальных результатов измерения температуры (рис. 7), на порядок отличались друг от друга. Подобное различие можно объяснить присутствием глицерина в каналах ПС, способствующее увеличению теплопередачи из области облучения в пластину.

Механизм нетеплового формирования МО в объеме пластин ПС

В настоящее время о процессах формирования МО в объеме пластин ПС, пропитанных полярными жидкостями, без разрушения и модификации кремнеземного каркаса под действием лазерного излучения нечего неизвестно. О результатах исследований подобного рода, проведенных ранее, ничего не сообщалось и, вполне вероятно, что они не проводились.

Природа формирования МО в объеме пластины ПС, пропитанной глицерином, также как и в работе [10] связана с массопереносом веществ, заполняющих каналы ПС, под действием постоянного электрического поля. Это электрическое поле обязано своим возникновением поляризации молекул глицерина и воды. Поляризация молекул и распределение зарядов в области фокусировки лазерного пучка представляют собой процесс пространственной ориентации молекул под действием переменного электрического поля с частотой оптического диапазона. В таком случае возможные процессы поляризации и массопереноса должны быть ярко выражены в области перетяжки лазерного пучка, так как именно в этой области амплитудное значение напряженности максимально.

Предположение о поляризации многоатомных молекул под действием лазерного излучения, сфокусированного в объем вещества, неслучайно. Известны работы по исследованию взаимодействия мощного лазерного излучения с атомами [15], а также по возмущению атомного спектра переменным электромагнитным полем [16]. В них деформация электронных оболочек атомов, обусловленная электронной поляризацией, описывалась динамическим эффектом Штарка. Напряженность электрического поля в работе [16] составляла $E \sim 10^5$ В/см (для сравнения: напряженность поля на орбите электрона в атоме водорода, находящегося в основном состоянии $E_a = 5 \cdot 10^9$ В/см) [16]. С другой стороны, для поляризации молекул жидких кристаллов, обладающих высоким значением дипольного момента, необходимо лазерное излучение сравнительно низкой интенсивности (менее 1 Вт) [17–19].

Вещества, которые использовались в настоящей работе, являются многоатомными молекулами, обладающими значительным дипольным моментом. Для осуществления поляризации таких молекул нет необходимости применять мощное лазерное излучение, поскольку связь между атомами и электронами в многоатомной молекуле ниже, чем в электронной оболочке отдельно взятого атома [20]. При слабом поглощении, то есть при нетермическом воздействии на вещество, действие излучением способно приводить к поляризации веществ и изменению физических характеристик в области перетяжки лазерного пучка – области с максимальной величиной напряженности переменного электрического поля с частотой оптического диапазона [21].

Полагая, что процесс формирования МО обусловлен массопереносом веществ под действием постоянного электрического поля, которое обязано своим возникновением поляризации молекул глицерина, находящихся в каналах кремнеземного каркаса, были оценены амплитудные значения напряженности переменного и вторичного постоянного электрического поля. Амплитудное значение напряженности переменного электрического поля, определенное с помощью усреднения вектора Пойнтинга по времени для плоской электромагнитной волны, вычислялось из выражения [22]

$$E_0 = \left(\frac{2q_0}{\varepsilon_0 c} \sqrt{\frac{\mu}{\varepsilon}} \right)^{1/2} \approx \left(\frac{2q_0}{\varepsilon_0 c} \right)^{1/2}, \quad (9)$$

где $c \approx 3 \cdot 10^8$ м/с – скорость света в вакууме, а $\varepsilon_0 \approx 8,85 \cdot 10^{-12}$ Ф/м – электрическая постоянная. При $P_{об} = 0,12$ Вт, $r \sim 12,5$ мкм, $q_0 \approx 2,45 \cdot 10^4$ Вт/см² значение напряженности переменного электрического поля при формировании МО в пластине пропитанного ПС составляло величину $E_0 \approx 5,03 \cdot 10^4$ В/см.

Значение напряженности вторичного постоянного электрического поля в области фокусировки, рассчитанное с учетом концентрации веществ, заполняющих каналы, и их поляризующей способности, могло бы быть равным $E_2 \approx 3,78 \cdot 10^3$ В/см. Найденное значение E_2 соизмеримо с напряженностью электрических полей, создаваемых при исследовании электродиффузии различных ионов и красителей в ПС [9, 23].

Области возможного использования результатов проекта

Создание локально модифицированных областей, физические, в том числе и оптические свойства, которых отличны от свойств оптически прозрачного материала, в объеме и на поверхности которого происходит их формирование, находит широкое применение. Это технологии по созданию МОЭ, лабораторий на чипах, микрохимических реакторов, фоточувствительных оптических элементов, различных микроэлементов оптоэлектронных приборов и сенсоров, всевозможных МО для волоконной оптики, фотоники и оптоинформатики. Сюда же можно отнести создание различных растров микролинз; волноводов различной формы и размеров, способных канализировать падающее излучение; плазмонных волноводов, по которым могут распространяться поверхностные электромагнитные волны. Запись, хранение и удаление информации на оптических носителях из стекла также продолжает интенсивно развиваться. Здесь в качестве основных задач выступает достижение максимальной скорости записи – перезаписи информации, минимальный размер МО, влияющий на объем общей памяти носителя, а также качества записываемой МО и ее стабильности при длительном хранении в различных условиях.

Технологии создания всех вышеперечисленных микроэлементов широко развиваются в настоящее время и применяются при создании передовых информационных оптоэлектронных и оптических устройств. Методы локальной модификации, в том числе и кристаллизации, стеклообразных материалов постоянно совершенствуются и развиваются. Все большее значение при этом приобретают технологии, основанные на взаимодействии лазерного излучения со стеклообразными материалами, слабо или частично поглощающими падающее излучение. В связи с этим, перспективным направлением является исследование процессов локальной лазерно-индуцированной модификации в объеме и на поверхности стеклообразных материалов, а также разработка новых методов формирования МО, основанных на полученных результатах.

Заключение

Формирование МО в объеме пластин ПС, пропитанных глицерином, при температуре в области облучения, не превышающей 100 °С, указывает на нетермический характер процесса их возникновения. Природа формирования МО, вполне вероятно, связана с процессом массопереноса веществ, частично заполняющих каналы ПС, поляризацией глицерина в области облучения и возникновением постоянного вторичного электрического поля. Под действием напряженности такого постоянного поля и может происходить диффузионное перемещение веществ из прилежащих областей в область с высокой степенью напряженности, то есть в область фокусировки лазерного излучения.

Повышение либо длительности облучения, либо мощности падающего излучения приводит к повышению различия в показателях преломления МО и пластины ПС на их границе, что и приводит к интенсивным структурно-фазовым изменениям в области воздействия и разрушению образца.

Регистрация изменения мощности лазерного пучка, проходящего сквозь пластину ПС, пропитанную глицерином, в процессе ее облучения позволяет осуществлять обратную связь при формировании МО, а также контролировать оптические характеристики и размеры самой МО. Принцип обратной связи, реализованный таким способом, позволит производить обработку пластин ПС в режиме, предотвращающем разрушение.

Результатом работы стало создание нового метода модификации оптических свойств в объеме стеклообразных материалов с применением лазерных технологий. Новый метод формирования МО в объеме пластин ПС позволяет создавать трехмерные структуры

различного размера и формы в объеме стекла под действием лазерного излучения с длиной волны, слабо поглощаемой пластиной ПС. Преимуществом подобного метода формирования МО является простота реализации, а также возможность оптического контроля характеристик формируемой микрообласти при ее создании.

Публикации

1. Сергеев М.М. Формирование модифицированных областей под действием лазерного излучения в объеме пропитанных глицерином силикатных пористых стекол // Физика и химия стекла. 2013. Т. 39. № 5. С. 691–703.
2. Костюк Г.К., Сергеев М.М., Вейко В.П., Яковлев Е.Б., Антропова Т.В., Анфимова И.Н. Описание метода приведено в авторском свидетельстве «Способ изготовления микрооптического растра»; № 2012154939, дата подачи: 18.12.2012.

Литература

1. Nukaga O. et al. Lateral nano-channel fabricated in fused silica by femtosecond laser irradiation and wet etching // 14th International Conference on Miniaturized Systems for Chemistry and Life Sciences (3–7 October 2010, Groningen). Vol. 2. Groningen, 2010. P. 1199–1201.
2. Bellouard Y. et al. Monolithic three-dimensional integration of micro-fluidic channels and optical waveguides in fused silica // Materials Research Society. 2004. Vol. 782. P. A 3.2.1–A 3.2.6.
3. Городецкий М.Л. Основы теории оптических микрорезонаторов. М., 2010. – 203 с.
4. Yamabe T., Murotani H., Nakamoto K., Wakaki M. Composition change of glass materials in microlenses formed by CO₂ laser // Journal of Laser Micro / Nanoengineering. 2007. Vol. 2. No 1. P. 40–43.
5. Teng Y. et al. Controllable space selective precipitation of copper nanoparticles in borosilicate glasses using ultrafast laser irradiation // Journal of Non-Crystalline Solids. 2011. Vol. 357. P. 2380–2383.
6. Hanada Y., Sugioka K., Midorikawa K. Highly sensitive optofluidic chips for biochemical liquid assay fabricated by 3D femtosecond laser micromachining followed by polymer coating // Lab on a Chip. 2012. Vol. 12. P. 3688–3693.
7. Яковлев Е.Б. Особенности поведения стекол и стеклообразных материалов при быстром нагревании. СПб., 2004. – 88 с.
8. Молчанова О.С. Натриевоборосиликатные и пористые стекла. М., 1961. – 163 с.
9. Мешковский И.К. Композиционные оптические материалы на основе пористых матриц. СПб., 1998. – 332 с.
10. Костюк Г.К. и др. Лазерно-индуцированное локальное изменение оптических свойств боросиликатных стекол // Физика и химия стекла. 2013. Т. 39. № 3. С. 395–407.
11. Костюк Г.К. и др. Структурные изменения в пористом стекле под действием лазерного излучения при «холодном» и термическом уплотнении // Стекло и керамика. 2012. № 12. С. 1–5.
12. Костюк Г.К., Сергеев М.М., Яковлев Е.Б. Природа модифицированных областей в объеме стекла, индуцированных лазерным излучением с длиной волны, слабо поглощаемой стеклом // Перспективные материалы. 2013. № 9. С. 43–53.
13. Рахманкулов Д.Л., Кимсанов Б.Х., Чанышев Р.Р. Физические и химические свойства глицерина. М., 2003. – 200 с.
14. Вейко В.П., Либенсон М.Н., Червяков Г.Г., Яковлев Е.Б. Взаимодействие лазерного излучения с веществом: Силовая оптика. М., 2008. – 312 с.
15. Зон Б.А. Взаимодействие лазерного излучения с атомами // Соросовский образовательный журнал. 1998. № 1. С. 84–88.
16. Делоне Н.Б. Возмущение атомного спектра в переменном электромагнитном поле // Соросовский образовательный журнал. 1998. № 5. С. 90–95.
17. Крайнов В.П. Ориентация и фокусировка молекул полем лазерного излучения // Соросовский образовательный журнал. 2000. Т. 6. № 4. С. 90–95.
18. Романов П.В. Пороговые эффекты в жидких кристаллах // Соросовский образовательный журнал. 2001. Т. 7. № 1. С. 96–101.

19. *Simoni F., Francescangeli O.* Effects of light on molecular orientation of liquid crystals // *Journal of Physics: Condensed Matter*. 1999. Vol. 11. P. R439–R487.
20. *Михайлов Г.П., Бурштейн Л.Л.* Современные теории дипольной поляризации молекулярных конденсированных систем // *Успехи физических наук*. 1961. Т. 74. № 1. С. 3–30.
21. *Еремин И.Е., Еремина В.В.* Моделирование поляризационных свойств конденсированных диэлектрических сред // *Моделирование систем*. 2005. № 1 (9). С. 41–55.
22. *Савельев И.В.* Курс общей физики. Т. 2. М., 1988. – 496 с.
23. *Мешковский И.К., Клим О.В.* Эффекты влияния лазерного излучения на электродиффузию молекулярных ионов в стержне пористого стекла // *Письма в Журнал технической физики*. 1997. Т. 23. № 10. С. 4–8.



Трофимов Александр Олегович
Факультет фотоники и оптоинформатики
Кафедра оптоинформационных технологий и материалов
E-mail: exeptional777@mail.ru

ВЛИЯНИЕ СТРУКТУРЫ, ХИМИЧЕСКОГО СОСТАВА И СПОСОБА ПОЛУЧЕНИЯ ПЛЕНОК НА ЭЛЕКТРОХРОМНЫЕ ОКНА

А.О. Трофимов

Научный руководитель – профессор Е.А. Колобкова

Введение

На сегодняшний день электрохромные (ЭХ) свойства органических и неорганических полимеров привлекают большое внимание благодаря возможности их применения для создания энергосберегающих окон или других устройств для модуляции светового потока. Скрупулезный анализ индивидуального потребления энергии выявляет удивительные вещи. Например, до 70 % потерь тепла в доме происходит через окна и двери. Стекло пропускает в обе стороны весь оптический и инфракрасный спектр, уносящий или приносящий тепло до 0,45 кВт на 1 м² остекления. Различные отражающие и поглощающие пленки на стеклах окон могут ограничить один из векторов этого потока. Однако это не дает существенного результата, когда перепады температур внутри и вне помещения достигают 60 °С. Шторы и занавески также не решают эту проблему. Выход один – окно нужно превратить в регулируемый оптический и инфракрасный фильтр, с помощью которого можно оптимизировать баланс притока и оттока энергии с учетом индивидуального ощущения комфорта или из целей экономии энергии.

Эту задачу с успехом может решить простенький процессор. Если твердое вещество абсолютно прозрачно, то в нем нет свободных или локализованных электронов способных изменить свою энергию, то есть поглотить или отразить квант света. Если в такую среду внести (инжектировать), например, атом водорода, который тоже прозрачен (энергия ионизации, то есть отрыва электрона, лежит в далекой ультрафиолетовой области), энергия связи электрона с протоном понизится на величину обратно пропорциональную диэлектрической проницаемости среды.

При соответствующем подборе среды энергия связи в таком водородоподобном атоме может оказаться соизмеримой с энергией квантов света в видимой области спектра, и каждый такой атом, поглотивший квант света, станет так называемым центром окраски – F-центром (farben centrum). При достаточной плотности F-центров вещество начинает окрашиваться, то есть изменять свою поглощающую, а при определенных условиях и отражательную

способность. Если из такого легированного вещества экстрагировать легирующую примесь – атомы водорода, – то оно обесцвечивается. Иными словами, принципиально регулируемый светофильтр возможен, что и подтверждается экспериментально.

Прежде всего оконный светофильтр должен быть не просто регулируемым, а электрорегулируемым, то есть он должен управляться от поворота ручки реостата. Он должен изменять свой цвет в привычной шкале серого – от прозрачного до черного. Одновременно с ростом коэффициента поглощения в нем должна расти компонента отражения инфракрасной составляющей потока энергии, иначе через окно будет происходить обычный теплообмен, как у печи с комнатой через горячую стенку. Кроме того, он должен быть работоспособен при любых климатических условиях и перепадах температур и иметь ресурс, соизмеримый с современными металлопластиковыми стеклопакетами.

В поисках таких процессов и материалов обратились к оксидам металлов. В некоторые из них легко инжектируются протоны и электроны (как бы разобранный атом водорода) при электрохимической катодной поляризации и также легко экстрагируются при анодной, что, собственно, и используется, например, в железоникелевых аккумуляторах. Но оксид никеля зеленый и непрозрачный. Перебрав все остальные оксиды, остановились на единственно приемлемом – триоксиде вольфрама WO_3 . При любом способе нанесения на подложку в нем наблюдаются обратимые изменения цвета при знакопеременной электрохимической поляризации под тонким слоем электролита. Но WO_3 окрашивается в синий цвет, отражательная ИК-компонента появляется только при высокой степени окрашивания, то есть высокой концентрации протонов и электронов, инжектированных в оксид из электролита и подложки соответственно. Тем не менее другого, лучшего материала нет.

Помимо WO_3 хорошими ЭХ свойствами обладают тонкие аморфные слои MoO_3 , кристаллические нанослои Nb_2O_5 , соль РВ (прусский голубой), органические полимеры полианилин и его производные. Многочисленные исследования показали, что аморфный триоксид вольфрама ($\alpha-WO_3$) является лучшим неорганическим ЭХ материалом. Поэтому на сегодняшний день WO_3 – неотъемлемый фрагмент всех ЭХ приборов (в том числе ЭХ окон). Очевидно, что способы получения пленок вносят свой вклад в особенности структуры и свойств ЭХ материала. Основными традиционными методами получения пленок $\alpha-WO_3$ являются вакуумное и магнетронное распыления, а также электрохимический метод катодного осаждения [1].

ЭХ системы состоят из пяти функциональных единиц: двух проводящих электричество прозрачных стеклянных подложек, катодоокрашиваемого ЭХ рабочего электрода, анодоокрашиваемого или проводящего противоиэлектрода и электролита. Такие системы способны обратимо изменять светопропускание в широкой области спектра, в частности в видимой, при приложении низких напряжений. Несмотря на заметные достижения в применении электрохромизма для контроля за солнечной энергией в приложении к архитектуре, автомобильному транспорту и авиации с целью уменьшения слепящего эффекта и нагрева солнечными лучами, существует много проблем, которые до конца не решены.

Строение пленок $\alpha-WO_3$, полученных термическим испарением в вакууме, представляет интерес как с точки зрения влияния их структуры и состава на ЭХ свойства, так и общего понимания строения аморфных оксидов переходных металлов, имеющих октаэдрический ближний порядок в кристаллическом состоянии. В связи с отсутствием дальнего порядка в аморфной пленке, важной является информация о ближнем порядке и об особенностях взаимного расположения крупных структурных группировок в пленке при разных способах ее получения.

Строение ЭХ WO_3 пленок, полученных методами термического испарения в вакууме, было рассмотрено в работах [2, 3]. Оно анализировалось методами рентгеновской и электронной дифракции, комбинационного рассеяния света (КРС) [4] и атомной (АТСМ) и электронной просвечивающей (ТЭМ) микроскопией [6]. На основании анализа пленки $\alpha-WO_3$ радиальной функцией распределения можно предположить, что полученная методами термического испарения в вакууме или катодно-реактивного испарения при низких давлениях газовой смеси, она содержит октаэдры WO_6 , соединенные вершинами. Данные ТЭМ были

интерпретированы, как формирование кристаллитов с размерами до 10 А, причем при нагреве их размеры увеличиваются. Предположение о существовании в структуре α -WO₃ нанокристаллических образований гексагональной структуры также следовало из интерпретации данных КРС. Плотность пленок составляла 6,0 г/см³, что на 15 % ниже плотности кристаллического оксида вольфрама.

Изучение пленок α -WO₃, полученных методом электроосаждения из водного раствора солей, показало, что они по своим характеристикам сильно отличаются от пленок, полученных безводными вакуумными методами. Они рентгеноаморфны и сильно гидратированы, а их плотность в два раза меньше плотности кристаллического оксида вольфрама (7,16 г/см³). После завершения процесса осаждения в объеме пленки сохраняется высокая концентрация молекул воды, однако эффективность их окрашивания – 50 см²/С на длине волны 650 нм – не хуже эффективности окрашивания α -WO₃ пленок, полученных вакуумными методами [5].

На основании комплекса данных были сформированы две модели структуры аморфных α -WO₃ пленок. Первая основывалась на представлении о существовании молекулярных трехмерных образований [WO₃]₃ [5]. Вторая (кластерно-пористая модель) отражала способность геля адсорбировать молекулы газа. Структуру пленок образуют кластеры разного размера, соединенные цепями W-O-W. Поверхность кластеров покрыта ОН-группами, а между ними находятся поры, заполненные молекулами Н₂О, причем большая поверхность пор химически активна. Эта модель является расширением модели случайной непрерывной сетки стекла с точечными дефектами.

Прошедшие десятилетия внесли определенные коррективы в представления о координационной форме вольфрама, входящего в структуру α -WO₃. В частности, было обнаружено, что ЭХ свойства обусловлены возникновением двойной связи W=O [5], что противоречит мнению о преимущественно или исключительно октаэдрической координации вольфрама [7–10].

Модели механизма окрашивания были рассмотрены в работах [1, 5]. Обратимое окрашивание α -WO₃ под действием катодного тока в электрохимической ячейке с протонсодержащим электролитом основано на двойной инжекции в объем пленки: электронов из прозрачного электрода с концентрацией близкой к концентрации атомов вольфрама и того же числа протонов из электролита. Инжектированные электроны занимают незаполненные d -орбитали W⁵⁺, которые становятся центрами окраски ЭХ материалов. Оптическое возбуждение таких W⁵⁺ с избыточным электроном инициирует переход электронов на уровни с более высокими энергиями, обуславливая широкую полосу поглощения с $\lambda_{\text{макс}} = 1,4$ мкм. Заряд избытка электронов компенсируется в объеме пленки мобильными протонами (или щелочными ионами), лежащими около атомов кислорода, которые формируют мостиковые и немостиковые связи W-O-W и W=O типов. При внедрении электронов в α -WO₃ происходит изменение его свойств. В связи с целым комплексом параметров, которые необходимо учитывать, существует очевидная неопределенность при формулировке общей концепции ЭХ эффекта в α -WO₃. Она распространяется и на природу оптических переходов, ответственных за механизм окрашивания.

В настоящем исследовании предложены новые данные и их интерпретация по термическому поведению пленок α -WO₃, получаемых по золь-гель технологии. В частности, рассматривается процесс сушки (переход: золь-гель – твердый аморфный полупроводник) и его влияние на получение рабочего золь для формирования ЭХ электрода.

Процесс приготовления ЭХ пленок с помощью золь-гель метода является наиболее прогрессивным и перспективным как с точки зрения простоты технологии, а, следовательно, стоимости получаемого покрытия, так и с точки зрения совершенства структуры ЭХ слоев, их химической и структурной однородности, характерной для химической сборки, а также возможности модификаций химического состава.

Методы исследования

В работе были синтезированы ЭХ пленки WO_3 на основе порошка WC марки CRC030-040 (Германия), размер зерна 0,7–1,0 мкм. Технология получения экспериментальных образцов прозрачных пленок оксида вольфрама проходила в несколько этапов:

- Получение пероксовольфрамовой кислоты (ПВК) ($WO_3 \cdot nH_2O_2 \cdot nH_2O$): в качестве исходных материалов были использованы WC и перекись водорода (30–35 %); очистка раствора после реакции осуществлялась центрифугированием и фильтрацией; процесс упаривания раствора ПВК. Растворы подверглись термической обработке при 70–80 °C в сушильном шкафу. В результате реакции поликонденсации была получена полипероксовольфрамовая кислота (ППВК).

- Сушка упаренного раствора. В работе было опробовано два метода сушки: капельная на оконном стекле при температуре до 450–850 °C и вакуумная сушка (криогенная и роторная).

Измерения и математическая обработка данных дифференциальной сканирующей калориметрии (ДСК) проводились на дифференциальном сканирующем калориметре STA 449F1 Jupiter фирмы Netzsche (рис. 1). Диапазон измерений температур составил от 50 до 700 °C.

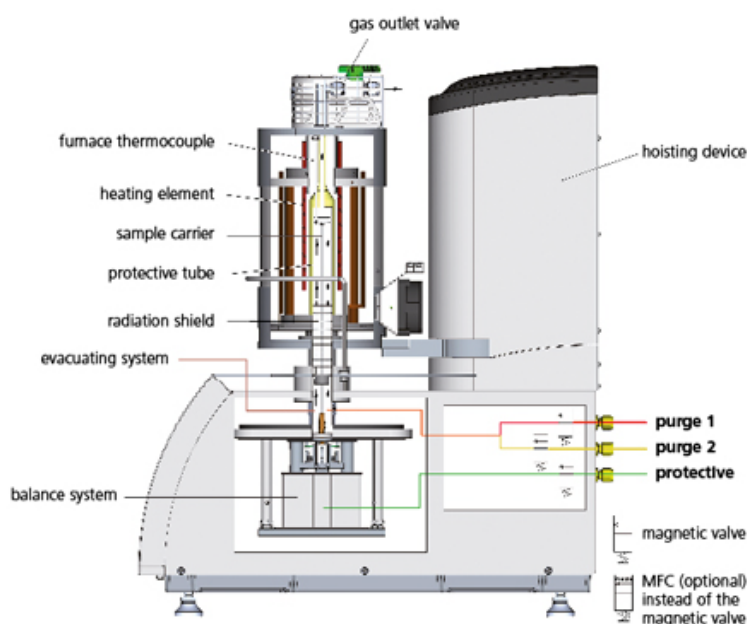


Рис. 1. Принципиальная схема дифференциального сканирующего калориметра STA 449F1 Jupiter [11]

Для определения кристаллической структуры или аморфности образцов был использован рентгеновский дифрактометр (рис. 2).

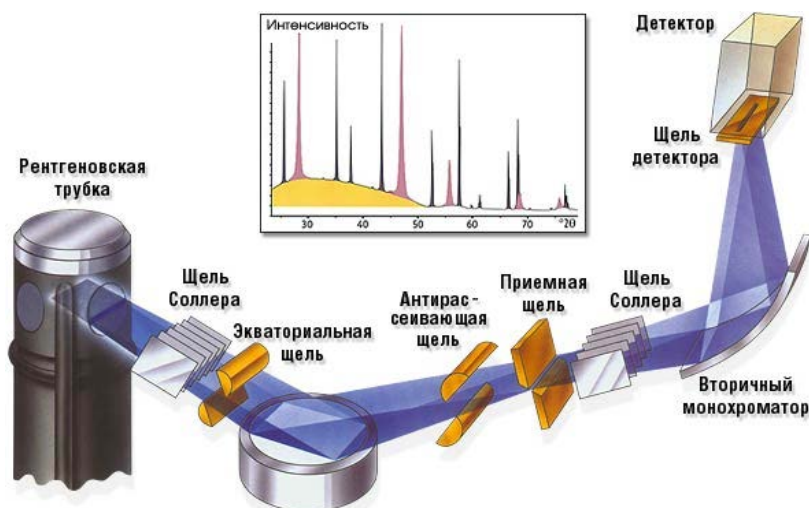


Рис. 2. Принципиальная схема дифрактометра «RigakuUltimaIV»

Определение структуры пленки в зависимости от условий синтеза была осуществлена методами колебательной спектроскопии в зависимости от предоставленных условий синтеза. ИК-Фурье спектры измерялись на приборе ФСМ 1202 (рис. 3) в диапазоне $400\text{--}4000\text{ см}^{-1}$.

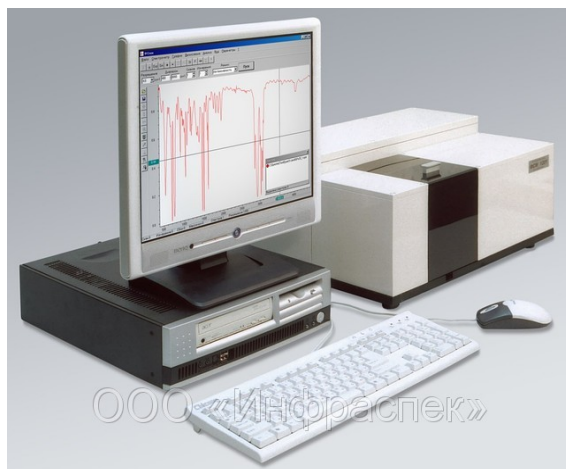


Рис. 3. Внешний вид ФСМ 1202 [13]

Результаты

В работе рассмотрены особенности первой стадии процесса получения ЭХ пленок методом золь-гель технологии, а именно влияние температурно-временного режима сушки предварительного продукта. Технология нанесения пленок $\alpha\text{-WO}_3$ включает две стадии:

- формирование предварительного продукта (стеклообразного $\alpha\text{-WO}_3$);
- растворение предварительного продукта, создание однородной суспензии, нанесение ее на электропроводящий слой и создание однородной прозрачной мезопористой твердой пленки.

Предварительный продукт представляет собой тонкие хрупкие прозрачные пленки желтого цвета, в состав которых входят вольфрам, водород и кислород. Прозрачность его находится в широком диапазоне ($E_g=2,5\text{ эВ}$), рентгеноаморфность и хрупкость свидетельствуют о стеклообразном состоянии. Это стекло, полученное по золь-гель технологии, состав которого может быть записан как $\text{WO}_3 \times x\text{H}_2\text{O}$. Известно, что в зависимости от времени сушки при постоянной температуре в этом стекле могут присутствовать как структурно-связанные ОН-группы, в том числе образующие водородные связи, так и молекулярная группа H_2O . Энергия разрыва таких групп и уход их из состава стеклообразного материала различны. Они происходят при разных температурах и зависят от вида катиона, с которым осуществляется связь ОН-групп.

Дифференциальная сканирующая калориметрия. На рис. 4 представлены кривые, полученные методом ДСК для четырех разных режимов приготовления продукта первой стадии. Очевидно, что условия сушки весьма значительно влияют на концентрацию различных видов ОН-групп и, соответственно, на наблюдаемые термоэффекты. Рассмотрены основные из них, характерные для всех случаев предварительной сушки.

Анализируя спектры ДСК (рис. 4) можно выявить следующие закономерности:

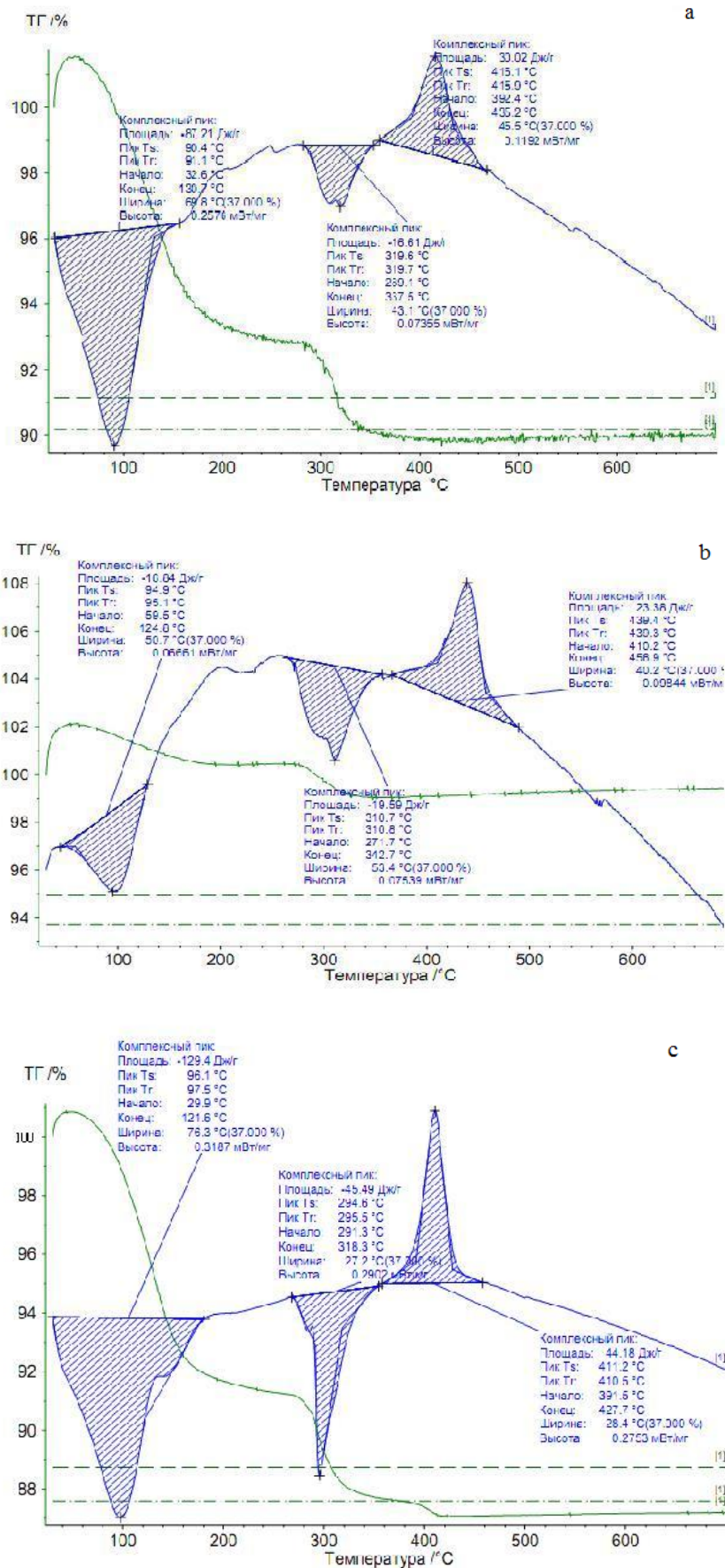
1. В диапазоне температур $T=30\text{--}180\text{ }^\circ\text{C}$ наблюдается интенсивный эндоэффект, связанный с потерей веса. По-видимому, это уход несвязанной молекулярной воды из пористой структуры.

2. Второй эндоэффект наблюдается при температуре в $300\text{ }^\circ\text{C}$. Он сопровождается потерей массы порядка 2–4 масс %) и связан с потерей части ОН-групп.

3. При температурах $410\text{--}440\text{ }^\circ\text{C}$ (максимум) наблюдается экзотермический эффект (кристаллизация), причем удельная энтальпия процесса зависит от исходного состава. При этом также происходит мало заметное уменьшение массы (менее 0,5 %).

Общий вид кривой ДСК характерен для традиционных стеклообразных материалов, полученных методом высокотемпературного плавления, однако наблюдаются дополнительные

термоэффекты, обусловленные потерей воды, присутствующей в порах пленок $\alpha\text{-WO}_3$. Основные экзо и эндоэффекты отражены в Табл.1.



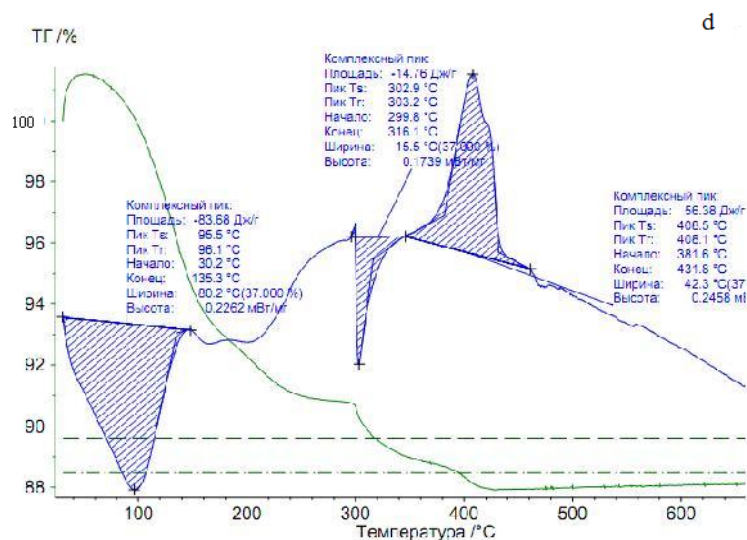


Рис. 4. ДСК пленок WO_3 , полученные при разных условиях сушки:
 а – капельный метод; б – вакуумно-ротаторная сушка; в – сушка в бьюксе (толстый слой);
 д – вакуумная (криогенная) сушка

Таблица 1

Данные ДСК пленок, полученных по разным методикам и их сопоставление по растворимости в спирте

№п/п	Эндо эффект (Дж/г)	Потери веса (%)	Экзоэффект: энергия/температура максимума (Дж/г) /°C	Особенности сушки продукта /растворение в спирте
1.	87/16	10	30/416	Капельная /хорошо растворяется
2	83/14	12	58/410	вакуумная роторная / не растворяется
3	129/45	12	44/410	криогенная вакуумная/ плохо растворяется
4	16/19	4 (2/2)	23/410	в бьюксе в толстом слое/ растворяется

ДСК не дает однозначного ответа, почему материал, получаемый по капельному методу, хорошо растворяется в спирте, в то время как объекты, полученные по методу вакуумной роторной и криогенной сушки, не растворяются. У пленок, полученных разными методами сушки, наблюдаются одинаковые термоэффекты. Никаких особенностей, которые могли бы указать на структурные различия хорошо и плохо растворимых пленок, не наблюдается. Из этого можно сделать предположение, что растворимость связана не с концентрацией молекул воды или групп W-O-H, а со структурой пор, формирующихся в результате сушки. По-видимому, капельный способ сушки приводит к формированию легко растворимой слоистой структуры. Если вспомнить, что химический состав пленки можно условно записать, как $WO_3 \cdot xH_2O$, где $0 \leq x \leq 1$, то напрашивается некоторая аналогия со строением стеклообразных щелочных метафосфатов.

Рентгеновская дифракция. Методом рентгеновской дифракции были проведены исследования полученных пленок на рентгеноаморфность. Вне зависимости от способа сушки для всех пленок характерна рентгенограмма, представленная на рис. 5а. Дифрактограмма подтверждает рентгеноаморфность пленок, полученных золь-гель методом, независимо от условия сушки.

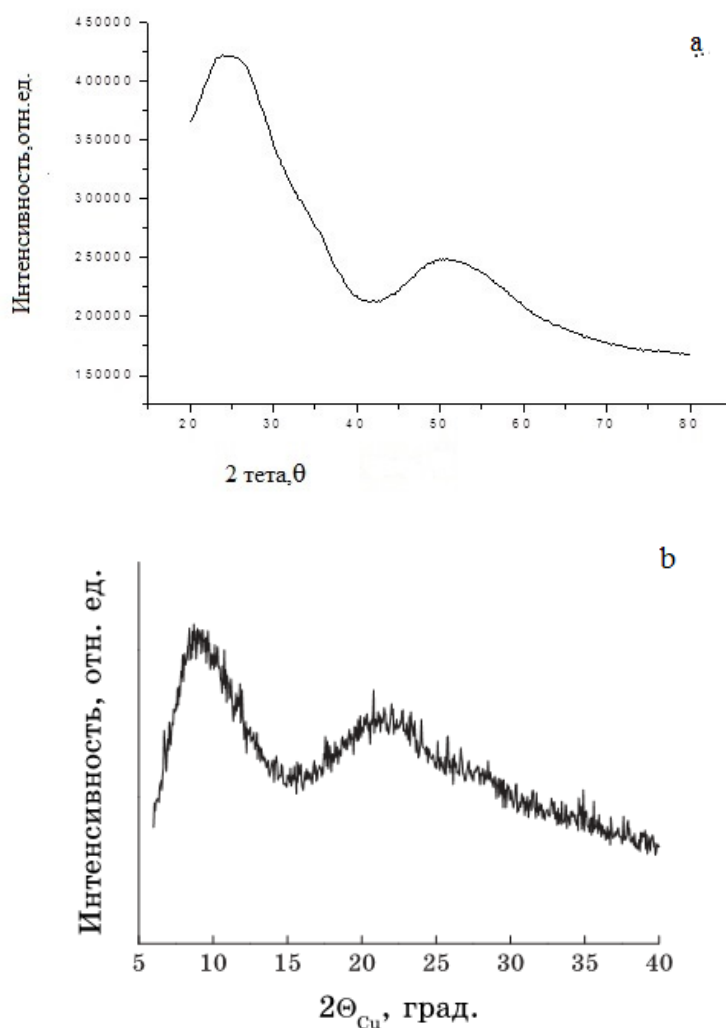


Рис. 5. Рентгеновская дифракция на пленках полученных: *a* – по золь-гель технологии; *b* – катодным осаждением из раствора [5]

Колебательная спектроскопия (инфракрасная спектроскопия и комбинационное рассеяние света). Структура пленки в зависимости от условий синтеза изучена методами колебательной спектроскопии. ИК-Фурье спектры представлены на рис. 6, 7. Очевидно, что основными отличиями являются разные интенсивности полос 550, 710, 830, 950, 1380, 1600 и 3400 см^{-1} , причем две последние, соответствующие деформационному и валентному колебаниям H_2O , уменьшают свою интенсивность при увеличении температуры сушки. Это находится в соответствии с уменьшением содержания молекулярной воды в порах при уменьшении размеров самих пор. Полосы 550, 700 и 850 см^{-1} обусловлены преимущественно колебаниями вольфраматных группировок. Следует напомнить, что интерпретация ИК-спектров «воды» является сложной и не до конца решенной проблемой. С учетом низких температур (45–85 °С) было бы логичным допустить, что все изменения в ИК-спектрах обусловлены изменением в состоянии и концентрации «водной» компоненты. Если сравнить ИК-спектры пленок, полученных золь-гель методом, и пленок, полученных методом анодного осаждения (рис. 7), то очевидно значительное перераспределение интенсивностей основных полос в диапазоне колебаний вольфраматных групп. При катодном осаждении самыми интенсивными являются полосы 950 и 790 см^{-1} , в то время как у полученных по золь-гель технологии – 850, 720, 550 и 600 см^{-1} .

До сих пор не окончена дискуссия о координации вольфрама в стеклообразном состоянии. Результаты анализа самых чувствительных структурных методов – КРС и рентгеновской спектроскопии поглощения вблизи края поглощения (XANES) – не приводят к согласующимся выводам.

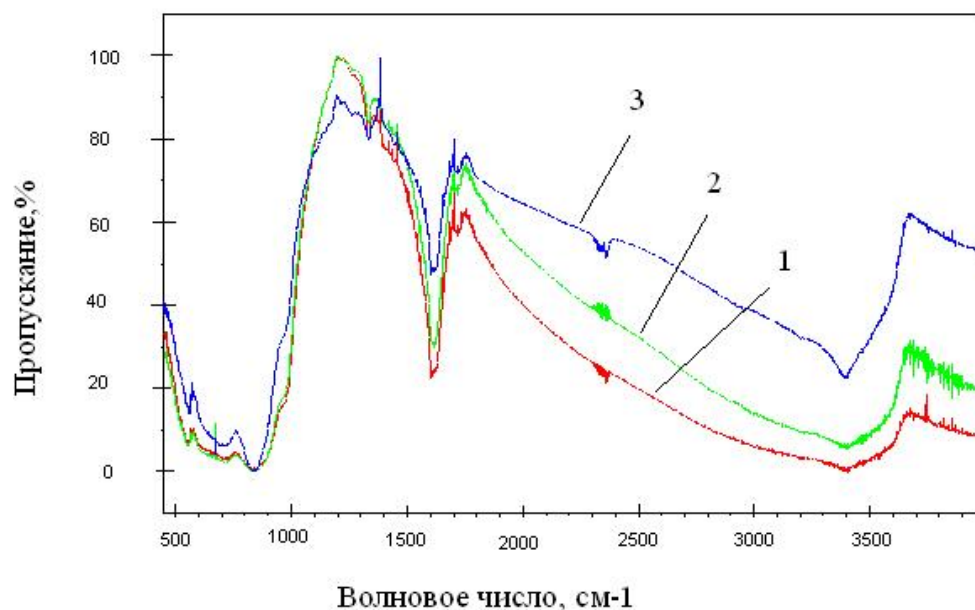


Рис. 6. Влияние температуры сушки на ИК-спектры: 1 – ($T=65^{\circ}\text{C}$); 2 – ($T=45^{\circ}\text{C}$); 3 – ($T=85^{\circ}\text{C}$)

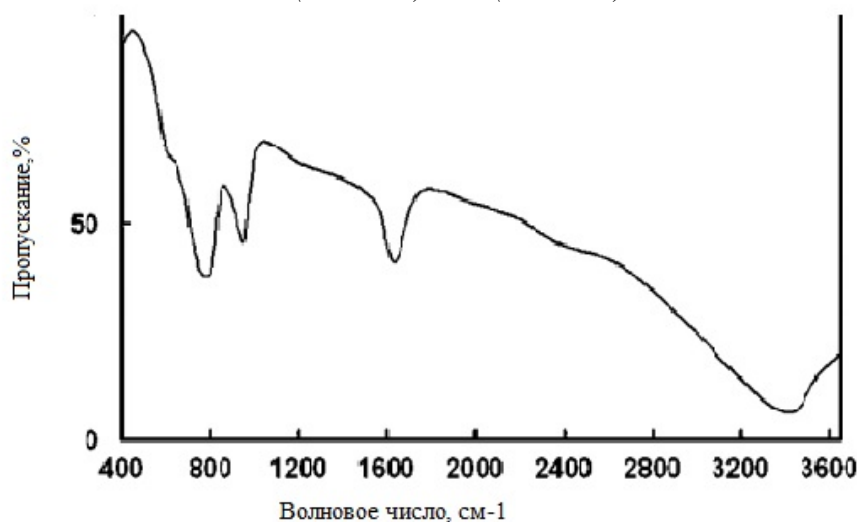


Рис. 7. ИК-спектр пленки, полученной анодным осаждением [3]

Известно [13], что группировки WO_x , ($x=4,6$) могут быть рассмотрены как молекулярные единицы внутри решетки. Их колебания дают информацию о химической связи и симметрии группы. Для структур, содержащих октаэдры WO_6 , характерны симметричные и асимметричные валентные колебания, имеющие частоты в диапазонах $850\text{--}800\text{ см}^{-1}$ и $700\text{--}600\text{ см}^{-1}$, вне зависимости от того, формируют ли октаэдры мостиковые связи (например, WO_3) (рис. 8) или не мостиковые (упорядоченные перовскиты) [12]. Для тетраэдрических группировок характерны полосы, соответствующие валентным колебаниям в диапазонах $900\text{--}970$ и $800\text{--}850\text{ см}^{-1}$ (например, Na_2WO_4 , рис. 9).

Анализ спектров КРС пленок $\alpha\text{-WO}_3$ (рис. 9), полученных при различных способах сушки, однозначно указывает на преимущественное присутствие в их составе вольфраматных тетраэдров. Самой интенсивной в спектре КР-света является полоса 970 см^{-1} , обусловленная полностью симметричным колебанием $\nu_1(A_1)$ вольфраматного тетраэдра, связанного в цепочечные структуры. Полоса 570 см^{-1} относится к мостиковому колебанию W-O-W. В диапазоне $700\text{--}900\text{ см}^{-1}$ находится группа полос: 720 , 827 , 850 и 890 см^{-1} . Согласно нашей интерпретации, полосы 850 и 827 см^{-1} связаны со снятием вырождения с асимметричного колебания $\nu_3(F)$ [WO_4]. Полосы 720 и 890 см^{-1} можно отнести к колебаниям вольфраматных октаэдров, которые также в определенной концентрации присутствуют в сетке стекла и формируют совместно с тетраэдрами единую сетку. Концентрация вольфраматных октаэдров, по-видимому, влияет на

структуру образующихся пор, что в свою очередь определяет скорость растворимости продукта в спирте.

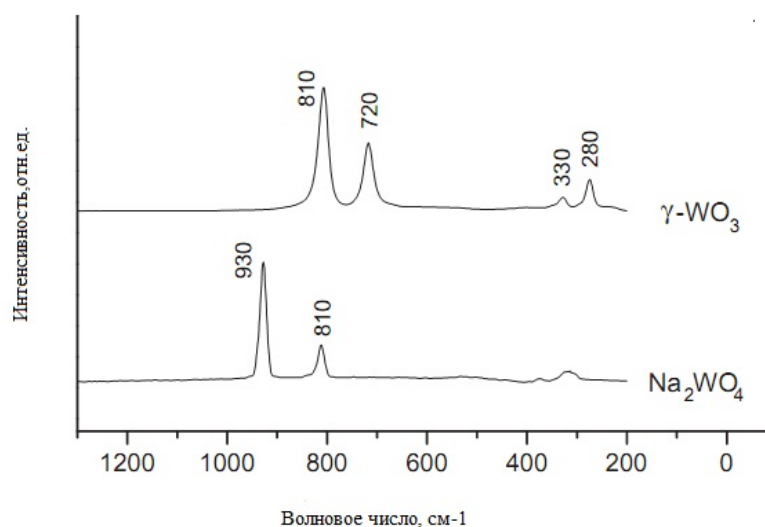


Рис. 8. Спектр комбинационного рассеяния кристаллического WO_3 в гексагональной модификации (октаэдры) и вольфрамата натрия (изолированные тетраэдры)

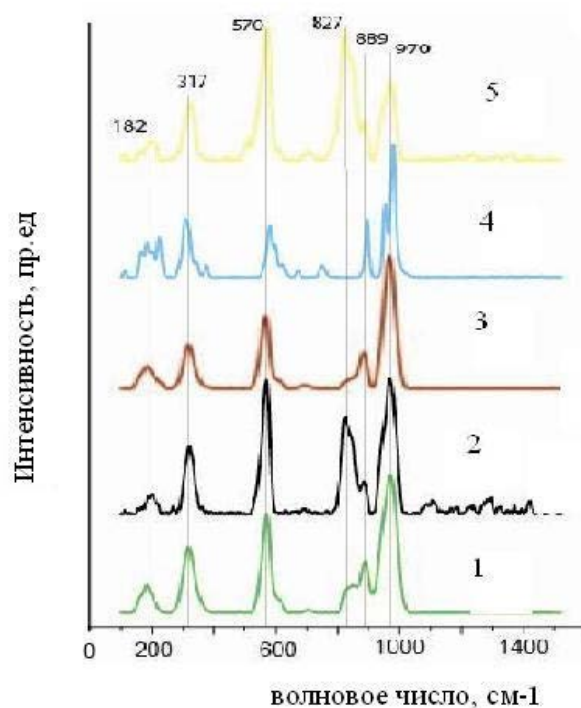


Рис. 9. Спектр комбинационного рассеяния пленок, полученных при разных способах сушки: 1 – капельная сушка $T=65^\circ C$; 2 – капельная сушка $T=85^\circ C$; 3 – вакуумная роторная сушка; 4 – добавление в раствор 10 мол. % уксусной кислоты, капельная сушка; 5 – капельная сушка (исходный реагент – металлический вольфрам)

Выводы

Прозрачность в широком диапазоне ($E_g = 2,5\text{эВ}$), рентгеноаморфность и хрупкость материала свидетельствуют о стеклообразном состоянии предварительного продукта, то есть это стекло, полученное по золь-гель технологии, состава $WO_3 \cdot xH_2O$. Данные ДСК показали большую чувствительность структуры к особенностям режима сушки предварительного продукта полипероксвольфрамовой кислоты.

На основании анализа данных колебательной спектроскопии сделано заключение о присутствии в структуре аморфного прозрачного стеклообразного полупроводника $\alpha-WO_3$

преимущественно тетраэдрических группировок [WO₄]. Способ синтеза *a*-WO₃ влияет как на структуру пор, что определяет концентрацию водной компоненты, так и на соотношение числа мостиковых и немостиковых связей W=O и W-O-W.

Рассмотренные вопросы технологии получения ЭХ пленок *a*-WO₃ тесно связаны с особенностями структуры продуктов сушки для получения из них зольных спиртовых растворов. Сложность интерпретации структуры продуктов сушки стеклообразного *a*-WO₃ предполагает углубление исследований в этом направлении. Подтверждение особенностей структуры продуктов сушки целесообразно связать в дальнейшем с исследованием ЭХ характеристик пленок *a*-WO₃ и ЭХ окон на их основе.

ЭХ пленки на основе WO₃ могут применяться в ЭХ дисплеях, оптических модуляторах, окнах с регулируемым светопропусканием, зеркалах заднего обзора в автомобилях и т. п. Оксид вольфрама перспективен также как материал для газовых сенсоров H₂, NO₂, NO, NH₃, H₂S, в большинстве которых используется в форме тонкой пленки.

Публикации

1. *Trofimov A.O., Kolobkova E.V., Sohovich E.V., Zemko V.S.* Electrochromic film based on *a*-WO₃ obtained by sol-gel // Proceedings of the International conference. Nanomaterials: Applications and properties (16–21 September 2013, Alushta, the Crimea, Ukraine). 2013. Vol. 2. No 4. 04NAESP11 (3 pp).

Литература

1. *Краснов Ю.С., Колбасов Г.Я., Волков С.В.* Поликластерная структура и электрохромизм пленок оксида вольфрама // Наносистемы, наноматериалы, нанотехнологии. 2008. Т. 6. № 3. С. 845–853.
2. *Раманс Г.М., Патмалниекс А.А.* Структура и морфология аморфных пленок триоксида вольфрама и молибдена // Электрохромизм. Рига, 1985. С. 67–82.
3. *Ramans G.M., Gabrusenoks J.V., Veispals A.A.* Structure of tungstic acids and amorphous and crystalline WO₃ thin films // Physica status solidi (a). 1982. Vol. 74. No 1. P. 41–44.
4. *Бродский М.Х.* Комбинационное рассеяние света в аморфных полупроводниках // Рассеяние света в твердых телах. М., 1979. С. 230–289.
5. *Krasnov Yu.S., Kolbasov G.Ya.* Electrochromism and reversible changes in the position of fundamental absorption edge in cathodically deposited amorphous WO₃ // Electrochimica Acta. 2004. Vol. 49. P. 2425–2433.
6. *Arnoldussen T.C.* A model for electrochromic tungsten oxide microstructure and degradation // Journal of The Electrochemical Society. 1982. Vol. 128. № 1. P. 117–123.
7. *Shaltout I., Tang Y., Braunstein R., Shaisha E.E.* FTIR spectra and some optical properties of tungstate-tellurite glasses // Journal of Physics and Chemistry of Solids. 1996. Vol. 57. No 9. P. 1223–1230.
8. *Dimitrov V., Arnaudov M., Dmitriev Y.* IR-spectral study of the effect of WO₃ on the structure of tellurite glasses // Monatshefte für Chemie / Chemical Monthly. 1984. Vol. 115. P. 987–991.
9. *Sekiya T., Mochida N., Ogawa S.* Structural study of WO₃ TeO₂ glasses // Journal of Non-Crystalline Solids. 1994. Vol. 176. No 2–3. P. 105–115.
10. *Charton P., Gengembre L., Armand P.* TeO₂-WO₃ glasses: infrared, XPS and XANES structural characterizations // Journal of Solid State Chemistry. 2002. Vol. 168. No 1. P. 175–183.
11. <http://www.netzsch-thermal-analysis.com/ru/produkty-resheniya/termogravometrija-differencialnaja-skanirujushchaja-kalorimetrija/sta-449-f1-jupiter-sovmeshchennyi-tg-dsk.html>.
12. *Cape T.W., Maroni V.A., Cunningham P.T. et al.* Raman and i.r.-emission studies of some tungstate- and molybdate-containing melts // Spectrochimica Acta. 1976. Vol. 32A. No 5. P. 1219–1223.



Ульянцев Владимир Игоревич
Факультет информационных технологий и программирования,
Кафедра компьютерных технологий
E-mail: ulyantsev@rain.ifmo.ru, vl.ulyantsev@gmail.com

ПРИМЕНЕНИЕ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ О ВЫПОЛНИМОСТИ БУЛЕВОЙ ФОРМУЛЫ ДЛЯ ПОСТРОЕНИЯ УПРАВЛЯЮЩИХ КОНЕЧНЫХ АВТОМАТОВ ПО СЦЕНАРИЯМ РАБОТЫ

В.И. Ульянов

Научный руководитель – профессор А.А. Шалыто

Введение

В последние годы для решения разнообразных задач все чаще применяется автоматное программирование [1]. В рамках этого подхода поведение программ описывается с помощью детерминированных конечных автоматов. Важным достоинством автоматных программ является возможность их автоматической верификации [2], что существенно для систем с повышенными требованиями к надежности.

Важную роль для этих систем играет влияние человеческого фактора на процесс разработки. В связи с этим актуальна задача разработки методов автоматизации построения конечных автоматов, которые позволят снизить это влияние.

Одним из способов автоматизированного построения конечных автоматов является их построение по экспертным данным [3], которым искомый автомат должен соответствовать. Альтернативный метод связан с применением эволюционных алгоритмов [4, 5, 6]. Он подразумевает введение функции приспособленности, определенной над множеством автоматов.

В известных работах, предлагающих методы построения автоматов по экспертным данным, к искомой модели предъявляется требование непротиворечивости: не должно быть двух переходов, исходящих из одного состояния управляющего автомата и одновременно выполнимых при некоторой комбинации события и входных переменных. В некоторых системах к автоматам управления предъявляется требование полноты: любой комбинации события и входных переменных должен соответствовать переход в каждом состоянии. Однако существующие методы построения автоматов не удовлетворяют этому требованию.

В настоящем проекте проводится разработка, реализация и экспериментальное исследование метода построения управляющих автоматов, удовлетворяющих не только требованию непротиворечивости, но и требованию полноты. Метод основан на решении задачи удовлетворения ограничений (*constraint satisfaction problem, CSP*). Пусть имеется набор переменных x_1, \dots, x_n , для каждой из которых задано множество допустимых значений. Задача удовлетворения ограничений заключается в подборе таких значений x_1, \dots, x_n , что выполняются все заданные ограничения на эти переменные. Примерами ограничений являются такие утверждения, как $2x_1 + x_2 = x_3$, $(x_1 = x_2) \Rightarrow (x_3 = 1)$.

Метод построения управляющих автоматов в качестве входных экспертных данных использует безошибочные сценарии работы программы. Метод показывает существенно более высокую производительность по сравнению с генетическими и эволюционными алгоритмами, которые традиционно применяются для решения указанной задачи. Производительность метода позволяет строить управляющие автоматы с десятками состояний, а размер входных данных сможет составлять тысячи событий и выходных воздействий. В качественном отношении метод позволяет строить автоматные программы, удовлетворяющие не только

требованию непротиворечивости, но и требованию полноты, что не позволяет делать ни один из существующих в настоящее время методов машинного обучения.

Постановка задачи

Управляющим конечным автоматом называется детерминированный конечный автомат, каждый переход которого помечен событием, последовательностью выходных воздействий и охранным условием, представляющим собой логическую формулу от входных переменных.

Автомат получает события от так называемых поставщиков событий (в их роли могут выступать внешняя среда, интерфейс пользователя и т. д.) и генерирует выходные воздействия для объекта управления. При поступлении события автомат выполняет тот соответствующий ему переход, для которого охранное условие оказывается истинным. При выполнении перехода генерируются выходные воздействия, которыми он помечен, и автомат переходит в соответствующее состояние. Отметим, что состояния такого автомата не делятся на допускающие и не допускающие (рис. 1).

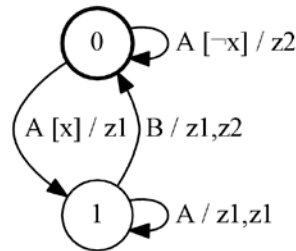


Рис. 1. Пример управляющего конечного автомата

Для данного автомата множество входных событий равно $\{A, B\}$, охранные условия зависят от единственной логической входной переменной x , множество выходных воздействий равно $\{z1, z2\}$. Далее состояние автомата с номером 0 будем считать начальным.

В качестве исходных данных для построения управляющего конечного автомата используется множество сценариев работы. Сценарием работы будем называть последовательность $T_1 \dots T_n$ троек $T_i = \langle e_i, f_i, A_i \rangle$, где e_i – входное событие, f_i – булева формула от входных переменных, задающая охранные условия, A_i – последовательность выходных воздействий. В дальнейшем тройки T_i будем называть элементами сценария.

Автомат, находясь в состоянии $state$, удовлетворяет элементу сценария T_i , если из $state$ исходит переход, помеченный событием e_i , последовательностью выходных воздействий A_i и охранным условием, тождественно равным f_i как булева формула. Автомат удовлетворяет сценарию работы $T_1 \dots T_n$, если он удовлетворяет каждому элементу данного сценария, находясь при этом в состояниях пути, образованного соответствующими переходами.

К искомой автоматной модели обязательно предъявляется требование непротиворечивости: не должно быть двух переходов, исходящих из одного состояния управляющего автомата и одновременно выполнимых при некоторой комбинации события и входных переменных. В некоторых системах к автоматам управления предъявляется требование полноты: любой комбинации события и входных переменных должен соответствовать переход в каждом состоянии.

Решается задача построения управляющего конечного автомата, удовлетворяющего следующим требованиям:

- число состояний автомата равно заданному числу S ;
- автомат удовлетворяет заданному множеству сценариев работы S_c ;
- автомат удовлетворяет требованию полноты;
- каждый переход автомата подтвержден хотя бы одним сценарием работы.

Без последнего требования можно эвристически дополнить автомат до полного, однако поведение автомата на добавленных переходах ничем не подтверждено.

Этапы выполнения предлагаемого метода

Предлагаемый алгоритм включает в себя пять основных этапов.

1. Построение дерева сценариев.
2. Построение графа совместимости вершин дерева сценариев.
3. Построение набора ограничений на целочисленные переменные, задающего требования к «раскраске» построенного графа и выражающего непротиворечивость и полноту системы переходов искомого автомата.
4. Запуск сторонней программы, решающей задачу удовлетворения построенным ограничениям (*constraint satisfaction problem – CSP*), или вызов соответствующего метода сторонней библиотеки.
5. Построение автомата по найденной выполняющей подстановке.

Построение дерева сценариев

Деревом сценариев назовем дерево, каждый переход которого помечен событием, булевой формулой и последовательностью выходных воздействий. Опишем алгоритм построения дерева сценариев по заданному множеству сценариев Sc .

Сначала дерево сценариев состоит из единственной вершины – корня. Затем по очереди добавим в него все сценарии работы из Sc . Для каждого из сценариев будем добавлять его элементы в дерево в порядке возрастания их номеров, начиная с первого. При этом будем хранить указатель на текущую вершину дерева v и номер i первого необработанного элемента сценария.

В начале процесса v указывает на корень дерева сценариев, а $i = 1$. На каждом шаге проверяется существование исходящего из вершины v ребра, помеченного событием e_i и логической формулой, задающей ту же булеву функцию, что и f_i . Если такое ребро не существует, то создается новая вершина дерева u , и в нее направляется ребро, помеченное тройкой $\langle e_i, f_i, A_i \rangle$. После этого u становится текущей вершиной, а значение i увеличивается на единицу.

Если такое ребро существует, то производится сравнение последовательности A_i и последовательности выходных воздействий A' , которой помечено рассматриваемое ребро. Если $A_i = A'$, то текущей становится вершина, в которую ведет рассматриваемое ребро дерева, а значение i увеличивается на единицу. Если же указанные последовательности не совпадают, то заданное множество сценариев Sc является противоречивым, поэтому работа алгоритма прерывается, и пользователю выводится соответствующее сообщение.

После завершения добавления в дерево всех сценариев производится проверка охранных условий. Для каждой вершины перебираются все пары исходящих из нее ребер. Если существуют пары, помеченные одним и тем же событием, а их охранные условия имеют общий выполняющий набор значений входных переменных, то множество сценариев предполагает недетерминированное поведение. Поэтому работа алгоритма прерывается, и пользователю выводится соответствующее сообщение.

Рассмотрим пример работы алгоритма построения дерева сценариев. Пусть набор сценариев состоит из следующих последовательностей:

- $\langle T; x; z1 \rangle, \langle T; \neg x; z2 \rangle, \langle T; \neg x; z3 \rangle, \langle T; x; z1 \rangle$;
- $\langle T; \neg x; z2 \rangle, \langle T; \neg x; z3 \rangle, \langle T; x; z1 \rangle, \langle T; x; z1 \rangle$;
- $\langle T; x; z1 \rangle, \langle T; x; z1 \rangle, \langle T; \neg x; z2 \rangle$;
- $\langle T; \neg x; z2 \rangle, \langle T; \neg x; z3 \rangle, \langle T; x; z1 \rangle, \langle T; \neg x; z2 \rangle$.

Данные сценарии работы содержат единственное входное событие T (например, оно соответствует тикю часов), единственную переменную x , два различных охранных условия x и $\neg x$, три различных последовательности выходных воздействий ($z1$), ($z2$) и ($z3$). Построенное описанным алгоритмом дерево сценариев приведено на 0.

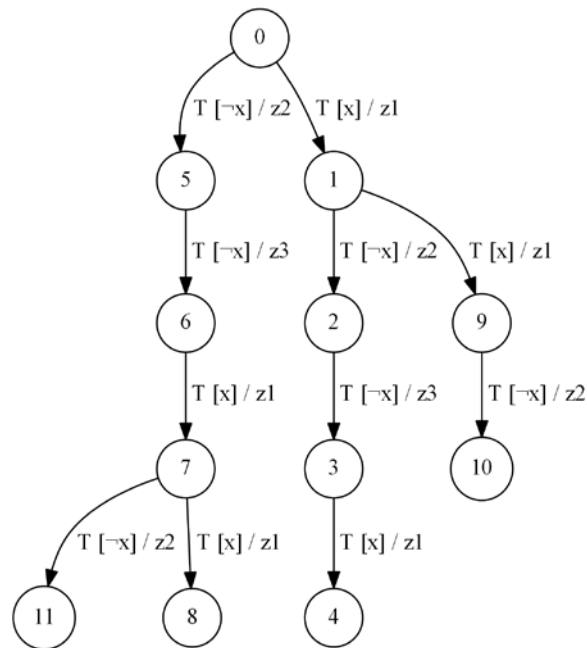


Рис. 2. Пример построения дерева сценариев

Построение графа совместимости вершин дерева сценариев

Для построения управляющего автомата необходимо «раскрасить» вершины дерева сценариев в заданное число цветов (равное числу состояний, которое задается в качестве одного из параметров алгоритма). При этом вершины одного цвета будут объединены в одно состояние результирующего автомата, а множество исходящих из него переходов будет строиться из объединения множеств ребер исходящих из вершин заданного цвета.

Для задания ограничений на раскраску построим так называемый граф совместимости вершин дерева сценариев. Множество вершин этого графа совпадает со множеством вершин дерева сценариев, поэтому в дальнейшем вершины графа и дерева различаться не будут. Ребра графа определяются следующим образом.

Вершины графа совместимости u и v соединены ребром (далее такие вершины будем называть несовместимыми), если существует последовательность пар $\langle e_1; values_1 \rangle \dots \langle e_k; values_k \rangle$ событий и наборов значений входных переменных, которая различает соответствующие вершины дерева.

Опишем алгоритм построения графа совместимости. Напомним, что множество вершин этого графа совпадает с множеством вершин дерева сценариев. Основной идеей данного алгоритма является метод динамического программирования [7, 8].

Для каждой вершины дерева сценариев v найдем все несовместимые с ней вершины. Обозначим $S(v)$ множество вершин, несовместимых с v . Будем вычислять значения функции $S(v)$ начиная с листьев дерева сценариев. Для каждого из листьев u множество $S(u)$ пусто по определению несовместимых вершин. Время работы этого алгоритма составляет $O(n^2)$ (где n – число вершин в дереве сценариев), так как каждая пара ребер дерева сценариев в процессе работы алгоритма будет рассмотрена не более одного раза. При этом такое время работы достижимо, если заранее для каждой пары формул вычислено, равны ли они как булевы функции и имеют ли общий выполняющий набор значений входных переменных.

В худшем случае время работы этапа обработки формул составляет $O(2^{2m} n^2)$, где за m обозначено максимальное число входных переменных, использующихся в одном охранном условии. На практике число m не превышает четырех.

Для дерева сценариев, приведенного на 0, описанный алгоритм построит граф совместимости, приведенный на 0. Сплошными линиями здесь выделены ребра графа совместимости, а штриховыми – исходные ребра дерева сценариев. Вершины с номерами 2 и 5 соединены с вершинами 0, 1, 7 и 9, так как их различает последовательность $\langle T; x = false \rangle$ –

пометки соответствующих ребер дерева сценариев различаются в выходных воздействиях. Других различающих последовательностей в приведенном дереве сценариев не существует.

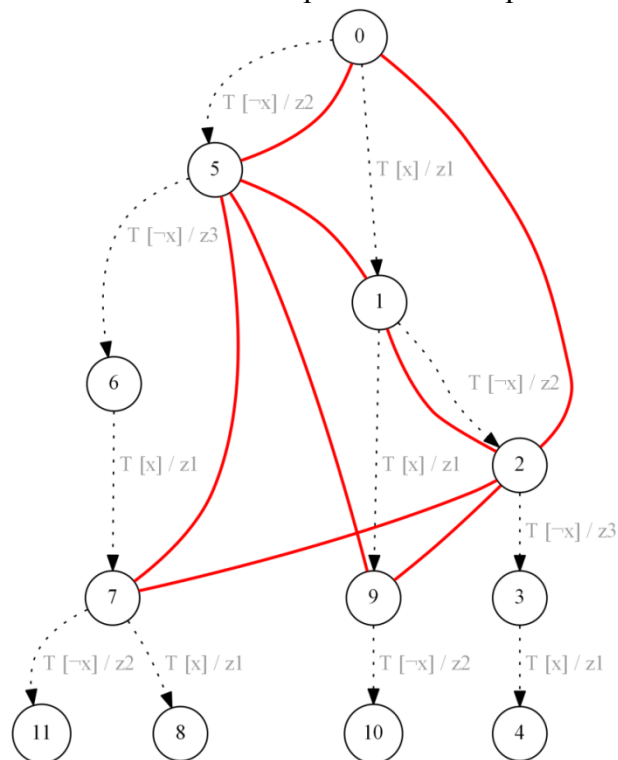


Рис. 3. Пример графа совместимости

Построение набора ограничений на целочисленные переменные

Следующим этапом работы алгоритма построения управляющих конечных автоматов по сценариям работы программы является построение набора ограничений на целочисленные переменные, задающего требования к раскраске графа совместимости и выражающего непротиворечивость и полноту системы переходов искомого автомата. Напомним, что на вход алгоритму подается число C состояний результирующего автомата.

Множество всех условий переходов, встречающихся в сценариях работы S_c , будем обозначать символом F . Множество входных событий e будем обозначать как E . Под F_e подразумеваем множество условий переходов, помеченных входным событием e .

Для построения набора ограничений создадим и будем использовать следующие целочисленные переменные.

1. Переменные x_v соответствуют цвету каждой вершины дерева сценариев v и принимают значения от 0 до $C - 1$. Напомним, что вершины одного цвета будут объединены в одно состояние результирующего автомата.

2. Переменные $y_{i,e,f}$ являются вспомогательными для построения ограничений, задающих непротиворечивость (детерминированность) искомого управляющего автомата, и хранят в себе информацию о его переходах. Данные переменные являются вспомогательными, так как наличие в результирующем автомате переходов не определяется исходя из их значений. Используются переменные для каждого состояния i результирующего автомата (значение от 0 до $C - 1$), каждого события e и каждого условия перехода f из множества F_e , встречающегося в сценариях. Каждая переменная принимает значения от 0 до $C - 1$ и соответствует номеру состояния, в которое ведет переход искомого автомата из состояния i по событию e и условию перехода f .

3. Переменные $z_{i,e,f}$ используются для задания требования полноты искомого автомата и принимают значения 0 или 1, то есть по своей сути являются логическими. Данные переменные задаются для каждого состояния i результирующего автомата (значение от 0 до $C - 1$), каждого события e , каждого условия перехода f из F_e , встречающегося в сценариях. Переменная $z_{i,e,f}$ равна 1, если существует вершина v в дереве сценариев, цвет которой равен i

($x_v = i$), и из нее ведет ребро, помеченное событием e и условием перехода f . В противном случае значение переменной равно 0. Таким образом, данные переменные хранят информацию о структуре переходов результирующего автомата, получающегося в результате объединения вершин дерева сценариев.

Составим набор ограничений на указанные переменные, задающий требования полноты и непротиворечивости искомого автомата.

1. $x_0 = 0$ – ограничение, задающее соответствие корня дерева сценариев начальному состоянию искомого автомата. В настоящем методе начальным состоянием автомата считается состояние с номером 0.

2. $x_v \neq x_u$ (для каждой несовместимой пары вершин дерева сценариев u и v , то есть соединенных ребром в графе совместимости) – ограничения, задающие непротиворечивость искомого автомата. Они гарантируют отсутствие различающих последовательностей, ведущих из одного состояния автомата. Число ограничений данного вида равно числу ребер графа совместимости, то есть в худшем случае таких ограничений может быть $O(n^2)$, где n – число вершин дерева сценариев.

3. $(x_v = i) \Rightarrow (x_u = y_{i,e,f})$ (для каждого цвета i и каждого ребра дерева сценариев vu , помеченного событием e и условием перехода f) – ограничения, задающие детерминированность искомого автомата. А именно, если вершине v присвоен цвет i , то цвет вершины u совпадает со значением переменной $y_{i,e,f}$, хранящей номер состояния автомата, в которое ведет переход из состояния i , помеченный событием e и условием перехода f . Число данных ограничений равно $C \cdot (n - 1)$.

4. $z_{i,e,f} = 1 \Leftrightarrow (x_{v_1} = i \vee \dots \vee x_{v_n} = i)$ (для каждого цвета i , входного события e , условия перехода f из F_e , встречающегося в заданных сценариях, и вершин $v_1 \dots v_n$ дерева сценариев, из которых ведет ребро, помеченное событием e и условием перехода f) – ограничения, необходимые для правильного задания значений переменных $z_{i,e,f}$. Количество данных ограничений оценивается как $O(C \cdot |E| \cdot |F|)$, где как $|E|$ обозначено число событий, а как $|F|$ – число различных условий перехода.

5. $\left(\sum_{f \in F_e} (z_{i,e,f} \cdot c(f)) = 0 \right) \vee \left(\sum_{f \in F_e} (z_{i,e,f} \cdot c(f)) = 2^m \right)$ (для каждого цвета i и каждого события

e) – ограничения, задающие требование полноты искомого автомата. Здесь как $c(f)$ обозначена функция, которая возвращает число выполняющих подстановок для булевой формулы f . При подсчете $c(f)$ считается, что булева формула зависит от всех m переменных, содержащихся в сценариях (например, $c(\text{true}) = 2^m$, а $c(x_1 \vee \neg x_2) = 2^{m-2}$). Сумма $\sum_{f \in F_e} (z_{i,e,f} \cdot c(f))$ равна числу

комбинаций значений входных переменных $values$, для которых существует переход из состояния i , помеченный событием e и условием перехода f таким, что выполняется $f(values)$. Условие полноты искомого автомата выражается тем, что или для любого значения входных переменных найдется переход, или ни для одного из значений переменных перехода не существует. Заметим, что условие того, что для любого значения входных переменных найдется переход, можно выразить как $\sum_{f \in F_e} (z_{i,e,f} \cdot c(f)) = 2^m$, так как считается, что требование

непротиворечивости уже выполнено, то есть все формулы, для которых выполняется $z_{i,e,f} = 1$, попарно не имеют общих выполняющих подстановок.

Приведенные ограничения пяти типов составляют набор, задающий требования непротиворечивости и полноты искомого автомата, удовлетворяющего сценариям S_c и содержащего C управляющих состояний.

Нахождение выполняющей подстановки для построенного набора ограничений и построение управляющего автомата

Следующим этапом работы алгоритма является нахождение выполняющей подстановки для построенного на предыдущем этапе набора ограничений на целочисленные переменные. Для решения задачи удовлетворения ограничений используется библиотека *Choco* [9] для

языка программирования *Java*. Она предоставляет удобный интерфейс пользователя для создания переменных и ограничений, а также реализует производительные методы решения задачи удовлетворения ограничений.

Если библиотека *Choco* не обнаружила набор значений переменных, удовлетворяющих построенным ограничениям, то будем считать, что по данному набору сценариев S_c невозможно построить управляющий автомат с заданным числом состояний C . Если же программа обнаружила выполняющий набор, то построим искомый автомат. Для этого на основании полученных значений переменных x_v определим цвет каждой вершины дерева сценариев. На 0 приведен пример раскраски дерева сценариев, представленного на 0.

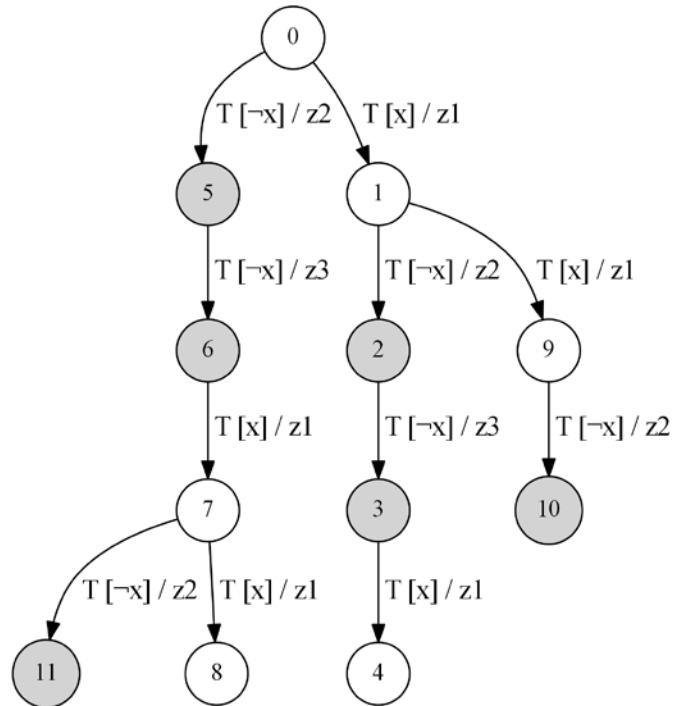


Рис. 4. Пример раскраски дерева сценариев

После этого объединим все вершины одного цвета в одно состояние автомата, а начальным состоянием положим то, которое соответствует цвету корня дерева сценариев. Множество исходящих из состояния переходов построим из объединения множеств ребер, исходящих из вершин заданного цвета. Например, после объединения вершин дерева, приведенного на 0, получим автомат, приведенный на 0.

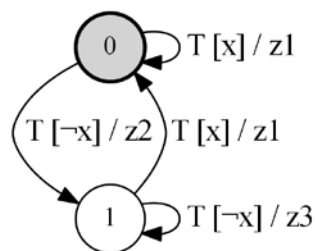


Рис. 5. Автомат, полученный после объединения вершин дерева сценариев

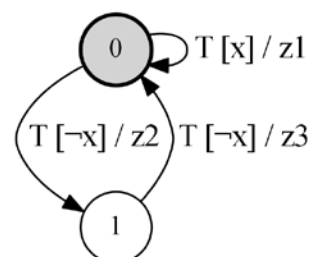


Рис. 6. Автомат, не обладающий свойством полноты

Отметим, что не исключено существование нескольких автоматов, удовлетворяющих сценариям множества S_c . Также отметим, что обозначенным сценариям работы удовлетворяет автомат, приведенный на 0. Он обладает свойством непротиворечивости, но не обладает свойством полноты. Его можно получить, если не добавлять ограничения четвертого и пятого типов.

Экспериментальное исследование

Работоспособность метода проверена на специально сгенерированных случайных входных данных. Искомый автомат содержал от четырех до десяти состояний, а суммарная длина тестов составляла от 800 до 1200 элементов сценариев работы. Результаты экспериментов показали наличие большого числа входных данных, для которых ранее предложенный метод находил управляющий автомат, не удовлетворяющий требованию полноты, в то время как разработанный метод справлялся с задачей, пусть и с меньшей производительностью.

Заключение

Настоящий проект направлен на повышение уровня автоматизации производства программного обеспечения – предложен метод построения управляющих автоматов по сценариям работы программы, основанный на методах решения задачи удовлетворения ограничений.

Была произведена разработка и реализация метода построения управляющих автоматов. Разработанный метод состоит из пяти этапов: построение дерева сценариев, графа совместимости, ограничений на целочисленные переменные, решение задачи удовлетворения построенных ограничений и, наконец, построение искомого автомата, в случае его существования. Предлагаемый подход оперирует с тремя типами целочисленных переменных и пятью типами ограничений на них. Разработка проводилась на языке программирования *Java*. Части из них зарегистрированы: получено свидетельство о регистрации программы для ЭВМ.

Качественное отличие предложенного метода от существующих заключается в предоставлении возможности построения управляющих автоматов, удовлетворяющих требованию полноты. Экспериментальные исследования показали, что метод показывает высокую производительность на различных задачах, справляясь с большинством из них менее чем за минуту работы персонального компьютера. Однако существуют задачи построения автоматов, удовлетворяющих требованию полноты, которые вычисляются несколько часов – построенные ограничения в данных случаях непросто разрешить даже современным производительным методам решения задачи удовлетворения ограничений.

Дальнейшие исследования могут проводиться в следующих направлениях: применение методов удовлетворения ограничений для решения других задач построения управляющих автоматов по экспертным данным; исследование возможности построения автоматов по экспертным данным, содержащим ошибки, вещественным сценариям работы; построение автоматов с учетом заданных темпоральных свойств.

Области возможного использования результатов проекта включают в себя системы управления встраиваемых систем и области с повышенными требованиями к качеству программного обеспечения, такие как авиационная промышленность, медицина, телекоммуникации.

Публикации

1. *Ulyantsev V., Tsarev F.* Extended finite-state machine induction using SAT-solver // Proceedings of the 14th IFAC Symposium “Information Control Problems in Manufacturing – INCOM’12”. Bucharest, 2012. P. 512–517.
2. *Ульянцев В.И., Царев Ф.Н.* Применение методов решения задачи о выполнимости булевой формулы для построения управляющих конечных автоматов по сценариям работы // Научно-

технический вестник информационных технологий, механики и оптики. 2012. № 1 (77). С. 96–100.

3. *Ulyantsev V., Tsarev F.* Extended finite-state machine induction using SAT-solver // Proceedings of the 10th International Conference on Machine Learning and Applications. Vol. 2. Honolulu, 2011. P. 346–349.

Литература

1. *Поликарпова Н.И., Шалыто А.А.* Автоматное программирование. 2-е изд. СПб., 2011. – 176 с.

2. *Кларк Э., Грамберг О., Пелед Д.* Верификация моделей программ. М., 2002. – 416 с.

3. *Поликарпова Н.И., Точилин В.Н., Шалыто А.А.* Метод сокращенных таблиц для генерации автоматов с большим числом входных переменных на основе генетического программирования // Известия РАН. Теория и системы управления. 2010. № 2. С. 100–117.

4. *Гладков Л.А., Курейчик В.В., Курейчик В.М.* Генетические алгоритмы. М., 2006. – 320 с.

5. *Царев Ф.Н.* Совместное применение генетического программирования, конечных автоматов и искусственных нейронных сетей для построения системы управления беспилотным летательным аппаратом // Научно-технический вестник СПбГУ ИТМО. Вып. 53. Автоматное программирование. СПб., 2008. С. 42–60.

6. *Koza J.* Genetic programming. On the programming of computers by means of natural selection. Cambridge, MA, 1998. – 813 p.

7. *Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К.* Алгоритмы. Построение и анализ. М., 2010. – 1296 с.

8. *Скиена С.* Алгоритмы. Руководство по разработке. СПб., 2011. – 720 с.

9. Choco: A java library for constraint satisfaction problems (CSP) and constraint programming (CP): <http://www.emn.fr/z-info/choco-solver/>.



Хасанов Алишер Артикалиевич
Факультет компьютерных технологий и управления
Кафедра систем управления и информатики
E-mail: a.khssnv@gmail.com

СИНТЕЗ СИСТЕМЫ УПРАВЛЕНИЯ РОБОТОМ ЧЕРЕЗ ИНТЕРНЕТ

А.А. Хасанов

Научный руководитель – профессор А.А. Бобцов

Цель и задачи:

- провести анализ существующих технических решений для управления роботом через сети общего назначения, использующие протоколы группы TCP/IP;
- провести анализ задачи управления мобильным роботом через сеть Интернет;
- провести анализ современных алгоритмов управления с учетом запаздывания в канале передачи информации;
- разработать программную архитектуру и сетевую инфраструктуру для реализации сетевой системы управления роботом;
- синтезировать алгоритм управления мобильным роботом.

Метод проведения работы

Работа выполнена с использованием методов гибридного и ситуационного управления, синтеза иерархичных систем управления, пропорционально-интегрально-дифференциального регулирования (ПИД), планирования движения в дискретном пространстве методом Дейкстры и DWA.

Результаты

В результате выполнения проекта создана сетевая информационная инфраструктура, приведенная на рис. 1. В ходе анализа поставленной задачи было выявлено, что разработанная инфраструктура обеспечивает необходимые компьютерные и сетевые сервисы для управления роботом пользователем и для автономной работы робота при необходимости коммуникаций с серверным оборудованием.

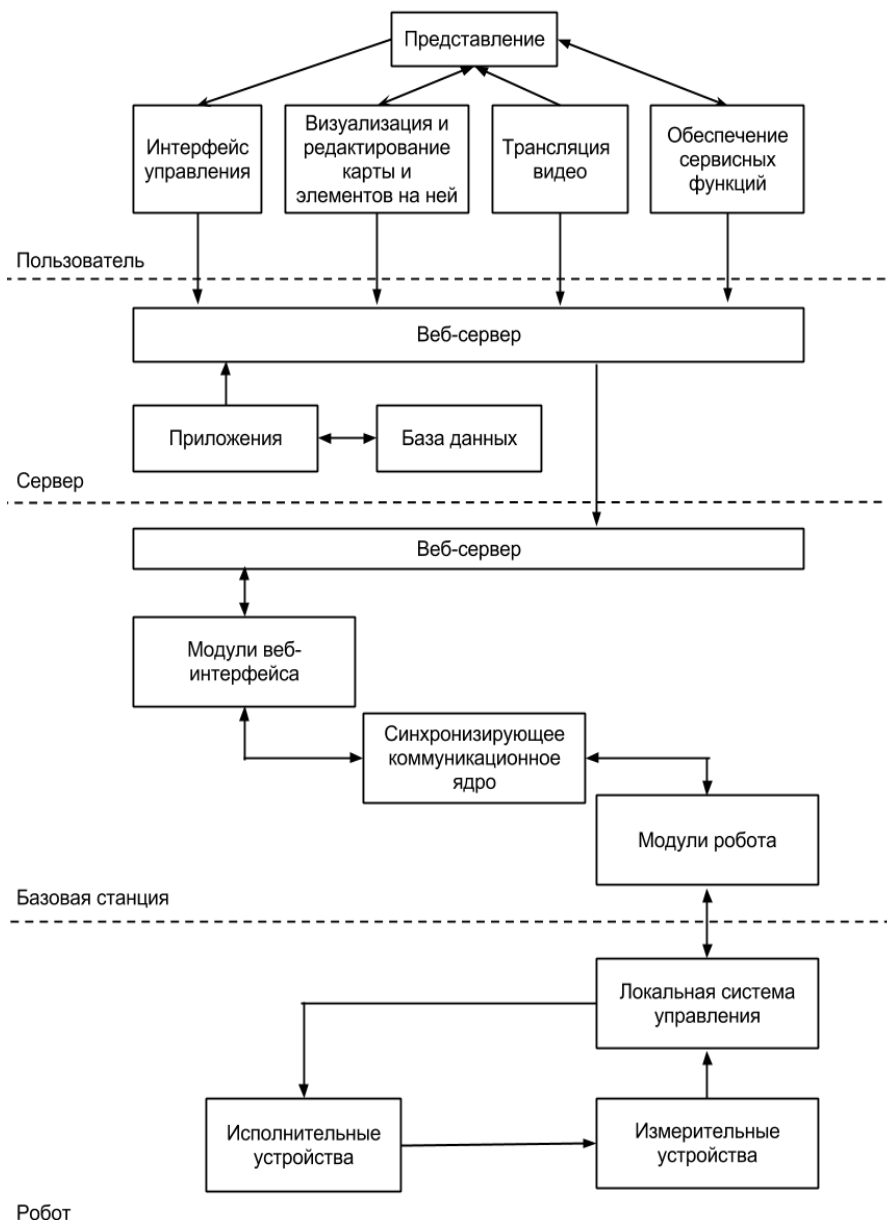


Рис. 1. Структура сетевой информационной системы

В ходе работы также был проведен анализ программных архитектур роботов, из которого выявлено, что для поставленной задачи достаточна архитектура 3Т, расширенная алгоритмами локального управления движением робота. Выбранный вариант представлен на рис. 2. Здесь показаны три основные компоненты программной архитектуры робота, которые можно назвать уровнями управления системы в силу явной иерархии в решении задач, распределенных между ними.

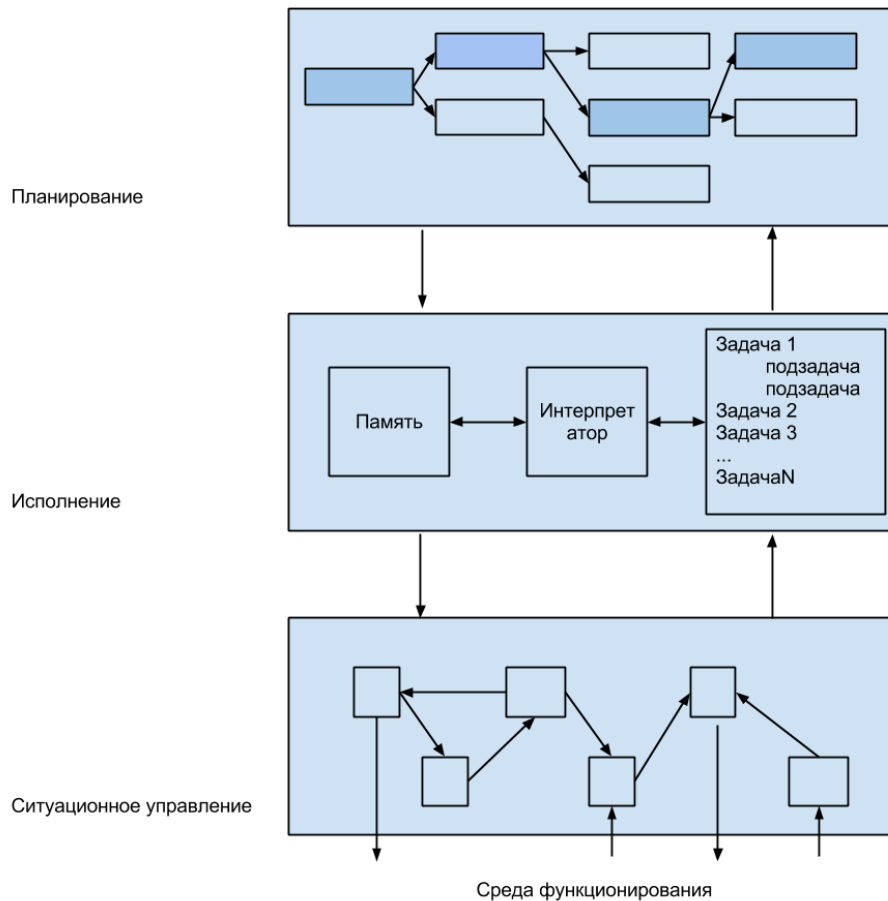


Рис. 2. Программная архитектура робота

Уровень планирования обеспечивает решение задачи нахождения пути и траектории от текущего положения мобильного робота до требуемого пользователем или другими компонентами информационной инфраструктуры. Здесь использованы методы планирования движения в известной среде, когда имеется информация о свободных и занятых препятствиях элементарных площадках в области функционирования робота. В силу представления среды в виде графа, а возможных положений робота – в виде узлов графа применен алгоритм Дейкстры. Для того чтобы учесть при движении робота движущиеся препятствия, в подсистеме локального планирования движения использован метод DWA [1].

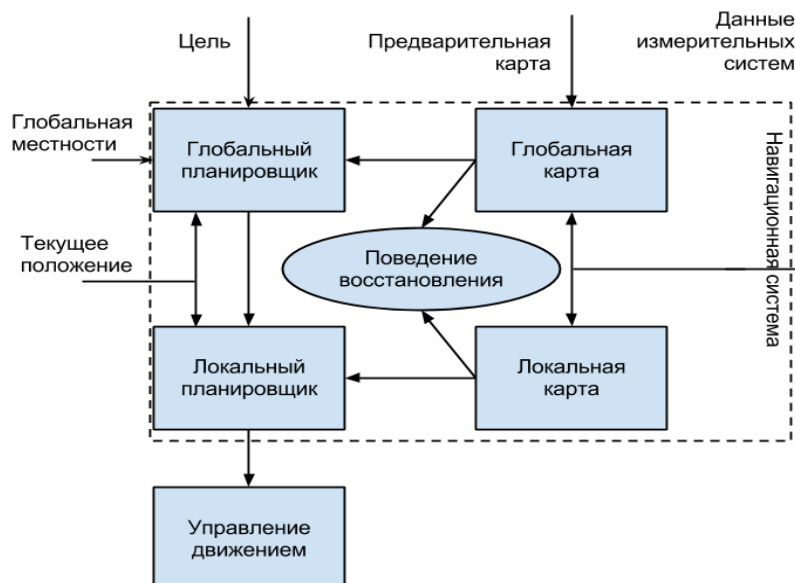


Рис. 3. Структура системы планирования движения

Уровень исполнения обеспечивает решение предзаписанных типовых задач. Уровень ситуационного управления – взаимодействие робота со средой, принимая решения как на уровне своей подсистемы, так и транслируя информацию и запрашивая решения от подсистем верхнего уровня.

Структура подсистемы планирования движением робота представлена на рис. 3. Из рисунка видно, что решение задач глобального и локального планирования разделено. Также показано, что при планировании используются не только известные данные о среде, но и текущие измерения сенсорной системы робота.

Были проанализированы алгоритмы ситуационного управления мобильным роботом и синтезирован эффективный алгоритм обхода препятствий, который использован в подсистеме на уровне ситуационного управления и предлагается для дальнейшего агентного алгоритма планирования движением. На рис. 4 представлена кинематическая схема мобильного робота.

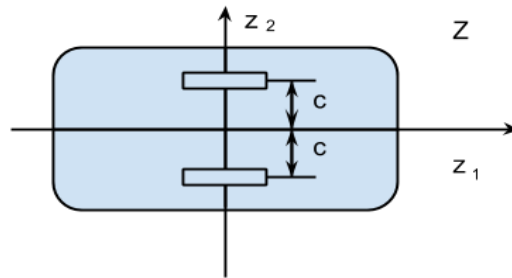


Рис. 4. Кинематическая схема мобильного робота с дифференциальным приводом:
 c – расстояния от колесных модулей до продольной оси робота

Робот имеет два активных колесных модуля, которые симметрично расположены на оси z_2 и фиксированы параллельно оси z_1 . То есть:

$$z^1 = \begin{vmatrix} 0 \\ c \end{vmatrix}, z^2 = -\begin{vmatrix} 0 \\ c \end{vmatrix}.$$

В ходе работы рассмотрен синтез системы управления для модели, входом которой являются линейные скорости колесных модулей. Из уравнений силомоментного управления [2] для мобильного робота с показанной на рис. 4 кинематической схемой выведены следующие уравнения для управления движением робота по скорости с использованием ПИД-регулирования:

$$\varphi_{Ж} = \arctg\left(\frac{u_1}{u_2}\right), \omega = K_{П}e(t) + K_{И}\int_0^t e(\tau)d\tau + K_{D}\dot{e}(t), e = \varphi_{Ж} - \varphi,$$

$$v = \sqrt{u_1^2 + u_2^2}, v_{П} = \frac{2v + \omega L}{2R}, v_{Л} = \frac{2v - \omega L}{2R}.$$

$$\dot{y}_1 = v\cos\varphi, \dot{y}_2 = v\sin\varphi.$$

Здесь: φ – угол текущего направления движения робота, $\varphi_{Ж}$ – желаемое направление его движения, v – скорость и ω – угловая скорость платформы мобильного робота, $v_{П}$ и $v_{Л}$ – линейные скорости правого и левого колесных модулей соответственно, e – ошибка направления движения робота, K_i ($i = П, И, D$) – коэффициенты ПИД-регулятора, u_1 и u_2 – силы действия колесных модулей на платформу мобильного робота, L – расстояние между колесными модулями, \dot{y}_i – скорость движения робота по соответствующей декартовой координате, R – радиус колес.

Для управления движением по траектории, представленной точками в пространстве функционирования робота, проанализированы ситуационные алгоритмы управления с жесткими переключателями (рис. 5) и гибридный алгоритм с использованием скользящего управления (рис. 6). Результаты работы этих алгоритмов (натурные эксперименты) представлены на рис. 7 и 8 соответственно.

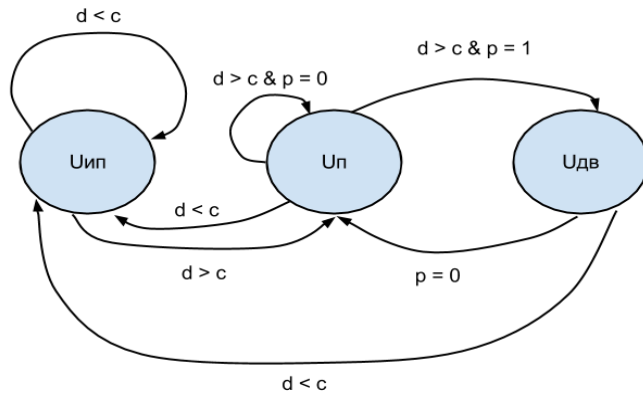


Рис. 5. Алгоритм с жесткими переключателями

Здесь: d – расстояние, измеряемое периферийными датчиками робота; c – постоянная, определяющая расстояние регистрации объекта как препятствия; $p = \{0, 1\}$ – переменная, отражающая соответствие направления движения робота вперед заданному; Уип – управление, реализующее функцию обхода препятствия; Уп – управление, реализующее функцию поиска направления (вращение по часовой стрелке вокруг собственной оси робота); Удв – управление, обеспечивающее движение робота вперед.

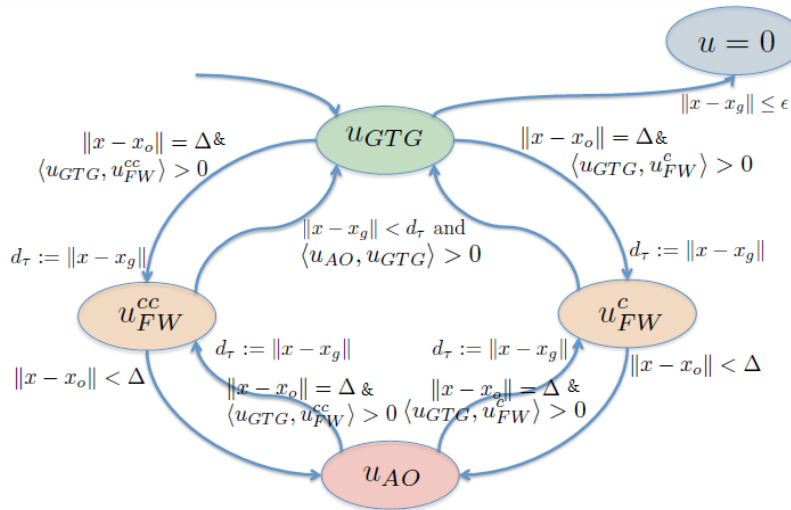


Рис. 6. Гибридный алгоритм ситуационного управления

На этом рисунке: U_{GTG} – управление, реализующее движение к цели, U_{OA} – управление, реализующее движение от препятствия, а U_{FW}^c и U_{FW}^{cc} – управления, реализующие движение вдоль препятствия по часовой и против часовой стрелки соответственно, Δ – расстояние от препятствия, при котором оно считается детектированным, x и x_0 – координаты мобильного робота и препятствия, соответственно, d_τ – расстояние, при котором необходимо обеспечить реакцию на препятствие, ϵ – малая константа одного порядка с x , необходимая для установления факта достижения цели как зоны, а не точки пространства, $\langle U, U \rangle$ – скалярное произведение векторов управляющих воздействий.

Синтезированный алгоритм обеспечивает движение вдоль траектории, пролегающей между зонами действия управляющего воздействия, обеспечивающего избегание препятствий U_{AO} и управляющего воздействия, обеспечивающего следование к цели – U_{GTG} .

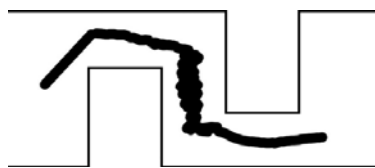


Рис. 7. Путь геометрического центра мобильного робота (в ходе движения точкой оставался след в виде круга) при использовании алгоритма на основе жестких переключателей

Далее, объединяя уравнения модели (1) и (3) получим

$$x(k+1) = A^{h(\tau_k)+1}x(k-h(\tau_k)) + \sum_{j=0}^{h(\tau_k)} A^j B[r(k-j) - k\hat{y}(k-h(\tau_k)-j)], \quad (4)$$

где \hat{y} – оценка показаний измерительного устройства, которая может быть использована при $\tau_k > \tau_0$.

В ходе работы на основе описанной системы управления и предложенной информационной архитектуры разработано программное обеспечение, реализующие управление мобильным роботом через сеть Интернет [4, 5].

Заключение

Разработана система управления мобильным роботом через сеть Интернет, ее программная и сетевая архитектуры. Разработан алгоритм управления мобильным роботом и проанализированы некоторые методы управления при использовании сетевых решений на основе группы протоколов TCP/IP в качестве основного канала передачи информации.

Области возможного использования результатов проекта

Результаты работы могут быть использованы при разработке систем управления мобильными роботами или сетевыми приложениями, предоставляющими пользовательский сервис для управления роботом. Проведенный в рамках проекта анализ позволил выявить эффективные архитектурные и инфраструктурные решения, которые могут быть использованы при разработке программного обеспечения, а синтезированный алгоритм – для реализации функций безопасного маневрирования или автономной навигации мобильного робота при управлении с запаздыванием в канале связи.

Публикации

1. Хасанов А.А., Крупенькин А.С. Веб-фреймворк для систем телемеханики “DCU-Control-Unit” // Сб. материалов XIV Конференции молодых ученых «Навигация и управление движением». СПб., 2012.
2. Хасанов А.А., Крупенькин А.С. “DCU-Control-Unit” – SDK для систем телемеханики с веб-интерфейсом // Сб. материалов конференции «Список – 2012». СПб., 2012.

Литература

1. Fox D., Burgard W., Thrun S. The dynamic windows approach to collision avoidance // IEEE Robotics and Automation Magazine. 1997. Vol. 4. No 1. P. 23–33.
1. Бурдаков С.Ф., Мирошник И.В., Стельмаков Р.Э. Системы управления движением колесных роботов. СПб., 2001. – 228 с.
2. Kim W.-J., Ji K., Srivasta A. Network-based control with real-time prediction of delayed / lost sensor data // IEEE Transactions on control systems technology. 2006. Vol. 14. No 1. P. 182 – 185.
3. Хасанов А.А., Крупенькин А.С. Веб-фреймворк для систем телемеханики “DCU-Control-Unit” // Сб. материалов XIV Конференции молодых ученых «Навигация и управление движением». СПб., 2012.
4. Хасанов А.А., Крупенькин А.С. “DCU-Control-Unit” – SDK для систем телемеханики с веб-интерфейсом // Сб. материалов конференции «Список – 2012». СПб., 2012.



Чивилихин Даниил Сергеевич
Факультет информационных технологий и программирования
Кафедра компьютерных технологий
E-mail: chivilikhin.daniil@gmail.com

РАЗРАБОТКА МЕТОДОВ ГЕНЕРАЦИИ УПРАВЛЯЮЩИХ КОНЕЧНЫХ АВТОМАТОВ НА ОСНОВЕ МУРАВЬИНЫХ АЛГОРИТМОВ

Д.С. Чивилихин

Научный руководитель – профессор А.А. Шалыто

Введение

В последнее время проводится все больше исследований в области поисковой инженерии программного обеспечения [1]. В этом подходе методы поисковой оптимизации применяются для автоматизированного построения программ для прикладных задач. Одним из классов таких задач является управление объектами со сложным поведением, для которых весьма эффективно использование автоматного программирования [2]. Его основное преимущество – возможность автоматизированной верификации автоматных программ при помощи метода Model Checking [3]. Основной проблемой при этом является трудность построения автоматов вручную. Для решения этой задачи широко применяются методы поисковой оптимизации, например генетические алгоритмы [4].

Одним из представителей методов роевого интеллекта [5], использующих принципы поведения коллективов живых существ для решения комбинаторных задач, являются муравьиные алгоритмы [6]. Цель настоящей работы – разработка метода построения конечных автоматов, основанного на муравьиных алгоритмах. Новизна предложенного подхода заключается в том, что ни один из методов роевого интеллекта до сих пор не применялся для построения конечных автоматов.

Постановка задачи

Предложенный метод позволяет строить в том числе управляющие конечные автоматы. Управляющий конечный автомат – это семерка $\langle S, X, \Sigma, \Delta, \delta, \lambda, s_0 \rangle$, где S – множество состояний, X – множество булевых входных переменных, Σ – множество входных событий, Δ – множество выходных воздействий, $\delta: S \times \Sigma \times 2^X \rightarrow S$ – функция переходов, $\lambda: S \times \Sigma \times 2^X \rightarrow \Delta^*$ – функция выходов, а $s_0 \in S$ – стартовое состояние. Пример управляющего конечного автомата приведен на рис. 1.

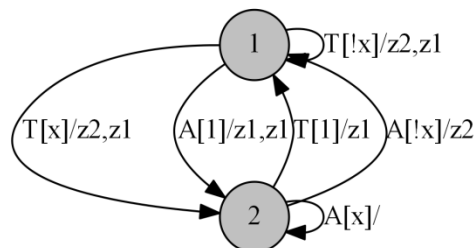


Рис. 1. Пример управляющего конечного автомата

Введем вещественную функцию приспособленности f автомата (ФП), определенную на множестве всех конечных автоматов, значения которой пропорциональны близости наблюдаемого поведения автомата к желаемому. Тогда задача построения автомата может быть сформулирована как задача оптимизации. Пусть задано число состояний N_{states} ,

множество входных событий Σ и множество выходных воздействий Δ целевого автомата. Требуется найти автомат x с достаточно большим значением ФП $f(x) \geq f_b$.

Недостатки существующих решений

Все существующие методы построения конечных автоматов можно условно разделить на два класса: методы, основанные на слиянии состояний, и методы, основанные на метаэвристических алгоритмах, в том числе на генетических.

Первый класс методов также разбивается на два подкласса: методы, использующие для слияния состояний некоторые эвристики [7], и методы, основанные на сведении задачи построения автомата к задаче о выполнимости булевой формулы (SAT) [8, 9]. Недостатком методов, использующих эвристики, является невысокая эффективность, в то время как методы, основанные на сведении к SAT, куда более эффективны. Общим существенным недостатком методов первого класса является то, что их можно применять только для построения автоматов по наборам тестов или сценариев. Они не используются для задач построения автоматов на основе моделирования.

У второго класса методов, основанных на генетических алгоритмах, есть два недостатка, один из которых – невысокая эффективность. Второй недостаток заключается в том, что для применения генетического алгоритма необходимо для каждого случая реализовывать оператор скрещивания особей, что является нетривиальной задачей.

Основные положения предложенного метода

Метод построения конечных автоматов основан на альтернативном сведении этой задачи к оптимизации на графе. Центральным понятием здесь является понятие мутации конечного автомата – небольшого изменения его структуры. Используется два типа мутаций: для случайного перехода меняется либо состояние, в которое он ведет, либо записанное на нем выходное воздействие. Пример применения мутаций приведен на рис. 2.

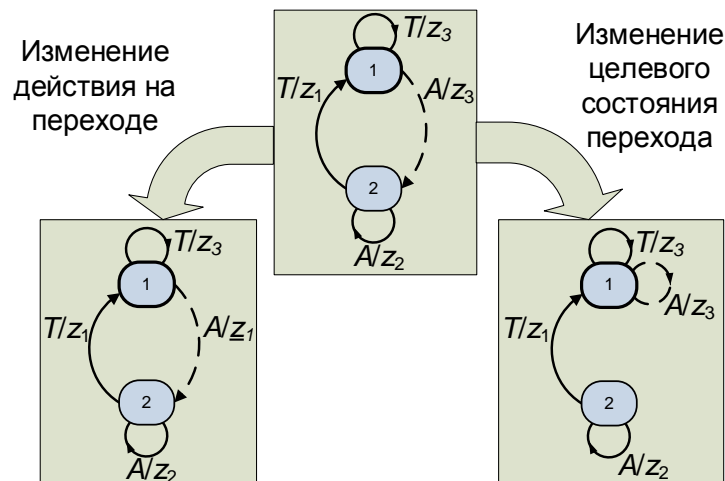


Рис. 2. Пример применения мутаций

Представление пространства поиска

Множество всех автоматов с заданными параметрами представляется в виде ориентированного графа G , который будем называть графом мутаций. Две вершины в графе G соединены ребром, если соответствующие им автоматы могут быть получены друг из друга с помощью одной операции мутации.

Кроме того, на каждом ребре (u, v) графа задается значение эвристической информации, вычисляемое по формуле:

$$\eta_{uv} = \max(\eta_{\min}, f(v) - f(u)),$$

где η_{\min} – небольшая положительная константа, например 10^{-3} . Пример небольшой части графа мутаций приведен на рис. 3.

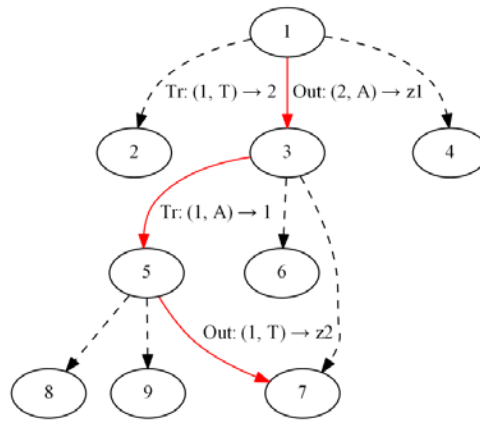


Рис. 3. Пример графа мутаций. «Tr» обозначает мутацию целевого состояния; «Out» – мутацию выходного воздействия

Алгоритм работы муравьев

В начале работы предложенного метода строится случайный конечный автомат с заданным числом состояний N_{states} , который добавляется в граф и становится его первой вершиной. Далее, пока не найдено решение или не исчерпаны выделенные алгоритму вычислительные ресурсы, в бесконечном цикле выполняются две операции: построение решений муравьями и обновление значений феромона.

Во время первой операции все муравьи начинают строить решения из вершины, соответствующей лучшему на данный момент решению. На каждой итерации муравьиной колонии каждый муравей обходит граф. Пусть муравей находится в вершине u , ассоциированной с автоматом A . Если у вершины u существуют инцидентные ей ребра, то муравей выполняет одно из следующих действий: построение новых решений либо выбор из существующих. Если у вершины u нет инцидентных ей ребер, то муравей всегда выполняет первое действие – построение новых решений.

При построении новых решений муравей выполняет N_{mut} случайных мутаций автомата A . Результатом выполнения каждой мутации является автомат A_{mut} . Далее, если в графе G нет вершины t , ассоциированной с автоматом A_{mut} , такая вершина создается и добавляется в граф. Наконец, в граф добавляется ребро (u, t) . После выполнения всех мутаций муравей переходит в вершину, соответствующую лучшему из построенных на этом шаге автоматов. Данное правило выбора проиллюстрировано на рис. 4.

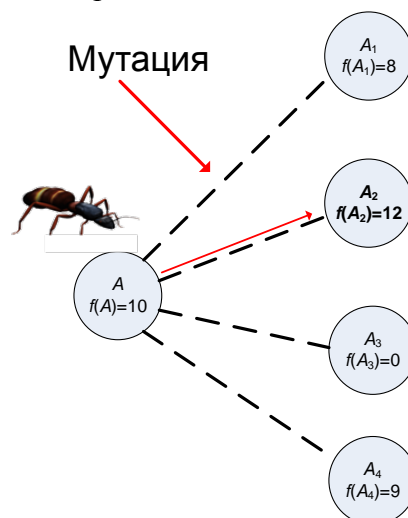


Рис. 4. Выбор новой вершины путем построения новых решений. Муравей производит несколько мутаций автомата A , получая автоматы A_1 – A_4 . В качестве следующей вершины выбирается вершина A_2 как соответствующая автомату с наибольшим значением функции приспособленности $f(A_2) = 12$

При выборе из существующих решений следующая вершина $v \in N_u$ выбирается из множества вершин N_u , инцидентных текущей вершине u с вероятностью:

$$p_{uv} = \frac{\tau_{uv}^\alpha \cdot \eta_{uv}^\beta}{\sum_{w \in N_u} \tau_{uw}^\alpha \cdot \eta_{uw}^\beta}, \quad \alpha, \beta \in [0; +\infty).$$

Последнее правило выбора проиллюстрировано на рис. 5.

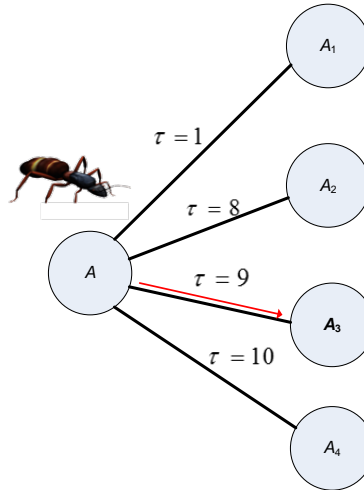


Рис. 5. Выбор новой вершины из числа существующих инцидентных вершин. В силу того, что формула выбора вершины является вероятностной, муравей может перейти не в вершину A_4 по ребру наибольшему значению феромона τ , а в вершину A_3

Каждый муравей может выполнить максимум n_{stag} шагов без увеличения своего значения ФП. После этого муравей будет остановлен. Аналогично, колония муравьев может выполнить максимум N_{stag} итераций без увеличения максимального значения ФП. После этого алгоритм перезапускается.

Обновление значений феромона

При обновлении значений феромона значение, которое муравей откладывает на ребрах своего пути, равно максимальному значению ФП всех автоматов, посещенных этим муравьем. Для каждого ребра (u, v) графа G хранится число τ_{uv}^{best} – наибольшее из значений феромона, когда-либо отложенных на этом ребре. Из пути каждого муравья выделяется отрезок от начала пути до вершины, содержащей автомат с наибольшим на пути значением ФП. Для всех ребер из этого отрезка обновляются значения τ_{uv}^{best} , а затем значения феромона на всех ребрах графа G обновляются по формуле:

$$\tau_{uv} = \max(\tau_{min}, \rho\tau_{uv} + \tau_{uv}^{best}),$$

где $\tau_{min} = 10^{-3}$ – фиксированная нижняя граница значений феромона, а $\rho \in [0,1]$ – скорость его испарения.

Экспериментальные исследования

Сравнение с генетическим алгоритмом

Одной из рассмотренных в рамках экспериментальных исследований задач является задача построения управляющих автоматов по тестам. Она состоит в построении автомата с фиксированным поведением и заданным набором обучающих тестовых примеров (тестов) T . Тест задается двумя последовательностями: последовательностью входов $In[i]$ и соответственно выходов $Ans[i]$. Последовательность $In[i]$ составлена из элементов множества входных событий Σ , а последовательность выходов $Ans[i]$ – из элементов множества выходов

Δ. Конечный автомат удовлетворяет тесту $T_i = \{In[i], Ans[i]\}$, если последовательность $Ans[i]$ является результатом передачи автомату на вход последовательности $In[i]$. Задача состоит в нахождении автомата с заданным числом состояний, удовлетворяющего всем тестам.

Используемая ФП основана на редакционном расстоянии между строками [10]. Каждая входная последовательность $In[i]$ подается на вход построенному автомату, при этом запоминается выходная последовательность $Out[i]$. Полученная выходная последовательность сравнивается с эталонной $Ans[i]$. Для оценки схожести полученных последовательностей вычисляется значение выражения:

$$f_1 = \frac{1}{|T|} \sum_{i=1}^{|T|} \left(1 - \frac{ED(Out[i], Ans[i])}{\max(\text{len}(Out[i]), \text{len}(Ans[i]))} \right),$$

где $|T|$ – число тестов в наборе, $\text{len}(s)$ – длина последовательности s , а $ED(s_1, s_2)$ – редакционное расстояние между последовательностями s_1 и s_2 . Окончательное выражение для ФП учитывает число переходов $n_{\text{transitions}}$ автомата:

$$f = \begin{cases} 10 \cdot f_1 + \frac{1}{M} (M - n_{\text{transitions}}), & \text{если } f_1 < 1 \\ 20 + \frac{1}{M} (M - n_{\text{transitions}}), & \text{если } f_1 = 1 \end{cases}, \quad M > n_{\text{transitions}}.$$

В каждом эксперименте сначала генерировался случайный конечный автомат и по нему строились тесты. Далее с помощью алгоритмов поисковой оптимизации строился удовлетворяющий тестам управляющий автомат. Процесс построения набора тестов состоит из этапов построения случайного управляющего автомата с заданным числом состояний N_{states} и построения по нему тестов суммарной длиной $150 \times N_{\text{states}}$. Сравнение велось с предложенным в [12] алгоритмом ГА-2+ЭС, который комбинирует генетический алгоритм и эволюционную стратегию. Особенностью генетического алгоритма является то, что в нем применяется специальный оператор кроссовера, учитывающий поведение автомата на тестах.

Для каждого сгенерированного автомата с 4–10 состояниями проводилось по 100 запусков предложенного метода и по 100 запусков алгоритма ГА-2+ЭС. В каждом запуске оба алгоритма работали вплоть до достижения целевого значения ФП. Графики средних значений числа вычислений ФП этих методов приведены на рис. 6.

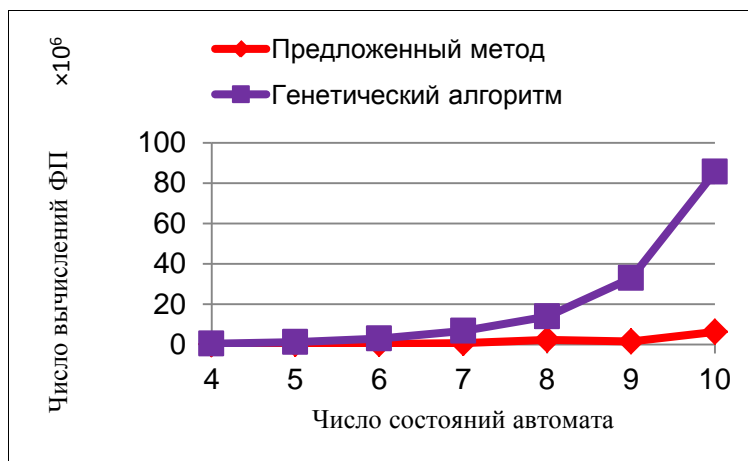


Рис. 6. Среднее число вычислений ФП для предложенного метода и генетического алгоритма при различных размерах целевых автоматов

Как видно из графиков, предложенный метод во всех случаях оказался в среднем в 3–20 раз эффективнее генетического алгоритма, причем это преимущество увеличивается с ростом числа состояний целевых автоматов. Статистическая значимость полученных результатов была подтверждена с помощью статистического теста ANOVA [7], показавшего во всех случаях результат $p < 10^{-4}$, что говорит о существенном различии в производительности сравниваемых алгоритмов.

Сравнение с классическим муравьиным алгоритмом

Рассмотрена также возможность применения к задаче построения автоматов классических муравьиных алгоритмов. В этом случае рассматривается так называемый граф конструирования, содержащий все возможные переходы автомата с заданными параметрами. Граф является полным, то есть каждая вершина связана ребром с каждой другой вершиной. Пример такого графа для автоматов из двух состояний с двумя входными событиями и единственным выходным воздействием приведен на рис. 7.

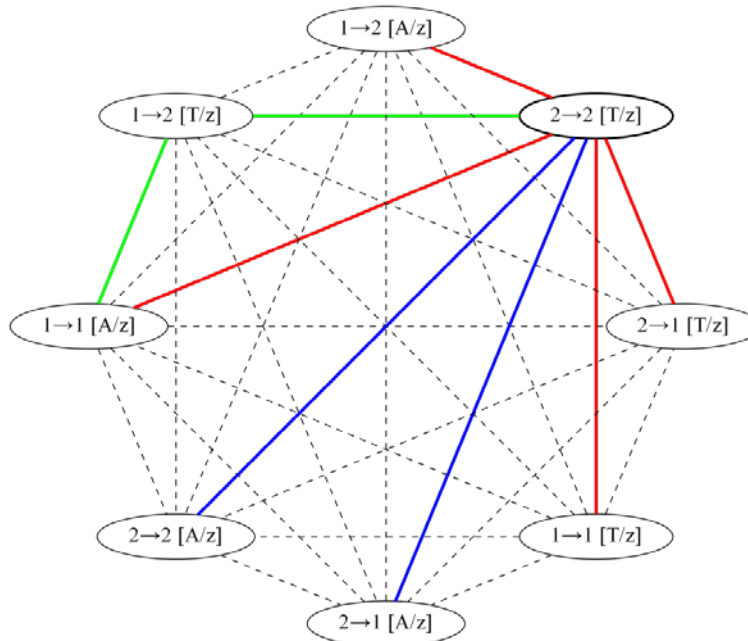


Рис. 7. Пример графа конструирования для построения автоматов с помощью классических муравьиных алгоритмов. Зеленым отмечены ребра, посещенные муравьем; красным – запрещенные; синим – разрешенные ребра

Сравнение предложенного метода с классическим муравьиным алгоритмом проводилось на примере простой задачи об «Умном муравье» с полем Санта Фе [2]. Экспериментальное сравнение алгоритмов выполнено для автоматов из $N_{\text{states}} = [5..10]$ состояний. Эксперимент для каждого числа состояний был повторен 1 000 раз. Каждый эксперимент продолжался либо до достижения оптимального решения, либо до исчерпания алгоритмом выделенных 30 тыс. вычислений ФП. Алгоритмы сравнивались по доле успешных запусков и среднему времени работы. Запуск алгоритма считается успешным, если в нем было найдено оптимальное решение задачи.

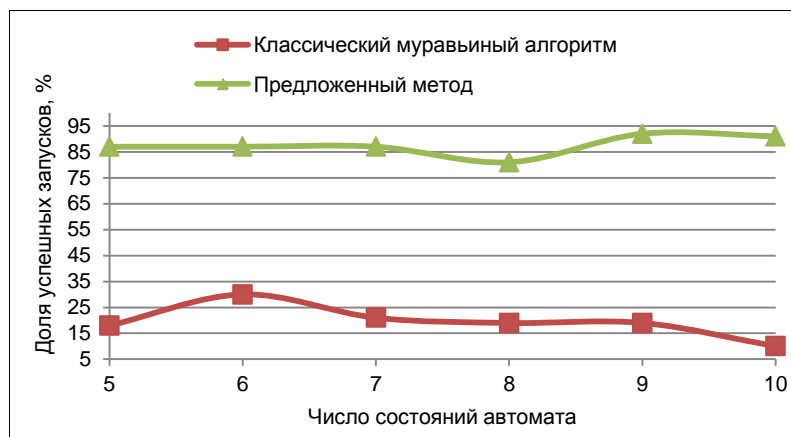


Рис. 8. График зависимости доли успешных запусков алгоритмов от числа состояний автомата

Результаты, представленные на рис. 8 и 9 позволяют утверждать, что метод, основанный на графе мутаций, существенно превосходит по производительности классический муравьиный алгоритм. Он находит оптимальное решение в 3–9 раз чаще, причем делает это в 30–400 раз быстрее.

Способ увеличения производительности предложенного метода

При разработке методов и проведении экспериментальных исследований был предложен подход, не имеющий прямого отношения к муравьиным алгоритмам, однако позволяющий увеличить их производительность. Необходимо при вычислении значения ФП автомата пометать переходы, которые при этом совершались. Если переход автомата не использовался при вычислении ФП, то поведение автомата при изменении такого перехода не может измениться, следовательно, значение ФП можно не вычислять заново.

Предложенный подход тестировался в применении к методу, основанному на графе мутаций и задаче об «Умном муравье». На рис. 10 приведены графики зависимости среднего числа вычислений ФП от числа состояний автомата с использованием и без использования пометок на переходах для поля Санта Фе. Как видно из графиков, выигрыш от использования пометок увеличивается с ростом числа состояний целевого автомата. Это объясняется тем, что вероятность совершить мутацию, не меняющую поведение автомата, увеличивается с ростом числа его состояний.

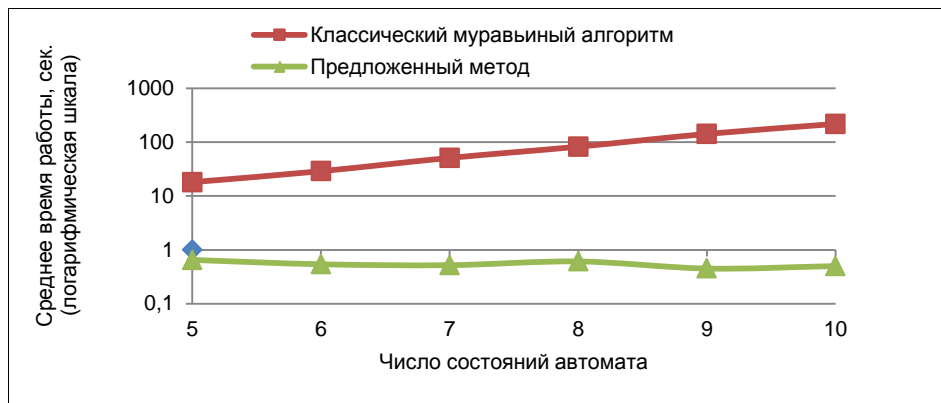


Рис. 9. График зависимости среднего времени работы алгоритмов от числа состояний автомата (логарифмическая шкала)

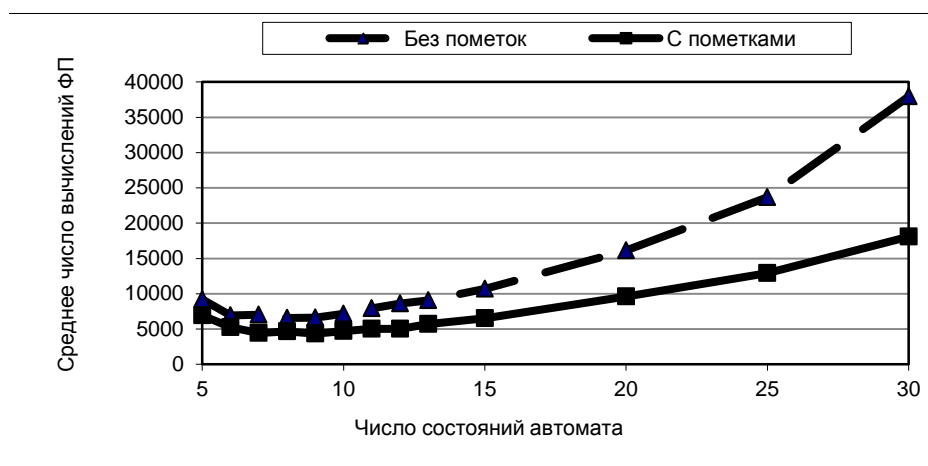


Рис. 10. Влияние пометок об использовании переходов на производительность предложенного алгоритма

Заключение

В работе рассмотрены существующие методы построения управляющих конечных автоматов, проведен их критический анализ. Разработан новый метод построения конечных

автоматов, основанный на муравьином алгоритме нового типа и представлении пространства поиска в виде графа мутаций.

Проведены экспериментальные сравнения предложенного метода с известными, в том числе с генетическим и классическим муравьиным алгоритмом. Полученные экспериментальные результаты свидетельствуют о том, что для рассмотренных задач предложенный метод существенно эффективнее. Показана статистическая значимость полученных результатов. Реализован метод дальнейшего ускорения разработанного варианта, основанный на введении пометок об использовании переходов конечных автоматов.

Результаты проведенной работы можно использовать для автоматизации разработки автоматных программ. Это позволит сократить время, необходимое для генерации управляющих автоматов для программ рассматриваемого класса. Реализация предлагаемого метода может быть оформлена в виде программного средства и коммерциализирована.

Публикации

1. *Chivilikhin D., Ulyantsev V.* MuACOsm – A New Mutation-Based Ant Colony Optimization Algorithm for Learning Finite-State Machines / In Proceedings of the fifteenth Genetic and Evolutionary Computation Conference (GECCO'13), Christian Blum (Ed.). ACM, New York, NY, USA, 2013, pp. 511–518
2. *Chivilikhin D., Ulyantsev V., Shalyto A.* Solving Five Instances of the Artificial Ant Problem with Ant Colony Optimization / In Proceedings of the 2013 IFAC Conference on Manufacturing Modelling, Management and Control (MIM'13). Saint-Petersburg, Russia, 2013. Vol. 7, part 1, pp. 1043–1048

Литература

1. *Harman M.* Software engineering meets evolutionary computation // *Computer*. 2011. Vol. 44. No 11. P. 31–39.
2. *Поликарпова Н.И., Шалыто А.А.* Автоматное программирование. СПб., 2011. – 176 с.
3. *Вельдер С.Э., Лукин М.А., Шалыто А.А., Яминов Б.Р.* Верификация автоматных программ. СПб., 2011. – 244 с.
4. *Mitchell M.* An introduction to genetic algorithms. Cambridge, MA, 1996. – 209 p.
5. *Bonabeau E., Dorigo M., Theraulaz G.* Swarm intelligence: From natural to artificial systems. New York; Oxford, 1999. – 320 p.
6. *Dorigo M., Stützle T.* Ant colony optimization. Cambridge, MA; London, 2004. – 319 p.
7. *Lang K.J., Pearlmutter B.A., Price R.A.* Results of the Abbadingo One DFA learning competition and a new evidence-driven state merging algorithm // *Grammatical Inference (4th International Colloquium, ICGI-98, Ames, Iowa, USA, July 12–14, 1998)*. Proceedings. Lecture Notes in Computer Science. 1998. Vol. 1433. P. 1–12.
8. *Heule M., Verwer S.* Exact DFA identification using SAT solvers // *Grammatical Inference: Theoretical Results and Applications (10th International Colloquium, ICGI-2010, Valencia, Spain, September 13–16, 2010)*. Proceedings. Lecture Notes in Computer Science. 2010. Vol. 6339. P. 66–79.
9. *Ulyantsev V., Tsarev F.* Extended finite-state machine induction using SAT-solver // *Proceedings of the 14th IFAC Symposium “Information Control Problems in Manufacturing (INCOM’12)”*. Buharest, Romania. 2012. P. 512–517.
10. *Царев Ф.Н.* Методы построения конечных автоматов на основе эволюционных алгоритмов: Дис. ... канд. техн. наук. СПб., 2012. – 196 с. (http://is.ifmo.ru/disser/tsarev_disser.pdf).
11. *Miller R.G.* Beyond ANOVA: Basics of applied statistics. Chapman and Hall/CRC. 1997. – 336 p.
12. *Chivilikhin D., Ulyantsev V., Shalyto A.* Solving five instances of the artificial ant problem with ant colony optimization // *Proceedings of the 2013 IFAC Conference on Manufacturing Modelling, Management and Control (MIM’13)*. 2013. Vol. 7. P. 1. P. 1043–1048.



Шитов Денис Дмитриевич
Факультет оптико-информационных систем и технологий
Кафедра оптико-электронных приборов и систем
E-mail Qdenisq@gmail.com

ИССЛЕДОВАНИЕ И РАЗРАБОТКА МЕТОДИКИ НАСТРОЙКИ ЦВЕТОПЕРЕДАЧИ ВИДЕОИНФОРМАЦИОННЫХ СИСТЕМ ЦВЕТОВОГО АНАЛИЗА

Д.Д. Шитов

Научный руководитель – А.Н. Чертов, доцент

Введение

На сегодняшний день оптико-электронные системы технического зрения (ОЭС ТЗ) широко применяются в промышленности, науке и повседневной жизни. Как правило, цветопередача таких систем вполне достаточна для визуализации изображения на экране. Однако нередко возникает необходимость определения точных цветовых координат отображаемого объекта в каждой точке его поверхности.

В производственной отрасли с помощью ОЭС ТЗ контролируют процесс производства в автоматическом режиме, который не требует постоянного наблюдения оператора; в научной сфере они позволяют проводить цветовой анализ исследуемых объектов для получения необходимых сведений. В горнодобывающей промышленности ОЭС ТЗ применяются для обогащения руд твердых полезных ископаемых. Даже в пищевой промышленности они нашли свое применение. Так, качество растительного масла определяют именно по его цвету, а специализированные системы производят эту операцию более оперативно и зачастую качественнее, чем человек. Цвет лакокрасочной, текстильной продукции также контролируется при помощи систем технического зрения [1].

Таким образом, ОЭС ТЗ нашли широкое применение, но до сих пор остается ряд проблем, которые не позволяют исключить необходимость участия человека в процессе. Одной из них является некорректная цветопередача таких систем. Несоответствие цветовых координат снимаемого объекта или сцены и цветовых координат изображения, получаемого оптико-электронной системой, являются причиной ошибок при принятии решения, погрешностей результатов цветового анализа [2].

Для устранения ошибок и повышения качества цветопередачи существует ряд методик по настройке ОЭС ТЗ. Их главная цель – настройка цветопередачи таким образом, чтобы получаемая картина в наибольшей степени соответствовала картине, воспринимаемой наблюдателем. Следует отметить, что данная цель в силу психофизиологических свойств человеческого зрения осуществима лишь в некотором приближении [3]. Но, даже не учитывая этот факт, предлагаемые на сегодняшний день методики по настройке цветопередачи в большинстве случаев не способны обеспечить необходимую точность измерения и цветового анализа. Более того, разработка принципиально новой или усовершенствованной методики с увеличенной эффективностью позволит значительно расширить границы применения ОЭС ТЗ.

Цель и задачи работы

Целью данной работы является создание специализированной методики оперативной подстройки параметров цветопередачи ОЭС ТЗ под особенности конкретной задачи и условия обнаружения и/или анализа.

Для достижения указанной цели необходимо решить следующие задачи:

- изучить физиологию цветового зрения;

- исследовать существующие методики настройки цветопередачи ОЭС ТЗ;
- разработать теоретические основы процесса настройки цветопередачи видеоинформационных систем;
- реализовать разработанный алгоритм и тестирование его в различных условиях освещения.

Методы исследования

В работе применяются:

- элементы классической теории цвета и его измерения, в том числе модели представления и методы преобразования цвета;
- итерационный метод построения сетки на основе критерия Делоне на заданном наборе точек.

Достигнутые результаты

Разработка теоретических основ процесса настройки цветопередачи видеоинформационной системы

Концепция предлагаемой методики настройки цветопередачи заключается в последовательном преобразовании всех точек цветового пространства, построенного по показаниям сигнала камеры. Для построения цветового пространства используется аддитивная цветовая модель RGB, которая является наиболее удобной для цветовоспроизведения и не вызывает трудностей при линейных преобразованиях [4].

Преобразование точек происходит на основе анализа сигнала ОЭС ТЗ при наблюдении тестовой колориметрической таблицы в заданных условиях освещенности рабочей зоны. На данный момент такие таблицы используются в качестве эталонных образцов для оценки качества цветопередачи цветных телевизионных систем наблюдения. Они представляют собой таблицы, содержащие ячейки различных цветов с известными цветовыми координатами, которые для предлагаемого метода будут являться «опорными цветами». Использование такой таблицы позволяет произвести анализ показаний сигнала камеры во всех областях цветового пространства.

Процесс настройки цветопередачи согласно предлагаемому методу можно проводить в несколько этапов:

1. Выделение в цветовом пространстве RGB точек, соответствующих сигналам ОЭС ТЗ при наблюдении тестовой колориметрической таблицы (рис. 1).
2. Аналогичное выделение «опорных точек» в цветовом пространстве RGB, основанное на заранее известных цветовых координатах эталонных образцов.
3. Создание матрицы преобразования всех точек пространства ОЭС ТЗ в эталонное цветовое пространство.

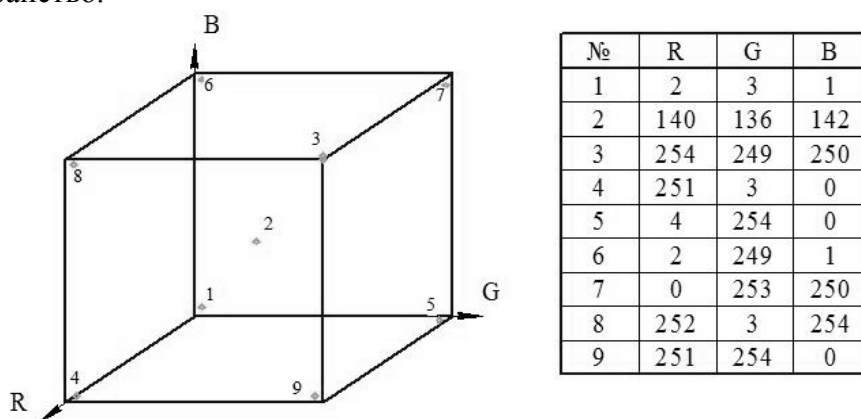


Рис. 1. Положение «опорных точек» в цветовом пространстве RGB

Заключительный этап настройки ОЭС ТЗ является самым трудоемким и сложным и требует подробного рассмотрения. Матрица преобразования имеет количество элементов, равное количеству цветов в цветовом пространстве, которые различает ОЭС ТЗ. Причем

каждый элемент содержит в себе информацию о цветовых координатах соответствующей точки пространства и о необходимом изменении этих координат.

В разрабатываемом методе расчет изменения координат каждой точки пространства производится путем анализа смещения «опорных цветов» от необходимого положения в пространстве. Предлагается поочередно выделять в цветовом пространстве элементарные объемные фигуры (треугольные пирамиды), вершинами которых являются точки, соответствующие сигналам ОЭС ТЗ при наблюдении тестовой колориметрической таблицы. Далее следует поочередно производить преобразование каждой вершины так, чтобы ее цветовые координаты стали равны координатам соответствующего «опорного цвета», и рассчитывать, как изменятся координаты всех точек, заключенных в пирамиде (рис. 2). Такой метод позволяет значительно упростить расчет матрицы преобразования, так как при движении одной вершины можно считать, что остальные вершины пирамиды остаются статичными. Таким образом, задача сводится к определению смещения всех точек внутри пирамиды при изменении координат одной из ее вершин.

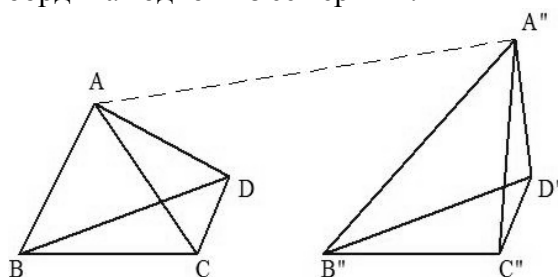


Рис. 2. Деформация пирамиды при смещении одной вершины

Любое изменение координат вершины пирамиды можно представить в виде суммы двух типов движения: сжатия (растяжения) и сдвига.

Сжатие (растяжение) является типом движения, при котором все точки пирамиды, за исключением статичного основания, перемещаются в направлении, перпендикулярном плоскости основания. В таком случае все точки, лежащие на одной и той же высоте от основания, изменят свои цветовые координаты на одно и то же значение.

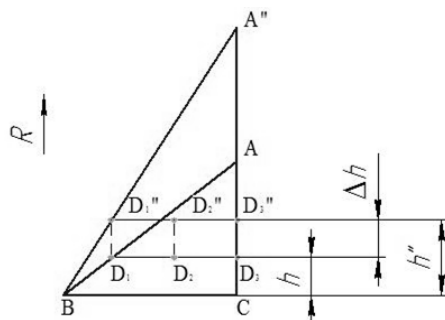


Рис. 3. Пояснительный рисунок к процессу «сжатие (растяжение)»

Для пояснения метода ниже рассмотрено преобразование треугольника ABC вместо пирамиды $ABCD$. Несмотря на подобное упрощение, все формулы применимы и для точек внутри объемной фигуры. Рис. 3 показывает, как изменяется положение точек треугольника ABC при растяжении. Вершина $A(R_a; G_a; B_a)$ изменяет свои координаты $A''(R_a''; G_a; B_a)$, основание BC остается неподвижным. При этом точки D_1, D_2, D_3 смещаются в D_1'', D_2'', D_3'' соответственно и изменяют свою координату по оси R на одну и ту же величину Δh . Введем обозначения: h – высота точки D_1 над основанием; h'' – высота смещенной D_1'' ; $k=h''/h$ – коэффициент растяжения, который остается *const* для всех точек треугольника ABC .

Таким образом:

$$k = \frac{R_a'' - R_c}{R_a - R_c} \quad (1)$$

Искомая координата точки D'' по оси R :

$$R_d'' = R_c + h'' = R_c + h \cdot (k - 1) = R_c + (R_d - R_c) \cdot (k - 1) \rightarrow$$

$$R_d'' = R_c + (R_d - R_c) \cdot \left(\frac{R_a'' - R_c}{R_a - R_c} - 1 \right) = R_c + \frac{(R_d - R_c) \cdot (R_a'' - R_a)}{R_a - R_c}. \quad (2)$$

Полученное выражение справедливо для всех точек, находящихся внутри и на сторонах треугольника ABC .

Сдвиг представляет собой движение всех точек пирамиды, за исключением основания, в плоскости, параллельной неподвижному основанию. Рассмотрим треугольник ABC со статичным основанием BC , вершина A перемещается в A'' вдоль оси G (рис. 4).

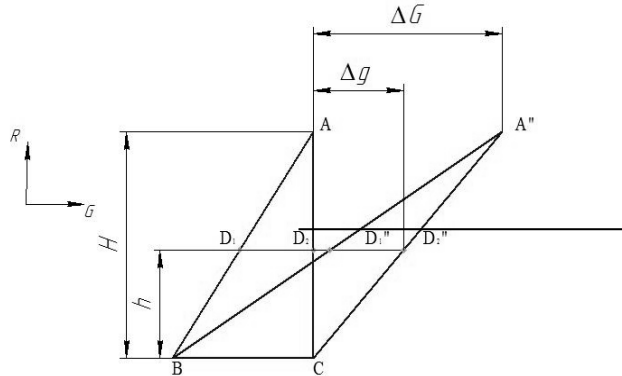


Рис. 4. Пояснительный рисунок к процессу «сдвиг»

Площади треугольников ABC и $A''BC$ равны, так как имеют общее основание BC и равные высоты H . Треугольник AD_1D_2 подобен ABC и относится к нему, как отношение $(H-h)/H$. Треугольник $AD_1''D_2''$ подобен $A''BC$ и относится к нему, как отношение высот $(H-h)/H$, следовательно, площади треугольников AD_1D_2 и $AD_1''D_2''$ равны. Из предыдущего заключения следует, что $D_1D_2 = D_1''D_2''$ и точки D_1 и D_2 изменяют свою координату по оси G на одну и ту же величину Δg . Введем обозначение: ΔG – величина смещения вершины A ; h – длина перпендикуляра, опущенного из точки D на основание BC ; $p = h/H$ – коэффициент сдвига, который одинаков для всех точек на отрезке D_1D_2 .

Из факта подобия треугольников CD_2D_2'' и CAA'' следует:

$$\frac{D_2D_2''}{AA''} = \frac{CD_2}{CA} \rightarrow \Delta g = p \cdot \Delta G. \quad (3)$$

Тогда координата точки D_2'' по оси G :

$$G_d'' = G_d + \Delta g = G_d + p \cdot \Delta G = G_d + \frac{h \cdot \Delta G}{H} = G_d + \frac{(R_d - R_c) \cdot (G_a'' - G_a)}{(R_a - R_c)}. \quad (4)$$

Формула (4) применима и для нахождения координаты по оси B с внесением соответствующих изменений. Используя формулы (2) и (4) можно произвести пересчет координат любой точки внутри пирамиды.

Все расчеты были выполнены для случая, когда основание BC параллельно плоскости GOB . В остальных случаях требуется производить расчет в новой системе координат с прежним центром, но повернутой вокруг него так, что плоскость неподвижного основания будет параллельной одной из плоскостей, образованной координатными осями.

На основании предложенной методики реализован алгоритм выделения и преобразования точек, заключенных в одном тетраэдре, на языке программирования Matlab. Блок-схема алгоритма приведена на рис. 5.

Весь алгоритм делится на несколько логически обособленных частей, каждая из которых выполняет свою определенную функцию. Первая часть описывает функцию выделения всех точек, находящихся внутри и на гранях тетраэдра, вершины которого предварительно заданы. Результатом ее работы является двумерный массив, состоящий из координат точек

и отсортированный по возрастанию одной из координат. В свою очередь эта функция разбивается на несколько более простых частей. Это вызвано тем, что при рассмотрении сечения тетраэдра, параллельного одной из координатных плоскостей, возможен случай, когда само сечение будет представлять собой треугольник либо некую четырехугольную фигуру (рис. 6).

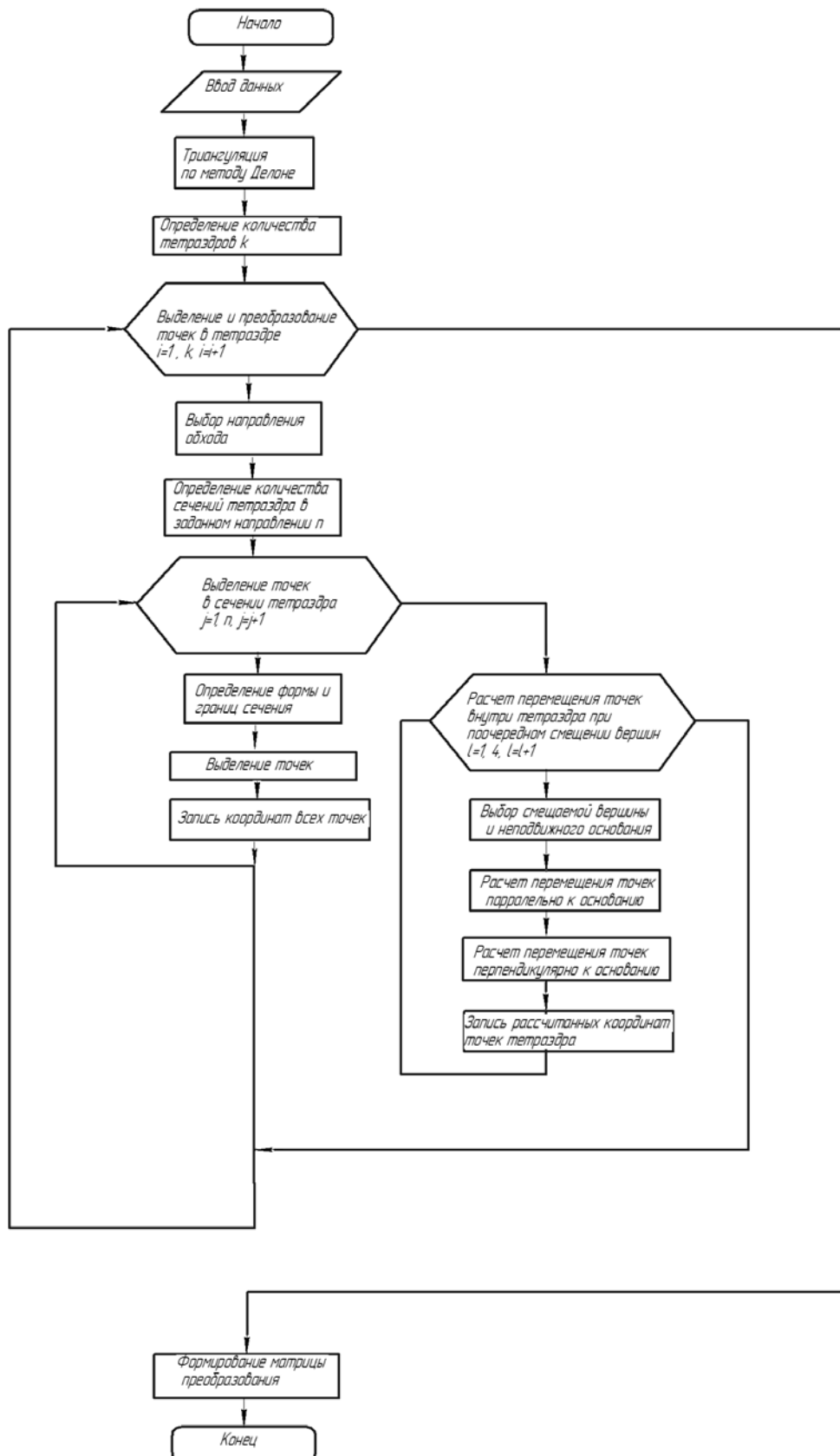


Рис. 5. Блок-схема алгоритма настройки цветопередачи

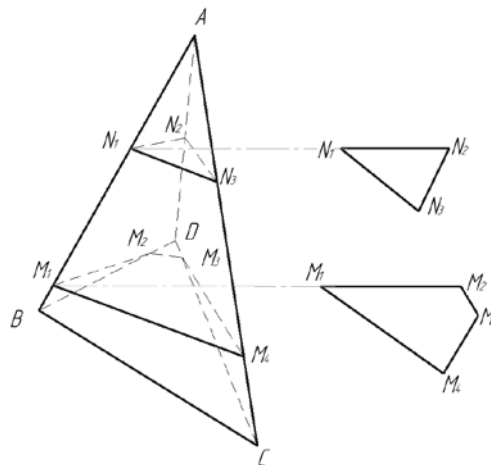


Рис. 6. Различные варианты сечений тетраэдра

Поэтому выделение точек во всей пирамиде разделено на три части. Предлагаемый алгоритм позволяет рассчитать матрицу преобразования точек пространства, замкнутого внутри одного тетраэдра, тогда как задачей является произвести подобный расчет для всех точек пространства. Для этого требуется разбить все цветовое пространство на некоторое количество тетраэдров и затем последовательно проделать процедуру расчета над каждым тетраэдром в отдельности. Фактически задача сводится к построению некоторой сетки, узлами которой являются «опорные цвета». Процесс построения сетки обычно называется дискретизацией или триангуляцией (даже если речь идет о трех измерениях) [5].

Все методы триангуляции по принципу построения можно разбить на две большие группы: прямые и итерационные. В прямых методах сетка строится за один этап, причем ее топология (иначе говоря, граф связей между узлами) и координаты всех узлов известны изначально. В итерационных методах сетка строится последовательно; на каждом шаге добавляется один или несколько элементов, причем изначально не известны ни координаты узлов, ни топология сетки. Кроме того, координаты узлов и топология могут меняться прямо в процессе построения. В данной работе используется именно итерационный метод построения сетки, а точнее метод на основе критерия Делоне по заданному набору точек. Как следует из названия раздела, исходными данными алгоритма является некий набор точек, которые должны стать узлами будущей триангуляции.

Опишем более подробно содержание этапов разработанного алгоритма:

1. Формирование множества U -набора заданных узлов.
2. Создание так называемой «суперструктуры», представляющей собой произвольный выпуклый многогранник с треугольными гранями, такой, что все заданные узлы лежат внутри него. Вершинами многогранника могут быть как элементы U , так и дополнительные узлы. В дальнейшем до определенного этапа с этими узлами обращаются, как и со всеми остальными. В качестве суперструктуры может быть использован и тетраэдр.

3. Формирование множества G узлов сетки, куда переносятся все узлы U , использованные как вершины суперструктуры (если такие есть).

4. Если в качестве суперструктуры использован тетраэдр, производится переход к п. 5; иначе на основе узлов суперструктуры формируется триангуляция Делоне. Если в качестве суперструктуры использован правильный многогранник (октаэдр или икосаэдр), то это можно сделать следующим образом: выбрать произвольный узел из U , перенести его в G и путем вставки ребер между этим узлом и всеми вершинами многогранника сформировать сетку из n тетраэдров, где n – число граней, равное 8 или 20 соответственно. Эта сетка будет являться триангуляцией Делоне.

5. Поиск для всех тетраэдров сетки центров и радиусов описанной сферы.

6. Выбор произвольного узла q из множества U и перенос его в G , затем удаление всех тетраэдров, для которых q попадает внутрь описанной сферы. При этом в сетке образуется полость в виде многогранника, в общем случае имеющего форму звезды. При этом любой луч, исходящий из q , должен пересекать границу этого многогранника в единственной точке. Если

обнаруживаются тетраэдры, для которых q лежит в плоскости одной из граней (это возможно в случае неоднозначности триангуляции Делоне, если, например, пять или больше точек лежат на одной сфере), то их тоже необходимо удалить. Заметим, что фактически ребро (или грань) удаляется только в том случае, если удаляются все смежные с ним тетраэдры; при этом ребра и грани суперструктуры не удаляются никогда. Новые тетраэдры образуются путем вставки ребер между q и вершинами этого многогранника. Доказано, что при этом получается триангуляция Делоне.

7. Нахождение для новообразованных тетраэдров центра и радиуса описанной сферы.

8. Если множество U не пусто, то происходит переход к п. 6, иначе – к п. 9.

9. Удаление из сетки всех тетраэдров, среди вершин которых есть вспомогательные узлы, использовавшиеся для построения суперструктуры. В результате получается сетка, построенная только на заданных узлах (G).

Пример работы алгоритма представлен на рис. 7. Он позволяет гарантированно строить триангуляцию Делоне для произвольного набора точек, причем граница сетки будет представлять собой (в общем случае невыпуклый) многогранник с треугольными гранями, опирающимися на наиболее удаленные от центра триангуляции узлы.

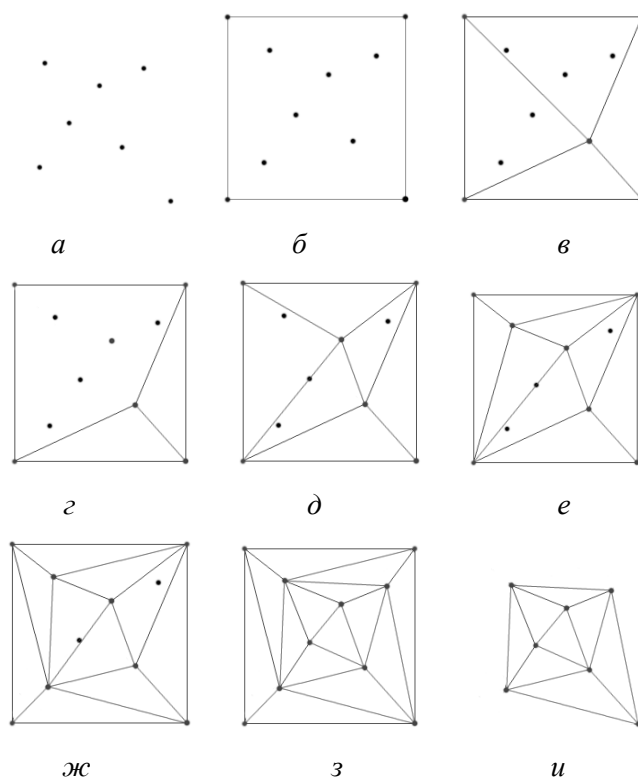


Рис. 7. Двумерный вариант алгоритма на основе критерия Делоне: а – заданный набор точек; б – постройка суперструктуры (квадрат); в – постройка исходной триангуляции Делоне; г – выбор нового узла и удаление треугольников, для которых выбранный узел лежит внутри описанной окружности; д – соединение ребрами новой вершины с углами образовавшегося многоугольника; е – выбор следующего узла и дальнейшее повторение процедуры; ж – обработка еще одной точки; з – сетка построена; и – удаление из сетки всех треугольников, среди вершин которых были вспомогательные узлы суперструктуры

Итак, описаны теоретические основы новой методики настройки цветопередачи видеоинформационных систем. Разработан и реализован алгоритм выделения и преобразования точек цветового пространства RGB в соответствии с положением «опорных точек» в нем, включающий в себя предварительную триангуляцию цветового пространства, последующее выделение точек внутри образованных сегментов (тетраэдров), расчет функций преобразования для выделенных точек и последующее создание матрицы преобразования.

Экспериментальные исследования

Экспериментальные исследования на установке проводились с целью выявить границы применимости разработанного алгоритма. Было проведено несколько серий исследований для определения зависимостей эффективности работы алгоритма от условий освещения, выбора видеокамеры, предварительной настройки цветопередачи камеры.

В первой серии в качестве фотоприемного устройства была выбрана камера фирмы ЭВС с матрицей VEI-545. Предварительная настройка цветопередачи не проводилась. Зона анализа освещалась одновременно естественным (дневной) и искусственным (люминесцентные лампы) светом. Следует отметить, что камера была расположена таким образом, что свет от окна попадал прямо в объектив, вследствие этого полученным изображениям характерна некоторая засвеченность.

Ниже (рис. 8) приведены исходное изображение тестовой таблицы и оно же после выполнения настройки цветопередачи.

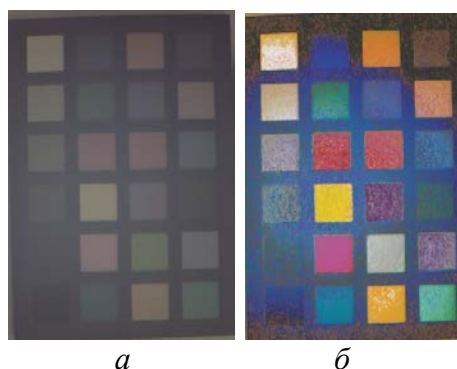


Рис. 8. Изображение тестовой таблицы: а – исходное; б – после настройки

Видно, что полученный результат не отвечает даже минимальным требованиям качества. Приходится заключить, что разработанный алгоритм работает с практически обесцвеченными изображениями некорректно.

Следующий шаг – оценка работы алгоритма в других условиях освещения, более благоприятных для эффективной настройки цветопередачи. В качестве источника излучения были выбраны две линейки светодиодов. Измерения проводились в темной комнате, с тем чтобы исключить влияние посторонних источников излучения. На этот раз источники излучения были расположены за камерой, следовательно, прямой свет от них в камеру не попадал. Предварительная настройка цветопередачи камеры не проводилась, как и в первой серии измерений. Результаты измерений представлены на рис. 9.

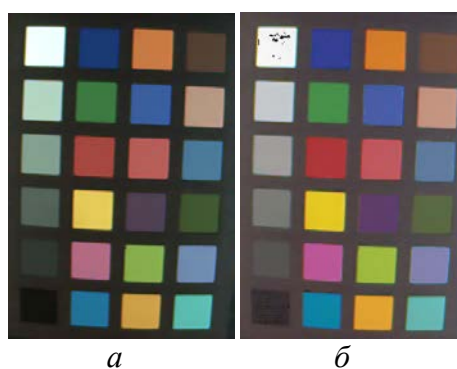


Рис. 9. Изображение тестовой таблицы в благоприятных условиях освещения: а – исходное; б – после настройки

Координаты цветов ячеек тестовой таблицы в исходном и обработанном изображениях, а также эталонные их значения представлены в Табл. 1.

Анализ данных из Табл. 1 показывает, что погрешность цветопередачи на исходном изображении в среднем варьируется от 10 до 30 градаций в каждом канале, в некоторых случаях достигая 60. При такой цветопередаче камеры невозможно говорить о цветовом анализе изображения.

Таблица 1

Результат цветокоррекции в благоприятных условиях освещения

№	Исходные координаты цвета ячеек на изображении			Координаты цвета ячеек на изображении после настройки			Эталонные координаты цвета ячеек		
	R	G	B	R	G	B	R	G	B
1	85	62	51	115	82	65	115	82	68
2	183	152	131	196	148	129	194	150	130
3	76	122	152	95	124	160	98	122	157
4	61	89	52	89	105	67	87	108	67
5	114	139	178	132	130	176	133	128	177
6	92	188	175	99	185	179	103	189	170
7	199	121	71	211	125	38	214	126	44
8	58	96	158	79	90	166	80	91	166
9	183	95	85	191	87	100	193	90	99
10	81	64	84	94	60	108	94	60	108
11	138	175	76	163	186	62	157	188	64
12	203	159	73	226	163	47	224	163	46
13	26	58	117	51	63	150	56	61	150
14	62	122	63	67	146	71	70	148	73
15	164	72	66	175	54	60	175	54	60
16	232	203	92	229	202	38	231	199	31
17	174	106	132	187	87	152	187	86	149
18	32	116	151	8	135	164	8	133	161
19	244	255	255	245	244	244	243	243	242
20	182	212	191	200	198	202	200	200	200
21	132	161	144	160	159	157	160	160	160
22	87	99	87	126	121	115	122	122	121
23	49	55	54	90	83	80	85	85	85
24	17	19	10	52	47	47	52	52	52

После настройки по описанному алгоритму средняя погрешность цветопередачи не превышает пять градаций, а максимальная погрешность – десять. Полученный результат свидетельствует о значительном повышении качества цветопередачи камеры, делая возможным ее использование для цветового анализа снимаемых объектов.

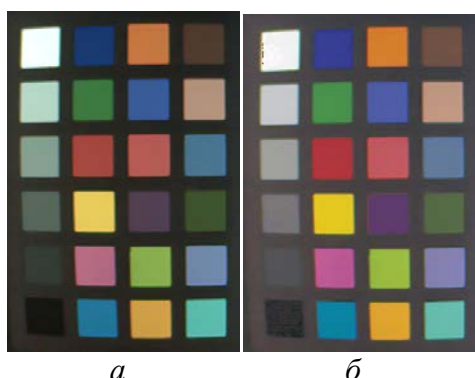


Рис. 10. Изображение тестовой таблицы после предварительной настройки с использованием программного обеспечения: а – исходное; б – после настройки

В следующей серии исследований рассматривается изображение, полученное на той же камере, что и в предыдущей серии в тех же условиях. Отличием является наличие предварительной настройки цветопередачи камеры, реализуемой программным обеспечением, идущим в комплекте с камерой. Результаты измерений представлены на рис. 10.

Координаты цветов ячеек тестовой таблицы в исходном и обработанном изображениях, а также эталонные их значения представлены в Табл. 2. В этой серии измерений исходное изображение хоть и лучше, чем в предыдущем варианте, однако погрешность цветопередачи в некоторых случаях достигает значительной величины в 35 градаций, а в среднем составляет 15. Обработанное изображение дает результат, аналогичный полученному в прошлой серии, средняя погрешность вновь не превышает 5 градаций.

Таблица 2

Результат цветокоррекции с предварительной обработкой изображения

№	Исходные координаты цвета ячеек на изображении			Координаты цвета ячеек на изображении после настройки			Эталонные координаты цвета ячеек		
	R	G	B	R	G	B	R	G	B
1	79	60	46	112	81	72	115	82	68
2	182	146	125	193	153	131	194	150	130
3	70	116	146	95	122	157	98	122	157
4	64	83	41	84	109	72	87	108	67
5	114	135	170	135	127	176	133	128	177
6	90	187	170	104	188	170	103	189	170
7	200	120	64	215	126	43	214	126	44
8	59	91	150	82	93	165	80	91	166
9	178	91	86	192	89	95	193	90	99
10	79	66	83	93	60	110	94	60	108
11	135	173	74	157	188	67	157	188	64
12	205	160	72	225	164	43	224	163	46
13	26	57	114	57	62	146	56	61	150
14	60	122	59	70	149	72	70	148	73
15	167	71	63	178	53	64	175	54	60
16	234	205	90	231	200	30	231	199	31
17	179	106	130	184	88	140	187	86	149
18	29	118	154	7	131	162	8	133	161
19	251	255	255	244	243	243	243	243	242
20	180	215	198	196	202	203	200	200	200
21	134	161	145	162	161	158	160	160	160
22	86	105	88	119	115	120	122	122	121
23	52	60	51	83	83	87	85	85	85
24	27	21	14	50	48	45	52	52	52

В следующей серии измерений настройка цветопередачи проводилась на видеокамере Basler. В Табл. 3 отображены координаты цветов ячеек тестовой таблицы в исходном и обработанном изображениях, а также их эталонные значения. На основе визуальной оценки исходного изображения, представленного на рис. 11, можно прийти к выводу, что настройка цветопередачи была выполнена недостаточно качественно, и изображение получается «слишком ярким».

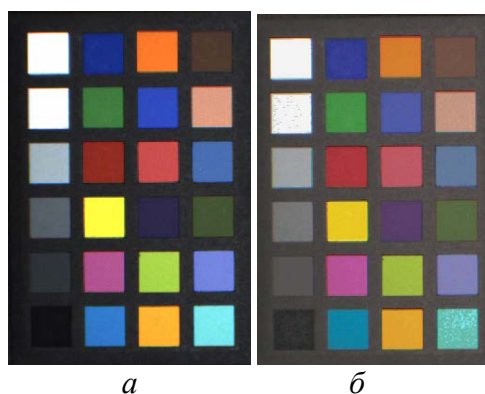


Рис. 11. Результат цветокоррекции с предварительной настройкой экспозиции на видеокамере: а – исходное; б – после настройки

Таблица 3

Результат цветокоррекции с предварительной настройкой экспозиции и точки белого на видеокамере

№	Исходные координаты цвета ячеек на изображении			Координаты цвета ячеек на изображении после настройки			Эталонные координаты цвета ячеек		
	R	G	B	R	G	B	R	G	B
1	76	59	51	116	87	74	115	82	68
2	224	156	133	192	150	131	194	150	130
3	75	110	178	95	122	155	98	122	157
4	65	86	55	87	106	66	87	108	67
5	117	132	213	132	127	174	133	128	177
6	121	234	238	106	187	172	103	189	170
7	255	124	46	214	126	44	214	126	44
8	47	73	196	78	91	172	80	91	166
9	206	76	74	192	86	94	193	90	99
10	46	40	74	91	64	106	94	60	108
11	165	220	75	155	186	69	157	188	64
12	255	184	48	223	162	45	224	163	46
13	28	48	143	54	61	146	56	61	150
14	68	128	66	64	148	71	70	148	73
15	147	43	34	169	56	63	175	54	60
16	255	255	63	230	199	32	231	199	31
17	182	84	143	186	83	146	187	86	149
18	50	120	205	8	132	160	8	133	161
19	254	254	254	243	243	242	243	243	242
20	254	254	254	243	243	242	200	200	200
21	166	183	203	158	157	163	160	160	160
22	85	96	102	114	120	120	122	122	121
23	47	48	52	86	84	81	85	85	85
24	17	16	21	53	53	54	52	52	52

Проведенная цветокоррекция с предварительной настройкой экспозиции видеокамеры не способна устранить ошибку предыдущей настройки, поэтому некоторые цвета отображаются в той же степени некорректно, как и на исходном изображении. Наиболее показательными являются цвета № 19 и № 20. На исходном изображении и на обработанном они «сливаются» в один цвет.

Итак, видно, что для эффективной настройки цветопередачи необходимо проводить предварительную настройку баланса яркости.

Далее (рис. 12) приведены изображения, полученные при фотографировании с камеры планшета iPad и результат настройки цветопередачи.

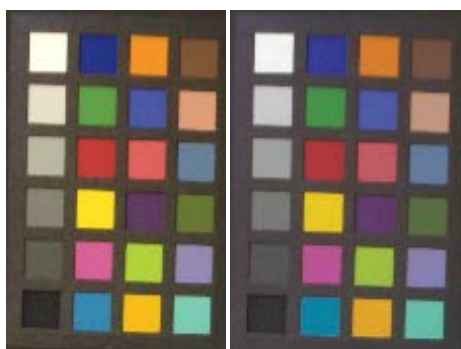


Рис. 12. Результат цветокоррекции изображения, полученного с iPad

Различие между исходным изображением и полученным не являются существенными, за исключением того, что в ходе настройки цветопередачи была устранена «желтизна», присущая исходному изображению.

Таким образом, в ходе экспериментальных исследований была проведена оценка качества работы алгоритма по настройке цветопередачи видеоинформационных систем в различных условиях освещения при использовании различных видеоинформационных ОЭС и различных методик предварительной настройки цветопередачи.

Было выявлено, что для условий недостаточной или избыточной освещенности алгоритм дает результат, который не отвечает требованиям качества, необходимым для проведения цветовых измерений. Также установлено, что для улучшения результата работы алгоритма желательно проводить предварительную настройку экспозиции.

Области возможного использования

Основная область применения результатов проекта – опико-электронные приборы и комплексы цветового анализа, например для:

- горнодобывающей промышленности – создание высокопроизводительных комплексов сортировки полезных ископаемых, селективируемых на основе цветового анализа;
- пищевой и фармацевтической отраслей промышленности – контроль качества продукции.

Заключение

В соответствии с поставленными целью и задачами были достигнуты следующие результаты:

- разработана новая теоретическая модель настройки цветопередачи ОЭС ТЗ, отличающаяся от остальных подходом к решению проблемы. Модель реализована в программной среде Matlab в виде алгоритма выделения и преобразования точек цветового пространства RGB в соответствии с положением в нем «опорных точек». Он состоит из предварительной триангуляции цветового пространства, последующего выделения точек внутри образованных сегментов (тетраэдров), расчета функций преобразования для выделенных точек и последующего создания матрицы преобразования;
- проведена оценка качества работы алгоритма по настройке цветопередачи видеоинформационных систем в различных условиях освещения при использовании различных приемников излучения, а также методик предварительной настройки цветопередачи. В результате оценки выявлены недостатки в работе алгоритма, требующие устранения.

Дальнейшим направлением работы является выявление недоработок в реализованном алгоритме, их устранение, внедрение в реализованную программу по настройке цветопередачи дополнительных функций, сводящих к минимуму работу пользователя по облегчению и ускорению процесса настройки.

Научная новизна работы заключается в создании абсолютно нового подхода к решению проблемы настройки цветопередачи видеоинформационных систем и создании новой методики настройки, базирующейся на отличных от типовых принципах.

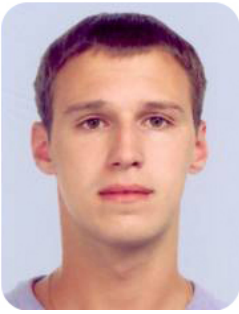
Практическая значимость данной работы заключается в возможности применения описанной методики в системах цветового анализа для повышения качества цветопередачи и соответственно эффективности работы таких систем. Они довольно широко применяются в горно-добывающей промышленности для обогащения руд твердых полезных ископаемых. Также их используют для контроля качества различной продукции.

Публикации

1. *Алехин А.А., Горбунова Е.В., Чертов А.Н., Шитов Д.Д.* Об описании объектов анализа для оптико-электронных систем цветовой идентификации // Научно-технический журнал «Приборостроение». 2012. № 12. С. 65–66.
2. *Горбунова Е.В., Шитов Д.Д.* Настройка цветопередачи оптико-электронных систем технического зрения по методу «опорных цветов» // Научно-технический вестник информационных технологий, механики и оптики. 2012. Т. 81. № 5. С. 153–154.
3. *Шитов Д.Д.* Особенности настройки цветопередачи оптико-электронных систем технического зрения по методу «опорных цветов» // Труды X Международной конференции «Прикладная оптика – 2012» (15–19 октября 2012 г., секция 7 «Передача, восприятие и измерение цвета»). СПб., 2012. С. 270–274.
4. *Шитов Д.Д., Горбунова Е.В., Чертов А.Н.* Свидетельство о государственной регистрации программы для ЭВМ № 2013615782 «Коррекция цветопередачи камеры в выделенной области цветового пространства».

Литература

1. *Алехин А.А., Горбунова Е.В., Чертов А.Н., Шитов Д.Д.* Об описании объектов анализа для оптико-электронных систем цветовой идентификации // Научно-технический журнал «Приборостроение». 2012. № 12. С. 65–66.
2. *Годен Ж.* Колориметрия при видеообработке. М., 2008. – 328 с.
3. *Домасев М.В., Гнатюк С.П.* Цвет, управление цветом, цветовые расчеты и измерения. СПб., 2009. – 224 с.
4. *Шитов Д.Д.* Особенности настройки цветопередачи оптико-электронных систем технического зрения по методу «опорных цветов» // Труды X Международной конференции «Прикладная оптика – 2012» (15–19 октября 2012 г., секция 7 «Передача, восприятие и измерение цвета»). СПб., 2012. С. 270–274.
5. *Галанин М.П., Щеглов И.А.* Разработка и реализация алгоритмов трехмерной триангуляции сложных пространственных областей: итерационные методы: http://www.keldysh.ru/papers/2006/prep09/prep2006_09.html.



Щеглов Константин Андреевич
Факультет компьютерных технологий и управления
Кафедра вычислительной техники
E-mail: Scheglov.konstantin@gmail.com

**РАЗРАБОТКА И ОБОСНОВАНИЕ ТРЕБОВАНИЙ К ПОСТРОЕНИЮ СИСТЕМЫ,
РЕАЛИЗУЮЩЕЙ ЗАЩИТУ ОТ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА
К СОЗДАВАЕМЫМ ФАЙЛОВЫМ ОБЪЕКТАМ,
ОСНОВАННОЙ НА ИСКЛЮЧЕНИИ СУЩНОСТИ «ОБЪЕКТ ДОСТУПА»
ИЗ РАЗГРАНИЧИТЕЛЬНОЙ ПОЛИТИКИ**

К.А. Щеглов

Научный руководитель – профессор А.А. Ожиганов

Цель проекта

Разработка и обоснование требований к построению системы, реализующей защиту от несанкционированного доступа к создаваемым файловым объектам, основанной на исключении сущности «объект доступа» из разграничительной политики.

Задачи проекта

Основу защиты от несанкционированного доступа к обрабатываемой на компьютере информации составляет реализация разграничительной политики доступа к ресурсам, в первую очередь, к файловым объектам. Эта политика реализуется диспетчером доступа, перехватывающим и анализирующим все запросы от субъектов к объектам. В зависимости от решаемой задачи защиты доступ к обрабатываемой на компьютере информации может разграничиваться между пользователями и/или приложениями (процессами). В зависимости от способов задания правил доступа методы контроля подразделяют на дискреционный и мандатный.

С целью реализации корректной разграничительной политики доступа к ресурсам должны быть сформулированы, обоснованы и далее реализовываться требования к построению безопасной системы, направленные на предотвращение утечки прав доступа $R = \{x, w, r, d\}$ (выполнение, чтение, запись, удаление). Они выполняются администратором безопасности при настройке средств защиты информации. Актуальность подобных работ в современных условиях обуславливается кардинальным пересмотром модели нарушителя. Например, в исследовании [1] показано (рис. 1), что наибольшую опасность для компаний сегодня представляют собственные сотрудники (инсайдеры).

Это требует пересмотра и базовых требований к построению безопасной системы, в частности из разграничительной политики доступа должен быть исключен пользователь – не должна использоваться широко реализуемая в современных ОС и приложениях сущность «владение» и т. д.

Актуальность решаемой в проекте задачи подтверждается широким практическим применением предложенных автором принципов и методов контроля доступа к создаваемым файлам.

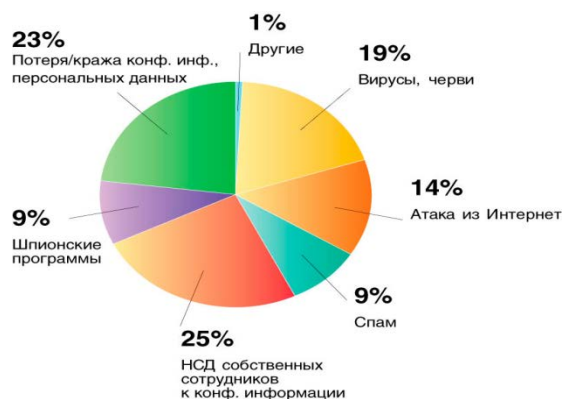


Рис. 1. Наиболее актуальные современные угрозы информационной безопасности

Первичным при назначении разграничений прав доступа в известных методах контроля доступа является объект. На первый взгляд это логично, так как именно объект, предназначенный для хранения обрабатываемой информации и должен защищаться от несанкционированного доступа. Однако файловые объекты принципиально различаются. Они могут быть подразделены на статичные (в первую очередь системные) и создаваемые пользователями в процессе работы системы. Принципиальная разница между этими группами объектов – в части задания разграничительной политики доступа – огромна и состоит в том, что системные объекты присутствуют на компьютере на момент настройки администратором прав доступа субъектов к объектам, а создаваемых еще попросту нет – они появляются позже в процессе работы пользователей на компьютере. Резонно возникает вопрос: как же к ним разграничивать доступ, если их еще нет? Ведь это именно те объекты (прежде всего файлы), которые нуждаются в защите от несанкционированного доступа, поскольку именно они и содержат обрабатываемую на компьютере защищаемую информацию.

Известными методами контроля доступа статичные и создаваемые файловые объекты никак не различаются. Поэтому задача задания прав доступа к создаваемым файлам решается опосредованно, путем включения администратором в систему папок (своего рода «контейнеров»), доступ к которым для субъектов он и разграничивает, в том числе и по созданию в них новых файлов. При этом соответствующими разграничениями пользователь «принуждается» создавать новые файловые объекты (файлы) только в тех папках (контейнерах), которые специально для этих целей созданы. Создаваемые же файловые объекты наследуют заданные администратором разграничения доступа от соответствующих включающих объектов (папок).

Как видим, файл вообще «исчезает» как сущность из разграничительной политики доступа. Используются две сущности: субъект доступа и специально созданный администратором контейнер – объект доступа (папка). Это обуславливает сложность задачи администрирования, которая многократно возрастает в случае, если в качестве субъекта доступа рассматривать не только учетную запись (пользователя), но и непосредственно процесс (приложение), без чего эффективную защиту в современных условиях не создать. Для различных процессов потребуется создание своих папок-контейнеров с последующим разграничением прав доступа к ним. Подобное решение имеет и иные принципиальные недостатки. Например, при наличии неразделяемых системой и приложениями папок не может быть реализована корректная разграничительная политика доступа, что сводит на нет мандатную схему контроля доступа.

Принципы и методы контроля доступа к создаваемым файлам

Принципы контроля доступа к создаваемым файлам

Впервые данные принципы контроля доступа опубликованы автором в работе [2] и состоят в следующем:

1. Сущность «объект» должна быть исключена из схемы реализации разграничительной политики доступа к создаваемым файловым объектам по причине ее отсутствия на момент задания прав доступа.

2. Создаваемый субъектом доступа файловый объект (файл) однозначно характеризуется (уровень конфиденциальности, критичность приложения и т. д.) субъектом, создавшим его. В качестве субъекта доступа, в зависимости от решаемой задачи защиты, следует рассматривать либо сущность «пользователь», либо «процесс», либо пару сущностей «пользователь-процесс».

3. При реализации разграничительной политики доступа (назначении правил доступа) должны использоваться две сущности: идентифицируемый субъект (учетная информация – идентификатор либо метка безопасности), создавший объект и идентифицируемый субъект, запрашивающий доступ к созданному объекту.

4. Создаваемый (модифицируемый) файловый объект (файл) при записи диспетчером доступа должен включать в себя (в зависимости от реализации в качестве атрибута либо непосредственно в «теле» файла) учетную информацию субъекта, создавшего/модифицировавшего его.

5. При запросе доступа к созданному файлу диспетчер доступа, используя изначально заданные правила (матрица доступа либо правила сравнения меток безопасности), должен анализировать учетную информацию субъекта, создавшего этот файл (является принадлежностью созданного файла), и учетную информацию субъекта, запросившего доступ к этому файлу. Далее диспетчер должен либо разрешать, либо отказывать субъекту в запрошенном доступе.

6. В зависимости от решаемой задачи защиты и реализуемого метода контроля доступа в качестве учетной информации могут выступать: имя пользователя (учетная запись), полнопутьное имя исполняемого файла, характеризующее субъект «процесс», метка безопасности.

Разработанные и апробированные методы контроля доступа к создаваемым файлам

Реализацию разработанных методов контроля доступа рассмотрим на апробированном решении комплексной системы защиты информации «Панцирь+» для ОС Microsoft Windows.

Метод дискреционного контроля доступа

При задании администратором разграничительной политики доступа субъекты доступа из интерфейса, представленного на рис. 2, наделяются тремя сущностями: исходный идентификатор пользователя, эффективный идентификатор пользователя и полнопутьное имя (может использоваться маска) исполняемого файла процесса (приложения). В этом интерфейсе задаются те субъекты, которые примут участие в разграничительной политике (те, кем будет контролироваться и разграничиваться доступ к созданным файлам).

Правила доступа задаются из интерфейса, приведенного на рис. 3 (здесь субъекты доступа отображаются присвоенными им при создании именами, см. рис. 2).

При создании контролируемым субъектом нового файла этот файл размечается средством контроля доступа (диспетчером доступа). Им наследуется учетная информация субъекта доступа – имя пользователя (учетной записи) и полнопутьное имя процесса (имя исполняемого файла процесса), создавшего этот файл. Данная информация размещается в атрибутах созданного файла. Аналогичным образом размечается и неразмеченный ранее файл при его модификации контролируемым субъектом.

При запросе доступа к любому файлу средство контроля доступа анализирует наличие, а при наличии – содержимое унаследованной файлом учетной информации создавшего его субъекта доступа. Это осуществляется посредством считывания и анализа атрибутов файла, к которому запрошен доступ. При наличии разметки анализируются заданные правила доступа, в результате чего диспетчер доступа предоставляет запрошенный субъектом доступ либо отказывает в нем, признавая тем самым его несанкционированным. Отображение разметки созданных файлов контролируемыми субъектами проиллюстрировано на рис. 4.

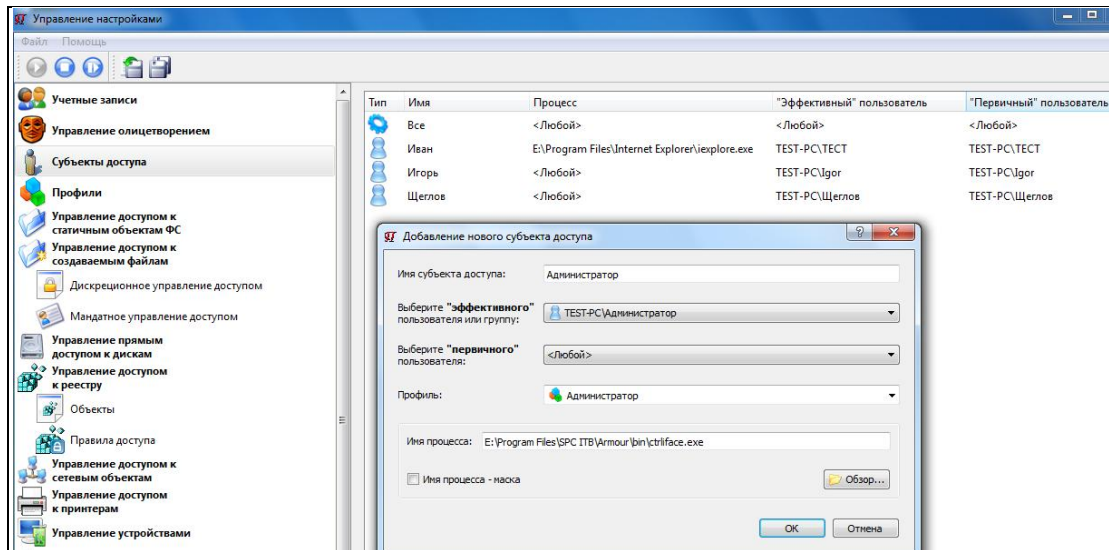


Рис. 2. Задание и отображение в интерфейсе субъектов доступа

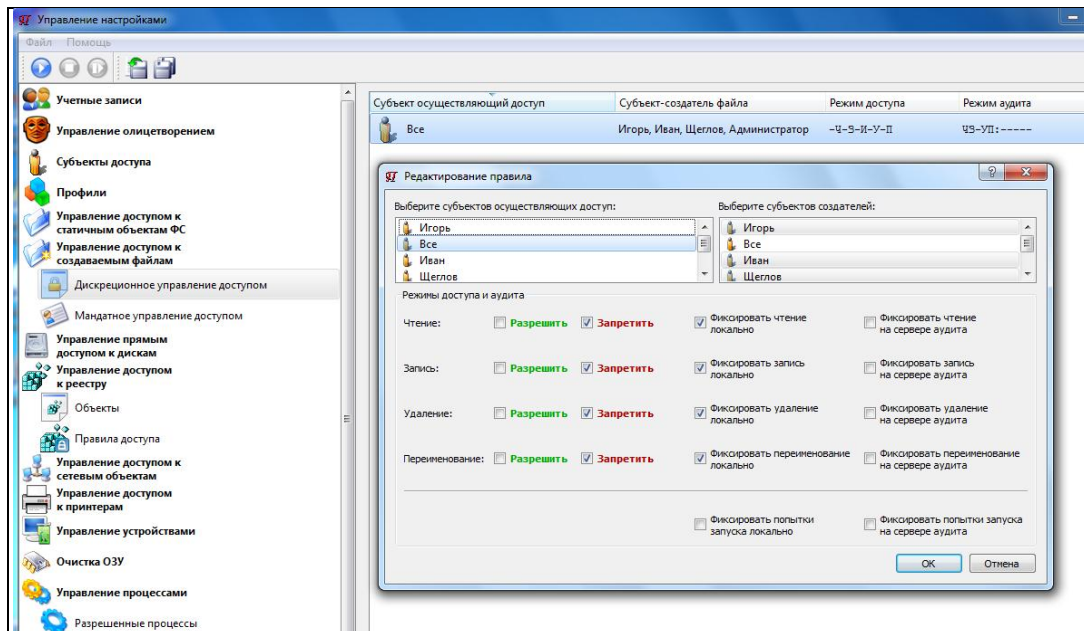


Рис. 3. Задание и отображение в интерфейсе правил доступа

Имя	Тип	Размер	Дата и время	Субъект	Процесс
111	Папка с файлами		13.08.2012 12:20:28		
1.jpg	jpg Файл	19 Кб	20.08.2012 11:58:09	TEST-PC\Administrator	E:\Program Files\SPC ITB\Armour\bin\utils\mdtexplore.exe
2.jpg	jpg Файл	12 Кб	20.08.2012 11:58:09	TEST-PC\Igor	E:\Program Files\Internet Explorer\iexplore.exe
3.jpg	jpg Файл	11 Кб	29.05.2012 16:32:53	TEST-PC\Administrator	E:\Program Files\SPC ITB\Armour\bin\utils\mdtexplore.exe
4.jpg	jpg Файл	42 Кб	29.05.2012 16:33:55	TEST-PC\Administrator	E:\Program Files\SPC ITB\Armour\bin\utils\mdtexplore.exe

Рис. 4. Отображение разметки созданных файлов при дискреционном контроле доступа

Замечание. В интерфейсе (рис. 3), отсутствует возможность разрешить выполнение создаваемых файлов. Их запуск «по умолчанию» запрещен средством защиты. На этом основывается защита от запуска вредоносных программ.

Как видим, в разграничительной политике отсутствует сущность «объект доступа». Все разграничения задаются исключительно между субъектами. Доступ к создаваемым ими файлам контролируется и разграничивается.

Метод мандатного контроля доступа

В данном случае разграничительная политика основана на назначении субъектам (пользователям) уровней доступа (меток безопасности). Она реализуется из интерфейса,

представленного на рис. 5. При сохранении файла контролируемым пользователем (с присвоенным уровнем доступа) файлом автоматически будет наследоваться уровень доступа создавшего его пользователя. Наследование будет происходить и в том случае, если контролируемый пользователь модифицирует неразмеченный ранее файл. При последующем обращении к контролируемому (размеченному) файлу доступ к нему будет контролироваться – разрешаться/запрещаться в соответствии с заданными правилами мандатного доступа, основанными на арифметическом сравнении числовых значений меток безопасности (уровней доступа). Правила доступа устанавливаются в системе «по умолчанию» – это возможность чтения/записи пользователем файла одного с ним уровня и возможность чтения файла более низкого уровня доступа (категории). Исполнение создаваемого файла запрещено.

Как видим, для реализации разграничительной политики необходимо лишь назначить уровни доступа пользователям. Не требуется каким-либо образом размечать (назначать метки безопасности) какие-либо объекты доступа. Сущность «объект доступа» исключена из разграничительной политики.

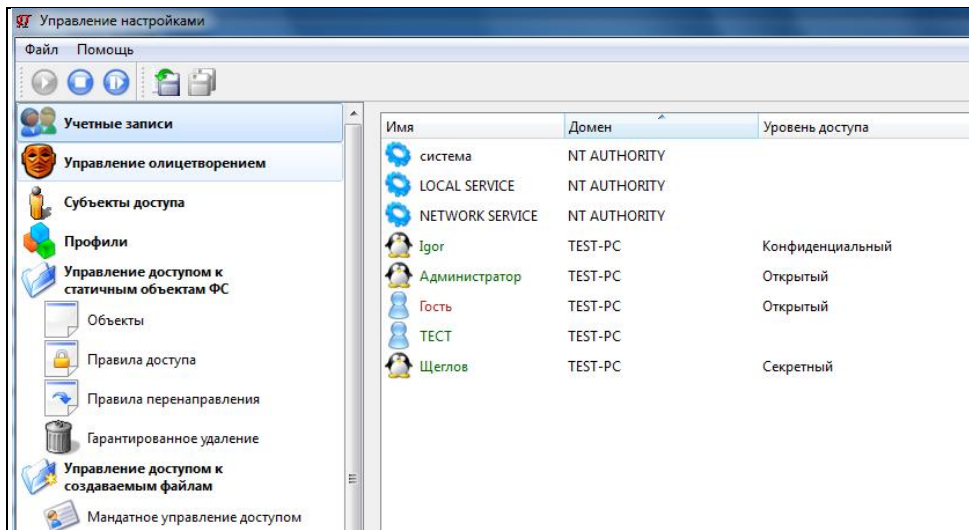


Рис. 5. Отображение заведенных в системе пользователей с заданными для них уровнями доступа

Использованные методы исследования

Методы исследования базируются на построении и исследовании следующих моделей контроля доступа к создаваемым файловым объектам.

Если считать, что множество $C = \{C1, \dots, Cl\}$ линейно упорядоченное множество субъектов доступа, а $R = \{R1, \dots, Rm\}$ конечное множество прав доступа (r – чтение, w – запись, d – удаление, x – исполнение и т. д., 0 – отсутствие прав доступа) субъекта Ci к объекту, созданному субъектом Cj ; $i=1, \dots, l, j=1, \dots, l$, то матрица доступа M , используемая для реализации разграничительной политики методом контроля доступа с принудительным управлением потоками информации (пользователь исключен из схемы администрирования) имеет следующий вид (условимся в строках матрицы указывать учетную информацию субъектов, запрашивающих доступ к объектам, а в столбцах – учетную информацию субъектов, унаследованную созданными объектами):

$$M = \begin{matrix} & C1 & C2 \dots \dots \dots Cl \\ \begin{matrix} C1 \\ C2 \\ \cdot \\ \cdot \\ \cdot \\ Cl-1 \\ Cl \end{matrix} & \begin{bmatrix} r,w,d & w & 0 \\ r & r,w,d & 0 \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ x & 0 & r \\ 0 & w & r,w,d \end{bmatrix} \end{matrix}$$

В любой момент времени система описывается своим текущим состоянием $Q = (C, C, M)$, $M[C, C]$ – ячейка матрицы, которая содержит набор прав доступа. Будем обозначать $C_i(R)C_j$ разрешением права доступа субъекту C_i к объекту, созданным субъектом C_j , $i=1, \dots, l$; $j=1, \dots, l$, где $R = \{x, w, r, d\}$: x – исполнение, w – запись, r – чтение, d – удаление.

Общее основополагающее требование к безопасности системы может быть сформулировано следующим образом: «Для заданной системы состояние $Q_0 = (C_0, C_0, M_0)$ следует считать безопасным относительно некоторого права R , если не существует применимой к Q_0 последовательности действий, в результате выполнения которых субъектом C_0 приобретает право R доступа к объекту, созданному субъектом C_0 , исходно отсутствующее в ячейке матрицы $M_0[C_0, C_0]$ ». Если же право R , отсутствующее в ячейке матрицы $M_0[C_0, C_0]$, приобретает субъектом C_0 , то следует говорить, что произошла утечка права R , а система небезопасна относительно права R .

В рамках проводимого исследования должны быть сформулированы и обоснованы требования, предотвращающие утечку права R из множества $R = \{x, w, r, d\}$, применительно к разработанным методам контроля доступа к создаваемым файловым объектам.

Основные достигнутые результаты

Разработаны и строго обоснованы требования к построению безопасной системы, основанной на реализации методов контроля доступа к создаваемым файловым объектам, обязательные для выполнения администраторами безопасности при реализации разграничительной политики доступа к объектам файловой системы.

Основополагающие требования к корректности реализации контроля доступа к файловым объектам

Поскольку сформулированные основополагающие требования применимы к реализации контроля доступа как к статичным, так и к создаваемым файловым объектам, будем здесь рассматривать контроль доступа субъектов $C = \{C_1, \dots, C_l\}$ к объектам $O = \{O_1, \dots, O_k\}$. В данном случае будем считать, что число субъектов и объектов доступа совпадает.

Модель дискреционного контроля доступа. Основополагающие требования к построению безопасной системы

Лемма 1. Пользователям запрещается иметь права исполнения x из папок-контейнеров, созданных администратором для сохранения ими файлов в процессе функционирования системы. При реализации данного требования система безопасна относительно права исполнения x .

Доказательство. Недопустимо одновременно (для одного и того же объекта – папки-контейнера) предоставлять право на запись w и исполнение x . Это обусловлено тем, что пользователь сможет записать и выполнить несанкционированную программу, что приведет к утечке права исполнения x , предполагающего при принудительном управлении потоками создание программ исключительно администратором. Лемма доказана.

Следствие. К разрешенным правам доступа к создаваемым файловым объектам относятся $R = \{w, r, d\}$.

Аксиома 1. При назначении разграничительной политики «по умолчанию» должны быть установлены права доступа: $C_i(w, r, d)O_j$, $i=1, \dots, l$, $j=1, \dots, k$. Данное правило обуславливает задание диагональной «канонической» матрицы доступа, характеризуемой условием: $C_i(w, r, d)C_j$, $i=1, \dots, l$, $j=1, \dots, k$; $C_i(0)C_j$, $j \neq i$, $i=1, \dots, l$, $j=1, \dots, k$.

Лемма 2. Система безопасна относительно права записи w и права чтения r если она реализует каноническую модель доступа.

Доказательство. В системе, реализующей каноническую матрицу доступа, невозможно генерирование иных потоков, кроме $C_i(w, r, d)C_i$, $i=1, \dots, l$, априори не вызывающих утечку права $R = \{w, r, d\}$, так как обработка информации субъектами полностью изолирована. Лемма доказана.

Аксиома 2. Контроль доступа в безопасной системе реализован корректно только в том случае, если настройками диспетчера доступа может быть реализована каноническая матрица доступа.

Следствие. Ввиду того что утечка права R может быть вызвана добавлением права R к канонической матрице доступа, далее будем формулировать требования к корректности расширения канонической матрицы доступа правом R .

Лемма 3. При расширении канонической матрицы доступа правом чтения (r): $Ci(r)Oj$, при уже разрешенном в матрице праве чтения r : $Ck(r)Oi$ одновременно с этим должно разрешаться право чтения r : $Ck(r)Oj$, где $i \neq j \neq k$, $i=1, \dots, l$; $j=1, \dots, l$; $k=1, \dots, l$. Это предотвращает возможность несанкционированного обмена информацией между субъектами $Ck[P]Cj$ из-за образующейся при этом утечки права чтения (r): $Ck(r)Oj$.

Доказательство. Обратимся к рис. 6, где проиллюстрирована утечка права чтения r , происходящая при невыполнении требования, сформулированного Леммой 3.

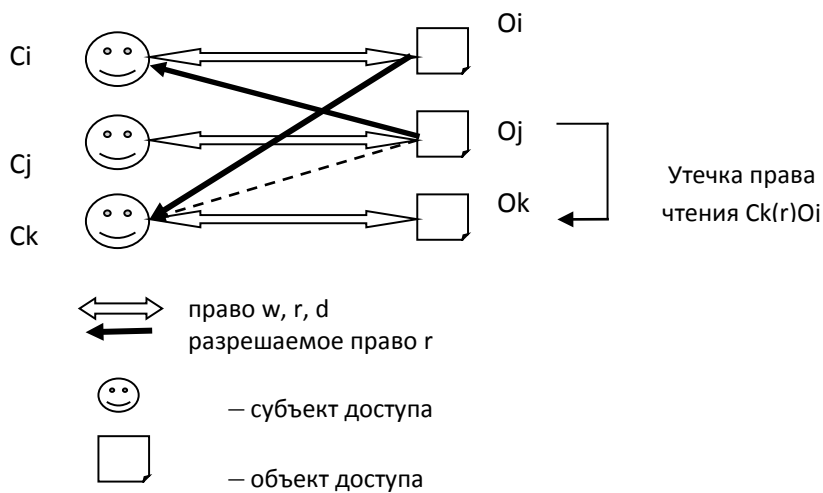


Рис. 6. Иллюстрация к Лемме 3

Из рисунка видим, что в случае разрешения права $Ci(r)Oj$ при исходно заданном праве $Ck(r)Oi$ с учетом того, что Ci априори обладает правом записи $Ci(w)Oi$, организуется информационный поток чтения от Oj в Ck путем реализации следующей последовательности действий: $Ci(r)Oj$, $Ci(w)Oi$, $Ck(r)Oi$. Если право $Ck(r)Oj$ не разрешено, то присутствует утечка права чтения $Ck(r)Oj$. Следовательно, появляется возможность несанкционированного обмена информацией между субъектами $Ck[P]Cj$. Из рис. 6 также видно, что сформулированное Леммой 3 правило распространяется на все случаи, задаваемые условием: $i \neq j \neq k$, $i=1, \dots, l$; $j=1, \dots, l$; $k=1, \dots, l$. Лемма доказана.

Лемма 4. При расширении канонической матрицы доступа правом записи w : $Ci(w)Oj$, при уже разрешенном в матрице праве записи w : $Cj(w)Ok$ одновременно должно разрешаться право записи w : $Ci(w)Ok$, где $i \neq j \neq k$, $i=1, \dots, l$; $j=1, \dots, l$; $k=1, \dots, l$, что предотвращает возможность несанкционированного обмена информацией между субъектами $Ci[P]Ck$ из-за образующейся при этом утечки права записи w : $Ci(w)Ok$.

Доказательство. Обратимся к рис. 7, на котором проиллюстрирована утечка права записи (w), происходящая при невыполнении требования, формулируемого Леммой 4.

Из рисунка следует, что в случае разрешения права $Ci(w)Oj$ при исходно заданном праве $Cj(w)Ok$ с учетом того, что Cj априори обладает правом чтения: $Cj(r)Oj$, организуется информационный поток записи от Ci в Ok путем реализации следующей последовательности действий: $Ci(w)Oj$, $Cj(r)Oj$, $Cj(w)Ok$. Если право $Ci(w)Ok$ не разрешено, то присутствует утечка права записи $Ci(w)Ok$ и, как следствие, возможность несанкционированного обмена информацией между субъектами $Ci[P]Ck$. Из рис. 7 также видно, что сформулированное

Леммой 4 правило распространяется на все случаи, задаваемые условием: $i \neq j \neq k$, $i=1, \dots, l$; $j=1, \dots, l$; $k=1, \dots, l$. Лемма доказана.

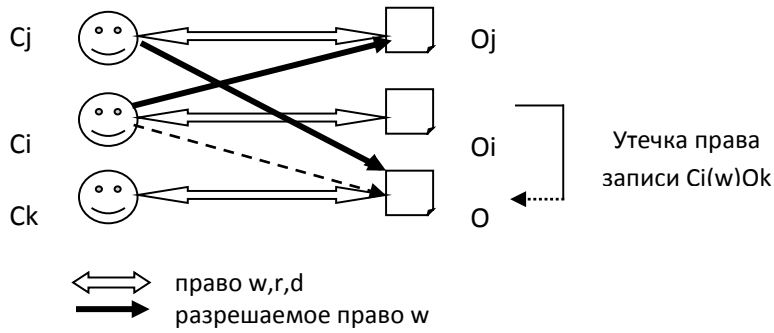


Рис. 7. Иллюстрация утечки права записи w

Относительно удаления. Данное право d тесно связано с правами записи и модификации (естественно, что в моделях нами рассматривается некий усеченный набор базовых прав доступа). Можно рассмотреть два способа записи: «запись» как право, позволяющая, кроме создания нового, модифицировать и удалять существующие файлы, и можно рассмотреть право «добавление» – право, позволяющее осуществлять запись новых файлов без возможности модификации и удаления ранее созданных. Естественно, что по своему функциональному назначению и возможностям применения это совершенно различные права доступа.

Замечание. Ранее под правом записи w мы рассматривали право модификации существующих файлов и записи новых без возможности удаления существующих (в моделях право записи и право удаления нами были разделены). Исходя из этого, сформулируем требование к безопасной системе. Здесь опять же возникает вопрос, а следует ли рассматривать модификацию как удаление? Ведь модификация не приводит к удалению объекта, но может привести к нарушению доступности к информации, хранящейся в данном объекте.

Если рассматривать право записи w как право модификации существующих файлов и записи новых без возможности удаления существующих, а модификацию существующих файлов не рассматривать как удаление, то утечки права удаления d не происходит и дополнительных требований к построению безопасной системы формулировать не требуется. Если возможность модификации файла, входящей в право записи w , рассматривать как удаление (искажение информации в определенном смысле можно приравнять к ее удалению), то дополнительно может быть сформулировано следующее требование, направленное на предотвращение утечки права удаления d : при расширении канонической матрицы доступа правом записи w : $Ci(w)Oj$ дополнительно должно включаться право удаления $Ci(d)Oj$.

Рассмотрим пример реализации правил, формулируемых Леммами 3 и 4, при построении корректной разграничительной политики доступа к файловым объектам.

Пусть исходная матрица доступа $M \delta I$ имеет следующий вид:

$$M \delta I = \begin{matrix} & O1 & O2 & O3 & O4 & O5 \\ \begin{matrix} C1 \\ C2 \\ C3 \\ C4 \\ C5 \end{matrix} & \left[\begin{array}{ccccc} r, w, d & 0 & 0 & 0 & 0 \\ 0 & r, w, d & r & 0 & 0 \\ w & 0 & r, w, d & 0 & 0 \\ 0 & 0 & 0 & r, w, d & 0 \\ 0 & 0 & 0 & 0 & r, w, d \end{array} \right] & . \end{matrix}$$

В нее требуется внести правило $C4(r, w)O3$. При этом, выполняя требования Лемм 3 и 4, получаем следующую матрицу доступа, $M \delta 2$:

$$M_{\partial 2} = \begin{matrix} & O1 & O2 & O3 & O4 & O5 \\ \begin{matrix} C1 \\ C2 \\ C3 \\ C4 \\ C5 \end{matrix} & \left[\begin{array}{ccccc} r,w,d & 0 & 0 & 0 & 0 \\ 0 & r,w,d & r & r & 0 \\ w & 0 & r,w,d & 0 & 0 \\ w & 0 & r,w & r,w,d & 0 \\ 0 & 0 & 0 & 0 & r,w,d \end{array} \right] \end{matrix}$$

Вывод. Выполнение сформулированных требований к реализации разграничительной политики доступа к ресурсам позволяет построить безопасную систему, так как предотвращает возможность утечки права $R = \{x, w, r, d\}$.

Модель мандатного контроля доступа. Основополагающие требования к построению безопасной системы

Как отмечалось выше, известные правила, используемые на практике, направленные на защиту от понижения категории обрабатываемой информации, имеют следующий вид:

1. Субъект C имеет доступ к объекту O в режиме «чтение» в случае, если выполняется условие $M_c < = M_{co}$.
2. Субъект C имеет доступ к объекту O в режиме «запись» и «удаление» в случае, если выполняется условие $M_c = M_{co}$.

Рассмотрим матрицу контроля доступа для мандатного метода. Правила сравнения меток безопасности могут быть отображены в виде матрицы доступа M_m , рассматриваемые права доступа $R = \{r, w, d\}$: w – запись, r – чтение, d – удаление:

$$M_m = \begin{matrix} & O1(M1) & O2(M2) & \dots & Ol(Ml) \\ \begin{matrix} C1(M1) \\ C2(M2) \\ \cdot \\ \cdot \\ C_{l-1}(M_{l-1}) \\ Cl(Ml) \end{matrix} & \left[\begin{array}{cccc} r,w,d & r & & r \\ 0 & r,w,d & & r \\ \dots & \dots & \dots & \dots \\ 0 & 0 & & r \\ 0 & 0 & & r,w,d \end{array} \right] \end{matrix}$$

Оценим корректность правил мандатного контроля доступа, в части возможности построения безопасной системы.

Лемма 5. Система, реализующая метод мандатного контроля доступа, безопасна относительно права записи w .

Доказательство. Право записи в мандатном контроле доступа к создаваемым файловым объектам задается следующим образом: $C_i(w)O_i, i=1, \dots, l; C_i(w=0)O_j, j \neq i, i=1, \dots, l, j=1, \dots, l$, что априори не может привести к утечке права записи w . Лемма доказана.

Лемма 6. Система, реализующая метод мандатного контроля доступа, безопасна относительно права чтения r .

Доказательство. Обратимся к Лемме 3, в которой сформулировано корректное правило управления чтением: «При расширении канонической матрицы доступа правом чтения r : $C_i(r)O_j$, при уже разрешенном в матрице праве чтения r : $C_k(r)O_i$ одновременно с этим должно разрешаться право чтения r : $C_k(r)O_j$, где $i \neq j \neq k, i=1, \dots, l; j=1, \dots, l; k=1, \dots, l$, что предотвращает возможность несанкционированного обмена информацией между субъектами $C_k[P]C_j$ из-за образующейся при этом утечки права чтения r : $C_k(r)C_j$ ». Как следует из представленной выше матрицы M_m , требования, сформулированные Леммой 3, для метода мандатного контроля доступа выполняются. Доказательство Леммы 6 сводится к доказательству Леммы 3. Лемма доказана.

Говоря о мандатном методе контроля доступа нельзя не затронуть проблемы разметки включающих объектов. Дело в том, что созданные для сохранения пользователями информации различных категорий конфиденциальности папки-контейнеры, размечаемые администратором, располагаются в некоторых включающих объектах (диск, каталог). Уж совсем особый вариант – назначение своего логического диска для хранения информации каждой категории конфиденциальности – рассматривать не будем. Однако включающие объекты также следует как-то размечать (присваивать им метки безопасности), вопрос – как?

Лемма 7. Иерархическому объекту (папке), включающему в себя размечаемые включающие объекты (папки), при использовании в системе меток безопасности $M_i, i = 1, \dots, k$ должна присваиваться метка безопасности $M_{вкл}$, исходя из следующего правила

$$M_{вкл} > \min\{M_i, i = 1, \dots, k\}, \quad (1)$$

при условии, что чем меньше значение метки, тем выше уровень конфиденциальности. В этом случае отсутствует утечка прав записи и чтения.

Доказательство. Иллюстрация корректного правила назначения метки безопасности иерархическим включающим файловым объектам представлена на рис. 8.

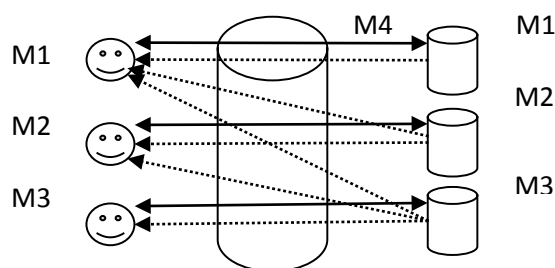


Рис. 8. Иллюстрация к Лемме 7

При реализации правила (1) назначения меток безопасности иерархическим включающим файловым объектам субъект имеет право только чтения иерархического включающего файлового объекта и не имеет возможности создавать в нем новые объекты. Он может осуществлять «обзор» размеченных папок, созданных для хранения файлов с категоризированной информацией и создавать файлы исключительно в предназначенных для этой цели включающих объектах. Таким образом, реализуется основополагающее правило контроля доступа, корректность которого в части утечек прав записи и чтения доказывается Леммами 5 и 6. Лемма доказана.

Следствие. При реализации мандатного контроля доступа, основанного на наследовании метки безопасности от включающего файлового объекта, для реализации корректного правила назначения метки безопасности иерархическим включающим файловым объектам целесообразно включить иерархическую метку «обзор включающих объектов» $M_{об}$, исходя из следующего правила: $M_{об} = M_{вкл} > \min\{M_i, i = 1, \dots, k\}$, при условии, что чем меньше значение метки, тем выше уровень конфиденциальности. Метка $M_{об}$ должна присваиваться всем иерархическим включающим файловым объектам (папкам), в которых не требуется создания файлов для хранения конфиденциальной информации.

Следствие. При условии включения в схему контроля доступа иерархической метки «обзор включающих объектов» $M_{об}$ должны быть реализованы следующие правила наследования меток безопасности между включающими и включаемыми файловыми объектами (папками):

- «главные» иерархические включающие объекты – логические диски – должны размечаться меткой $M_{об}$ в обязательном порядке;
- включаемый объект (папка) наследует метку безопасности от включающего объекта (папка) $M_{вкл} = M_{об}$ в том случае, если на включаемый объект (папка) не установлена собственная метка безопасности (которая устанавливается из условия $M_{вкл} < M_{об}$);

– если на включаемый объект (папка) установлена метка безопасности $M_{вкл}$, то включаемые в нее объекты – папки и файлы, наследуют данную метку, если на них не установлено собственных меток безопасности.

Лемма 8. Мандатный контроль доступа, основанный на наследовании метки безопасности от включающего файлового объекта, корректно реализует правило контроля удаления (рис. 9).

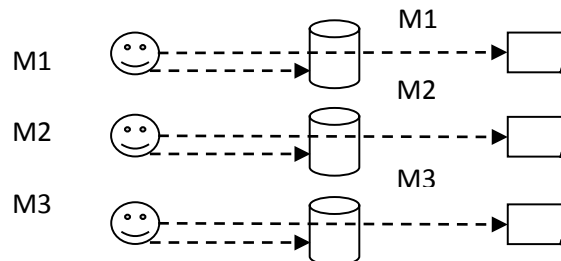


Рис. 9. Возможные правила контроля удаления для мандатного контроля доступа, основанного на наследовании метки безопасности от включающего файлового объекта

Доказательство. Любой субъект S имеет доступ к объекту O в режиме «удаление» только в случае, если выполняется условие $M_s = M_o$. Это справедливо и в части включающего объекта (папки), который размечается, и в части вложенных объектов (файлов), наследующих метку включающего объекта. Иллюстрация корректного правила удаления для случая, если иерархическому включающему файлового объекту назначить метку, исходя из условия (1), приведена на рис. 10.

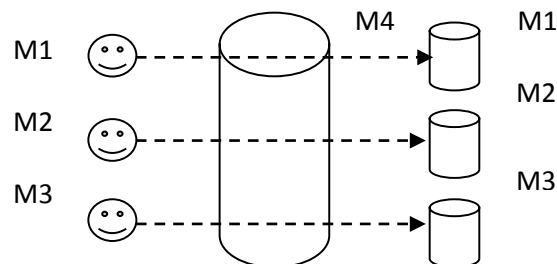


Рис. 10. Правило контроля удаления при иерархии размеченных файловых объектов

Поскольку ни один субъект не может удалить иерархический объект ввиду выполнения условия (1), правило контроля удаления в системе с иерархическими включающими объектами, представленное на рис. 10, сводится к корректному правилу удаления, проиллюстрированному на рис. 9. Лемма доказана.

Важнейшим является требование Аксиомы 2: «Контроль доступа в безопасной системе реализован корректно только в том случае, если настройками диспетчера доступа может быть реализована каноническая матрица доступа». Без его реализации о корректности контроля доступа и, следовательно, об эффективности защиты говорить не приходится. Речь идет о создаваемых файловых объектах.

На практике это требование связано с наличием файловых объектов неразделяемых системой и приложениями. В современных ОС и приложениях, как правило, присутствуют папки, не предназначенные для их разделения между субъектами доступа. Ярким примером тому служат папки временного хранения файлов «Temp». Чтобы обеспечить нормальную работу системы и таких приложений, папки, в которые помещаются временные файлы, необходимо разрешить для доступа всем субъектам вне зависимости от назначаемых им прав доступа. И это не единственный пример неразделяемых файловых объектов, называемых объектами общего доступа. Заметим, что коллективный доступ должен быть предоставлен различным пользователям и к ряду системных файлов.

Таким образом, существуют папки, для которых невозможно разграничить доступ пользователям (при этом либо пользователи, которым запрещен доступ к рассматриваемым папкам, не смогут работать с приложением, либо им не сможет предоставляться необходимый сервис). Действительно, если обозначить группу объектов файловой системы (каталогов), не разделяемых системой между пользователями, как O_n , то каноническая матрица доступа M_k примет следующий вид:

$$M_k = \begin{matrix} & O_1 & O_2 \dots O_n \dots O_l \\ \begin{matrix} C_1 \\ C_2 \\ \vdots \\ C_{l-1} \\ C_l \end{matrix} & \begin{bmatrix} r,w,d & 0 & r,w,d & 0 \\ 0 & r,w,d & r,w,d & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & r,w,d & 0 \\ 0 & 0 & r,w,d & r,w,d \end{bmatrix} \end{matrix} .$$

Рассмотрим мандатный метод контроля доступа в данных предположениях. Пусть между субъектами, которым соответственно присвоены метки M_1 и M_2 (обрабатывают информацию различных уровней конфиденциальности), какая-либо папка не может быть разделена. К этой папке требуется обеспечить право записи/чтения для обоих субъектов. Поскольку одному объекту (папке) может быть присвоена только одна метка безопасности (в нашем случае M_1 или M_2), соответственно получаем возможность реализации одного из двух правил контроля записи/чтения, проиллюстрированных на рис. 11.

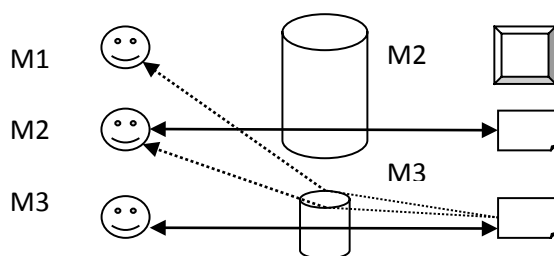
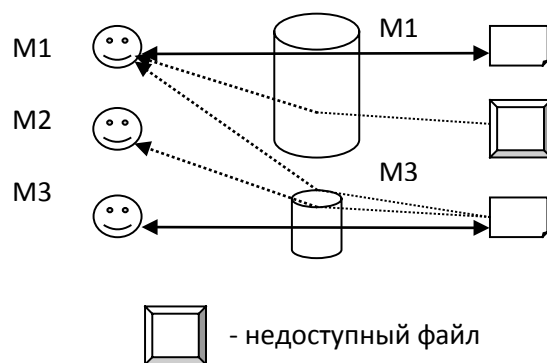


Рис. 11. Иллюстрация возможных правил контроля записи/чтения: а – неразделяемой папке присвоена метка M_1 ; б – неразделяемой папке присвоена метка M_2

Естественно, что наличие подобных файловых объектов общего (коллективного) доступа не позволяет выполнить основополагающего требования к корректности реализации контроля доступа к объектам (реализовать каноническую модель доступа), а об обработке на одном компьютере информации различных категорий конфиденциальности в этом случае говорить просто не приходится.

Вывод. В современных вычислительных системах мандатный контроль доступа не может быть корректно реализован из-за наличия в них неразделяемых ОС и приложений файловых объектов.

Требования к корректности реализации контроля доступа к создаваемым файловым объектам

Матрица доступа в данном случае приобретает совершенно иной вид, так как из разграничительной политики доступа исключена сущность «объект».

Требования к построению безопасной системы во многом совпадают с аналогичными требованиями для контроля доступа к статичным файловым объектам с поправкой на то, что в разграничительной политике доступа отсутствует сущность «объект» доступа (то есть матрица доступа имеет иной вид). Вместе с тем ввиду специфики контроля, когда не контролируется доступ на запись к статичным объектам – папкам, в которых могут создаваться файлы различными субъектами, появляются дополнительные требования.

Сначала приведем требования, разработанные к методу дискреционного контроля доступа, совпадающие (с соответствующей поправкой) с требованиями, сформулированными выше для метода дискреционного контроля доступа к статичным файловым объектам.

Лемма 9. При расширении канонической матрицы доступа разрешенным правом записи $w: C_j(w)C_i, i \neq j$, при исходно разрешенном праве записи $w: C_i(w)C_k, i \neq k$ должно быть разрешено право записи $w: C_j(w)C_k, j \neq k$, что предотвратит утечку права записи $w: C_j(w)C_k$ (рис. 12).

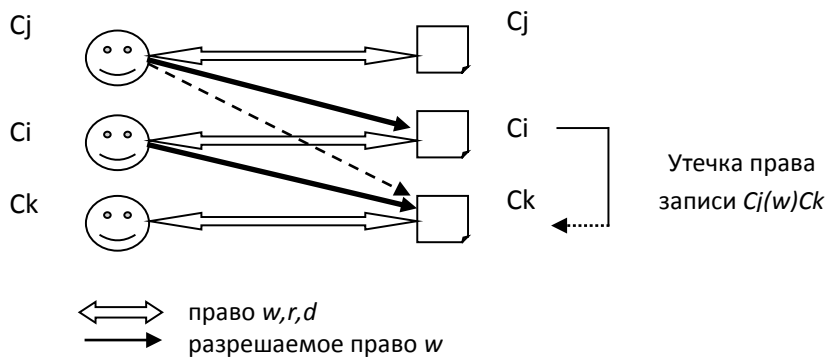


Рис. 12. Иллюстрация утечки права записи w

Доказательство. Лемма доказывается аналогично Лемме 3.

Лемма 10. При расширении канонической матрицы контроля доступа разрешенным правом чтения $r: C_j(r)C_i, i \neq j$, при исходно разрешенном праве чтения $r: C_i(r)C_k, i \neq k$ должно быть разрешено право чтения $r: C_j(r)C_k, j \neq k$, что предотвратит утечку права чтения $r: C_j(r)C_k$ (рис. 13).

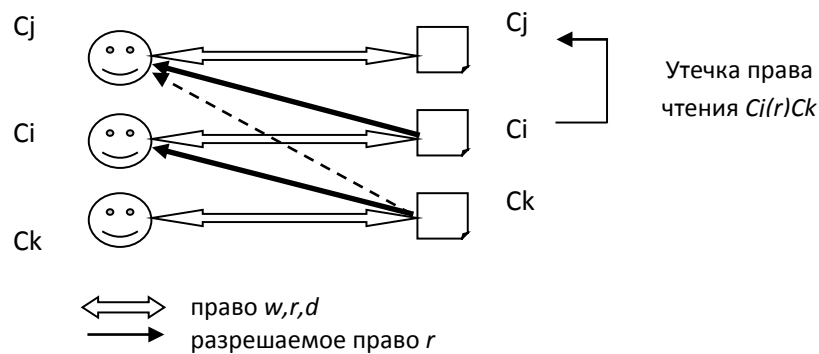


Рис.13. Иллюстрация утечки права чтения r

Доказательство. Лемма доказывается аналогично Лемме 4.

Дополнительные требования

Лемма 11. Система безопасна относительно права удаления d при реализации следующего правила управления: любой включающий объект (папка) может быть удален любым субъектом при условии отсутствия включенных в него объектов, в первую очередь файлов.

Доказательство. Метод дискреционного контроля доступа к создаваемым файловым объектам не предполагает реализации разграничительной политики доступа к статичным объектам – папкам, в которые включаются (записываются) создаваемые файлы. Следовательно, любой субъект может удалить подобную папку и все включенные в нее файлы. Это может привести к утечке права удаления d . При запрете удаления папки, при наличии включенных в нее файлов, с учетом реализации очевидного права $Ci(d)Ci$ (право $Cj(d)Ci$, $i \neq j$ не имеет практического смысла) утечка права удаления d становится невозможной. Лемма доказана.

Замечание. Правило удаления, сформулированное в Лемме, справедливо и при реализации канонической матрицы доступа.

Лемма 12. Система безопасна относительно права исполнения x при реализации следующего правила: любой создаваемый в системе файл не должен иметь право на исполнение x пользователями.

Доказательство. Разрешение пользователям права исполнения x созданного в процессе функционирования системы файла недопустимо, поскольку установка в системе программного обеспечения является исключительно функцией администратора безопасности. На этом основывается защита от вредоносных программ. Ввиду этого единственно возможным способом защиты от утечки права исполнения x является предотвращение исполнения созданных пользователями исполняемых файлов. Лемма доказана.

Вывод. Выполнение сформулированных требований к реализации разграничительной политики доступа к создаваемым файловым объектам позволяет построить безопасную систему, так как предотвращает возможность утечки права $R = \{x, w, r, d\}$.

Мандатный контроль доступа

Особенностью контроля доступа к создаваемым файлам является разметка (назначаются метки безопасности) именно файлов. Осуществляется это автоматически при их создании посредством наследования файлом метки от создавшего его субъекта доступа. Включающие элементы (папки) никоим образом не размечаются – им метка безопасности не присваивается.

Лемма 13. Мандатный контроль доступа к создаваемым файловым объектам, основанный на наследовании метки безопасности от субъекта доступа, корректно реализует основополагающее правило контроля доступа.

Доказательство. С учетом того, что включающий объект (папка) никоим образом не является сущностью, используемой при задании разграничительной политики, правило контроля доступа сводится к схеме, представленной на рис. 14. Причем наличие/отсутствие включающего объекта, в том числе и объекта, не разделяемого системой и приложениями, не сказывается на корректности разграничительной политики. В этом случае основополагающее правило мандатного метода контроля доступа при иерархической структуре файловых объектов реализуется корректно. Лемма доказана.

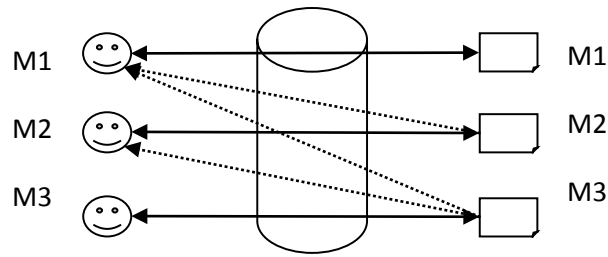


Рис. 14. Схема мандатного контроля доступа при иерархической структуре файловых объектов

Лемма 14. Мандатный контроль доступа, основанный на наследовании метки безопасности от субъекта доступа, некорректно реализует правило контроля удаления.

Доказательство. Иллюстрация возможных правил контроля удаления для мандатного контроля доступа, основанного на наследовании метки безопасности от субъекта доступа, создающего объект, представлена на рис. 15.

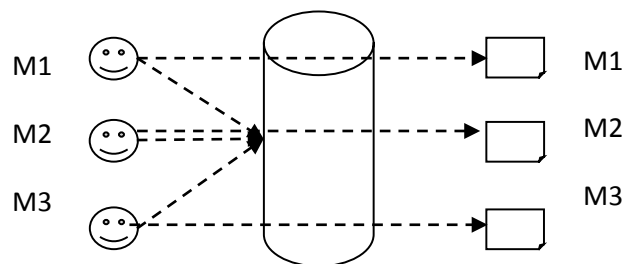
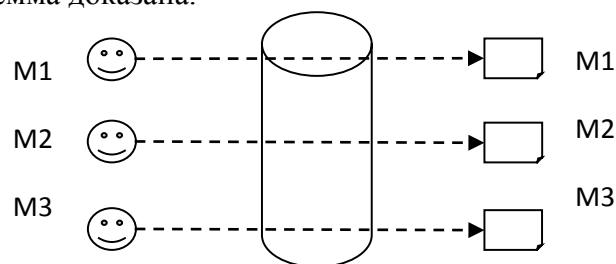


Рис. 15. Иллюстрация некорректности реализации правила контроля удаления

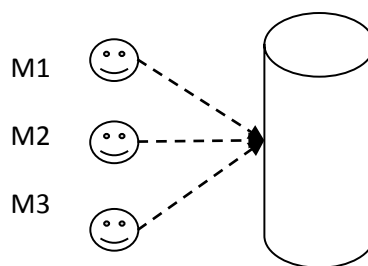
Поскольку доступ к включающим объектам никак не разграничивается, любой субъект доступа вне зависимости от его метки безопасности может удалить (и переименовать) любой включающий объект. При этом удалятся все файлы, включенные в этот объект, вне зависимости от унаследованных ими меток безопасности. Лемма доказана.

Лемма 15. Корректность реализации удаления для мандатного контроля доступа, основанного на наследовании метки безопасности от субъекта доступа, обеспечивается при реализации следующего правила удаления: включающий объект может быть удален (и переименован) любым субъектом доступа, вне зависимости от его метки безопасности, только при условии отсутствия в нем включаемых объектов.

Доказательство. При реализации данного правила удаления правила, представленные на рис. 15, сводятся к тем, что представлены на рис. 16. Правило удаления включаемых объектов, представленное на рис. 16а, не что иное, как корректное правило контроля удаления (см. рис. 15). Субъект может удалить объект только при выполнении условия $M_c = M_o$. Корректность же реализации этого правила обеспечивается включением правила удаления включающего объекта (см. рис. 16б): возможность удаления любым пользователем при отсутствии в нем включаемых объектов. Лемма доказана.



а



б

Рис. 16. Иллюстрация корректных правил удаления для мандатного контроля доступа:
 а – правило удаления включаемых объектов;
 б – правило удаления включающего объекта

Вывод. Мандатный контроль доступа к создаваемым файловым объектам при выполнении сформулированных требований к реализации разграничительной политики доступа позволяет построить безопасную систему в общем случае.

Области возможного использования результатов проекта

Использование полученных результатов определяется областью применения средств защиты информации от несанкционированного доступа, в состав которых в обязательном порядке включаются механизмы контроля доступа к файловым объектам, то есть во всех приложениях, где обрабатывается конфиденциальная и критически важная информация. Реализация разграничительной политики доступа к ресурсам, в первую очередь к файловым объектам, является основой защиты компьютерной информации.

Публикации

1. Щеглов К.А. Принципы контроля доступа к создаваемым файловым объектам // Сб. трудов молодых ученых и сотрудников кафедры ВТ. Вып. 3. СПб., 2012. С. 85–86.
2. Щеглов К.А., Щеглов А.Ю. Принцип и методы контроля доступа к создаваемым файловым объектам // Вестник компьютерных и информационных технологий. 2012. № 7. С. 43–47.
3. Щеглов К.А., Щеглов А.Ю. Принцип и метод дискреционного контроля доступа к создаваемым файловым объектам // Вопросы защиты информации. 2012. Вып. 96. № 1. С. 30–38.
4. Щеглов К.А., Щеглов А.Ю. Принцип и метод мандатного контроля доступа к создаваемым файловым объектам // Вопросы защиты информации. 2012. Вып. 96. № 1. С. 40–44.
5. Щеглов К.А., Щеглов А.Ю. Модель контроля доступа к создаваемым файловым объектам // Известия вузов. Приборостроение. 2012. Т. 55. № 10. С. 37–40.
6. Щеглов К.А., Щеглов А.Ю. Защита от вредоносных программ методом контроля доступа к создаваемым файловым объектам // Вестник компьютерных и информационных технологий. 2012. № 8. С. 46–51.
7. Щеглов К.А., Щеглов А.Ю. Методы идентификации и аутентификации пользователя при доступе к файловым объектам // Вестник компьютерных и информационных технологий. 2012. № 10. С. 47–51.
8. Щеглов К.А., Щеглов А.Ю. Контроль доступа к статичным файловым объектам // Вопросы защиты информации. 2012. Вып. 97. № 2. С. 12–20.
9. Щеглов К.А., Щеглов А.Ю. Защита от атак со стороны приложений, наделяемых вредоносными функциями. Модели контроля доступа // Вопросы защиты информации. 2012. Вып. 99. № 4. С. 31–36.
10. 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.
11. Щеглов К.А., Щеглов А.Ю. Система защиты от запуска вредоносных программ // Вестник компьютерных и информационных технологий. 2013. № 5. С. 38–43.

12. Щеглов К.А., Щеглов А.Ю. Защита от атак на уязвимости приложений. Модели контроля доступа // Вопросы защиты информации. 2013. Вып. 101. № 2. С. 36–43.

Литература

1. URL: <http://www.securitycode.ru/company/news/SC-analytic-2011//>.
2. Щеглов К.А. Принципы контроля доступа к создаваемым файловым объектам // Сб. трудов молодых ученых и сотрудников кафедры ВТ. Вып. 3. СПб., 2012. С. 85–86.

Сборник работ студентов Университета ИТМО, победителей конкурса грантов правительства Санкт-Петербурга / Под. ред. В.О. Никифоров – СПб.: Университет ИТМО, 2014 – 230 с.

СБОРНИК ТРУДОВ СТУДЕНТОВ УНИВЕРСИТЕТА ИТМО,
ПОБЕДИТЕЛЕЙ КОНКУРСА ГРАНТОВ
ПРАВИТЕЛЬСТВА САНКТ-ПЕТЕРБУРГА

Под редакцией проф. В.О. Никифорова

Дизайн обложки А.А. Кушнаренко

Редакторы:

А.В. Бутина, И.А. Лапшина, А.В. Рябчикова

Лицензия ИД № 00408 от 5.11.2013

Подписано в печать 30.05.2014

Заказ 3130 Тираж 100

ISBN 978-5-7577-0469-2