

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК

Выпуск 23

**ВЫСОКИЕ ТЕХНОЛОГИИ
В ОПТИЧЕСКИХ
И ИНФОРМАЦИОННЫХ СИСТЕМАХ**



**САНКТ-ПЕТЕРБУРГ
2005**

Научно-технический вестник СПбГУ ИТМО. Выпуск 23. Высокие технологии в оптических и информационных системах. / Главный редактор д.т.н., проф. В.Н. Васильев. – СПб: СПбГУ ИТМО, 2005. 175 с.

В выпуск включены статьи, освещающие современный уровень развития высоких технологий построения оптических и интеллектуальных информационных систем.

Издание выпуска поддержали:

- тематический план-задание Федерального агентства по образованию;
- ведомственная целевая программа «Развитие научного потенциала высшей школы» Федерального агентства по образованию;
- Федеральная целевая научно-техническая программа Федерального агентства по науке и инновациям «Исследования и разработки по приоритетным направлениям развития науки и техники» на 2002–2006 гг.;
- гранты Российского фонда фундаментальных исследований;
- гранты Санкт-Петербурга в сфере научной и научно-технической деятельности.

ISBN 5-7577-0194-3

© Санкт-Петербургский государственный университет
информационных технологий, механики и оптики,
2005

**РАЗВИТИЕ ФИЗИКО-ХИМИЧЕСКОЙ КОНЦЕПЦИИ
НАНОТЕХНОЛОГИИ ОРИЕНТАЦИИ ЖИДКИХ КРИСТАЛЛОВ**

Е.А. Коншина

Введение

Актуальность работы связана с интенсивным развитием оптических информационных технологий, в которых широко используются сегодня жидкие кристаллы (ЖК). Одним из важных технологических процессов в изготовлении всевозможных ЖК устройств, включая дисплеи, является технология создания однородной ориентации молекулярных осей ЖК по всему образцу. Жидкие кристаллы очень чувствительны к анизотропии поверхностных свойств ориентирующего слоя. Метод натирания полиимидных ориентирующих слоев для создания направления ориентации молекул ЖК, широко используемый в производстве ЖК-дисплеев, в том числе и большого размера, имеет ряд недостатков. Этот метод включает в себя нескольких стадий: нанесение слоев из жидких растворов, последующую их термическую обработку, а также стадию натирания поверхности для создания анизотропии. Генерация пыли и статического электричества в результате натирания полимера, которое в производственном процессе проводится на специальных машинах, является существенным недостатком этой технологии. Поэтому развитие альтернативных технологий получения ориентирующих слоев и создания анизотропии поверхностных свойств таких слоев представляет научный и промышленный интерес. Одним из известных альтернативных натиранию направлений генерации поверхностной анизотропии является фотохимический способ [1].

Другим направлением, в котором может быть получен ориентирующий слой и генерация угла подвеса директора, является технология химического осаждения паров (CVD-technology) с помощью плазмы [2, 3]. CVD-процесс получения ориентирующих слоев имеет ряд преимуществ по сравнению с традиционно используемыми технологиями. Этот метод основан на деструкции паров углеводородов в плазме тлеющего разряда и осаждении продуктов деструкции на подложки. Он не требует сложного вакуумного технологического оборудования и допускает его масштабирование. Широкий выбор относительно дешевого жидкого и газообразного углеводородного сырья, используемого для получения ориентирующих слоев, открывает большие возможности химической и физической модификации поверхностных свойств ориентирующих слоев в одном процессе. Ориентирующие слои, полученные CVD-технологией, представляют собой тонкие пленки аморфного гидrogenизированного углерода (a-C:H). В зависимости от условий получения в плазме они могут иметь полимероподобную (polymer-like carbon) или алмазоподобную (diamond-like carbon) структуру и свойства. В последние годы исследование свойств подобных ориентирующих слоев привлекает внимание наших корейских коллег [4–8]. Однако приоритет в развитии этого направления получения и исследования физико-химических свойств ориентирующих слоев на основе a-C:H принадлежит российским ученым [2,3, 9–19].

**Способ получения ориентации нематических
жидких кристаллов с помощью CVD-технологии**

Для получения ориентирующих слоев, обеспечивающих гомогенную или планарную (направление директора параллельно поверхности) ориентацию нематических

жидких кристаллов (НЖК), был предложен оригинальный способ, основанный на CVD-технологии [9]. Способ получения ориентирующих а-С:Н слоев включает в себя создание углеводородной плазмы тлеющего разряда и осаждение образующихся в плазме углеводородных радикалов на стеклянные подложки с прозрачным проводящим электродом. Слои осаждают на подложки, расположенные между анодом и катодом и находящиеся под отрицательным потенциалом, при углах наклона относительно оси системы в интервале $5-30^\circ$ и температуре окружающей среды. Для получения ориентирующих слоев могут быть использованы любые газообразные и жидкие углеводороды, в том числе толуол, бензол, ацетилен и октан [2, 9, 10, 13]. Принципиальная схема устройства для получения ориентирующих слоев показана на рис 1.

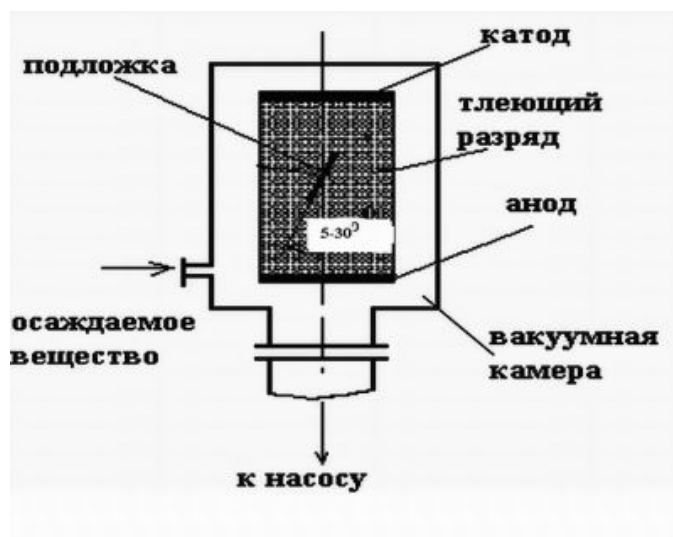


Рис. 1. Принципиальная схема устройства для осаждения ориентирующих слоев с помощью CVD-технологии

Существенным преимуществом ориентирующих слоев, полученных CVD-технологией, по сравнению с близкими им по химической природе полимерными слоями, является их химическая стойкость к воде и органическим растворителям, а также термическая стойкость. Слои а-С:Н, полученные из паров толуола при давлении 0.03–0.09 Па и подводимой мощности 1.6–2.0 Вт, были прозрачны в видимой области и имели показатель преломления 1,55–1,63 и коэффициент экстинкции менее 0, 03 на длине волны 632,8 нм [10].

Способ позволяет получать а-С:Н слои, ориентирующие НЖК гомеотропно (направление директора перпендикулярно поверхности). Для этого продукты деструкции углеводородов осаждают на изолированную подложку, что создает условия для процесса полимеризации в газовой фазе. Подложки заряжаются до потенциала анода, что исключает воздействие ионов в процессе формирования слоев. Полученные в этих условиях слои обеспечивает стабильную гомеотропную ориентацию ЖК [10, 13, 16]

Для измерения угла наклона директора молекул ЖК (Θ) нами был использован оптический метод, основанный на определении характеристического угла Ψ_x , соответствующего положению локального экстремума в разности оптического пути для двух нормальных оптических мод. Угол наклона директора определяли путем записи пропускания света, падающего при различных углах поворота ячейки Ψ в интервале от $+30^\circ$ до -30° . В качестве источника света был использован He-Ne лазер с длиной волны 632.8 нм. ЖК ячейку помещали между скрещенными поляризаторами таким образом, чтобы ее оптическая ось лежала в плоскости вращения и составляла угол 45° с плоскостью поляризации падающего света. Угол Ψ определяли как среднее значение несколь-

ких повторных измерений ($\Delta\Psi=0.5^0$), полученных обработкой кривых пропускания с помощью компьютера. Зависимость Θ от Ψ , приведенная на рис. 2, рассчитывали по формуле

$$\sin \Theta = \frac{1}{2} (n_o + n_e) \sin \Psi, \quad (1)$$

где n_o и n_e – показатели преломления обыкновенного и необыкновенного лучей для ЖК.

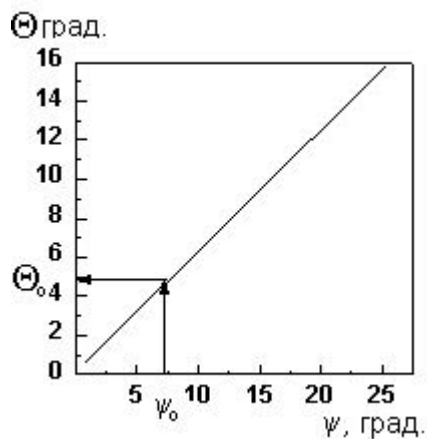


Рис. 2. Зависимость угла наклона директора ЖК Θ от характеристического угла Ψ для планарной ориентации ЖК смеси №1282 ($n_o=1.504$, $n_e=1.672$, $T=20^0$)

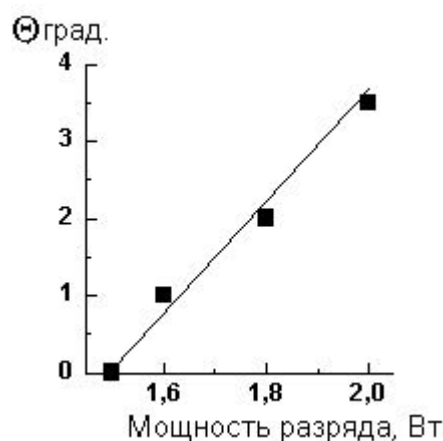


Рис. 3. Зависимость угла наклона директора Θ молекул ЦБ ЖК от мощности тлеющего разряда на постоянном токе для а-С:Н слоев, полученных из паров толуола

Особенностью а-С:Н слоев является их способность планарно ориентировать ЖК, обеспечивая близкие к нулю угол наклона директора. Это приводит к визуализации различного рода дефектов поверхности и влияет на качество ориентации. Увеличение мощности разряда не значительно повышает угол до $3,5^0$ (рис. 3). Получению однородной планарной ориентации нематических и смектических ЖК способствует создание небольшого угла наклона директора молекул относительно подложки в отсутствие внешнего поля, приложенного к электродам. Искусственное создание регулярной текстуры поверхности электродного слоя на основе окислов индия и олова ($In_2O_3+SnO_2$) перед осаждением а-С:Н путем ее натирания в одном направлении привело к увеличению угла наклона директора до $7-14^0$ и одновременно позволило улучшить однородность азимутальной ориентацией молекул [13].

Существенные изменения угла наклона директора и в целом характера ориентации молекул ЦБ были получены в результате обработки поверхности тонких слоев

а-С:Н после их получения с помощью плазмы кислорода. С увеличением времени экспозиции при постоянной мощности тлеющего разряда (4Вт мин) угол наклона директора скачкообразно изменялся от 2° до 90° и при дальнейшем воздействии оставался равным 90° [2]. Приготовленные таким образом слои вызвали гомеотропную ориентацию смеси на основе молекул цианобифенила (ЦБ). Это свидетельствует о кардинальном изменении механизма взаимодействия на границе ЦБ–а-С:Н, причиной которого является модификация физико-химического состояния ориентирующей поверхности в кислородной плазме.

Исследование микроструктуры поверхности ориентирующих тонких слоев

Микрорельеф поверхности тонких ориентирующих слоев а-С:Н, обеспечивающих планарную ориентацию ЖК, исследовался с помощью электронной микроскопии [11]. Эти исследования показали, что а-С:Н слои повторяют рельеф поверхности подложки. Наклонное расположение подложек не приводит к явно выраженной анизотропии рельефа поверхности а-С:Н слоев (рис. 4а), в отличие от слоев СеО, полученных косым напылением (рис. 4б). Эти исследования показали, что в отличие от слоев СеО, в основе ориентации которых лежит анизотропно-упругая деформация молекул ЖК, планарная ориентация молекул ЖК с помощью слоев а-С:Н происходит в результате физико-химического взаимодействия.

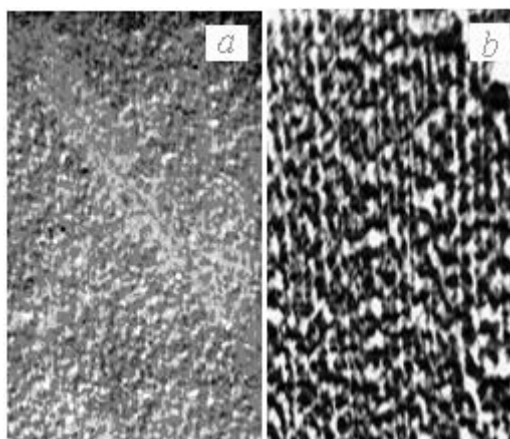


Рис. 4. Микрофотографии поверхности косонапыленного слоя СеО (а) и слоя а-С:Н (b) (20000^x)

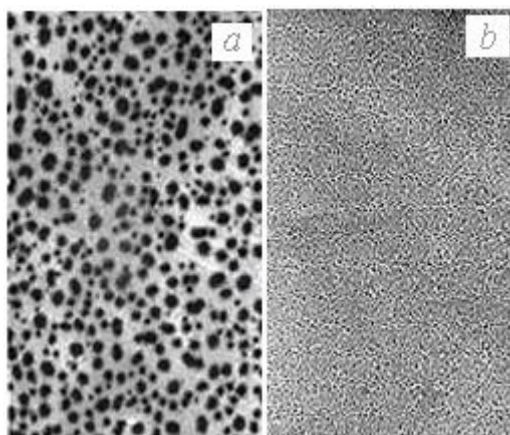


Рис. 5. Микрофотографии поверхности а-С:Н, декорированной островковой пленкой Ag толщиной 50 Å (50000^x).

Как видно на рис. 5, поверхность а-С:Н слоев имеет электрический рельеф, который хорошо визуализируется методом декорирования [12]. Электрически активными центрами на поверхности а-С:Н являются отдельные π -связанные кластеры и их сегрегации размером от 2 нм до 40 нм. Структуру поверхности а-С:Н можно представить как «архипелаг островков», состоящий из отдельных π -кластеров и их сегрегаций, в которых π -электроны делокализованы. Между ними находится электрически неактивное пространство (рис. 5a). Промежуточные области между частицами серебра на поверхности а-С:Н соответствуют структурам, содержащим СН, СН₂ и СН₃ группы в sp^3 состоянии, которые способствуют локализации отдельных π -кластеров. В тонких слоях а-С:Н (рис. 5a) расстояние между отдельными частицами серебра изменялось от 3 до 30 нм. При увеличении толщины а-С:Н слоя до ~120 нм расстояние между декорирующими частицами уменьшилось до 2–6 нм (рис. 5b), что можно объяснить влиянием объемных свойств на электрический рельеф поверхности.

Результаты микроскопических исследований поверхности а-С:Н слоев свидетельствуют о наличии электрически активной наноструктуры поверхности ориентирующих пленок. Отсутствие явно выраженной анизотропии слоев при наличии ориентирующих свойств позволяют сделать предположение о действии сил дисперсионной природы на границе раздела фаз ЖК – а-С:Н, связанных с присутствием π -связанных углеродных кластеров. Следует заметить, что не было обнаружено никаких существенных различий между ориентирующими а-С:Н слоями, полученными при наклонном и нормальном расположении подложек в плазме.

Исследование ориентации НЖК на поверхности а-С:Н слоев методом многократного нарушенного полного внутреннего отражения

Метод ИК спектроскопии полного внутреннего отражения (ПВО) дает не только информацию о химическом составе и структуре поверхности твердого тела, но и об адсорбированных на ней молекулах. Поэтому этот метод интересен не только для исследования особенности структуры ориентирующей поверхности, но и структуры ориентированного тонкого слоя ЖК на границе с ней. В случае ПВО электромагнитное поле, взаимодействующее с поглощающей средой, существует у отражающей поверхности в трех измерениях, а измерение потерь отражения дает информацию об анизотропии среды. Для исследования планарной и гомеотропной ориентации НЖК на основе ЦБ с помощью ориентирующих слоев а-С:Н, полученных CVD-технологией, мы использовали метод многократного нарушенного ПВО [13–16].

Для измерений была собрана ЖК ячейка типа сэндвич толщиной ~10 мкм, состоящая из германиевой призмы (элемент МНПВО) и плоской стеклянной пластины (рис. 6a). Призма давала 12 отражений при угле падения 45°. В ориентированной тонкой пленке НЖК преимущественное направление осей молекул для планарной ориентации параллельно оси Y, а для гомеотропной ориентации параллельно оси Z (рис. 6b).

Механизмы межмолекулярного взаимодействия на границе раздела фаз «ориентирующая поверхность – ЦБ» в рассмотренных случаях обусловлены особенностями молекулярной структуры этих слоев. Планарная ориентация ЦБ на поверхности слоя а-С:Н связана с взаимодействием бифенильных колец молекул с π -связанными кластерами на поверхности ориентирующего слоя, что заставляет директор ЖК ориентироваться параллельно оси Y. Природа сил, действующих в этом случае, имеет дисперсионный характер и связана с коллективными взаимодействиями π -электронов бифенильных колец и π -кластеров на поверхности. Исследования структуры а-С:Н слоев с помощью ИК спектроскопии МНПВО [14, 15] позволяют предположить, что гомеотропная ориентация обусловлена взаимодействием концевых алкильных групп молекул

ЦБ с CH_3 концевыми группами, входящими в структуру ориентирующего слоя (рис. 6b).

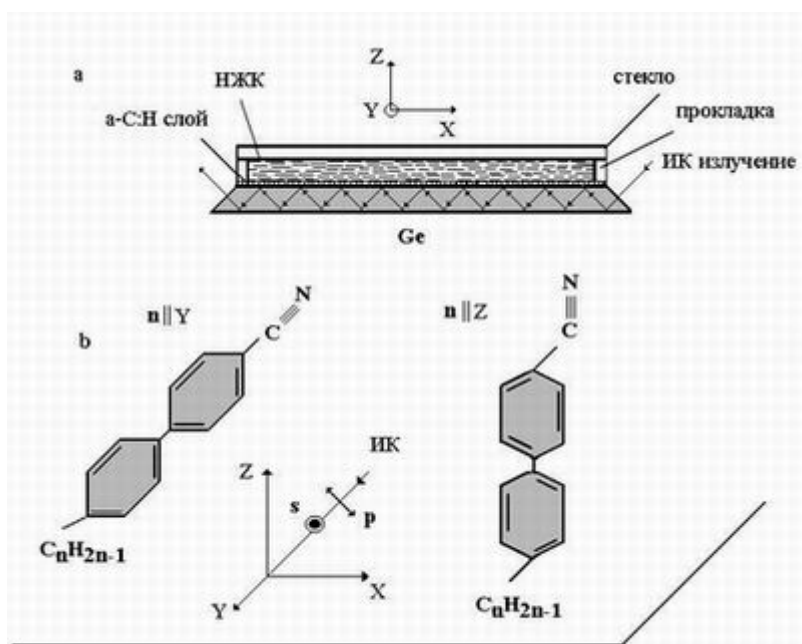


Рис 6. а) ЖК-ячейка на базе элемента МНПВО для измерения ИК спектров; б) расположение бифенильного ядра молекул ЦБ для случая планарной ($n \parallel Y$) и гомеотропной ($n \parallel Z$) ориентации

На рис. 7 приведены ИК спектры для s - (R_s) и p -поляризации (R_p) в случае планарной (а) и гомеотропной (б) ориентации ЦБ ЖК на поверхности а-С:Н слоев, записанные в интервале частот $4000\text{-}900\text{см}^{-1}$ на спектрофотометре РЕ-621 с приставкой МНПВО.

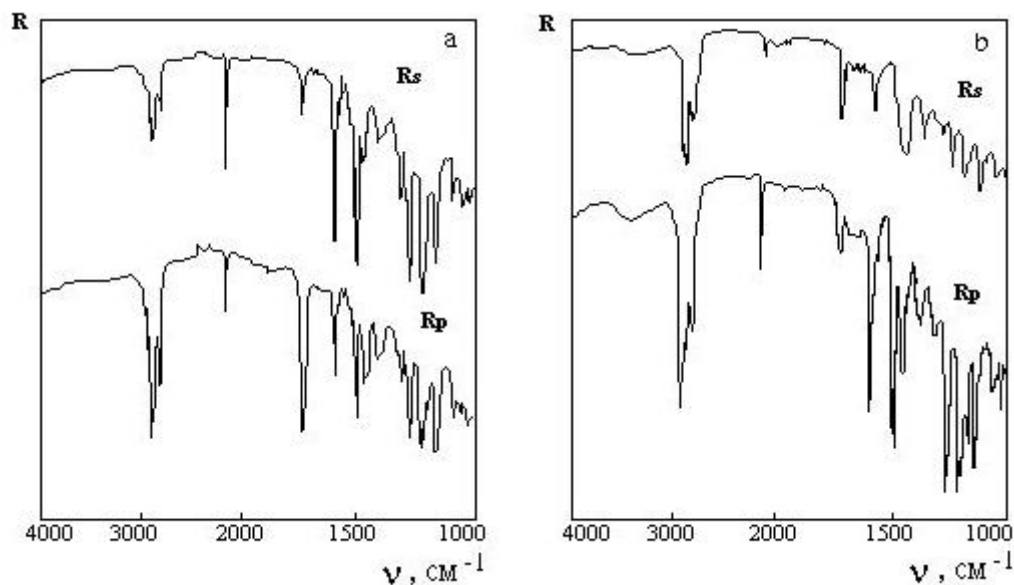


Рис. 7. Поляризационные ИК спектры МНПВО тонких ЦБ пленок, ориентированных планарно (а) и гомеотропно (б) с помощью слоев, полученных CVD технологией

В спектроскопии МНПВО величина измеренного отражения равна

$$R = (1 - \alpha d_e) N, \quad (2)$$

где N – число отражений, α – коэффициент поглощения, d_e – эффективная толщина. Дихроичное отношение ИК полос поглощения молекул ЦБ рассчитывали как отношение коэффициента поглощения ИК излучения для s -поляризации (α_s) к коэффициенту поглощения для p -поляризации (α_p). При угле падения ИК излучения $\beta=45^\circ$ глубина проникновения излучения $d_{ep} = 2d_{es}$, а дихроичное отношение равно

$$\alpha_s/\alpha_p = \frac{2(1 - \sqrt[N]{R_s})}{(1 - \sqrt[N]{R_p})} \quad (3)$$

Глубину проникновения ИК излучения рассчитывали по формуле

$$d_p = \frac{\lambda_1}{2\pi(\sin^2 \beta - n_{21}^2)^{1/2}}, \quad (4)$$

где длина волны $\lambda_1 = \lambda/n_1$ и показатель преломления $n_{21} = n_2/n_1$. В нашем случае $n_1 = 4.0$ (показатель преломления Ge), а $n_2 = 1.6$ (показатель преломления ЖК смеси ЦБ). Параметр ориентационного порядка S определяли по формуле

$$S = \frac{|D_p - 2D_s|}{D_p + 2D_s}, \quad (5)$$

где D_p и D_s – оптические плотности в полосах поглощения спектра МНПВО для p - и s -поляризации ИК излучения. В идеальном монокристалле $D_s = 0$ и $S = 1$. Для планарно-ориентированных молекул ЦБ на поверхности а-С:Н величина S соответствовала 0.63 ± 0.68 . В случае гомеотропной ориентации пленок ЦБ на поверхности слоя а-С:Н $S = 0.43 - 0.45$.

Наибольшая величина дихроичного отношения наблюдается для полос, которые колеблются вдоль направления электрического вектора падающего ИК излучения. Дихроичные отношения полос поглощения и степень ориентационного порядка в тонких пленках НЖК были рассчитаны с помощью спектров, приведенных на рис. 7 для случая параллельной ориентации ЦБ на поверхности слоя а-С:Н. Наибольшие значения дихроичного отношения, равного 4.3–5.3, наблюдали для полос поглощения, соответствующих колебаниям $C \equiv N$ связи (2220 см^{-1}), а также $C-C$ связей в фенильном (1610 см^{-1} и 1500 см^{-1}) и бифенильных (1240 см^{-1}) кольцах. Направления колебаний этих молекулярных связей параллельны директору НЖК. При гомеотропной ориентации ЦБ на поверхности слоя отношение для тех же самых полос $\alpha_s/\alpha_p < 1$ [15].

Энергия межфазного взаимодействия на границе твердое тело - жидкий кристалл

Для исследования взаимодействия на границе «ЖК – твердое тело» были сделаны оценки поверхностной энергии слоев а-С:Н по результатам измерений краевых углов смачивания θ [13,17–19]. При расчете поверхностной энергии был использован подход, согласно которому полярные компоненты поверхностной энергии могут быть представлены электронно-донорными и электронно-акцепторными центрами одной фазы. Количественно это выражается в виде значений γ^- и γ^+ в единицах поверхностной энергии. Эти взаимодействия относят к короткодействующим силам, которые можно описать уравнением

$$\gamma_{SR} = 2(\gamma^+ + \gamma^-)1/2. \quad (6)$$

Вещества, не имеющие γ^+ и γ^- , т.е. не содержащие полярные группы, называют неполярными, и при их взаимном контакте $\gamma_{SR} = 0$. Неполярная компонента включает взаимодействия, охватывающие дисперсионные, ориентационные и индукционные си-

лы согласно теории Лифшица-Ван-дер-Ваальса (γ_{LW}). Основное уравнение, связывающие полярные и неполярные взаимодействия со значением краевого угла θ на тройственной границе твердое тело–жидкость–газ, имеет следующий вид:

$$(\cos \theta + 1)\gamma_L = 2(\gamma_S^{LW} \gamma_L^{LW})^{1/2} + 2(\gamma_S^+ \gamma_L^-)^{1/2} + 2(\gamma_S^- \gamma_L^+)^{1/2}, \quad (7)$$

индексы L и S обозначают жидкую и твердую фазу, соответственно. Суммарная поверхностная энергия для любой фазы равна

$$\gamma = \gamma_{LW} + \gamma_{SR}. \quad (8)$$

Неизвестные компоненты поверхностной энергии твердого тела γ^{LW} , γ^+ и γ^- определяли путем расчета $\theta_{расч}$ в заданных границах $\gamma^{LW}=20-50$ мДж/м², $\gamma^+=0-30$ мДж/м² и $\gamma^-=0-40$ мДж/м² для каждой жидкости. Далее $\theta_{расч}$ сравнивали с $\theta_{изм}$ и выбирали минимальное значение функции качества F :

$$F = 1/M \sum [\theta_{расч}^i - \theta_{изм}^i]^2, \quad (9)$$

где i – номер жидкости для угла $\theta_{изм}^i$, M – общее число применяемых жидкостей. F является критерием точности экспериментального определения $\theta_{изм}^i$ и правильности расчета компонент поверхностной энергии.

Кроме планарно (а-С:Н) и гомеотропно (РРО) ориентирующих слоев нами были исследованы также параметры поверхностной энергии слоя поливинилового спирта (РVА), который наносили на поверхность полированных стеклянных подложек методом центрифугирования раствора. Следует отметить, что наименьшие краевые углы для ЦБ были получены на слоях а-С:Н, что свидетельствует о более сильном сцеплении на границе раздела фаз в этом случае [19]. Величина энергии межфазного взаимодействия ЦБ возрастет с увеличением поверхностной энергии твердого тела γ_S , как показано на рис. 8.

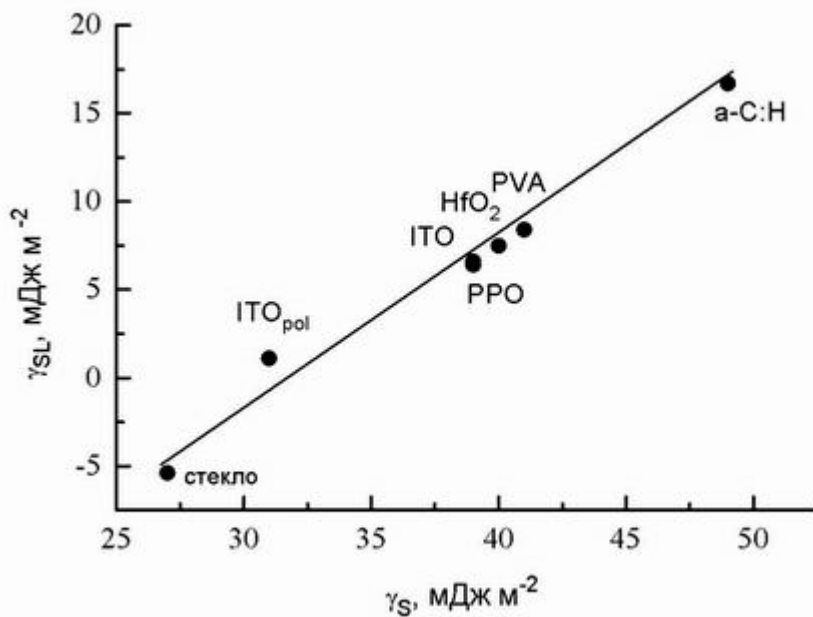


Рис. 8. Зависимость энергии межфазного взаимодействия на границе «ЖК – твердое тело» γ_{SLC} от поверхностной энергии γ_S для стекла, пленок $\text{I n}_2\text{O}_3\text{-SnO}_2$ (ITO), HfO_2 и ориентирующих слоев PVA, а-С:Н и PPO

Анализируя результаты, приведенные на рис. 8, и характер ориентационного взаимодействия исследуемых поверхностей с ЦБ ЖК [13], можно отметить следующие особенности. Чистая поверхность стеклянной подложки имеет наименьшее значения γ_{SLC} .

Хаотическая ориентация молекул ЖК в случае $\gamma_S > \gamma_{LC}$ свидетельствует о слабом взаимодействии на границе ЦБ–стекло. Осаждение на поверхность стекла прозрачного проводящего электрода ИТО привело к росту γ_{SLC} до значения 6.6 мДж/м². В этом случае повышение γ_{SLC} по сравнению с полированной стеклянной подложкой может быть вызвано действием сил упругой деформации между ЖК и шероховатой поверхностью этого слоя. Полировка поверхности ИТО батистом вызывала существенное понижение энергии γ_{SLC} до величины 1.1 мДж/м². Нанесение на поверхность ИТО дополнительного диэлектрического слоя окисла HfO₂, имеющего, как и электродный слой, поликристаллическую структуру, приводило к незначительному росту величины γ_{SLC} до 7.5 мДж/м².

Значения энергии межфазного взаимодействия ЦБ с поверхностью ориентирующих слоев PPO, PVA, и а-С:Н, имеющих органическую природу, были равны 6.4, 8.4 и 16.7 мДж/м², соответственно. Наименьшее значение γ_{SLC} соответствует PPO слою, ориентирующему молекулы ЦБ гомеотропно. Наибольшая величина энергии взаимодействия соответствует слою а-С:Н, который ориентирует молекулы ЦБ планарно с углом наклона близким к нулю градусов. Слой PVA параллельно ориентирующий молекулы ЦБ с углом наклона $\sim 10^0$ имеет промежуточное значение γ_{SLC} . Оценка величины энергии межфазного взаимодействия дает количественную характеристику физико-химического взаимодействия на границе раздела фаз.

Заключение

Полученные результаты исследования ориентации ЖК с использованием CVD технологии являются научным заданием для проведения фундаментальных и прикладных исследований в области развития основ нанотехнологии ориентации нематических и смектических жидких кристаллов для различных ЖК устройств, включая дисплеи. В качестве основных задач таких исследований предполагается:

- проведение оригинальных комплексных экспериментальных и теоретических исследований в области физики конденсированных сред с использованием новейших оптических методов;
- изучения физико-химического нанорельефа поверхности ориентирующих тонких слоев а-С:Н;
- исследование природы взаимодействия жидких кристаллов с нанорельефом поверхности а-С:Н;
- исследование влияния физико-химических свойств поверхности а-С:Н на ориентацию жидких кристаллов и электрооптические характеристики ЖК устройств.

В результате проведения комплексных исследований ожидается получение ряда оригинальных результатов:

- развитие фундаментальных представлений о наноструктуре тонких слоев а-С:Н толщиной от 10 до 100 нм с алмазоподобной и полимероподобной структурой;
- оценка эффективности генерации анизотропии наноструктуры поверхности а-С:Н различными способами;
- создание теории физико-химического взаимодействия ЖК с наноструктурой поверхности аморфных углерода;
- получение экспериментальных зависимостей электрооптических свойств ЖК устройств от характера физико-химического взаимодействия на границе раздела «твердое тело – ЖК»;
- создание научной базы для разработки нанотехнологии ориентации нематических и смектических ЖК на основе CVD процесса для изготовления ЖК устройств.

Развитие физико-химической концепции нанотехнологии ориентации жидких кристаллов направлено на разработку нанотехнологии ориентации ЖК с использовани-

ем CVD технологии. Результаты исследования наноструктуры поверхности a-C:H и влияния на нее процессов синтеза и последующей обработки поверхности будут иметь приоритетный характер. Реализация проекта будет способствовать повышению эффективности и конкурентоспособности производства ЖК устройств на мировом уровне.

Литература

1. O'Neill M., S.M. Kelly. Photoinduced surface alignment for liquid crystal displays. // *J. Phys. D: Appl. Phys.* 2000. V. 33. R67-R84.
2. Коншина Е.А. Взаимодействие между нематическим жидким кристаллом и аморфными углеводородными ориентирующими слоями. // *Кристаллография.* 1995. Т.40. №6. 1074–1076.
3. Е.А. Konshina. Alignment of nematic LCs on the surface of amorphous hydrogenated carbon. // *Proc. SPIE.* 1995. Vol. 2731. 20–24.
4. Ahn H.J., Rho S.J., Kim K.C., Kim J.B., Hwang B.H., Park C.J., Baik H.K. Ion-beam induced liquid crystal alignment on diamond-like carbon and fluorinated diamond-like carbon thin films.// *Jap. J. of Appl. Phys. Part 1.* 2005. V. 44. № 6A. 4092–4097.
5. C.J. Park, J.Y. Hwang, H.K. Kang, D.S. Seo, H.J. Ahn, J.B. Kim, K.C. Kim, H.K. Baik. Liquid crystal alignment effects and electro optical performances of twisted nematic-liquid crystal display using ion-beam alignment method on nitrogen-doped diamond like carbon thin film layer. // *Jap. J. of Appl. Phys.* 2005. V. 44. № 3. 1352–1355.
6. Jo Y.-M., Hwang J.-Y., Seo D.-S., Rho S.-J., Baik H.-K. Liquid crystal aligning capabilities and eo characteristics of the TN-LCD with ion-beam exposure on a new diamond-like carbon thin film layer. // *Mol. Cryst. & Liq. Cryst.* 2004. V. 410. 319–327.
7. Hwang J.-Y., Jo Y.-M., Seo D.-S. Jang J. Liquid crystal alignment capability by the UV alignment method in a-C:H thin films. // *Jap. J. of Appl. Phys., Part 2: Letters.* 2003. V. 42, № 2A. L114–L116.
8. Rho S.J., Lee D.-K., Baik H.K., Hwang J.-Y., Jo Y.-M., Seo D.-S. Investigation of the alignment phenomena using a-C:H thin films for liquid crystal alignment materials. *Thin Solid Films.* 2002. V. 420–421, 259–262.
9. Патент РФ №2073902 Коншина Е.А. (1997) Способ однонаправленной ориентации жидких кристаллов.
10. Е.А. Konshina, V.A. Tolmachiev, A.I. Vangonen, and A.P. Onokhov. Novel alignment layers produced by CVD technique from hydrocarbon plasma. // *Proc. SPIE.* 1997. V. 3015. 52–60.
11. Исаев М.В., Коншина Е.А., Онохов А.П., Туровская Т.С. Влияние структуры поверхности конденсированных слоев на ориентацию жидких кристаллов. // *ЖТФ.* 1995. Т.65. № 10. 175–179.
12. Коншина Е.А., Туровская Т.С. Исследование микроструктуры поверхности тонких слоев a-C:H, ориентирующих жидкие кристаллы. // *ЖТФ.* 1998. Т.68. № 1. 106–108.
13. Коншина Е.А., Толмачев В.А., Вангонен А.И., Фаткулина Л.А. Исследование свойств плазменно-полимеризованных слоев и влияния их на ориентацию нематических жидких кристаллов. // *Оптический журнал.* 1997. Т.64. №5. 88–95.
14. A.I. Vangonen, E.A. Konshina. ATR-IR spectroscopy study of NLC orientation at the plasma-polymerized layers surface. // *Mol. Cryst. & Liq. Cryst.* 1997. V.304, 507-512.
15. Коншина Е.А., Вангонен А.И. Исследование ориентации тонких пленок нематика методом ИК спектроскопии МНПВО. // *Оптический журнал.* 1998. Т.65. №7. 39–42.
16. Коншина Е.А., Толмачев В.А., Вангонен А.И. Гомеотропная ориентация нематического жидкого кристалла на поверхности плазменно-полимеризованного октана. // *Кристаллография.* 1998. Т.43. №1. 107–110.

17. Коншина Е.А., Толмачев В.А. Краевые углы и поверхностная энергия слоев аморфного гидрогенизированного углерода. // Коллоидный журнал. 1996. Т.58. №4. 489–493.
18. Толмачев В.А., Коншина Е.А.. Оценка полярных компонентов при определении поверхностной энергии твердого тела. // Коллоидный журнал. 1998. Т.60. №4. 569–573.
19. Толмачев В.А., Коншина Е.А. Оценка энергии взаимодействия на границе раздела «твердое тело – жидкий кристалл». // Оптический журнал. 1998. Т.65. №7. 34–38.

ФАЗОВЫЙ РАСПАД И КРИСТАЛЛИЗАЦИЯ ЛИТИЕВОАЛЮМОСИЛИКАТНЫХ СТЕКОЛ С ДОБАВКОЙ ZrO_2

И.П. Алексеева, О.С. Дымшиц, В.А. Ермаков, А.А. Жилин, В.И. Петров,
М.Я. Центер

Методом спектроскопии комбинационного рассеяния (КР) света исследованы процессы фазового разделения, а также состав и структура нанокристаллов, образующихся в процессе объемной кристаллизации литиевоалюмосиликатных стекол в зависимости от условий их термообработки. В результате исследования обнаружено, что при термообработке исходных стекол с добавкой диоксида циркония в качестве катализатора объемной кристаллизации происходит фазовое разделение с образованием нанокристаллов тетрагональной ZrO_2 . При дальнейшей термообработке происходит частичная кристаллизация матрицы стекла с образованием нанокристаллов, имеющих структуру β -кварца и состав, промежуточный между $LiAlSi_2O_6$ и SiO_2 .

Введение

При термообработке литиевоалюмосиликатных стекол, содержащих диоксиды титана или циркония, в них происходят сложные физико-химические процессы фазового разделения и объемной кристаллизации, формирующие наноразмерную фазу β -кварцевого твердого раствора, при этом образуются стеклокристаллические материалы, так называемые ситаллы. Интерес к детальному исследованию кристаллической структуры ситаллов литиевоалюмосиликатной системы связан с их практически важной особенностью – чрезвычайно малым, близким к нулю или даже отрицательным коэффициентом термического расширения [1].

Исследования, проведенные с помощью рентгеновских методов, показали, что образующиеся нанокристаллы имеют структуру высокотемпературного β -кварца (SiO_2), в котором часть тетраэдров SiO_4 замещена тетраэдрами AlO_4 [2]. Наиболее интересным соединением этого класса является β -эвкрипит ($LiAlSiO_4$), в котором количество тетраэдров обоих типов – SiO_4 и AlO_4 – одинаково. Структурные группы SiO_4 и AlO_4 образуют чередующиеся слои, так что при этом в кристалле присутствуют только связи Si-O-Al, а связи Si-O-Si и Al-O-Al отсутствуют [3]. Кристаллы β -эвкрипитита относятся к довольно редкому типу материалов с отрицательным коэффициентом термического расширения в широком интервале температур.

Химический состав образующихся нанокристаллов в литиевоалюмосиликатных ситаллах спектроскопическими методами ранее детально не изучался, однако с использованием рентгеновских методов было установлено, что получающиеся кристаллы представляют собой твердые растворы, имеющие структуру β -кварца и состав, промежуточный между β -эвкрипитом и β -кварцем [4]. Определение состава твердых растворов и установление их связи с коэффициентом теплового расширения представляет не только общезначимый, но и практический интерес.

Задачей настоящей работы является исследование состава и структуры литиевоалюмосиликатных ситаллов в зависимости от длительности и температуры термообработки исходных стекол с диоксидом циркония в качестве катализатора объемной кристаллизации.

Методика исследований включает в себя широкое использование метода спектроскопии комбинационного рассеяния (КР) света, который уже показал свою высокую информативность при исследовании областей неоднородности стекол различных классов и ситаллов в диапазоне размеров единицы и десятки нм, возникающих в процессе фазового разделения и последующей кристаллизации. В частности, этим методом была исследована природа низкочастотной полосы (так называемого бозонного пика) в спектрах КР стекол различных классов – литиевоалюмосиликатных с добавками TiO_2 [5], магниевых- и цинковоалюмосиликатных с добавками TiO_2 [6,7], пирофосфатных [8] и

натриевообратных [9]. На основании этих исследований был сделан общий вывод, заключающийся в том, что происхождение бозонного пика всегда связано с областями неоднородности, обусловленными флуктуациями структуры и химического состава стекол. При этом было показано, что в литиевоалюмосиликатных стеклах бозонный пик связан с упругими колебаниями областей неоднородности, которые способны к эволюции при термообработке, превращаясь в процессе фазового распада в кристаллические частицы, содержащие катализатор объемной кристаллизации.

Результаты и обсуждение

Исследовались спектры КР литиевоалюмосиликатного стекла состава $12\text{Li}_2\text{O}\cdot 15\text{Al}_2\text{O}_3\cdot 73\text{SiO}_2$ с добавкой 3,5 мол. % ZrO_2 в качестве катализатора объемной кристаллизации. Стекло изготовлялось при температуре варки 1650°C . Тепловая обработка образцов заключалась в прогреве на стадии предситаллизации при 750°C в широком временном интервале, а также в дополнительной высокотемпературной термообработке, приводящей к кристаллизации матрицы.

Для возбуждения спектров КР использовался аргоновый лазер ЛГН-404А с излучениями с длинами волн $\lambda = 514.5$ и 488.0 нм и мощностью порядка 100-150 мВт. Регистрация спектров осуществлялась двойным монохроматором ДФС-52М (ЛОМО) с фотомножителем ФЭУ-79. Управление монохроматором осуществлялось компьютером. Использовалась геометрия рассеяния под 90° .

При изучении объемной кристаллизации литиевоалюмосиликатных стекол, содержащих диоксид циркония, методами рентгенофазового анализа (РФА) и рентгеновского малоуглового рассеяния (РМУ) было выявлено две стадии этого процесса [2]. Основная черта первой из них заключалась в фазовом разделении исходного стекла с выпадением и последующей кристаллизацией нанокристаллов ZrO_2 с радиусами, изменяющимися в пределах 2,5–4,0 нм. На второй стадии происходила кристаллизация из стекла твердых растворов со структурой β -кварца. Размеры кристаллов не превышали 30–50 нм, при этом материал остается оптически прозрачным. Химический состав твердых растворов ранее не изучался. Объемное содержание нанокристаллов возрастало с увеличением длительности термообработки, однако даже и в конечном продукте – ситалле всегда присутствовала остаточная стеклофаза в количестве порядка 30 % [1] в виде прослоек между частицами закристаллизованной матрицы, связывая их в прочный материал.

На рис. 1 приведены спектры КР исходного (а) и термообработанных (б-д) при 750°C стекол и ситалла (е). Спектр исходного стекла содержит ряд широких полос с частотами 75, 470, 600, 800, 970 и 1070 см^{-1} , из которых наиболее интенсивна поляризованная полоса 470 см^{-1} . Все полосы, за исключением самой низкочастотной, отвечают колебаниям структурных группировок, образующих сетку исходного стекла. Низкочастотная полоса 75 см^{-1} является сложной и состоит из двух компонентов с частотами 81 и 68 см^{-1} , проявляющихся в поляризационных измерениях [10]. На ранней стадии термообработки стекла в течение 28 и 72 часов в спектрах КР (рис. б, в) возникают и усиливаются полосы колебаний с частотами 148, 280, 310, 448 (видна только в перпендикулярной поляризации) и 643 см^{-1} , соответствующие колебаниям кристаллов тетрагонального ZrO_2 [11]. В низкочастотной области на фоне полосы 75 см^{-1} появляется более узкий максимум, отвечающий колебанию с частотой 62 см^{-1} , который затем смещается к 52 см^{-1} и значительно усиливается по интенсивности. Поляризационные измерения показали, что вновь возникающая полоса колебания с частотой 62 см^{-1} имеет степень деполяризации ρ , близкую к 0. Из измерений в высокочастотной области отметим монотонное ослабление интенсивности полосы 970 см^{-1} .

При дальнейшей термообработке от 170 до 636 ч и при завершении ситаллизации образца в спектрах КР (рис. 1, *з-е*) наблюдаются заметные изменения, касающиеся широкой полосы 470 см^{-1} . Она несколько сужается, в основном за счет уменьшения низкочастотного крыла, и на ее фоне возникает и усиливается узкая полоса 480 см^{-1} . Подобные изменения происходят и в области слабой широкой полосы, отвечающей колебанию с частотой 1070 см^{-1} . С началом кристаллизации значительно увеличился уровень светорассеяния в образце, что затруднило регистрацию низкочастотного КР (рис. 1, *з*). Однако при завершении этого процесса в низкочастотной области отчетливо видны две полосы колебаний с частотами 50 и 106 см^{-1} (рис. 1, *д, е*), первая из которых полностью поляризована ($\rho = 0$), а для второй $\rho \approx 0,3$. Относительный объем кристаллической фазы в образцах, спектры которых представлены на рис. 1, *з-е*, составляет 10, 60 и 70 %, соответственно (по данным РФА).

Перейдем к обсуждению полученных результатов. Рассмотрим сначала низкочастотную область спектра. Возникновение низкочастотных полос КР может быть связано с упругими колебаниями областей неоднородности в матрице стекла. В многокомпонентных стеклах области неоднородности с размерами порядка единиц нанометров, обусловленные структурно-химическими флуктуациями, могут возникать как в процессе варки, так и охлаждения расплава или термообработки.

Сравнение спектров КР исходного стекла с добавкой ZrO_2 (рис. 1, *а*) и стекла того же состава без катализатора кристаллизации [12] показывает совпадение их полос во всей области колебательных частот выше 100 см^{-1} .

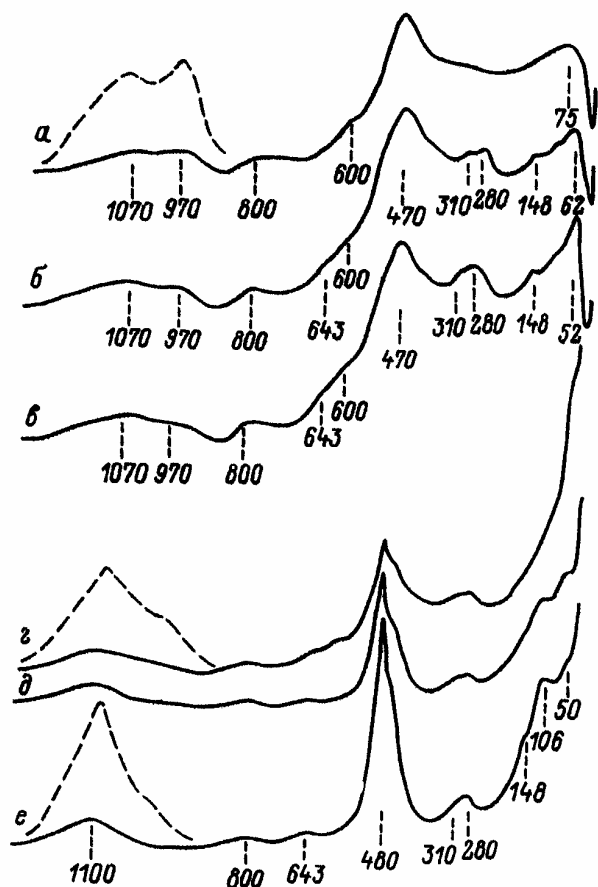


Рис. 1. Спектры КР стекла состава $12\text{Li}_2\text{O} \cdot 15\text{Al}_2\text{O}_3 \cdot 73\text{SiO}_2$ с добавкой 3,5 мол. % ZrO_2 на различных стадиях термообработки при $750\text{ }^\circ\text{C}$: а – исходное стекло, б – 28 час., в – 72 час., з – 170 час., д – 636 час., е – ситалл. Штриховой линией показана часть спектра при большем усилении

Спектры КР образцов, прошедших начальную тепловую обработку (рис. 1, б, в), отражают процесс фазового распада с образованием частиц кристаллического ZrO_2 при практически неизменной структуре матрицы. Возникновение смещающейся по частоте от 62 к 52 см^{-1} полосы естественно интерпретировать как КР на упругих колебаниях этих частиц. Оценка их размеров может быть сделана по формуле (1) [13]

$$\omega_{l,n}^{S,T} = \frac{\eta_{l,n}^{S,T} v_l}{2\pi R c}, \quad (1)$$

где $\omega_{l,n}^{S,T}$ – частоты сферических (S) и крутильных (T) колебаний, R – радиус частиц, $\eta_{l,n}^{S,T}$ – коэффициенты, зависящие от отношения поперечной v_t и продольной v_l скоростей звука, c – скорость света в вакууме, l – угловой момент, равный 0 или 2, $n = 1, 2, \dots$ – номер колебания. На практике обычно наблюдаются одно сферическое ω_{02}^S и одно крутильное (поверхностное) ω_{21}^T колебания, которым, как было установлено в [14], отвечают полностью поляризованная и деполаризованная полосы КР, соответственно.

Расчет по [13] коэффициентов $\eta_{l,n}^{S,T}$ с учетом значений скоростей звука в кристалле ZrO_2 [15], $v_t = 3,74 \cdot 10^5$ и $v_l = 7,84 \cdot 10^5$ см/с привел к $\eta_{21}^T = 5,1$, $\eta_{02}^S = 5,9$. Используя значение частоты упругих колебаний частиц ZrO_2 , равное 50 см^{-1} (кривые в-е), из (1) получаем оценку размеров частиц $2R = 4,7$ нм, что совпадает с данными рентгеновских методов [2]. При таком размере оценка частоты крутильного колебания ω_{21}^T дает значение 21 см^{-1} . Отсутствие в спектрах полосы, отвечающей этому колебанию, которое должно иметь максимальную амплитуду на поверхности частиц, свидетельствует по [13] о тесном контакте частиц с окружающей матрицей. На стадии термообработки, соответствующей началу фазового распада (рис. 1, б), частота упругих колебаний частиц ZrO_2 равна 62 см^{-1} , что приводит к средним размерам частиц $2R = 3,8$ нм. Большая ширина высокочастотных полос КР ZrO_2 говорит о невысокой степени упорядоченности внутри образующихся частиц, которая не улучшается с термообработкой.

На стадии объемной кристаллизации стекла (рис. 1, д, е) отношение интенсивностей полос низкочастотного и высокочастотного КР частиц ZrO_2 значительно уменьшается, что может быть связано только со сближением упругих постоянных этих частиц и окружающей матрицы. Наблюдаемое уменьшение трудно понять, если предположить, что частицы локализируются преимущественно в остаточной стеклофазе [2], даже при вполне возможном изменении ее состава по сравнению с исходным. Действительно, скорости звука в кварцевом стекле и в способных к ситаллизации стеклах не превышают значений $v_l = 6,5 \cdot 10^5$ см/с [16], что существенно отличается от аналогичных данных для ZrO_2 . В то же время оценка величин v_l из упругих постоянных для β -эвкрипта и β -кварца приводит к значениям $8,5 \cdot 10^5$ и $7,6 \cdot 10^5$ см/с соответственно. В этот же интервал попадает и $v_l = 7,8 \cdot 10^5$ см/с для ZrO_2 . Таким образом, наблюдаемое уменьшение относительной интенсивности полосы упругих колебаний нанокристаллов ZrO_2 в условиях частичной кристаллизации матрицы может быть объяснено тем, что кристаллы β -кварцевого твердого раствора растут вокруг кристаллов ZrO_2 . При этом дальнейший рост последних оказывается затрудненным как по данным КР (неизменность положения низкочастотной полосы), так и по данным РМУ [2]. Это может быть обусловлено не только возможным затруднением диффузии ионов циркония в кристаллической части матрицы, но и ее упругими свойствами.

Описанные выше результаты значительно отличаются от данных, полученных нами ранее для стекол того же состава с диоксидом титана [14]. Скорость звука в TiO_2 больше скоростей звука как в стеклах, так и в β -эвкриптите, поэтому в низкочастотных

спектрах КР на всех стадиях присутствуют две полосы, отвечающие упругим колебаниям ω_{21}^T и ω_{02}^S . При этом на стадии кристаллизации матрицы отношение интенсивностей полос ω_{21}^T и ω_{02}^S даже возрастает [14]. Известно также, что кристаллизация стекла в этом случае не препятствует дальнейшему росту частиц [17]. Эти данные свидетельствуют в пользу того, что кристаллы матрицы образуются между частицами титаносодержащей фазы и не препятствуют их росту. Различное расположение частиц ZrO_2 и TiO_2 относительно кристаллизующейся матрицы указывает на разные механизмы их действия как катализаторов мелкодисперсной кристаллизации. Результаты этого исследования частично опубликованы в [10].

Из особенностей низкочастотных спектров КР стекол с ZrO_2 на стадиях их предситализации и окончательного формирования ситалла (рис. 1, *д, е*) отметим исчезновение полосы бозонного пика с частотой 75 см^{-1} и возникновение новой $\sim 106\text{ см}^{-1}$.

Рассмотрим теперь высокочастотную ($200\text{--}1200\text{ см}^{-1}$) область спектров КР, последовательные изменения в которой свидетельствуют о появлении кристаллов тетрагонального ZrO_2 и, затем, о кристаллизации β -эвкриптитовых твердых растворов. Экспериментально было установлено, что по мере термообработки отношение интенсивности полос 480 и 470 см^{-1} , отвечающих полностью симметричным колебаниям тетраэдров SiO_4 в кристаллической фазе и остаточной стеклофазе соответственно, растет значительно медленнее, чем отношение объемов этих фаз (рис. 1, *з-д*).

Из теоретического рассмотрения процесса объемной кристаллизации следует, что при возникновении зародышей новой фазы соблюдается принцип минимума поверхностной энергии на границе фаз. Это означает, в частности, хорошую смачиваемость зародышей средой [18]. При возникновении и дальнейшем росте кристаллов энергетически выгодно их срастание с прослойками стекла и отсутствие напряжений и трещин. Сходство спектров кристаллической и аморфной частей матрицы, свидетельствующее о близости их состава и структуры, и хорошая смачиваемость кристалликов стеклянными прослойками позволяет высказать предположение о том, что стеклокристаллические материалы образуют не только единую механическую систему с высокой прочностью [18], но и являются, по-видимому, единой квантовой системой. При этом появляется возможность резонансных взаимодействий для фононов кристаллической и аморфной частей матрицы [19], которые обычно приводят к усилению слабой полосы за счет сильной при сохранении общей интегральной интенсивности.

В связи с этим для спектров, приведенных на рис. 1, *з-е*, были проведены измерения интегральной интенсивности сложной полосы $480\text{--}470\text{ см}^{-1}$ с использованием внутреннего эталона, в качестве которого была выбрана группа полос $280\text{--}310\text{ см}^{-1}$ ZrO_2 . Оказалось, что интегральная интенсивность этой полосы от момента начала кристаллизации стекла до его завершения практически не изменяется (с возможной погрешностью $\pm 10\%$). В ней происходит только некоторое перераспределение интенсивности между широкой и узкой компонентами.

Перейдем теперь к обсуждению состава нанокристаллов основной фазы в исследуемых ситаллах. Как упоминалось выше, при термообработке исходных стекол в них происходит объемная кристаллизация твердых растворов со структурой β -кварца и размерами кристаллических областей в несколько десятков нм. Близкий к нулю коэффициент теплового расширения ситаллов традиционно связывается с таким изменением длин осей в элементарной ячейке β -эвкриптита с ростом температуры, при котором объем элементарной ячейки уменьшается, что приводит к отрицательному коэффициенту теплового расширения для этого кристалла. Однако не следует забывать, что в ситаллах выделяются твердые растворы со структурой β -кварца, который сам имеет слабо положительный коэффициент теплового расширения. В [20] было показано, что вклад остаточной стеклофазы в термическое расширение ситаллов невелик из-за малого ее

содержания и малых значений коэффициента теплового расширения, так как по составу стеклофаза представляет собой кремнеземистое стекло с небольшими добавками катализатора и Al_2O_3 . Поскольку была обнаружена качественная связь между составом основной кристаллической фазы ситалла и его термомеханическими свойствами, представляется весьма важным проведение исследований количественного состава и структуры нанокристаллов твердых растворов $\text{LiAlSiO}_4\text{-SiO}_2$ в литиевоалюмосиликатных ситаллах в зависимости от состава исходного стекла, катализатора объемной кристаллизации (ZrO_2 и/или TiO_2), вводимых в состав стекла окрашивающих и других добавок, улучшающих потребительские свойства ситаллов, а также от условий термообработки. В настоящей работе сделан первый шаг в этом направлении.

Для проведения таких исследований спектроскопическими методами требуется знание спектров КР модельных соединений: кристаллов β -эвкриптита, β -кварца и кристаллов β -кварцевых твердых растворов промежуточных составов. Сравнение их со спектрами реальных образцов ситаллов различного состава позволяет интерпретировать полосы КР ситаллов и сделать заключение о составе выпадающей в них кристаллической фазы. В литературе имеются данные о спектрах КР β -эвкриптита (LiAlSiO_4) [3], β -кварца (SiO_2) [21,22] и кристаллов промежуточного состава $\text{LiAlSi}_2\text{O}_6$ [23]. Спектр КР β -эвкриптита содержит наиболее интенсивные полосы, отвечающие полностью симметричным колебаниям с частотами 498 и 1035 см^{-1} , а также ряд других, значительно менее интенсивных полос [3]. Аналогом полосы колебаний с частотой 498 см^{-1} в спектре КР β -кварца является полоса с частотой 464 см^{-1} , а полоса колебания β -кварца с частотой 1173 см^{-1} имеет малую интенсивность [3].

Кристаллы состава $\text{LiAlSi}_2\text{O}_6$ существуют в виде трех различных модификаций, спектры КР которых приведены на рис. 2 [23]. Как видно из рисунка, спектры α -сподумена (моноклинная пространственная группа $C2/c$) и β -сподумена (тетрагональная пространственная группа $P4_32,2$, $Z=4$) вследствие более низкой симметрии этих модификаций содержат большее число полос КР по сравнению со спектром высокосимметричной структуры β -кварцевого твердого раствора (гексагональная пространственная группа $P6_3,22$, $Z=1$). Наиболее интенсивные полосы в спектре кристалла со структурой β -кварца имеют частоты 480 и 1088 см^{-1} , которые рассматривались в [23] как аналоги полос КР β -кварца с частотами 464 и 1173 см^{-1} , соответственно. При этом наблюдаемая полоса колебания с частотой 1088 см^{-1} значительно уширена по сравнению с аналогичной полосой β -кварца, что связывалось со случайным расположением тетраэдров SiO_4 в кристалле твердого раствора состава $\text{LiAlSi}_2\text{O}_6$. Анализ приведенных спектров показал, что отношение интегральных интенсивностей высокочастотной и более низкочастотной полос в спектрах КР кристаллов несет информацию об их составе. Действительно, это отношение для кристалла $\text{LiAlSi}_2\text{O}_6$ со структурой β -кварца равно 0,54, тогда как для β -кварца оно не превышает 0,03. Поэтому можно ожидать, что для кристаллов с промежуточным составом отношение интенсивностей окажется внутри указанного интервала.

Сравнивая полученный нами спектр КР ситалла (рис. 1, e) со спектром кристаллов $\text{LiAlSi}_2\text{O}_6$ в структуре β -кварца (рис. 2, A), мы видим их хорошее соответствие. Низкочастотная полоса в обоих спектрах имеет одно и то же положение 480 см^{-1} . Однако высокочастотная полоса в спектре КР ситалла (1100 см^{-1}) попадает в интервал между положениями аналогичных полос в спектрах β -кварца (1173 см^{-1}) и кристалла $\text{LiAlSi}_2\text{O}_6$ (1088 см^{-1}). Кроме того, отношение интенсивностей высокочастотной и низкочастотной полос в спектре КР ситалла также имеет промежуточное значение, равное 0,33. Это означает, что в исследованном нами ситалле кристаллизуются нанокристаллы со структурой β -кварца и составом, обогащенным SiO_2 по сравнению с $\text{LiAlSi}_2\text{O}_6$, т.е. промежуточным составом между $\text{LiAlSi}_2\text{O}_6$ и SiO_2 .

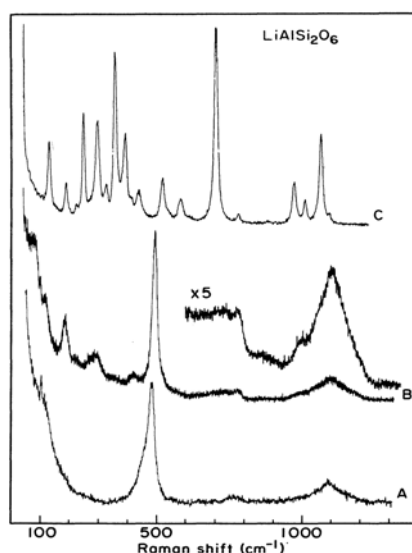


Рис. 2. Спектры КР $\text{LiAlSi}_2\text{O}_6$: (А) – кристалл со структурой β -кварца, (В) - β -сподумен и (С) - α -сподумен. Рисунок из статьи [23]

Заключение

При исследовании литиевоалюмосиликатных стекол с добавками ZrO_2 в качестве катализатора объемной кристаллизации и ситаллов на их основе методами обычного и низкочастотного КР было обнаружены проявления фазового распада с образованием нанокристаллов тетрагонального ZrO_2 и определены размеры этих нанокристаллов. Показано, что кристаллизация матрицы происходит вокруг частиц ZrO_2 . Полученные результаты сопоставлены с данными рентгеновских методов. Показано, что в исследованных стеклах и ситалле происходит кристаллизация матрицы с образованием нанокристаллов со структурой β -кварца и химическим составом, обогащенным SiO_2 по сравнению с $\text{LiAlSi}_2\text{O}_6$.

Литература

1. Бережной А.И. Ситаллы и фотоситаллы. М., 1966. 340 с.
2. Алексеева И.П., Голубков В. В., Карапетян Г.О., Чуваева Т.И. Изучение процесса кристаллизации литиевоалюмосиликатного стекла с добавками ZrO_2 // Физ. и хим. стекла. 1979. Т. 5. № 3. С. 296–302.
3. Sprengard R., Binder K., Brandle M., Fotheringham U., Sauer J., Pannhorst W. On the interpretation of the experimental spectrum of β -eucryptite $\text{LiAlSi}_4\text{O}_4$ from atomistic computer modeling // J. Non-Cryst. Solids. 2000. V. 274. P. 264–270.
4. Астахова В.В., Жилин А.А., Филатов С.К., Чуваева Т.И. Рентгенографическое исследование термического расширения β -эвкриптита и ситалла на его основе // Неорганические материалы. 1987. Т. 23. № 5. С. 841–845.
5. Chuvaeva T.I., Dymshits O.S., Petrov V.I., Tsenter M.Ya., Zhilin A.A., Golubkov V.V. Low-frequency Raman scattering and small-angle X-ray scattering of glasses inclined to phase decomposition. // J. Non-Cryst. Solids. 1999. V. 243. № 2–3. P. 244–250.
6. Chuvaeva T.I., Dymshits O.S., Petrov V.I., Tsenter M.Ya., Shashkin A.V., Zhilin A.A., Golubkov V.V. Low-frequency Raman scattering of magnesium aluminosilicate glasses and glass-ceramics // J. Non-Cryst. Solids. 2001. V. 282. № 2–3. P. 306–316.
7. Golubkov V.V., Dymshits O.S., Petrov V.I., Shashkin A.V., Tsenter M.Ya., Zhilin A.A., Kang U. Small-angle X-ray scattering and low-frequency Raman scattering study of liquid

- phase separation and crystallization in titania-containing glasses of the $\text{ZnO-Al}_2\text{O}_3\text{-SiO}_2$ system // *J. Non-Cryst. Solids*. 2005. V. 351. P. 711–721.
8. Баранов А.В., Голубков В.В., Петров В.И. Исследование строения пирофосфатных стекол методами комбинационного и рентгеновского малоуглового рассеяния // *Физ. и хим. стекла*. 1999. Т. 25. № 4. С. 385–401.
 9. Baranov A.V., Perova T.S., Petrov V.I., Vij J.K., Nielsen O.F. Nature of the Boson peak in Raman spectra of sodium borate glass systems: influence of structural and chemical fluctuations and intermolecular interactions // *J. Ram. Spectrosc.* 2000. V. 31. №. 8–9. P. 819–825.
 10. Бобович Я.С., Жилин А.А., Петров В.И., Центер М.Я., Чуваева Т.И. Проявление процесса фазового распада в литиевоалюмосиликатных стеклах, содержащих диоксид циркония, в спектрах комбинационного рассеяния света // *Опт. и спектр.* 1992. Т. 72. № 6. С. 1356–1362.
 11. Thackeray D.P.C. The Raman spectrum of zirconium dioxide // *Spectrochim. Acta A*. 1974. V. 30. №.2. P. 549–550.
 12. Бобович Я.С., Петров В.И., Центер М.Я. Проявления субмикронеоднородной структуры ситаллизирующихся многокомпонентных силикатных стекол в спектрах КРС // *Опт. и спектр.* 1990. Т. 68. № 4. С. 792–795.
 13. Tamura A., Higeta K., Ichinokava T. Lattice vibration and specific heat of a small particle. // *J. Phys. C: Solid State Phys.* 1982. V. 15. N. 24. P. 4975-4991.
 14. Петров В.И., Бобович Я.С. Комбинационное рассеяние на акустических фононах субмикрорекристаллов TiO_2 в стеклах // *Опт. и спектр.* 1989. Т. 67. № 3. С. 619–621.
 15. Акустические кристаллы. Справочник / Под ред. М. П. Шаскольской. М., 1982. 632 с.
 16. Champagnon B., Boukenter A., Duval E., Mai C., Vigier G., Rodek E. Early stages of nucleation of Zerodur glass. Very low frequency Raman scattering and small angle X-ray scattering investigations // *J. Non-Cryst. Solids*. 1987. V. 94. №2. P. 216–221.
 17. Алексеева И.П., Голубков В.В., Чуваева Т.И. Исследование кинетики фазового распада в стеклах системы $\text{Li}_2\text{O-Al}_2\text{O}_3\text{-SiO}_2\text{-TiO}_2$ методом РМУ // *Физ. и хим. стекла*. 1981. Т. 7. № 1. С. 47–54.
 18. Филипович В.Н. Начальные стадии кристаллизации стекол и образование ситаллов. / В сб.: *Стеклообразное состояние*. М.-Л.: Изд. АН СССР, 1963. С. 9–24.
 19. Mönch W. Semiconductor surfaces and interfaces. / Berlin, Hong Kong: Springer-Verlag. 1993. 366 p.
 20. Артюх Е.П., Астахова В.В., Жилин А.А., Филатов С.К., Чуваева Т.И. Терморентгенографическое исследование твердых растворов со структурой β -кварца // *Неорганические материалы*. 1992. Т. 28. № 2. С. 408–415.
 21. Bates J.B., Quist A.S. Polarized Raman spectra of β -quartz // *J. Chem. Phys.* 1972. V. 56. N 4. P. 1528–1533.
 22. Etchepare J., Merian M., Smetankine L. Vibrational normal modes of SiO_2 I. α and β quartz // *J. Chem. Phys.* 1974. V. 60. № 5. P. 1873–1876.
 23. Sharma S.K., Simons B. Raman study of crystalline polymorphs and glasses of spodumene composition quenched from various pressures // *American Mineralogist*. 1981. V. 66. P. 118–126,

РАЗРАБОТКА И ИССЛЕДОВАНИЕ НОВЫХ АКТИВНЫХ СРЕД ДЛЯ ТВЕРДОТЕЛЬНЫХ UP-КОНВЕРСИОННЫХ И ИК ЛАЗЕРОВ

А.М. Ткачук, С.Э. Иванова, И.К. Разумова, В.М. Рейтеров, Е.А. Пономарева, В.Ю. Егоров, Н.В. Никоноров

Модифицированным методом Бриджмена-Стокбаргера.-Степанова. выращены ориентированные кристаллы YLF:РЗИ высокого оптического качества и разработана технология выращивания кристаллов двойного фторида натрия-иттрия $\text{Na}_{0.4}\text{Y}_{0.6}\text{F}_{2.2}:\text{Er}^{3+}$ ($\text{NYF}:\text{Er}^{3+}$), активированных ионами Er^{3+} . Исследованы спектроскопические характеристики, кинетика люминесценции, концентрационное и температурное тушение люминесценции в серии кристаллов $\text{NYF}:\text{Er}^{3+}$ (Er 0.5-15%) при селективном лазерном возбуждении. Показано, что люминесценция с уровня $^4\text{S}_{3/2}$ значительно тушится с повышением температуры и с увеличением концентрации эрбия. Люминесценция с уровней $^4\text{G}_{11/2}$, $^2\text{G}(\text{H})_{9/2}$, $^4\text{F}_{9/2}$ и $^4\text{I}_{9/2}$ тушится, в основном, в результате многофононных безызлучательных переходов, а концентрационное тушение люминесценции с уровней $^4\text{I}_{11/2}$ и $^4\text{I}_{13/2}$ не наблюдается.

Изучены процессы формирования инверсной населенности рабочих уровней трехмикронного лазерного перехода в кристаллах двойных фторидов, активированных эрбием, под действием непрерывной накачки излучением InGaAs лазерных диодов (0.967-0.982 мкм). Экспериментально исследованы зависимости населенности уровней $^4\text{I}_{11/2}$ и $^4\text{I}_{13/2}$ от концентрации активаторных ионов при различной мощности селективной накачки. Сделан вывод о перспективности использования активированных эрбием кристаллов двойных фторидов в качестве активных сред для лазеров видимого и ИК диапазона.

1. Введение

В настоящее время, несмотря на огромное количество известных лазерных кристаллов, наиболее широкое практическое применение в лазерах с диодной накачкой нашли активированные неодимом кристаллы алюмо-иттриевого граната $\text{Nd}:\text{Y}_3\text{Al}_5\text{O}_{12}$ (YAG), иттриевого ванадата $\text{Nd}:\text{YVO}_4$ и двойного фторида лития-иттрия $\text{Nd}:\text{LiYF}_4$ (YLF).

По сравнению с известными активными средами для твердотельных ИК лазеров на основе кислородсодержащих соединений (гранаты различного состава, алюминаты, ванадаты, вольфраматы, силикатные стекла и др.) активные среды на основе кристаллов двойных фторидов имеют следующие преимущества и особенности:

- широкая область прозрачности (от 0.15 до 7.5 мкм);
- узкий фононный спектр (предельная частота фононов во фторидах составляет 400–560 см^{-1} , в гранатах 890 см^{-1}), что обеспечивает, с одной стороны, низкие потери на тепловыделение при безызлучательной релаксации энергии возбуждения, с другой стороны, в этих матрицах РЗИ излучают в широком спектральном диапазоне – от ИК до УФ;
- кристаллы двойных фторидов имеют исключительно высокие термооптические характеристики, что делает их незаменимыми в лазерных системах высокой мощности с высокими требованиями к качеству выходного пучка.

Кроме того, кристаллы двойных фторидов технологичны, выращиваются при более низких температурах в графитовых тиглях, что существенно удешевляет их производство.

Кристаллы $\text{LiYF}_4:\text{RE}$ могут рассматриваться как перспективные материалы для визуализации ИК излучения, как лазерные материалы, не уступающие по выходным характеристикам кристаллам гранатов, а в некоторых случаях их превосходящие.

Расширение спектрального диапазона ИК лазеров с диодной накачкой возможно при использовании кристаллов двойных фторидов, активированных редкоземельными ионами (РЗИ): неодима, эрбия, тулия и иттербия, а также кристаллов, соактивированных иттербием и празеодимом, диспрозием, гольмием. Так, на кристаллах двойных фторидов LiYF_4 (YLF), BaY_2F_8 (BYF), активированных эрбием, тулием или гольмием с соактиваторами, возможно получить генерацию при накачке лазерными диодами в ИК

области на ряде длин волн от 0,65 до 3,9 мкм. В кристаллах двойных фторидов YLF и YbF, активированных иттербием и гольмием, эффективно идут процессы up-конверсии ИК энергии возбуждения в коротковолновую область спектра. Активированные эрбием кристаллы двойных фторидов, в том числе YLF:Er³⁺, хорошо известны как активные среды для многоцветных лазеров, генерирующих на целом ряде длин волн в видимой и ИК области спектра как при ламповой [1–13], так и при селективной лазерной и диодной накачке [14–28], включая преобразование излучения 0.85 мкм во вторую гармонику 0.425 мкм [11, 25–28] и накачку по up-конверсионным схемам [29–48], на них продемонстрирован эффект лавинной накачки [49]. Эффективность up-конверсионного возбуждения кристаллов с РЗИ успешно продемонстрирована на ряде активированных РЗИ кристаллов [50]. Наиболее удобными для накачки лазерными диодами (LD) являются кристаллы, активированные Er и кристаллы соактивированные Yb.

В работе исследованы спектроскопические характеристики и концентрационное тушение люминесценции в серии новых кристаллов Na_{0.4}Y_{0.6}F_{2.2}:Er³⁺ 0.5-15% и процессы формирования инверсной населенности рабочих уровней трехмикронного лазерного перехода под действием непрерывной накачки излучением InGaAs лазерных диодов (0.967-0.982 мкм) и при накачке в области 1.5 мкм.

2. Выращивание кристаллов

Кристаллы YLF:Er и Na_{0.4}Y_{0.6}F_{2.2}:Er³⁺ (0.5-15) ат. % были выращены методом Стокбаргера–Бриджмена из шихты стехиометрического состава, приготовленной твердофазным методом. Синтез шихты для выращивания кристаллов осуществлялся путем термического фторирования каждого из компонент с последующим плавлением смесей в атмосфере аргона с фторирующими добавками.

Исследование диаграммы плавкости системы NaF–YF₃ показало, что в области больших концентраций YF₃ существует конгруэнтно плавящаяся при 984°C фаза переменного состава со структурой флюорита. Максимум отвечает 60% YF₃. Максимальная область гомогенности флюоритовой фазы при 850–900°C составляет 52-65 мол.% YF₃. Именно эта бертолитовая фаза была использована для выращивания кристаллов Na_{0.4}Y_{0.6}F_{2.2}. Выращены кристаллы Na_{0.4}Y_{0.6}F_{2.2}:Er³⁺ (0.5-100) ат. % из шихты стехиометрического состава.

Монокристаллы YLF:Er и Na_{0.4}Y_{0.6}F_{2.2}:Er³⁺ хорошего оптического качества (монокристаллы, прозрачные) имели диаметр 8 мм, длину 40–50 мм. Проведен контроль их оптического качества. Все выращенные кристаллы были без свилей и пузырей, однако имели ярко выраженную ячеистую структуру, причину возникновения которой предстоит выяснить. По результатам рентгеноспектрального анализа, проводившегося на электронно-зондовом микроанализаторе Camebax фирмы САМЕКА, и спектрофотометрическим методом проведено определение концентрации активаторных ионов в контрольных образцах. Из выращенных кристаллов состава Na_{0.4}Y_{0.6}F_{2.2}:Er³⁺ (0.5, 1, 2, 5, 10, 15 и 100) изготовлены образцы для спектроскопических исследований.

3. Структура

Кристаллы двойного фторида лития-иттрия YLF представляют собой кристаллы с упорядоченной структурой типа шеелита. Пространственная группа симметрии C_{4h}, кристаллы имеют тетрагональную структуру с параметрами решетки $a=5.175$ и $c=10.740$ А, число молекул в элементарной ячейке $z=4$. При активации кристаллов YLF редкоземельными ионами последние занимают позиции ионов иттрия с точечной симметрией S₄. Редкоземельные ионы (RE³⁺) изоморфно и изовалентно замещают ионы Y³⁺, матрица YLF допускает активацию редкоземельными ионами первой половины

лантаноидного ряда в высоких концентрациях, вплоть до 100%. Концентрация редкоземельных ионов первой половины лантаноидного ряда в буле изменяется по длине кристалла в соответствии с концентрацией примеси в расплаве и зависит от длины кристалла и объема расплава, для ионов второй половины ряда коэффициент вхождения практически равен 1.

Кристаллы двойного фторида натрия-иттрия представляют собой кристаллы с неупорядоченной структурой типа твердых растворов и принадлежат к системе NaF-YF₃.

Кристаллы состава 2NaF-3YF₃ имеют формулу Na_{0.4}Y_{0.6}F_{2.2} [51]. Методами рентгенофазового анализа и рентгеноструктурного микроанализа при исследовании различных участков монокристаллов обнаружено незначительное изменение постоянной кристаллической решетки по длине монокристалла, для Na_{0.4}Y_{0.6}F_{2.2} параметр решетки *a* изменялся от 0.5503 до 0.5499 нм [52]. Показано, что соединения с флюоритовой структурой в системе NaF-YF₃ не имеют стехиометрической дистектики и являются твердыми растворами с формулой Na_{0.5-x}Y_{0.5+x}F_{2+2x} [53]. Границам устойчивости существования таких растворов в широком температурном интервале отвечают значения *x* в диапазоне 0.08–0.12. Кристаллы состава 2NaF-3YF₃, имеющие формулу Na_{0.4}Y_{0.6}F_{2.2} в дальнейшем будем называть NYF.

Кристаллическая структура NYF представляет собой модифицированную структуру флюорита (CaF₂), в которой 40% ионов Ca²⁺ замещено ионами Na⁺, а остальные 60% – ионами Y³⁺. Кристаллы имеют кубическую структуру с постоянной решетки *a* = 0.5503 нм, число молекул в элементарной ячейке *z* = 4. Каждый катион расположен в центре куба из восьми анионов, и каждый анион окружен тетраэдром из катионов (F⁻).

При активации кристаллов NYF редкоземельными ионами последние изоморфно замещают ионы Y³⁺ и занимают позиции с точечной симметрией S₄. Матрица NYF допускает активацию редкоземельными ионами первой половины лантаноидного ряда в высоких концентрациях, коэффициент вхождения неодима в кристаллы NYF:Nd³⁺ близок к единице, *K*_{Nd} = 0,85 [54], коэффициент вхождения эрбия *K*_{Er} = 1.

4. Спектроскопические характеристики и самотушение люминесценции в кристаллах NYF:Er

Оптические спектры кристаллов NYF:Er представляют собой широкие слабо структурированные полосы поглощения и излучения. В спектре поглощения кристаллов NYF:Er³⁺ имеются широкие полосы поглощения в интервалах 790–801 нм, 965–980 нм и 1500 нм, удобные для накачки лазерными диодами. Соответствующие экспериментальные значения пиковых поперечных сечений поглощения из основного состояния (GSA) для NYF:Er при T=300 К равны: $\sigma_a^p(800) = 0,48 \cdot 10^{-21} \text{ см}^2$, $\sigma_a^p(970) = 1,5 \cdot 10^{-21} \text{ см}^2$, $\sigma_a^p(1530) = 4,2 \cdot 10^{-21} \text{ см}^2$.

Результаты детального исследования оптических спектров кристаллов Na_{0.4}Y_{0.6}F_{2.2}:Er³⁺ при различных температурах (6, 12 и 300 К) и различных способах возбуждения опубликованы в работе [55].

На основании данных анализа спектров поглощения и люминесценции при низких (6 и 12 К) температурах определена структура штарковского расщепления уровней эрбия как «квазицентров», для которых характерно неоднородное уширение штарковских компонент. Из спектров поперечных сечений поглощения при T=300 К рассчитаны силы осцилляторов переходов из основного состояния эрбия на возбужденные мультиплеты, методом Джадда-Офельта определены параметры интенсивности Ω_i : $\Omega_2 = 1.65 \cdot 10^{-20}$, $\Omega_4 = 0.56 \cdot 10^{-20}$, $\Omega_6 = 1.01 \cdot 10^{-20} \text{ см}^2$. С этими значениями параметров интенсивности рассчитаны вероятности излучательных переходов, коэффициенты ветвления и сделаны оценки вероятностей многофононных безызлучательных переходов

(МФБП) для NYF:Er . Изучена кинетика затухания люминесценции с излучательных уровней эрбия при их селективном возбуждении наносекундными лазерными импульсами и определены внутрицентровые времена жизни излучательных уровней эрбия при различных температурах.

В настоящей работе детально экспериментально и теоретически исследованы процессы самотушения люминесценции в кристаллах NYF:Er^{3+} при низкой (6 и 12 К) и комнатной температуре. Исследована кинетика люминесценции серии кристаллов $\text{Na}_{0.4}\text{Y}_{0.6(1-x)}\text{Er}_x\text{F}_{2.2}$ ($x=0.005-0.15$) при их селективном лазерном возбуждении наносекундными импульсами перестраиваемых лазеров. Рассмотрено концентрационное и температурное тушение люминесценции с излучательных уровней эрбия ${}^4\text{I}_j$ ($j = 13/2, 11/2, 9/2$), ${}^4\text{F}_{9/2}$, ${}^4\text{S}_{3/2}$, ${}^2\text{G}(\text{H})_{9/2}$ и ${}^4\text{G}_{11/2}$. В работе на основании спектроскопических характеристик внутрицентровых переходов, полученных в [55], методом модельного квантово-механического расчета получены микропараметры миграции и самотушения люминесценции ионов Er^{3+} , сделаны оценки скоростей самотушения излучательных уровней в концентрационной серии кристаллов.

Рассмотрение возможных схем тушения люминесценции с уровней ${}^4\text{I}_j$ ($j=11/2$ и $13/2$) показало, что люминесценция с этих уровней в кристаллах NYF:Er^{3+} концентрационно не тушится. Тушение люминесценции с уровней ${}^4\text{G}_{11/2}$, ${}^2\text{G}(\text{H})_{9/2}$, ${}^4\text{F}_j$ ($j=9/2, 7/2, 5/2$), ${}^4\text{S}_{3/2}$, и ${}^4\text{I}_{9/2}$ возможно как в результате МФБП, так и при безызлучательном переносе энергии по кросс-релаксационным схемам.

Теоретические оценки скоростей переноса энергии использованы для анализа кинетики затухания люминесценции с уровней ${}^4\text{G}_{11/2}$, ${}^2\text{G}(\text{H})_{9/2}$, ${}^4\text{F}_{9/2}$, ${}^4\text{S}_{3/2}$, и ${}^4\text{I}_{9/2}$ при их селективном возбуждении. Получено хорошее соответствие экспериментальных и расчетных кинетических кривых и скоростей самотушения в зависимости от концентрации эрбия.

Показано, что люминесценция с уровня ${}^4\text{S}_{3/2}$ значительно тушится с повышением температуры и с увеличением концентрации. Для люминесценции с уровней ${}^4\text{G}_{11/2}$, ${}^2\text{G}(\text{H})_{9/2}$ и ${}^4\text{I}_{9/2}$ характерно тушение за счет МФБП, поскольку они отделены от нижележащих уровней энергетическими зазорами $\Delta E_{ji} \leq 1900 \text{ см}^{-1}$. Так, при низких концентрациях эрбия (до 5 %) уровни ${}^4\text{G}_{11/2}$, ${}^2\text{G}(\text{H})_{9/2}$ и ${}^4\text{I}_{9/2}$ тушатся в значительной степени в результате МФБП, а при увеличении концентрации до 15% скорость самотушения становится сопоставимой со скоростью МФБП. Самотушение люминесценции с уровня ${}^4\text{F}_{9/2}$ незначительно, люминесценция с уровней ${}^4\text{F}_j$ ($j= 7/2, 5/2$) потушена за счет МФБП. Из кинетических кривых затухания люминесценции с уровня ${}^4\text{G}_{11/2}$ при УФ возбуждении сделано предположение о возможности up-конверсии с этого уровня при безызлучательном взаимодействии возбужденных уровней эрбия. Детально результаты исследования самотушения люминесценции в кристаллах NYF:Er изложены в статье [56].

Таким образом, в работе показано, что в кристаллах NYF:Er^{3+} процессы нелинейного тушения и up-конверсии играют существенную роль, особенно при высоких уровнях накачки. Конкуренция этих процессов при вариации концентрации эрбия и плотности мощности накачки позволяет управлять населенностью на возбужденных уровнях эрбия. Эти обстоятельства создают предпосылки для получения многоцветной генерации кристаллов NYF:Er^{3+} и позволяют рассматривать их в качестве потенциальных активных сред для лазеров видимого диапазона. В дальнейшем полученные данные будут использованы при исследовании эффективности заселения возбужденных уровней эрбия при непрерывной селективной накачке излучением лазерных диодов и оптимизации состава кристаллов для получения генерации в ИК и видимой области спектра.

По совокупности излучательных характеристик сделано заключение о перспективности использования кристаллов NYF:Er^{3+} в качестве активных сред твердотельных лазеров с накачкой лазерными диодами и перестройкой длины волны излучения.

5. Исследование процессов переноса энергии в кристаллах двойного фторида лития-иттрия $\text{LiYF}_4:\text{Er}$ при селективной накачке лазерными диодами

5.1. Исследование излучательных характеристик кристаллов $\text{YLF}:\text{Er}$

Спектроскопические параметры $\text{YLF}:\text{Er}$ лазерной среды были определены на основании результатов детального исследования спектров поглощения кристаллов $\text{YLF}:\text{Er}$, определения параметров интенсивности методом Джадда–Офельта, расчета сил осцилляторов переходов в поглощении [57], что позволило рассчитать силы осцилляторов межмультиплетных переходов в кристаллах $\text{YLF}:\text{Er}$.

На рис. 1 (а, б) и 2б (кривая 2) приведены фрагменты экспериментальных спектров поперечных сечений поглощения кристаллов $\text{YLF}:\text{Er}$ в областях, пригодных для лазерной диодной накачки, записанные нами при комнатной температуре.

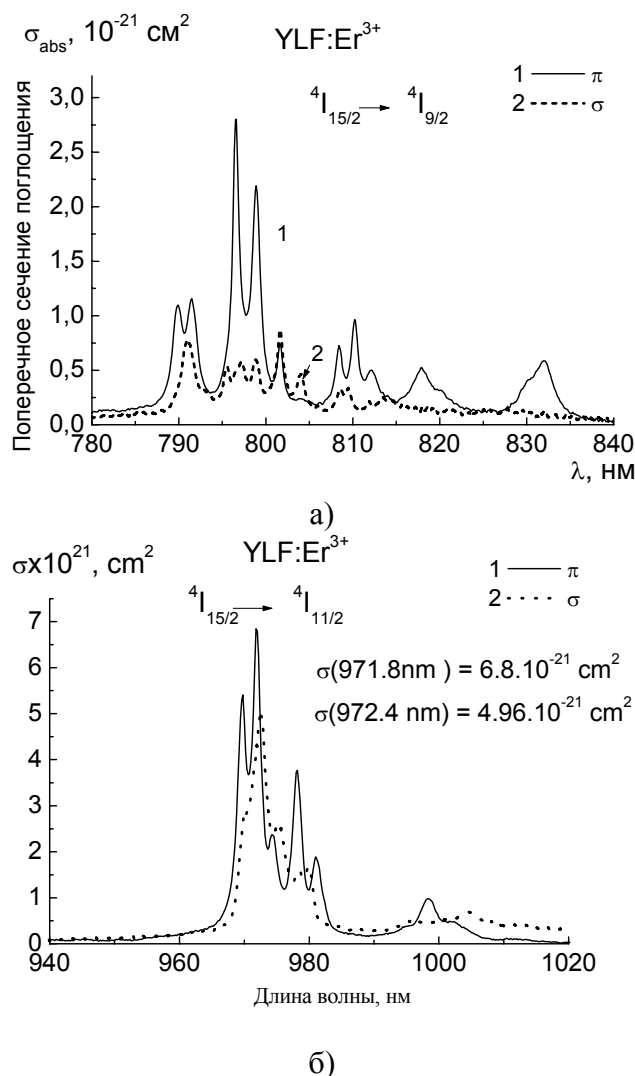
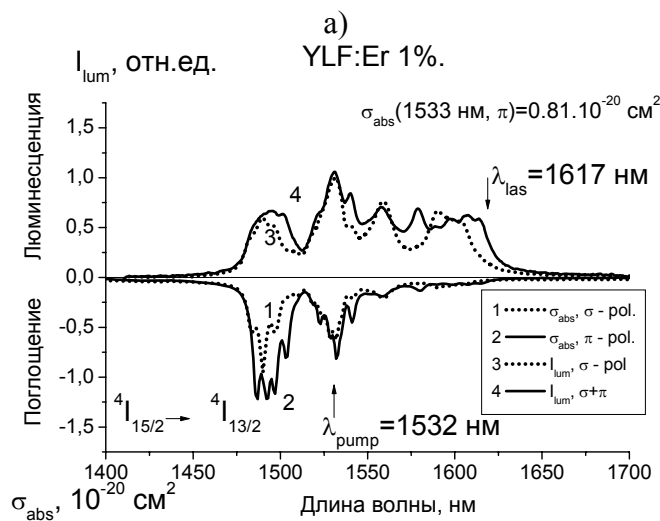
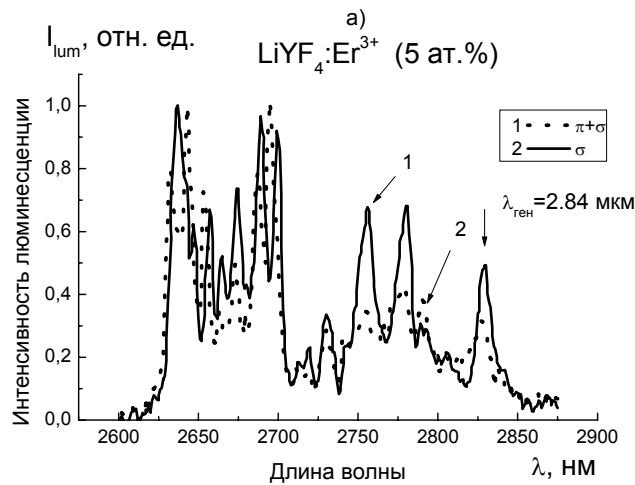


Рис. 1. Фрагменты спектров поглощения кристаллов $\text{YLF}:\text{Er}$ для π и σ -поляризаций в спектральной области, удобной для накачки LD, излучающими в области 800 и 975 нм

Нами исследованы излучательные характеристики кристаллов $\text{YLF}:\text{Er}$. Спектры люминесценции, соответствующие переходам ${}^4I_{11/2} \rightarrow {}^4I_{13/2}$ и ${}^4I_{13/2} \rightarrow {}^4I_{15/2}$, исследованы при непрерывной селективной накачке InGaAs лазерными диодами (LD), излучающими в области 972 мкм, и при накачке излучением лазера на $\text{Yb}:\text{Er}$ стекле, излучающего в области 1.53 мкм. На рис. 2а, 2б приведены фрагменты поляризованных спектров люминесценции кристаллов $\text{YLF}:\text{Er}$, соответствующие указанным переходам.



б)

Рис. 2. Оптические спектры кристаллов YLF:Er^{3+} . а) Люминесценция, переход ${}^4I_{11/2} \rightarrow {}^4I_{13/2}$, возбуждение LD 972 нм, поляризация σ (кривая 1) и $\pi+\sigma$ (кривая 2). б) Переход ${}^4I_{13/2} \rightarrow {}^4I_{15/2}$ в поглощении (внизу) и в излучении (кривые вверх), возбуждение LD 972 нм, поляризация σ - кривые 1, 3. π - кривая 2 и $\pi+\sigma$ (кривая 4)

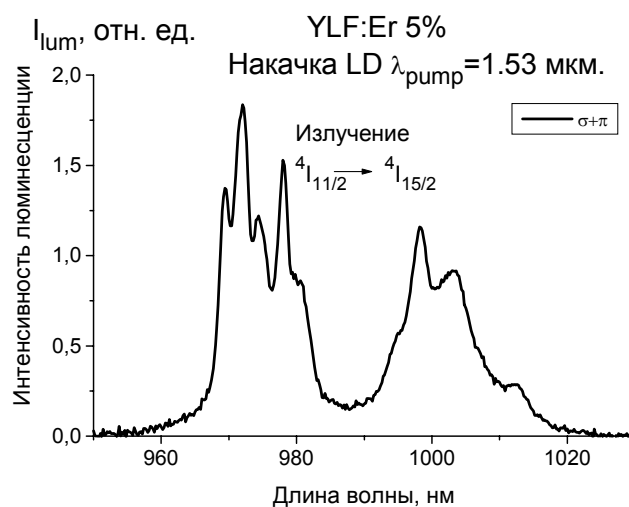


Рис. 3. Спектр люминесценции кристаллов YLF:Er^{3+} , переход ${}^4I_{11/2} \rightarrow {}^4I_{15/2}$ при ир-конверсионном возбуждении LD, излучающим в области 1.53 мкм (переход ${}^4I_{15/2} \rightarrow {}^4I_{13/2}$). $T=300^\circ\text{K}$

На рис. 2.6 показаны также σ - и π -поляризованные спектры поперечных сечений поглощения для перехода ${}^4I_{15/2} \rightarrow {}^4I_{13/2}$. Стрелками указаны длины волн накачки $\lambda_{\text{pump}} = 1.53$ мкм и генерации $\lambda_{\text{las}} = 1.617$ мкм, полученной на кристалле YLF:Er при накачке излучением лазера на Yb,Er стекле. Фрагмент спектра люминесценции кристаллов YLF:Er, соответствующий переходу ${}^4I_{11/2} \rightarrow {}^4I_{15/2}$, был записан при возбуждении излучением непрерывного волоконного лазера, излучающего в области 1.53 мкм (переход ${}^4I_{11/2} \rightarrow {}^4I_{15/2}$), он показан на рис. 3.

5.2. Концентрационные зависимости стационарных населенностей и динамика заселения уровней ${}^4I_{11/2}$ и ${}^4I_{13/2}$ эрбия в кристаллах YLF:Er при накачке CW излучением лазерных диодов

В активной среде возможность создания инверсной населенности рабочих уровней генерационных переходов во многом определяется скоростями процессов межуровневого взаимодействия, роль которых резко возрастает с ростом концентрации активатора. В частности, именно эти процессы ограничивают выходные лазерные характеристики, поскольку они определяют скорость заселения рабочих уровней лазерных переходов. С целью оптимизации состава активных сред на основе кристаллов YLF:Er нами проведено детальное исследование концентрационных зависимостей стационарных населенностей рабочих уровней ${}^4I_{11/2}$ и ${}^4I_{13/2}$ эрбия в условиях стационарной накачки.

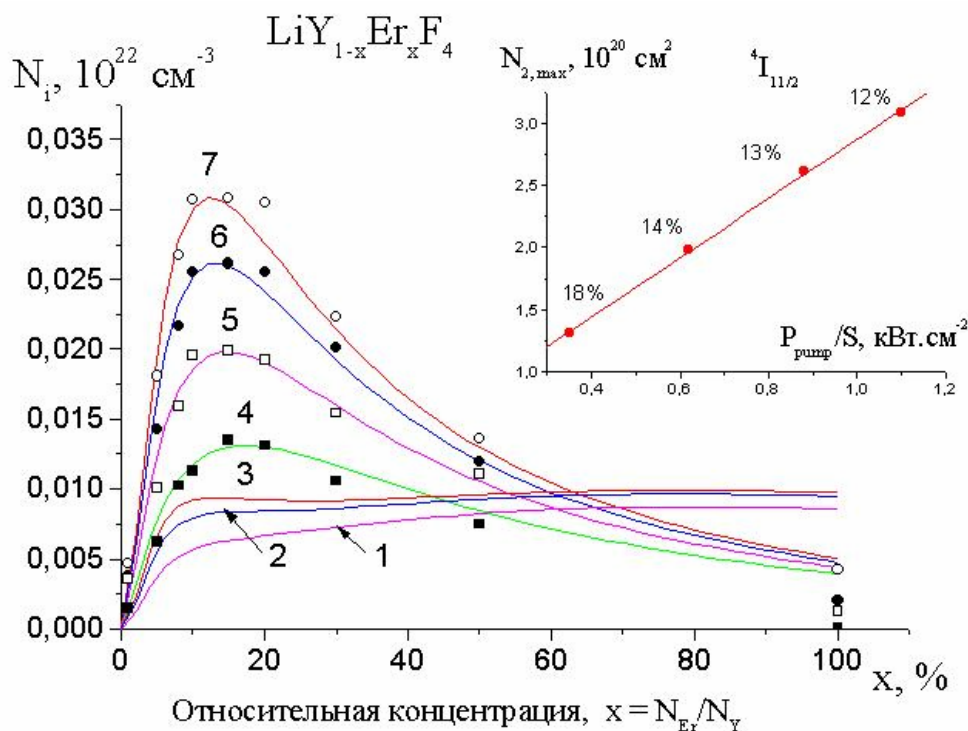


Рис. 4. Расчетные (сплошные кривые) и экспериментальные (точки) концентрационные зависимости стационарных населенностей $N_i = n_i N_Y$ рабочих уровней трехмикронного лазерного перехода ${}^4I_{11/2} \rightarrow {}^4I_{13/2}$ при непрерывной накачке излучением лазерных диодов в области 0.97 мкм. Кривые 1, 2, 3 – $N_i(x)$ для уровня ${}^4I_{13/2}$, кривые 4, 5, 6, 7 – $N_i(x)$ для уровня ${}^4I_{11/2}$. Плотность мощности накачки $P_{\text{pump}}/S = 350$ Вт.см $^{-2}$ (кривая 4), 620 Вт.см $^{-2}$ (кривые 1, 5), 880 Вт.см $^{-2}$ (кривые 2, 6) и 1100 Вт.см $^{-2}$ (кривые 3, 7). $T=300^\circ\text{K}$. На вставке – зависимость смещения максимума населенности ($N_{2,\text{max}}$) от плотности мощности накачки для уровня ${}^4I_{11/2}$

Экспериментально были исследованы поляризованные спектры люминесценции серии кристаллов $\text{LiY}_{1-x}\text{Er}_x\text{F}_4$ ($x=0.003-1$), соответствующие переходам с начального и

конечного уровней лазерного перехода ${}^4I_{11/2} \rightarrow {}^4I_{13/2}$ при селективной накачке. Относительные изменения населенностей исследуемых уровней контролировались по ИК спектрам стационарной люминесценции кристаллов в диапазонах длин волн, соответствующих переходам с уровня ${}^4I_{11/2}$: переходы ${}^4I_{11/2} \rightarrow {}^4I_{13/2}$ (2.7–2.8 мкм) и ${}^4I_{11/2} \rightarrow {}^4I_{15/2}$ (0.96–1.04 мкм), а также с уровня ${}^4I_{13/2}$: переход ${}^4I_{13/2} \rightarrow {}^4I_{15/2}$ (1.45–1.65 мкм). Селективное стационарное возбуждение уровня ${}^4I_{11/2}$ осуществлялось при накачке излучением $\lambda \sim 0.97$ мкм лазерных диодов. Результаты иллюстрируют кривые на рис. 4. На рис. 4 точками показаны концентрационные зависимости интенсивности люминесценции с уровней ${}^4I_{13/2}$ (кривые 1–3) и ${}^4I_{11/2}$ (кривые 4–6) при различных плотностях накачки в условиях стационарного, прямого, селективного LD возбуждения верхнего уровня ${}^4I_{11/2}$ лазерного трехмикронного перехода. Экспериментальные зависимости стационарной населенности уровней ${}^4I_{11/2}$ и ${}^4I_{13/2}$ от концентрации эрбия в кристаллах YLF:Er³⁺ при $x=0.003$ –1 получены при плотностях мощности накачки 350, 620, 880 и 1100 Вт/см², на рис. 4 они указаны точками. Сплошные кривые представляют собой расчетные зависимости стационарных населенностей $N_i = n_x N_Y$ рабочих уровней трехмикронного лазерного перехода ${}^4I_{11/2} \rightarrow {}^4I_{13/2}$, рассчитанные для условий непрерывной накачке излучением лазерных диодов в области 0.97 мкм.

Для выяснения динамики процессов заселения уровней эрбия при селективной ИК накачке была исследована теоретически и экспериментально кинетика люминесценции с возбужденных уровней эрбия при импульсной накачке уровня ${}^4I_{13/2}$ излучением 1.53 мкм моноимпульсного лазера.

Расчет проводился в рамках решения стационарных и кинетических уравнений баланса для шести нижайших уровней иона эрбия в кристаллах YLF с учетом наиболее вероятных излучательных и безызлучательных спонтанных переходов и процессов самотушения и up-конверсии. Все процессы, учтенные в расчетах, указаны на рис. 5 тонкими стрелками, α – процессы самотушения, γ – процессы up-конверсии. Значения констант переноса определены в рамках известных теорий переноса энергии по значениям микропараметров переноса [58], полученных методом модельного квантовомеханического расчета [59–61].

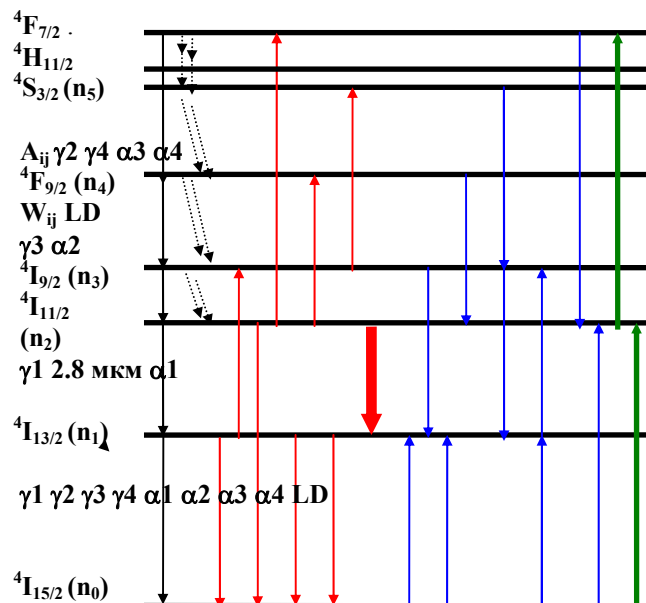


Рис. 5. Уровни энергии и схемы безызлучательного Er-Er взаимодействия в кристаллах YLF:Er³⁺. A_{ij} – вероятности излучательных переходов $i \rightarrow j$, W_{ij} – безызлучательная релаксация, α – константы самотушения, γ_i – константы up-конверсии.

Жирная стрелка – лазерный переход

Как видно из рис. 4, максимум интенсивности излучения с уровня ${}^4I_{11/2}$ лежит в области концентраций эрбия 8–15 %. Населенность уровня ${}^4I_{11/2}$ с увеличением концентрации и плотности мощности возрастает и достигает насыщения.

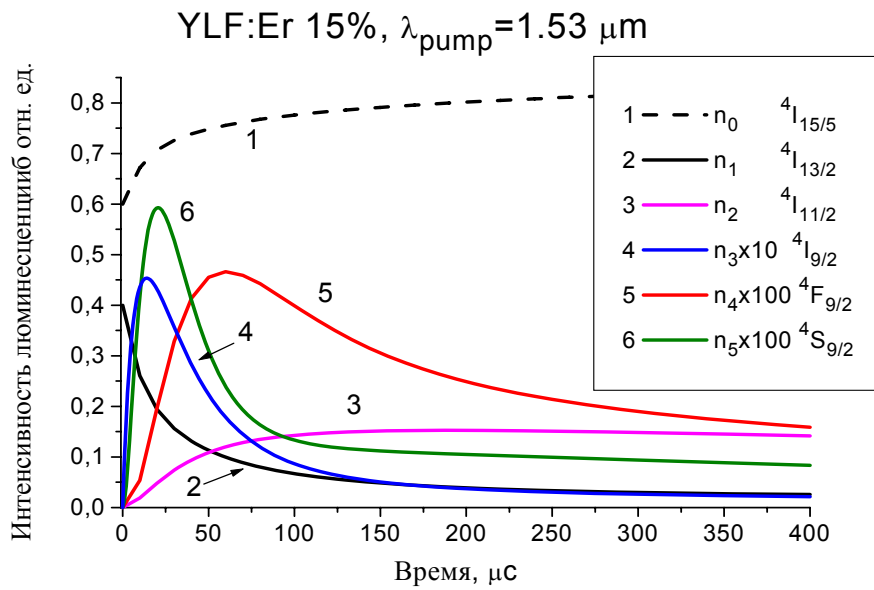


Рис. 6. Результаты расчета кинетических уравнений баланса для YLF:Er при ир-конверсионном возбуждении моноимпульсами 1.5 мкм лазера. Кривые 1–6 – относительная населенность уровней эрбия ${}^4I_{15/2}$ (основное состояние), ${}^4I_{15/2}$, ${}^4I_{13/2}$, ${}^4I_{11/2}$, ${}^4I_{9/2}$, ${}^4F_{9/2}$, ${}^4S_{3/2}$, соответственно

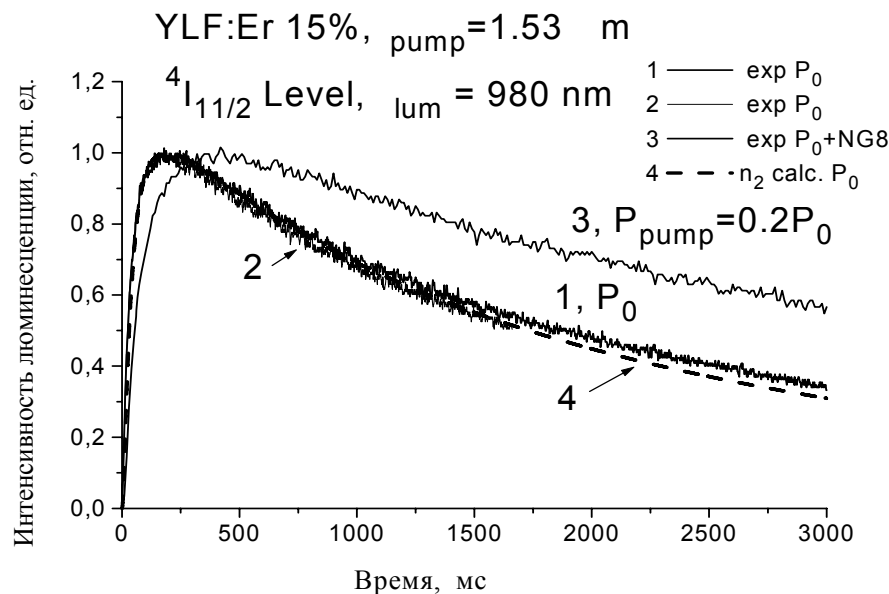


Рис. 7. Кинетика заселения кинетики заселения уровня ${}^4I_{11/2}$ при накачке моноимпульсами 1.5 мкм лазера уровня ${}^4I_{13/2}$. Кривые 1–3 – экспериментальные кинетики при различной плотности мощности накачки, кривая 4 – расчетная кривая населенности уровня ${}^4I_{13/2}$

Результаты расчета кинетических уравнений баланса для условий короткого импульса возбуждения (дельта-функция) уровня ${}^4I_{13/2}$ представлены на рис. 6. Расчет про-

веден с теми же самыми значениями микропараметров переноса, что и для схем безызлучательного переноса энергии, приведенных на рис. 5.

Как видно из расчетных кривых, при импульсной накачке возможно создание инверсной населенности между уровнями $^4S_{3/2}$ и $^4I_{9/2}$, что означает возможность получения генерации в области 1.7 мкм при up-конверсионной накачке лазерными диодами.

Экспериментальные зависимости интенсивности люминесценции с уровня $^4I_{11/2}$ в кристаллах YLF:Er (15%) при накачке, наносекундными импульсами 1.5 мкм лазера на Yb,Er стекле при различной плотности мощности накачки показаны на рис. 7.

Из сравнения расчетной кинетики заселения уровня $^4I_{11/2}$ с экспериментальными зависимостями следует, что одни и те же микропараметры переноса хорошо описывают процессы заселения возбужденных уровней эрбия в кристаллах YLF.

Выводы, сделанные на основании анализа расчетных и экспериментальных концентрационных зависимостей стационарных населенностей $N_i = n_i N_Y$ рабочих уровней 3-х микронного лазерного перехода $^4I_{11/2} \rightarrow ^4I_{13/2}$ (рис. 4), хорошо согласуются с экспериментальными данными по исследованию концентрационной зависимости эффективности непрерывной трехмикронной генерации кристаллов YLF:Er³⁺. Нами и в работе [19] получена максимальная эффективность генерации на переходе $^4I_{11/2} \rightarrow ^4I_{13/2}$ при непрерывной накачке уровня $^4I_{11/2}$ лазерными диодами на кристаллах YLF:Er³⁺ с концентрацией эрбия 15%. В работе [24] наибольшая эффективность генерации по наклону YLF:Er лазера $\eta_{sl} = 50\%$ получена также на кристаллах с концентрацией $N_{Er} = 15$ ат.% при накачке Ti - сапфировым лазером.

6. Заключение

Исследование спектроскопических характеристик кристаллов YLF:Er и NYF:Er показало, что непрерывную генерацию на самоограниченном переходе $^4I_{11/2} \rightarrow ^4I_{13/2}$ при накачке излучением лазерных диодов можно получить, если использовать для расселения конечного уровня лазерного перехода процессы безызлучательного переноса энергии – самогашения и up-конверсии. Скорость этих процессов увеличивается с ростом интенсивности накачки и концентрации эрбия по квадратичному закону. В результате населенность верхнего уровня $^4I_{11/2}$ увеличивается, а нижнего уровня $^4I_{13/2}$ – уменьшается, и при определенном соотношении плотности мощности накачки и концентрации активатора возможно достигнуть инверсию населенности на лазерном переходе $^4I_{11/2} \rightarrow ^4I_{13/2}$, несмотря на то, что время жизни конечного уровня в три раза больше времени жизни начального уровня лазерного перехода, т.е. на «самоограниченном» переходе. На основании полученных результатов сделан вывод о перспективности использования кристаллов двойных фторидов, активированных эрбием, для получения многоцветной генерации и генерации в области 2.8 мкм, что позволяют рассматривать их в качестве потенциальных активных сред для лазеров видимого и ИК диапазона.

Работа выполнена при поддержке Минобрнауки России, ЛОТ-21, проект 9.

Литература

1. Pollack S.A. // Proc. IEEE. 1963. V. 51. P. 1793
2. Kintz G.J., Allen R., Esterowitz L. // Techn. Digest of CLEO'95. 1995. P. 215. WE5-1
3. Chicklis E.P., Naiman C.S., Linz A. // Digest of Technical Papers VII IQEC, Montreal. 1972. P. 17.
4. Morosov A.M., Podkolzina I.G., Tkachuk A.M. et al. // Opt. Spektr. 1975. Vol. 39. P. 65.
5. Petrov M.V., Tkachuk A.M. // Opt. Spektr. 1978. Vol. 45. P. 147–155.

6. Korableva S.L., Livanova L.D., Petrov M.V., Tkachuk A.M. // *J. Techn. Physiki*. 1981. Vol. 51. P. 2572–2575
7. Petrov M.V., Tkachuk A.M., Feofilov P.P. // *Isv. Acad. Nauk SSSR. Ser. Phys.* 1981. V. 45 (3). P. 654–658
8. Kaminskii A.A., Sarkisov S.E., Seiranjan K.B. et al. // *Isv. Acad. Nauk SSSR. Ser. Neorg. Mater.* 1982. V. 18. P. 527.
9. Tkachuk A.M., Petrov M.V., Chilko A.V. // *Spektroskopija Crystallov*. L. Nayka. 1983. P. 106–123
10. Tkachuk A.M., Petrov M.V., Korableva S.L., Podkolzina I.G. // *Isv. Acad. Nauk SSSR, ser. phys.* 1988. Vol. 52(3). PP. 537-541
11. Hubert S., Meichenin D., Zhou B.W., Auzel F. // *J. of Luminescence*. 1991. Vol. 50. P. 7–15.
12. A.M.Tkachuk, V.V.Shumilin and V.V.Lazo // *Proc. 4th Int. Conf. on Laser Applications in Life Sciences*. SPIE. 1992. V. 1922. P. 263-271.
13. A.M. Tkachuk, V.V.Shumilin, V.V.Lazo, N.N.Smironov, V.F.Danilichev, A.F.Gatzu, D.V.Ganin. // *Proc. 4th Int. Conf. on Laser Applications in Life Sciences*, SPIE. 1992. Vol. 1922. P. 255-262
14. Kintz G.J., Allen R., Esterowitz L. // *Appl. Phys. Lett.* 1987. Vol. 50(22). P. 1553–1555
15. Auzel F., Hubert S., Meichenin D. // *Appl. Phys. Lett.* 1989. Vol. 54(8). P. 681–683
16. Stoneman R.C. Lynn J.G., Esterowitz L. // *IEEE J. Quantum Electron.* 1992. Vol. 28(4). P. 1041–1045
17. Schmaul B., Huber G., Clausen R. et al. // *Appl. Phys. Lett.* 1993. V. 62(6). P. 541-543.
18. Danger T., Koetke J., Brede R. et al. // *J. Appl. Phys.* 1994. V. 76(3). P. 1412-1422.
19. Jensen T., Dening A., Huber G., and Chai B.H.T. // *Optics Lett.* 1996. V. 21. P. 585–587.
20. Pollnau M., Luthy W., Weber H.P., Jensen T., Huber G., Cassanho A., Jenssen H.P., McFarline R.A. // *Opt. Lett.* 1996. Vol. 21. №1. P. 48–50.
21. Pollnau M., Spring R., Ghisler et. al. // *IEEE QE.* 1996. V. 32(4). P. 657–662.
22. Pollnau M., Spring R., Wittwer S. et al. // *J. Opt. Soc. Am. B* 1997. V. 14(4). P. 974–978.
23. Wyss Chr., Luthy W., Weber H.P. et al. // *SPIE. Biomedical System and Technologies II*. 1997. V. 3199. P. 206–214.
24. Wyss Chr., Luthy W., Weber H.P., Rogin P., Hulliger J. // *Opt. Commun.* 1997. V. 139. P. 215–218.
25. Wyss Chr., Luthy W., Weber Heinz P. // *IEEE QE* 1998. V. 34(6). P. 1041–1045.
26. Voss H., Massmann F. // *Techn. Digest CLEO'97*, 1997. P.74. CTuE-2
27. Mober P.E.-A., Heumann E., Huber G., Chai B.H.T. // *Appl. Phys. Lett.* 1998. V. 72(2). P. 139.
28. Voss H., Massmann F. // *OSA TOPS ASSL* 1997. V.10. P.217.
29. Pollack S.A., Chang D.B. // *J. Appl. Phys.* 1988. V. 64(6). P. 2885–2893
30. Lenth W., Silversmith, Macfarline R.M. // *Advances in Laser Spectroscopy*. 1988. V. 5. P. 5–12.
31. S.A., Chang D.B., Birnbaum M. // *Appl. Phys. Lett.* 1989. V. 54(6). P. 869–871
32. Tong F., Risk W.P., Macfarline R.M., Lenth W. // *Electronic Letters*. 1989. V. 25(20). P.1389–1391.
33. Macfarline R.M. // *Appl. Phys. Lett.* 1989. V. 54(23). P. 2301–2302.
34. Hebert T., Wannemacher R., Lenth W., Macfarline R.M. // *Appl. Phys. Lett.* 1990. V. 57(17). P. 1727–1729.
35. Hebert T., Risk W.P., Macfarline R.M., Lenth W., Jenssen J.P., Dube G. // *Proc. ASSL, OSA*. 1990. P. 379.
36. McFarline R.A., Robinson M., Pollack S.A. // *Proc. SPIE*. 1990. V. 1223. P. 294.
37. Macfarline R.M. // *Optics Lett.* 1991. V. 16(18). PP. 1397–1399.
38. Knowles D.S., Jenssen J.P. // *IEEE QE.* 1992. V. 28(4). PP. 1197–1208.

39. Macfarlane R.M., Whittaker E.A., Lenth W. // *Electronic Letters*. 1992. V. 28. P. 2136.
40. Xie P., Rand S.C. // *Optics Lett.* 1992. V. 17(17). P. 1198.
41. Xie P., Rand S.C. // *Appl. Phys. Lett.* 1990. V. 57(12). P. 1182.
42. Stephens R.R., Macfarlane R.M. // *Optics Lett.* 1993. V. 18. P. 34.
43. Xie P., Rand S.C. // *Appl. Phys. Lett.* 1993. V. 63. P. 3125.
44. Xie P., Rand S.C. // *J. Opt. Soc. Am. B*. 1994. V. 11(5). P. 901.
45. Brede R., Heumann E., Koetke J., Danger T., Huber G., Chai B. // *Appl. Phys. Lett.* 1993. V. 63(15). P. 2030–2031.
46. Heine F. // *Appl. Phys. Lett.* 1994. V. 65. P. 383.
47. Heine F., Heumann E., Mobert P., Huber G., Chai B. // *Techn. Digest of ASSL*. 1995. P. WD2.
48. Pinto J.F., Rosenblatt J.H., Esterowitz L. // *Electron. Lett.* 1994. V. 30. P. 1596.
49. Auzel F., Chen Y. // *J. Lumin.* 1995. V. 65. P.45.
50. Auzel F. // *Chem. Rev.* 2004. V. 104. P. 139–173.
51. Багдасаров Х.С., Каминский А.А., Соболев Б.П. // *Кристаллография*. 1969. Т.13. С. 779.
52. Заморянская М.В., Петрова М.А., Семенова Т.С. // *Неорганические материалы*. 1998. Т. 34б. № 6. С.752–757.
53. Заморянская М.В., Морозова Л.Г., Полетимова А.В. и др.// *ЖПС*. 1991. Т. 55. №6. С. 1010–1013.
54. Ткачук А.М., Полетимова А.В., Петрова М.А. и др. // *Опт. и спектр.* 1991. Т. 70, вып. 6. С. 1230–1235.
55. Ткачук А.М., Иванова С.Э., Joubert., М.-F., Guyot Y.// *Опт. и спектр.* 2004, Т. 97. № 2. С. 266–285.
56. Ткачук А.М., Иванова С.Э., Joubert., М.-F., Guyot Y. // *Опт. и спектр.* 2005. Т. 99. № 6. С. 969–986.
57. Ткачук А. М., Полетимова А.В., Петров М.В. // *Опт. и спектр.* 1985. Т. 59. № 5. С. 1136–1139.
58. Ткачук А.М., Клокишнер С.И., Петров М.В. // *Опт. и спектр.* 1985. Т. 59. № 4. С. 802–811.
59. Ткачук А.М., Клокишнер С.И. // *Опт. и спектр.* 1986. Т. 61. № 1. С. 84–90
60. Ткачук А.М., Клокишнер С.И. // *Опт. и спектр.* 1990. Т. 68. №. 4. С. 745–752.
61. Ткачук А.М. // *Опт. и спектр.* 1990. Т. 68. №. 6. С. 1324–1336.

ОСНОВЫ ТЕХНОЛОГИИ ГОЛОГРАФИЧЕСКОЙ СУБМИКРОННОЙ ФОТОЛИТОГРАФИИ

С.Н. Корешев, В.П. Ратушный

Созданы основы принципиально новой технологии проекционной субмикронной голографической фотолитографии. Разработана схема записи и восстановления голограмм-проекторов предельного разрешения, используемых без дополнительного объектива; сформулированы и научно обоснованы требования, предъявляемые к ее геометрическим параметрам и режимам экспонирования голограмм; экспериментально показана пригодность применения этой схемы в задачах проекционной литографии.

Введение

Современные тенденции развития дисплеев направлены не столько на уменьшение размера единичного пикселя изображения, сколько на увеличение общей площади устройств. Поэтому в сложившейся ситуации особо актуальной становится задача разработки методов высокоразрешающей проекционной фотолитографии, обеспечивающих одномоментное экспонирование относительно большого участка заготовки дисплея. Традиционный метод проекционной фотолитографии, основанный на использовании дорогостоящих проекционных объективов, обладает существенными ограничениями по площади одновременно экспонируемого участка фоторезиста, обусловленными невозможностью полной коррекции полевых аберраций проекционных оптических систем. При этом с уменьшением длины волны используемого излучения и, следовательно, с уменьшением характеристического размера отображаемых структур наблюдается устойчивая тенденция к уменьшению безаберрационного участка поля проекционных объективов. Так, экспонирующие устройства, использующие излучение с длиной волны, близкой к i -линии ртути, обладают полями диаметром в 10–15 мм. Устройства же, предназначенные для работы в излучении эксимерных лазеров с длинами волн 193 нм и 157 нм, обладают полями, не превышающими единиц миллиметров. Например, в [1] описаны объективы для проекционной фотолитографии на длинах волн 193 нм и 157 нм, обладающие двухмиллиметровыми полями.

Еще хуже будет складываться ситуация в более коротковолновых областях спектра. Там придется полностью перейти на зеркальные оптические системы, обладающие большими, по сравнению с обычно используемой в настоящее время рефрактивной оптикой, полевыми аберрациями. Производители микросхем компенсируют недостаточность мгновенных полей современных проекционных объективов путем использования сканирующего и пошагового экспонирования. При этом, как нетрудно подсчитать из приведенных данных, изготовление, например, процессора размером 50×50 мм² потребует выполнения более 600 операций экспонирования и совмещения рисунка. Здесь необходимо отметить, что точность совмещения рисунка при пошаговом экспонировании не должна быть хуже 0.1 характеристического размера отображаемой структуры. Поскольку операция совмещения в традиционном фотолитографическом оборудовании сложна и трудоемка, то возникает необходимость его замены альтернативными «большепольными» системами проекционной фотолитографии. При этом наиболее перспективной основой для разработки таких систем нам представляется метод голографии, характеризующийся возможностью формирования полностью свободных от аберраций, в том числе и полевых аберраций, действительных изображений зарегистрированного объекта.

Возможность использования принципов голографии при решении задач фотолитографии уже была рассмотрена рядом авторов на относительно ранних этапах развития голографии [2–4]. Однако результатов, пригодных для практического использования, в то время получено не было. На наш взгляд, основная причина этого состояла в использовании регистрирующих сред, наносимых из раствора и требующих «мокрой»

фотохимической обработки, приводящей к набуханию и последующей усадке среды. Сложившаяся ситуация изменилась в начале 90-х годов, когда специалисты швейцарской фирмы Holtronic Technologies решили указанные проблемы с помощью новой регистрирующей среды, а именно – фотополимерного материала компании Du Pont [5, 6]. При этом для реализации голографического варианта фотолитографического процесса ими была выбрана схема записи и восстановления голограмм «полного внутреннего отражения», предложенная Стетсоном [7]. Следует отметить, что в этой схеме используются как минимум два элемента, которые должны быть прозрачны для излучения рабочей длины волны – стеклянная призма и сам фотополимерный слой. Поэтому технические решения, разработанные швейцарской компанией Holtronic Technologies, пригодны для использования и используются лишь в относительно длинноволновом диапазоне УФ излучения на длине волны 364 нм, соответствующей УФ линии аргонового лазера. Тем не менее, работы специалистов Holtronic Technologies – первый пример успешного коммерческого применения методов голографии в высокоразрешающем фотолитографическом процессе.

Постановка задачи

С нашей точки зрения, для применения в коротковолновой голографической фотолитографии более перспективными являются регистрирующие среды на основе халькогенидного стеклообразного полупроводника (ХСП) [8]. Эти среды наносятся на подложки путем испарения в вакууме и позволяют получать высококачественные отражательные рельефно-фазовые голограммные оптические элементы [9]. Принципиальная возможность использования этих сред для получения рельефно-фазовых отражательных голограмм-проекторов была убедительно продемонстрирована нами в ходе НИР, поставленных и проведенных в ГОИ им. С.И. Вавилова в 2000–2005 гг. [10–13]. В настоящей статье представлены результаты следующего этапа работы, направленного на внедрение полученных нами ранее результатов и «know-how» в реальный, пусть даже и не самый прецизионный, фотолитографический процесс. Эта работа проводилась применительно к схеме проекционной голографической фотолитографии, реализуемой без дополнительного проекционного объектива [10]. В ее рамках были созданы основы принципиально новой технологии проекционной субмикронной голографической фотолитографии. Среди них – разработка схемы записи и восстановления голограмм-проекторов, характеризующейся минимальными требованиями к временной когерентности источника излучения, определение, обоснование и экспериментальная проверка требований, предъявляемых к режимам экспонирования голограмм, и демонстрация пригодности применения голограмм-проекторов в задачах проекционной фотолитографии.

Анализ требований, предъявляемых к временной когерентности источников излучения, используемых в схемах записи и восстановления прецизионных голограмм-проекторов высокого разрешения

Из анализа литературы следует, что, во-первых, малая степень временной когерентности излучения используемых для целей фотолитографии УФ лазеров, не превышающая 30 см даже в случае использования узкополосных спектральных фильтров с полосой пропускания 1 пм, приводит, в общем случае, к потерям разрешения, реализуемого в изображениях, восстанавливаемых методом голографии. Причем эти потери, в условиях недостаточной временной когерентности излучения, будут обусловлены как процессом записи, так и процессом восстановления голограммы. Во-вторых, практически все опубликованные в литературе технические решения, направленные на сниже-

ние требований к временной когерентности излучения, непригодны для целей голографической фотолитографии. Все вышеперечисленное предопределило необходимость определения и обоснования требований, предъявляемых к временной когерентности источников излучения, используемых в схемах записи и восстановления прецизионных голограмм-проекторов высокого разрешения, и разработки самой схемы, обеспечивающей достижение предельного разрешения в условиях малой временной когерентности.

При определении требований, предъявляемых к временной когерентности источников излучения, мы исходили из приведенной на рис. 1 обобщенной схемы записи голограммы точечного источника с параллельным опорным пучком. На этом рисунке X – ось координат, лежащая в плоскости регистрации голограммы. Отметим, что начало координат в данном случае совмещено с центром линейной апертуры голограмм. Буквами O и R обозначены расходящийся гомоцентрический объектный и параллельный опорный пучки лучей. α_r – угол падения опорного пучка лучей на плоскость регистрации голограммы, а α_o – угол, который составляет с осью Z прямая, соединяющая центр апертуры голограммы с источником объектной волны.

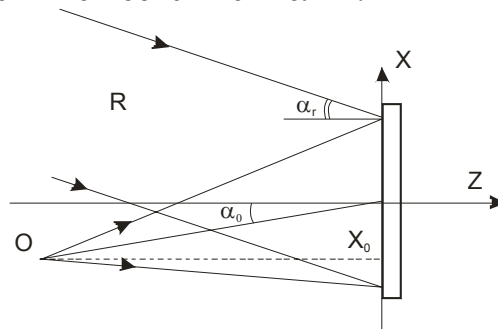


Рис.1. Обобщенная схема регистрации голограмм

С целью упрощения выкладок ограничимся рамками теории aberrаций голограмм третьего порядка и рассмотрением случая записи одномерной голограммы. Тогда, исходя из [13], напишем следующие выражения, описывающие распределения фаз объектной и опорной волн в плоскости апертуры голограммы:

$$\varphi_o(x) \approx \frac{2\pi}{\lambda} \left\{ \frac{x^2}{2R_o} - x \sin \alpha_o - \frac{1}{8R_o^3} [x^4 - 4x^3 x_o + 4x^2 x_o^2] \right\}, \quad (1)$$

$$\varphi_r(x) \approx \frac{2\pi}{\lambda} \{-x \sin \alpha_r\}, \quad (2)$$

где φ_o и φ_r – распределения фаз в плоскости голограммы объектной и опорной волн соответственно. При этом разность указанных распределений $\Delta\varphi(x)$ будет описываться следующим выражением:

$$\Delta\varphi(x) \approx \frac{2\pi}{\lambda} \left\{ \frac{x^2}{2R_o} - x(\sin \alpha_o - \sin \alpha_r) - \frac{1}{8R_o^3} [x^4 - 4x^3 x_o + 4x^2 x_o^2] \right\}. \quad (3)$$

Из (3) найдем элементарное приращение разности фаз, соответствующее малому изменению длины волны источника излучения $d\lambda$:

$$d\Delta\varphi = \frac{\partial \Delta\varphi}{\partial \lambda} d\lambda \approx -\frac{2\pi}{\lambda^2} d\lambda \left\{ \frac{x^2}{2R_o} - x(\sin \alpha_o - \sin \alpha_r) - \frac{1}{8R_o^3} [x^4 - 4x^3 x_o + 4x^2 x_o^2] \right\}. \quad (4)$$

Отсюда получим выражение, описывающее элементарное приращение разности длин оптических путей dl , соответствующее малому изменению длины волны источника излучения $d\lambda$.

$$dl \approx -\frac{d\lambda}{\lambda} \left\{ \frac{x^2}{2R_o} - x(\sin \alpha_o - \sin \alpha_r) - \frac{1}{8R_o^3} [x^4 - 4x^3x_o + 4x^2x_o^2] \right\}. \quad (5)$$

Если в (5) подставить $x = \pm D/2$, где D – размер линейной апертуры голограммы, и положить $x_o = x_{o \max}$, где $x_{o \max}$ – координата точки объекта, наиболее удаленной от оси Z , то его можно будет использовать для определения минимально необходимой длины когерентности источника излучения, предназначенного для формирования голографического поля. Из (5) следует, что требования к длине когерентности источника могут быть существенно снижены при расположении объекта вблизи оси Z , т.е. при $x_o \approx 0$ и, следовательно, при $\sin \alpha_o \approx 0$, а также в случае одноосевой схемы записи голограмм, т.е. при $\sin \alpha_o = \sin \alpha_r$. При указанных выше условиях выражение (5) может быть приведено к виду:

$$dl_n \approx \frac{d\lambda}{2\lambda} R_o \text{tg}^2 \omega \{1 - 0,25 \text{tg}^2 \omega\}, \quad (6)$$

где: dl_n – требуемая длина когерентности источника излучения, а ω – угловая апертура объектного пучка. Отметим, что, помимо одноосевой схемы, выражение (6) будет справедливо и во внеосевом случае, при условии предварительного внесения в опорный пучок разности хода лучей, равной $x \sin \alpha_r$. Такая разность хода лучей может быть обеспечена путем установки в опорный пучок, до его падения на плоскость регистрации голограммы, плоской дифракционной решетки с пространственной частотой, равной несущей пространственной частоте регистрируемой голограммы. При этом приведенная на рис. 1 и рассматриваемая нами обобщенная схема записи голограммы должна быть приведена к виду, представленному на рис. 2.

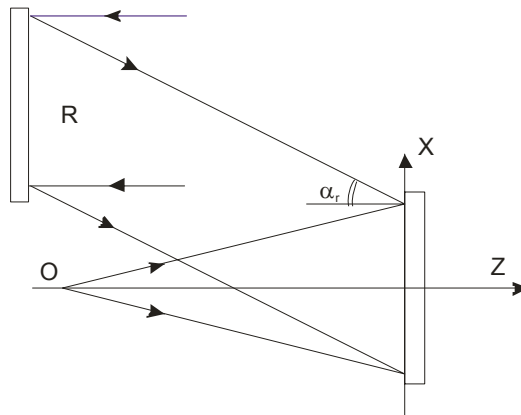


Рис. 2. Обобщенная схема регистрации голограммы с решеткой в опорном пучке

Из этого рисунка видно, что требуемая для данной схемы длина когерентности источника излучения определяется лишь числовой апертурой объектного пучка лучей и не зависит от угла падения опорной волны. Здесь необходимо отметить, что бесспорным условием практической реализации указанной схемы является выравнивание оптических длин пути света, приходящего в центр голограммы вдоль опорного и объектного плеч интерферометра. Напомним, что прием использования решеток на стадии восстановления голограмм для ахроматизации восстановленного изображения не нов и описан многими авторами, например [14]. В то же время, работы, в которых бы упоминалась возможность использования решеток для ахроматизации голографического поля на стадии записи голограмм, нам неизвестны.

Приведенное выше выражение (6) может быть использовано не только для определения требуемой длины когерентности излучения, но также и для определения максимально допустимого расстояния от объекта до плоскости записи голограммы, обеспечивающего при заданной спектральной ширине источника излучения достижение

заданного контраста интерференционных полос на всей линейной апертуре голограммы. Так, если положить допустимым для края апертуры голограммы сбой фазы интерференционных полос, соответствующих длинам волн $\lambda + \Delta\lambda$ и $\lambda - \Delta\lambda$, на 0,5 периода их следования, т.е. положить dI_n равным $0,25\lambda$, то из (6) получим:

$$R_0 \leq \frac{\lambda^2}{2d\lambda(1 - 0,25tg^2 \omega)tg^2 \omega} . \quad (7)$$

Напомним, что ранее нами было получено и опубликовано в [15] неравенство (8), обеспечивающее на стадии восстановления голограммы требуемую, т.е., лежащую в пределах 10% допуска, точность отображения элементов структуры объекта при использовании восстанавливающего излучения с конечной, равной $d\lambda$, спектральной шириной

$$R_0 \leq 0,68 \frac{\lambda^2 \left[1 - \frac{1}{2} \sin^2 \omega \right]^3}{d\lambda \sin^4 \omega} . \quad (8)$$

Из сравнения (7) и (8) можно сделать вывод о том, что требование сохранения контраста, описываемое выражением (7), обуславливает несколько более жесткие ограничения величины параметра R_0 . Проиллюстрируем это с помощью таблицы 1, содержащей результаты расчета максимально допустимой величины R_0 , выполненного с помощью выражений (7) - R_0 и (8) - R_0^* .

Таблица 1.

$\Delta\lambda$ (нм)	$\lambda = 488$ нм		$\lambda = 157$ нм	
	R_0 (мм)	R_0^* (мм)	R_0 (мм)	R_0^* (мм)
0,06	4	12	0,4	1,2
0,001	250	700	26	70

Необходимо отметить, что существует еще и третье ограничение, накладываемое на величину R_0 . Оно связано с необходимостью согласования линейной апертуры записываемой голограммы с угловой апертурой излучения, рассеиваемого объектом, т.е. с необходимостью записи на голограмме всего дифракционного поля объекта. Это ограничение играет решающую роль, т.е. становится определяющим, лишь в случаях большой временной когерентности источников излучения и малой площади записываемой голограммы.

Схема записи и восстановления голограмм-проекторов, обеспечивающая минимальные требования к длине когерентности используемых источников излучения

Исходя из сформулированных выше требований о минимальности разности хода интерферирующих пучков лучей и описанных нами ранее в работах [10–12] условий формирования безабберационного изображения, в ходе выполнения настоящей работы была разработана практическая схема записи и восстановления голограмм-проекторов предельного разрешения, содержащая плоскую дифракционную решетку в плечах формирования опорного и восстанавливающего пучков лучей. Она представлена на рис. 3. Поясним реализуемый в ней ход лучей. Излучение лазера 1 направляется на зеркало 2 и падает на частично прозрачное зеркало 3. Здесь оно разделяется на два пучка лучей – опорный пучок и объектный пучок. Опорный пучок отражается от зеркала 3, расширяется с помощью коллиматора 4, отражается от зеркал 5, 11 и направляется на плоскую дифракционную решетку 12. Излучение, дифрагирующее на решетке 12 в +1-ый порядок дифракции, направляется на плоскость регистрации голограммы – про-

ектора 10. Отметим, что излучение, дифрагирующее в сопряженный -1 порядок дифракции, на стадии записи голограмм не используется, а применяется лишь на стадии их восстановления в качестве параллельного восстанавливающего пучка лучей. Объектный пучок лучей проходит через зеркало 3, отражается от зеркала 6, расширяется с помощью коллиматора 7 и направляется на объект 9, представляющий собой бинарный амплитудный транспарант. Восстановление голограмм в данной схеме осуществляется излучением, дифрагирующим в -1 порядок дифракции на решетке 12 при перемещении стола 8, несущего на себе оправу объекта и кассету с голограммой – проектором, в положение 13. Применение решетки в данной схеме позволяет не только выровнять для всей поверхности регистрируемой голограммы оптические длины опорного и объектного плеч интерферометра, формирующего голографическое поле, но и существенно облегчает процесс восстановления работающих на отражение голограмм. Это осуществляется за счет использования при записи и восстановлении голограмм сопряженных порядков дифракции опорного излучения на решетке.

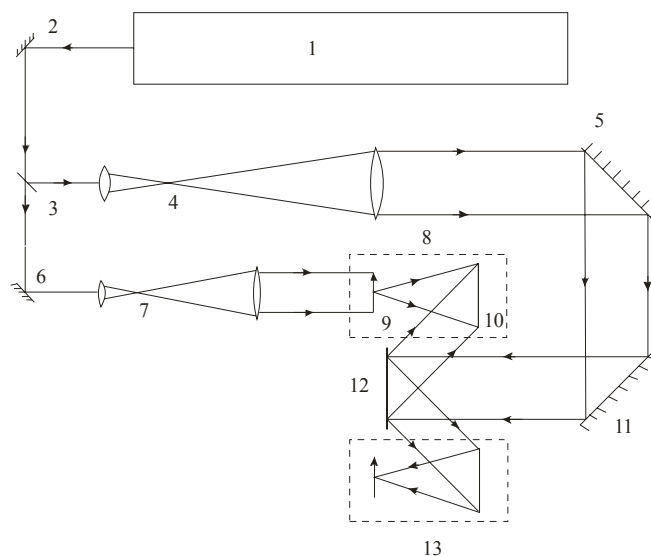


Рис.3. Схема записи и восстановления голограмм - проекторов

Экспериментальная проверка пригодности схемы и полученной с ее помощью голограммы–проектора к использованию в реальном литографическом процессе

Экспериментальная проверка пригодности схемы и полученной с ее помощью голограммы-проектора к использованию в реальном литографическом процессе проводилась нами в ходе изготовления четырехэлементного фотоприемника. диаметром 15 мм с минимальным шагом проектирования 1,3 мкм методом субмикронной фотолитографии, реализуемой на длине волны 488 нм с помощью отражательных голограмм, регистрируемых на слоях халькогенидного полупроводника [8]. Геометрические параметры схемы были выбраны равными: размер голограммы – 68×68 мм²; расстояние от голограмм до объекта 30 мм; углы падения параллельных опорного и восстанавливающего пучков на плоскость регистрации голограммы -70° (точность равенства модулей этих углов друг другу – 10,6"); минимальное расстояние от опорного и объектного точечных источников до центра голограммы – 1,1 км. Последнее требование обеспечивает точную фокусировку восстановленного изображения в той же плоскости, в которой находился объект при записи голограммы. Выбор угла падения опорного и восстанавливающего пучков лучей на голограмму определялся двумя факторами: во-первых, необходимостью обеспечения отсутствия виньетирования указанных пучков лучей оправой фотошаблона; во-вторых, необходимостью пространственного разделения восстанов-

ленного изображения фотошаблона и интермодуляционных помех голограммы, имеющих спектр пространственных частот, в два раза превышающий спектр объекта [16].

Регистрация голограмм по указанной схеме, как указывалось выше, осуществлялась на слоях халькогенидного стеклообразного полупроводника, нанесенного на высококачественные плоскопараллельные подложки, изготовленные из оптического стекла. Плоскостность рабочих поверхностей подложек лежала в пределах $N = 0,1$ и $\Delta N = 0,1$. Нанесение халькогенида на подложки осуществлялось методом распыления в вакууме. Толщина сформированных светочувствительных пленок равнялась 1 мкм. Их постэкспозиционное травление осуществлялось в жидкой среде. Величина оптимальной экспозиции и соотношение интенсивностей интерферирующих пучков при записи голограмм-проекторов подбирались исходя из условия отсутствия помех нелинейности в восстановленных изображениях. Для наших условий оптимальные величины экспозиции и соотношения интенсивностей опорного и объектного пучков лучей равнялись 10 Дж/см^2 и 6:1, соответственно. Регистрация действительного изображения фотошаблона, восстанавливаемого голограммой-проектором, осуществлялось напрямую, без дополнительных оптических систем, на слое позитивного фоторезиста Шипли AZ1350, толщиной 0,2–0,5 мкм, нанесенные поверх слоя хрома, осажденного на тонкие, толщиной 2 мм, стеклянные подложки. Проявление экспонированных слоев фоторезиста осуществлялось в слабом растворе едкой щелочи (KOH 2,5%). Травление пленки хрома производилось до стекла через полученную резистивную маску в растворе четырехводного сернокислого церия. Проведенные исследования показали, что наилучшее качество хромовых ламелей, составляющих структуру разводки изготавливаемого фотоэлемента, обеспечивается при экспозиционной дозе фоторезиста Шипли 40 Дж/см^2 и временах проявления около 1 минуты.

На рис. 4 представлены полученные с различным увеличением фотографии ламелей фотоэлемента, полученных методом голографической фотолитографии.



Рис.4. Изображения структур, полученных методом проекционной голографической фотолитографии

Отметим, что видимые на фотографиях расфокусированные серые пятна обусловлены недостатками оптической системы микроскопа, использовавшегося на стадии исследования качества полученных структур. Представленные фотографии свидетельствуют о достигнутом беспыльном резком восстановлении во всем поле зрения голограммы изображения структуры четырехкомпонентного фотоприемника, т.е. о пригодности к практическому использованию метода проекционной голографической субмикронной фотолитографии.

Заключение

В рамках теории aberrаций голограмм третьего порядка получены аналитические выражение, описывающие требования, предъявляемые к геометрическим параметрам

схем записи и восстановления голограмм-проекторов. На их основе разработана схема записи и восстановления голограмм-проекторов, позволяющая добиться дифракционно-ограниченного разрешения в восстановленном изображении в условиях относительно малой временной когерентности используемых источников. На примере процесса изготовления ламелей четырехкомпонентного фотоприемника разработаны основы технологии проекционной субмикронной голографической фотолитографии и показана пригодность этой технологии к практическому использованию для изготовления устройств оптоинформатики и электроники.

Литература

1. Webb J. All-calcium fluoride system uses 157nm light. // Laser Focus World. 2000. V.36. № 9. P. 87–92.
2. Beesley M. Holographic projection of microcircuit patterns. // Electron. Lett. 1968. V. 4. P. 49–53.
3. Levenson M., Johnson K., Hanchett V. Projection photolithography by wave-front con. // J. Opt Soc. Am. 1981. V. 71. P. 737–739.
4. Voschenkov A., Hanson P. Submicron resolution photolithography by spectral shaping. // IEEE Electron. Device Lett. 1982. V. 3. P. 208–209.
5. Ehbets P., Herzig H., Kuittinen M., Ciube F., Darbellay Y. High-carrier frequency fan-out gratings fabricated by total internal reflection holographic lithography. // Opt.Eng. 1995. Vol. 34. №8. P. 2377–2383.
6. Clube F., Gray S., Struchen D., Tisserand J., Malfoy S., Darbellay Y. Holographic microlithography. // Opt.Eng. 1995. V. 34. №9. P. 2724–2730.
7. Stetson K.A. Holography with total internally reflected light. // Appl.Phys.Lett. 1967. V. 11. P. 225–226.
8. Jusupov I.J., Mikhailov M.D. Investigation of the arsenic sulphide films for relief-phase holograms. // Proc. SPIE. 1989. V. 1238. P. 240–247.
9. Герке Р.Р., Корешев С.Н., Семенов Г.Б., Смирнов В.В. Голограммная оптика в "ГОИ им. С.И. Вавилова". // Оптический журнал. 1994. №1. С. 26–39.
10. Корешев С.Н., Ратушный В.П. Получение беспспекловых изображений двумерных объектов микроскопического масштаба методом голографии. // Оптика и спектроскопия. 2003. Т. 94. №1. С. 139–142.
11. Koreshev S.N., Ratushnyi V.P. holographic method for obtaining images with limiting high resolution for extreme shot-wave lithography problems. // Proc. SPIE. 2004. V. 5290. P. 221–232.
12. Корешев С.Н., Ратушный В.П. Использование метода голографии для получения изображений двумерных объектов при решении задач фотолитографии высокого разрешения. // Оптический журнал. 2004. Т. 71. №10. С. 32–39.
13. Champagne E.B. Nonparaxial imaging, magnification and aberration properties in holography. // JOSA. 1967. V.57. №1. P.51–55.
14. Оптическая голография. Пер. с англ. Под ред. Г.Колфилда. М.: Мир, 1982. 736 с.
15. Корешев С.Н. Выбор геометрических параметров схем записи и восстановления голограмм, частиц малого размера. // Оптика и спектроскопия. 2001. Т.90. №6. С. 1005–1011.
16. Кольер Р., Беркхард К., Лин Л. Оптическая голография. М.: Мир, 1973.. 686 с.

ОСНОВЫ ТЕХНОЛОГИИ ГОЛОГРАФИЧЕСКОЙ СУБМИКРОННОЙ ФОТОЛИТОГРАФИИ

С.Н. Корешев, В.П. Ратушный

Созданы основы принципиально новой технологии проекционной субмикронной голографической фотолитографии. Разработана схема записи и восстановления голограмм-проекторов предельного разрешения, используемых без дополнительного объектива; сформулированы и научно обоснованы требования, предъявляемые к ее геометрическим параметрам и режимам экспонирования голограмм; экспериментально показана пригодность применения этой схемы в задачах проекционной литографии.

Введение

Современные тенденции развития дисплеев направлены не столько на уменьшение размера единичного пикселя изображения, сколько на увеличение общей площади устройств. Поэтому в сложившейся ситуации особо актуальной становится задача разработки методов высокоразрешающей проекционной фотолитографии, обеспечивающих одномоментное экспонирование относительно большого участка заготовки дисплея. Традиционный метод проекционной фотолитографии, основанный на использовании дорогостоящих проекционных объективов, обладает существенными ограничениями по площади одновременно экспонируемого участка фоторезиста, обусловленными невозможностью полной коррекции полевых аберраций проекционных оптических систем. При этом с уменьшением длины волны используемого излучения и, следовательно, с уменьшением характеристического размера отображаемых структур наблюдается устойчивая тенденция к уменьшению безаберрационного участка поля проекционных объективов. Так, экспонирующие устройства, использующие излучение с длиной волны, близкой к i -линии ртути, обладают полями диаметром в 10–15 мм. Устройства же, предназначенные для работы в излучении эксимерных лазеров с длинами волн 193 нм и 157 нм, обладают полями, не превышающими единиц миллиметров. Например, в [1] описаны объективы для проекционной фотолитографии на длинах волн 193 нм и 157 нм, обладающие двухмиллиметровыми полями.

Еще хуже будет складываться ситуация в более коротковолновых областях спектра. Там придется полностью перейти на зеркальные оптические системы, обладающие большими, по сравнению с обычно используемой в настоящее время рефрактивной оптикой, полевыми аберрациями. Производители микросхем компенсируют недостаточность мгновенных полей современных проекционных объективов путем использования сканирующего и пошагового экспонирования. При этом, как нетрудно подсчитать из приведенных данных, изготовление, например, процессора размером 50×50 мм² потребует выполнения более 600 операций экспонирования и совмещения рисунка. Здесь необходимо отметить, что точность совмещения рисунка при пошаговом экспонировании не должна быть хуже 0.1 характеристического размера отображаемой структуры. Поскольку операция совмещения в традиционном фотолитографическом оборудовании сложна и трудоемка, то возникает необходимость его замены альтернативными «большепольными» системами проекционной фотолитографии. При этом наиболее перспективной основой для разработки таких систем нам представляется метод голографии, характеризующийся возможностью формирования полностью свободных от аберраций, в том числе и полевых аберраций, действительных изображений зарегистрированного объекта.

Возможность использования принципов голографии при решении задач фотолитографии уже была рассмотрена рядом авторов на относительно ранних этапах развития голографии [2–4]. Однако результатов, пригодных для практического использования, в то время получено не было. На наш взгляд, основная причина этого состояла в использовании регистрирующих сред, наносимых из раствора и требующих «мокрой»

фотохимической обработки, приводящей к набуханию и последующей усадке среды. Сложившаяся ситуация изменилась в начале 90-х годов, когда специалисты швейцарской фирмы Holtronic Technologies решили указанные проблемы с помощью новой регистрирующей среды, а именно – фотополимерного материала компании Du Pont [5, 6]. При этом для реализации голографического варианта фотолитографического процесса ими была выбрана схема записи и восстановления голограмм «полного внутреннего отражения», предложенная Стетсоном [7]. Следует отметить, что в этой схеме используются как минимум два элемента, которые должны быть прозрачны для излучения рабочей длины волны – стеклянная призма и сам фотополимерный слой. Поэтому технические решения, разработанные швейцарской компанией Holtronic Technologies, пригодны для использования и используются лишь в относительно длинноволновом диапазоне УФ излучения на длине волны 364 нм, соответствующей УФ линии аргонового лазера. Тем не менее, работы специалистов Holtronic Technologies – первый пример успешного коммерческого применения методов голографии в высокоразрешающем фотолитографическом процессе.

Постановка задачи

С нашей точки зрения, для применения в коротковолновой голографической фотолитографии более перспективными являются регистрирующие среды на основе халькогенидного стеклообразного полупроводника (ХСП) [8]. Эти среды наносятся на подложки путем испарения в вакууме и позволяют получать высококачественные отражательные рельефно-фазовые голограммные оптические элементы [9]. Принципиальная возможность использования этих сред для получения рельефно-фазовых отражательных голограмм-проекторов была убедительно продемонстрирована нами в ходе НИР, поставленных и проведенных в ГОИ им. С.И. Вавилова в 2000–2005 гг. [10–13]. В настоящей статье представлены результаты следующего этапа работы, направленного на внедрение полученных нами ранее результатов и «know-how» в реальный, пусть даже и не самый прецизионный, фотолитографический процесс. Эта работа проводилась применительно к схеме проекционной голографической фотолитографии, реализуемой без дополнительного проекционного объектива [10]. В ее рамках были созданы основы принципиально новой технологии проекционной субмикронной голографической фотолитографии. Среди них – разработка схемы записи и восстановления голограмм-проекторов, характеризующейся минимальными требованиями к временной когерентности источника излучения, определение, обоснование и экспериментальная проверка требований, предъявляемых к режимам экспонирования голограмм, и демонстрация пригодности применения голограмм-проекторов в задачах проекционной фотолитографии.

Анализ требований, предъявляемых к временной когерентности источников излучения, используемых в схемах записи и восстановления прецизионных голограмм-проекторов высокого разрешения

Из анализа литературы следует, что, во-первых, малая степень временной когерентности излучения используемых для целей фотолитографии УФ лазеров, не превышающая 30 см даже в случае использования узкополосных спектральных фильтров с полосой пропускания 1 пм, приводит, в общем случае, к потерям разрешения, реализуемого в изображениях, восстанавливаемых методом голографии. Причем эти потери, в условиях недостаточной временной когерентности излучения, будут обусловлены как процессом записи, так и процессом восстановления голограммы. Во-вторых, практически все опубликованные в литературе технические решения, направленные на сниже-

ние требований к временной когерентности излучения, непригодны для целей голографической фотолитографии. Все вышеперечисленное предопределило необходимость определения и обоснования требований, предъявляемых к временной когерентности источников излучения, используемых в схемах записи и восстановления прецизионных голограмм-проекторов высокого разрешения, и разработки самой схемы, обеспечивающей достижение предельного разрешения в условиях малой временной когерентности.

При определении требований, предъявляемых к временной когерентности источников излучения, мы исходили из приведенной на рис. 1 обобщенной схемы записи голограммы точечного источника с параллельным опорным пучком. На этом рисунке X – ось координат, лежащая в плоскости регистрации голограммы. Отметим, что начало координат в данном случае совмещено с центром линейной апертуры голограмм. Буквами O и R обозначены расходящийся гомоцентрический объектный и параллельный опорный пучки лучей. α_r – угол падения опорного пучка лучей на плоскость регистрации голограммы, а α_o – угол, который составляет с осью Z прямая, соединяющая центр апертуры голограммы с источником объектной волны.

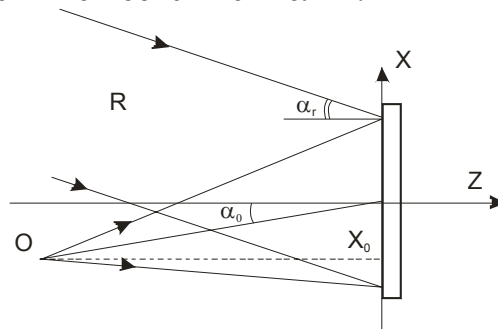


Рис.1. Обобщенная схема регистрации голограмм

С целью упрощения выкладок ограничимся рамками теории aberrаций голограмм третьего порядка и рассмотрением случая записи одномерной голограммы. Тогда, исходя из [13], напишем следующие выражения, описывающие распределения фаз объектной и опорной волн в плоскости апертуры голограммы:

$$\varphi_o(x) \approx \frac{2\pi}{\lambda} \left\{ \frac{x^2}{2R_o} - x \sin \alpha_o - \frac{1}{8R_o^3} [x^4 - 4x^3 x_o + 4x^2 x_o^2] \right\}, \quad (1)$$

$$\varphi_r(x) \approx \frac{2\pi}{\lambda} \{-x \sin \alpha_r\}, \quad (2)$$

где φ_o и φ_r – распределения фаз в плоскости голограммы объектной и опорной волн соответственно. При этом разность указанных распределений $\Delta\varphi(x)$ будет описываться следующим выражением:

$$\Delta\varphi(x) \approx \frac{2\pi}{\lambda} \left\{ \frac{x^2}{2R_o} - x(\sin \alpha_o - \sin \alpha_r) - \frac{1}{8R_o^3} [x^4 - 4x^3 x_o + 4x^2 x_o^2] \right\}. \quad (3)$$

Из (3) найдем элементарное приращение разности фаз, соответствующее малому изменению длины волны источника излучения $d\lambda$:

$$d\Delta\varphi = \frac{\partial \Delta\varphi}{\partial \lambda} d\lambda \approx -\frac{2\pi}{\lambda^2} d\lambda \left\{ \frac{x^2}{2R_o} - x(\sin \alpha_o - \sin \alpha_r) - \frac{1}{8R_o^3} [x^4 - 4x^3 x_o + 4x^2 x_o^2] \right\}. \quad (4)$$

Отсюда получим выражение, описывающее элементарное приращение разности длин оптических путей dl , соответствующее малому изменению длины волны источника излучения $d\lambda$.

$$dl \approx -\frac{d\lambda}{\lambda} \left\{ \frac{x^2}{2R_o} - x(\sin \alpha_o - \sin \alpha_r) - \frac{1}{8R_o^3} [x^4 - 4x^3x_o + 4x^2x_o^2] \right\}. \quad (5)$$

Если в (5) подставить $x = \pm D/2$, где D – размер линейной апертуры голограммы, и положить $x_o = x_{o \max}$, где $x_{o \max}$ – координата точки объекта, наиболее удаленной от оси Z , то его можно будет использовать для определения минимально необходимой длины когерентности источника излучения, предназначенного для формирования голографического поля. Из (5) следует, что требования к длине когерентности источника могут быть существенно снижены при расположении объекта вблизи оси Z , т.е. при $x_o \approx 0$ и, следовательно, при $\sin \alpha_o \approx 0$, а также в случае одноосевой схемы записи голограмм, т.е. при $\sin \alpha_o = \sin \alpha_r$. При указанных выше условиях выражение (5) может быть приведено к виду:

$$dl_n \approx \frac{d\lambda}{2\lambda} R_o \text{tg}^2 \omega \{1 - 0,25 \text{tg}^2 \omega\}, \quad (6)$$

где: dl_n – требуемая длина когерентности источника излучения, а ω – угловая апертура объектного пучка. Отметим, что, помимо одноосевой схемы, выражение (6) будет справедливо и во внеосевом случае, при условии предварительного внесения в опорный пучок разности хода лучей, равной $x \sin \alpha_r$. Такая разность хода лучей может быть обеспечена путем установки в опорный пучок, до его падения на плоскость регистрации голограммы, плоской дифракционной решетки с пространственной частотой, равной несущей пространственной частоте регистрируемой голограммы. При этом приведенная на рис. 1 и рассматриваемая нами обобщенная схема записи голограммы должна быть приведена к виду, представленному на рис. 2.

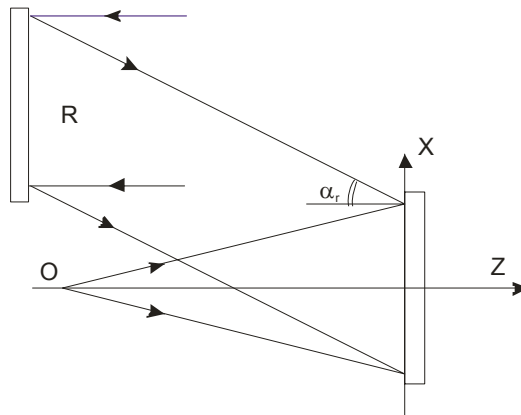


Рис. 2. Обобщенная схема регистрации голограммы с решеткой в опорном пучке

Из этого рисунка видно, что требуемая для данной схемы длина когерентности источника излучения определяется лишь числовой апертурой объектного пучка лучей и не зависит от угла падения опорной волны. Здесь необходимо отметить, что бесспорным условием практической реализации указанной схемы является выравнивание оптических длин пути света, приходящего в центр голограммы вдоль опорного и объектного плеч интерферометра. Напомним, что прием использования решеток на стадии восстановления голограмм для ахроматизации восстановленного изображения не нов и описан многими авторами, например [14]. В то же время, работы, в которых бы упоминалась возможность использования решеток для ахроматизации голографического поля на стадии записи голограмм, нам неизвестны.

Приведенное выше выражение (6) может быть использовано не только для определения требуемой длины когерентности излучения, но также и для определения максимально допустимого расстояния от объекта до плоскости записи голограммы, обеспечивающего при заданной спектральной ширине источника излучения достижение

заданного контраста интерференционных полос на всей линейной апертуре голограммы. Так, если положить допустимым для края апертуры голограммы сбой фазы интерференционных полос, соответствующих длинам волн $\lambda + \Delta\lambda$ и $\lambda - \Delta\lambda$, на 0,5 периода их следования, т.е. положить dI_n равным $0,25\lambda$, то из (6) получим:

$$R_0 \leq \frac{\lambda^2}{2d\lambda(1 - 0,25\text{tg}^2 \omega)\text{tg}^2 \omega} . \quad (7)$$

Напомним, что ранее нами было получено и опубликовано в [15] неравенство (8), обеспечивающее на стадии восстановления голограммы требуемую, т.е., лежащую в пределах 10% допуска, точность отображения элементов структуры объекта при использовании восстанавливающего излучения с конечной, равной $d\lambda$, спектральной шириной

$$R_0 \leq 0,68 \frac{\lambda^2 \left[1 - \frac{1}{2} \sin^2 \omega \right]^3}{d\lambda \sin^4 \omega} . \quad (8)$$

Из сравнения (7) и (8) можно сделать вывод о том, что требование сохранения контраста, описываемое выражением (7), обуславливает несколько более жесткие ограничения величины параметра R_0 . Проиллюстрируем это с помощью таблицы 1, содержащей результаты расчета максимально допустимой величины R_0 , выполненного с помощью выражений (7) - R_0 и (8) - R_0^* .

Таблица 1.

$\Delta\lambda$ (нм)	$\lambda = 488$ нм		$\lambda = 157$ нм	
	R_0 (мм)	R_0^* (мм)	R_0 (мм)	R_0^* (мм)
0,06	4	12	0,4	1,2
0,001	250	700	26	70

Необходимо отметить, что существует еще и третье ограничение, накладываемое на величину R_0 . Оно связано с необходимостью согласования линейной апертуры записываемой голограммы с угловой апертурой излучения, рассеиваемого объектом, т.е. с необходимостью записи на голограмме всего дифракционного поля объекта. Это ограничение играет решающую роль, т.е. становится определяющим, лишь в случаях большой временной когерентности источников излучения и малой площади записываемой голограммы.

Схема записи и восстановления голограмм-проекторов, обеспечивающая минимальные требования к длине когерентности используемых источников излучения

Исходя из сформулированных выше требований о минимальности разности хода интерферирующих пучков лучей и описанных нами ранее в работах [10–12] условий формирования безабберационного изображения, в ходе выполнения настоящей работы была разработана практическая схема записи и восстановления голограмм-проекторов предельного разрешения, содержащая плоскую дифракционную решетку в плечах формирования опорного и восстанавливающего пучков лучей. Она представлена на рис. 3. Поясним реализуемый в ней ход лучей. Излучение лазера 1 направляется на зеркало 2 и падает на частично прозрачное зеркало 3. Здесь оно разделяется на два пучка лучей – опорный пучок и объектный пучок. Опорный пучок отражается от зеркала 3, расширяется с помощью коллиматора 4, отражается от зеркал 5, 11 и направляется на плоскую дифракционную решетку 12. Излучение, дифрагирующее на решетке 12 в +1-ый порядок дифракции, направляется на плоскость регистрации голограммы – про-

ектора 10. Отметим, что излучение, дифрагирующее в сопряженный -1 порядок дифракции, на стадии записи голограмм не используется, а применяется лишь на стадии их восстановления в качестве параллельного восстанавливающего пучка лучей. Объектный пучок лучей проходит через зеркало 3, отражается от зеркала 6, расширяется с помощью коллиматора 7 и направляется на объект 9, представляющий собой бинарный амплитудный транспарант. Восстановление голограмм в данной схеме осуществляется излучением, дифрагирующим в -1 порядок дифракции на решетке 12 при перемещении стола 8, несущего на себе оправу объекта и кассету с голограммой – проектором, в положение 13. Применение решетки в данной схеме позволяет не только выровнять для всей поверхности регистрируемой голограммы оптические длины опорного и объектного плеч интерферометра, формирующего голографическое поле, но и существенно облегчает процесс восстановления работающих на отражение голограмм. Это осуществляется за счет использования при записи и восстановлении голограмм сопряженных порядков дифракции опорного излучения на решетке.

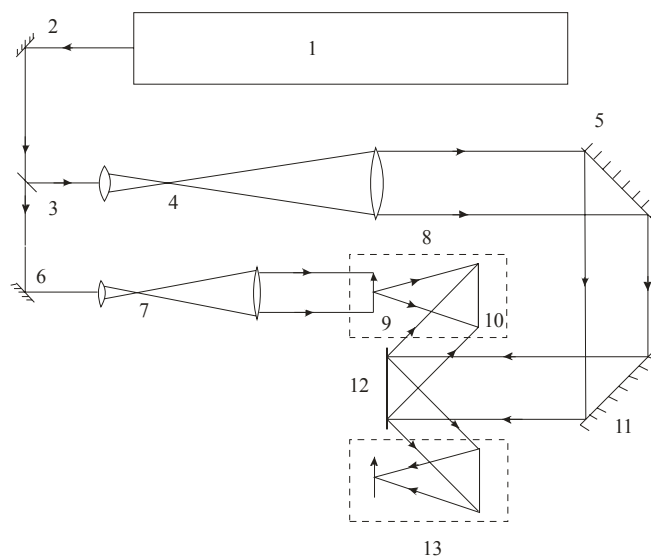


Рис.3. Схема записи и восстановления голограмм - проекторов

Экспериментальная проверка пригодности схемы и полученной с ее помощью голограммы–проектора к использованию в реальном литографическом процессе

Экспериментальная проверка пригодности схемы и полученной с ее помощью голограммы-проектора к использованию в реальном литографическом процессе проводилась нами в ходе изготовления четырехэлементного фотоприемника. диаметром 15 мм с минимальным шагом проектирования 1,3 мкм методом субмикронной фотолитографии, реализуемой на длине волны 488 нм с помощью отражательных голограмм, регистрируемых на слоях халькогенидного полупроводника [8]. Геометрические параметры схемы были выбраны равными: размер голограммы – 68×68 мм²; расстояние от голограмм до объекта 30 мм; углы падения параллельных опорного и восстанавливающего пучков на плоскость регистрации голограммы -70° (точность равенства модулей этих углов друг другу – 10,6"); минимальное расстояние от опорного и объектного точечных источников до центра голограммы – 1,1 км. Последнее требование обеспечивает точную фокусировку восстановленного изображения в той же плоскости, в которой находился объект при записи голограммы. Выбор угла падения опорного и восстанавливающего пучков лучей на голограмму определялся двумя факторами: во-первых, необходимостью обеспечения отсутствия виньетирования указанных пучков лучей оправой фотошаблона; во-вторых, необходимостью пространственного разделения восстанов-

ленного изображения фотошаблона и интермодуляционных помех голограммы, имеющих спектр пространственных частот, в два раза превышающий спектр объекта [16].

Регистрация голограмм по указанной схеме, как указывалось выше, осуществлялась на слоях халькогенидного стеклообразного полупроводника, нанесенного на высококачественные плоскопараллельные подложки, изготовленные из оптического стекла. Плоскостность рабочих поверхностей подложек лежала в пределах $N = 0,1$ и $\Delta N = 0,1$. Нанесение халькогенида на подложки осуществлялось методом распыления в вакууме. Толщина сформированных светочувствительных пленок равнялась 1 мкм. Их постэкспозиционное травление осуществлялось в жидкой среде. Величина оптимальной экспозиции и соотношение интенсивностей интерферирующих пучков при записи голограмм-проекторов подбирались исходя из условия отсутствия помех нелинейности в восстановленных изображениях. Для наших условий оптимальные величины экспозиции и соотношения интенсивностей опорного и объектного пучков лучей равнялись 10 Дж/см^2 и 6:1, соответственно. Регистрация действительного изображения фотошаблона, восстанавливаемого голограммой-проектором, осуществлялось напрямую, без дополнительных оптических систем, на слое позитивного фоторезиста Шипли AZ1350, толщиной 0,2–0,5 мкм, нанесенные поверх слоя хрома, осажденного на тонкие, толщиной 2 мм, стеклянные подложки. Проявление экспонированных слоев фоторезиста осуществлялось в слабом растворе едкой щелочи (KOH 2,5%). Травление пленки хрома производилось до стекла через полученную резистивную маску в растворе четырехводного сернокислого церия. Проведенные исследования показали, что наилучшее качество хромовых ламелей, составляющих структуру разводки изготавливаемого фотоэлемента, обеспечивается при экспозиционной дозе фоторезиста Шипли 40 Дж/см^2 и временах проявления около 1 минуты.

На рис. 4 представлены полученные с различным увеличением фотографии ламелей фотоэлемента, полученных методом голографической фотолитографии.



Рис.4. Изображения структур, полученных методом проекционной голографической фотолитографии

Отметим, что видимые на фотографиях расфокусированные серые пятна обусловлены недостатками оптической системы микроскопа, использовавшегося на стадии исследования качества полученных структур. Представленные фотографии свидетельствуют о достигнутом беспыльном резком восстановлении во всем поле зрения голограммы изображения структуры четырехкомпонентного фотоприемника, т.е. о пригодности к практическому использованию метода проекционной голографической субмикронной фотолитографии.

Заключение

В рамках теории aberrаций голограмм третьего порядка получены аналитические выражение, описывающие требования, предъявляемые к геометрическим параметрам

схем записи и восстановления голограмм-проекторов. На их основе разработана схема записи и восстановления голограмм-проекторов, позволяющая добиться дифракционно-ограниченного разрешения в восстановленном изображении в условиях относительно малой временной когерентности используемых источников. На примере процесса изготовления ламелей четырехкомпонентного фотоприемника разработаны основы технологии проекционной субмикронной голографической фотолитографии и показана пригодность этой технологии к практическому использованию для изготовления устройств оптоинформатики и электроники.

Литература

1. Webb J. All-calcium fluoride system uses 157nm light. // Laser Focus World. 2000. V.36. № 9. P. 87–92.
2. Beesley M. Holographic projection of microcircuit patterns. // Electron. Lett. 1968. V. 4. P. 49–53.
3. Levenson M., Johnson K., Hanchett V. Projection photolithography by wave-front con. // J. Opt Soc. Am. 1981. V. 71. P. 737–739.
4. Voschenkov A., Hanson P. Submicron resolution photolithography by spectral shaping. // IEEE Electron. Device Lett. 1982. V. 3. P. 208–209.
5. Ehbets P., Herzig H., Kuittinen M., Ciube F., Darbellay Y. High-carrier frequency fan-out gratings fabricated by total internal reflection holographic lithography. // Opt.Eng. 1995. Vol. 34. №8. P. 2377–2383.
6. Clube F., Gray S., Struchen D., Tisserand J., Malfoy S., Darbellay Y. Holographic micro-lithography. // Opt.Eng. 1995. V. 34. №9. P. 2724–2730.
7. Stetson K.A. Holography with total internally reflected light. // Appl.Phys.Lett. 1967. V. 11. P. 225–226.
8. Jusupov I.J., Mikhailov M.D. Investigation of the arsenic sulphide films for relief-phase holograms. // Proc. SPIE. 1989. V. 1238. P. 240–247.
9. Герке Р.Р., Корешев С.Н., Семенов Г.Б., Смирнов В.В. Голограммная оптика в "ГОИ им. С.И. Вавилова". // Оптический журнал. 1994. №1. С. 26–39.
10. Корешев С.Н., Ратушный В.П. Получение беспспекловых изображений двумерных объектов микроскопического масштаба методом голографии. // Оптика и спектроскопия. 2003. Т. 94. №1. С. 139–142.
11. Koreshev S.N., Ratushnyi V.P. holographic method for obtaining images with limiting high resolution for extreme shot-wave lithography problems. // Proc. SPIE. 2004. V. 5290. P. 221–232.
12. Корешев С.Н., Ратушный В.П. Использование метода голографии для получения изображений двумерных объектов при решении задач фотолитографии высокого разрешения. // Оптический журнал. 2004. Т. 71. №10. С. 32–39.
13. Champagne E.B. Nonparaxial imaging, magnification and aberration properties in holography. // JOSA. 1967. V.57. №1. P.51–55.
14. Оптическая голография. Пер. с англ. Под ред. Г.Колфилда. М.: Мир, 1982. 736 с.
15. Корешев С.Н. Выбор геометрических параметров схем записи и восстановления голограмм, частиц малого размера. // Оптика и спектроскопия. 2001. Т.90. №6. С. 1005–1011.
16. Кольер Р., Беркхард К., Лин Л. Оптическая голография. М.: Мир, 1973.. 686 с.

ОПТИЧЕСКОЕ ПЕРЕКЛЮЧЕНИЕ И ИР-КОНВЕРСИЯ В МАТЕРИАЛАХ С КВАНТОВЫМИ ЯМАМИ: ЭФФЕКТ ФОТОННОЙ ЛАВИНЫ Е.Ю. Перлин, А.В. Иванов, Р.С. Левицкий

Исследован процесс запуска фотонной лавины в глубоких квантовых ямах (КЯ). Пороговые интенсивности ИК света, вызывающего лавину, составляют сотни кВт/см². Показано, что энергии оптического переключения системы составляют десятки либо сотни фДж/мкм². Предложена новая перспективная модель переключения на основе эффекта фотонной лавины (ЭФЛ) в гетероструктурах типа I с глубокими КЯ. Рассчитаны вероятности элементарных процессов, формирующих лавину. Выполнено математическое моделирование и проведена оптимизация параметров гетероструктуры типа I с глубокими КЯ. Проведен сравнительный анализ моделей оптического переключения и ир-конверсии в наноструктурах различных типов. Предложен принципиально новый механизм многократной ир-конверсии в примесных системах и наноструктурах – так называемый оптический трамплин. Выполнен качественный анализ и даны предварительные оценки эффективности этого механизма. Показана его перспективность для решения практических задач.

1. Общие сведения об эффекте фотонной лавины

Практические задачи создания следующих поколений систем обработки и передачи данных обуславливают поиск новых физических механизмов управления светом с помощью света, в том числе механизмов оптического переключения. В зависимости от характера решаемой практической задачи на первый план могут выходить такие параметры, как время переключения системы между состояниями с принципиально различными оптическими и (или) электрическими свойствами, либо энергия светового импульса, необходимая для переключения, либо интенсивность управляющего излучения, либо возможность перестройки спектрального диапазона работы устройства и т.д.

Настоящая работа посвящена исследованию исключительно перспективных в прикладном плане механизмов сверхбыстрого оптического переключения твердотельных гетеросистем с различными типами электронной зонной структуры на основе так называемого эффекта фотонной лавины (ЭФЛ).

Эффект, получивший это название, был впервые обнаружен в 1979 году в работе [1], начиная с которой использование ЭФЛ стало одним из наиболее эффективных способов возбуждения коротковолновой люминесценции с помощью длинноволновой накачки. До настоящего времени ЭФЛ наиболее интенсивно изучался для систем примесных редкоземельных ионов (РЗИ) (см., например, [1–10]).

Поясним механизм фотонной лавины на примере простейшей трехуровневой схемы. Обозначим основное и два возбужденных состояния РЗИ как 1, 2 и 3. Состояние 2 обычно считается метастабильным. В отсутствие накачки заполнено только состояние 1. Пусть частота накачки ω близка к частоте перехода ω_{32} между состояниями 2 и 3, но далека от резонанса с частотой $\omega_{21} < \omega_{32}$ перехода между состояниями 1 и 2. При малых интенсивностях света j в системе ничего не происходит, так как могли бы идти лишь оптические переходы $2 \rightarrow 3$ между незаполненными состояниями. При увеличении j ситуация меняется. Пусть один из РЗИ каким-либо образом оказывается в состоянии 2. Поглощая квант $\hbar\omega$, он переходит в состояние 3. За счет кросс-релаксационного механизма ион возвращается из состояния 3 в состояние 2. При этом один из соседних РЗИ возбуждается из основного состояния 1 в состояние 2. Таким образом, в состоянии 2 оказываются уже два иона. Каждый из них, в свою очередь, может участвовать в таких же процессах, и т.д. В результате в состоянии 2 накапливается много электронов на различных ионах, возникает сильное поглощение на переходах $2 \rightarrow 3$ и, соответственно, высокая заселенность на уровне 3. В этом случае возможна люминесценция на переходе $3 \rightarrow 1$ на частоте $\Omega > \omega$. Ключевую роль в ЭФЛ играет конкуренция между релакса-

ционными потерями электронов в состоянии 2 и накоплением электронов в этом состоянии благодаря процессам поглощения света и кросс-релаксационного переноса возбуждения. Характерными свойствами ЭФЛ являются: а) четко выраженный пороговый характер явления – при $j \approx j_{th}$ скачком увеличиваются заселенности возбужденных состояний и поглощение света ω ; б) резкое возрастание времени τ_{eq} установления квазиравновесного распределения электронов в области пороговых интенсивностей света j_{th} . Значения j_{th} убывают с ростом концентрации РЗИ. Пороговый характер ЭФЛ позволил дать анализ этого явления в терминах теории фазовых переходов II рода Л.Д. Ландау [6].

В последнее десятилетие был предложен и практически реализован ряд различных схем лавинной *up*-конверсии в примесных системах (подробную библиографию см. в [7]). В частности, в [10] рассмотрена высокоэффективная лавинно-каскадная схема *up*-конверсии в восьмиуровневой модели трехзарядных ионов тулия в кристаллах YLF. Эта схема позволяет получать излучение с длиной волны $\lambda \sim 0.29$ мкм при накачке с $\lambda = 1.11$ мкм или $\lambda = 0.649$ мкм. Можно констатировать, что использование ЭФЛ позволяет весьма эффективным способом возбуждать коротковолновую люминесценцию с помощью длинноволновой накачки.

Фактически, когда при увеличении интенсивности накачки j эта величина переходит из области $j < j_{th}$ в область $j > j_{th}$, происходит переключение системы между состоянием (I) с незаселенными возбужденными состояниями и очень слабым поглощением излучения накачки и состоянием (II) с большими заселенностями возбужденных состояний и сильным поглощением накачки. Ясно, что после выключения накачки система возвращается из состояния II в состояние I за время порядка наиболее длинного из времен релаксации в электронной подсистеме. Введем времена τ_{eq} установления квазиравновесного распределения в электронной системе при ЭФЛ. Тогда энергию переключения системы можно определить следующим образом: $E_{sw} \approx \tau_{eq} j_{th}$.

Типичные времена τ_{eq} установления квазиравновесного распределения в электронной системе РЗ ионов при ЭФЛ составляют 1÷100 мс при пороговых плотностях энергии накачки, необходимых для включения лавинного механизма, $E_{sw} \sim 0.1 \div 10$ мкДж/мкм². Столь медленное протекание ЭФЛ в системах РЗИ связанное с малыми значениями сил осцилляторов для актуальных оптических переходов и большими временами жизни возбужденных электронных состояний, естественно, ограничивает круг возможностей практического использования этого явления в оптоэлектронике. Поэтому представляет интерес поиск твердотельных систем, где переключение системы осуществлялось бы за значительно более короткие времена с затратой меньшей энергии. Можно было предположить, что подходящей системой окажется квантовая яма (КЯ), где, с одной стороны, силы осцилляторов для переходов между подзонами размерного квантования являются величинами порядка единицы, а с другой стороны, достаточно коротки времена релаксации в электронной системе. В работах [11–13] была предложена полупроводниковая система с легированными КЯ, где переключение в режим лавинной *up*-конверсии может осуществляться за значительно более короткие времена с затратой меньшей энергии. Роль, которую в примесной системе играют межзонные кросс-релаксационные переходы, в легированной КЯ выполняют внутриямные межподзонные переходы оже-типа.

В системе, рассмотренной в [11, 12], сохраняются такие особенности ЭФЛ, как пороговый характер процесса и резкое увеличение времен τ_{eq} при интенсивностях накачки, близким к пороговым значениям. Фотоиндуцированная динамика в легированной КЯ описывается в [11, 12] системой нелинейных уравнений баланса для заселенностей трех нижних подзон 1, 2 и 3 (резонансные оптические переходы идут между подзонами 2 и 3). Анализ этих уравнений показал, что система обладает двумя стационар-

ными точками – стабильным узлом и нестабильным узлом. При бифуркационном значении интенсивности света $j = j_{th}$ эти две стационарные точки вырождаются в одну. В случае, когда вероятность оже-процесса W_A , при котором в результате столкновения электрона из подзоны 3 с электроном из подзоны 1 оба переходят в подзону 2, удается представить в виде $W_A \approx \gamma_A n_1 n_3$, имеем [12]:

$$j_{th} \approx \frac{W_{21}(n_0\gamma_A + W_{31} + W_{32})}{\sigma_{23}(n_0\gamma_A - W_{21} - W_{32})}, \quad (1)$$

где n_0 – полная (двумерная) концентрация носителей в КЯ, W_{ij} – скорости релаксационных переходов между i -й и j -й подзонами, σ_{ij} – сечения поглощения света на переходе между i -й и j -й подзонами. Формула (1) справедлива при $n_0\gamma_A > W_{ij}$, $\sigma_{12} = 0$. При конечных, но малых по сравнению с σ_{23} , значениях σ_{12} (формальная) бифуркационная интенсивность j_{th} оказывается комплексной: $\tilde{j}_{th} = j'_{th} + i j''_{th}$. Однако $j''_{th} \ll j'_{th}$ для актуальных значений σ_{ij} , W_{ij} , и n_0 . Поэтому при увеличении j вблизи j_{th} имеет место резкое возрастание квазиравновесных значений n_2 и n_3 и уменьшение n_1 . При увеличении σ_{12} изменение заселенностей n_i вблизи порога становится менее резким. Для времени τ_{eq} при $j = j_{th}$ в [12] получено выражение:

$$\tau_{eq} \sim \frac{1}{2W_{21}} \sqrt{\frac{n_0\gamma_A\sigma_{23}}{\sigma_{12}(W_{21} + 2W_{31} + W_{32})}}. \quad (2)$$

Видно, что τ_{eq} увеличивается с уменьшением скорости прихода электронов в состояние 2, где их количество может лавинообразно возрастать.

Как показали расчеты [11–16], благодаря ЭФЛ в легированных КЯ можно за времена ~ 1 –100 пс переключить материал из состояния I в состояние II. В состоянии I практически все электроны находятся в нижней подзоне размерного квантования, длинноволновый свет поглощается слабо. В состоянии II электроны заселяют вторую и третью подзоны, а в нижней подзоне их концентрация мала, так что возникает инверсия заселенностей между подзонами 2 и 1, а также 3 и 1. Поглощение света в состоянии II резко возрастает. Если симметрия квантовой ямы такова, что переходы 3→1 разрешены, то на этих переходах возможна фотолюминесценция с длиной волны, меньшей, чем у возбуждающего света. Плотность энергии E_{sw} , которую нужно затратить для переключения, оказывается очень малой: $E_{sw} \sim 10$ –100 фДж/мкм², что на 6–7 порядков ниже аналогичной величины для примесных систем с РЗИ.

Предложенная в работах [14, 15] схема ЭФЛ в системе с КЯ типа II существенно отличалась от рассмотренной в [11–12] и позволяла при умеренных интенсивностях накачки получить люминесценцию с длиной волны, в 3–5 раз большей, чем у возбуждающего света. При этом энергия переключения $E_{sw} \sim 1$ –10 пДж/мкм², т.е. на 4–5 порядков ниже, чем в системе РЗИ.

Как следует из сказанного выше, ЭФЛ возникает в ситуациях, когда поглощение света происходит на переходах из возбужденных электронных состояний. На первый взгляд, заведомо более эффективны схемы возбуждения, где оптические переходы идут из основного электронного состояния, в котором до начала действия импульса накачки находятся почти все примеси. Однако, парадоксальным образом, как показало исследование ЭФЛ в системах РЗИ (см., например, [1–10]), при $j > j_{th}$ каскадно-лавинные схемы с поглощением из нижних возбужденных состояний позволяют получить значительно бóльшую заселенность верхних возбужденных состояний, чем чисто каскадные схемы с поглощением из основного состояния. Нами показано, что аналогичная ситуация имеет место и в случае лавинной *np*-конверсии в системах с глубокими КЯ.

2. Принципы расчетов динамики процессов типа фотонной лавины

Задачи о расчете процессов типа фотонной лавины в локальных системах и наноструктурах состоят из нескольких взаимосвязанных частей.

А) Для каждой конкретной системы нужно построить модель процесса, т.е. выделить электронные или электрон-колебательные состояния, к квантовым переходам между которыми сводится кинетика лавины, а также выделить все существенные для ЭФЛ одно- и многочастичные переходы.

Б) Требуется рассчитать (или оценить по имеющимся экспериментальным данным) вероятности фотопереходов между различными электронными состояниями примесного центра или квантовой яме.

В) Нужно оценить скорости релаксационных процессов, к числу которых относятся излучательные и безизлучательные процессы, причем в большинстве актуальных случаев процессы второго типа играют доминирующую роль. Примером того, как это делается в случае примесных РЗИ, является оценка, выполненная в работе [10]. В случае материалов с КЯ основную роль в процессах внутри- и межподзонной релаксации играет взаимодействие с полярными оптическими фононами. Анализ усложняется тем, что необходимо учитывать взаимодействие как с конфайнментными, так и с интерфейсными колебательными модами. Этим вопросам посвящена довольно обширная литература (см. например, [14–19]). Сечения захвата носителей из зонных состояний в подзону размерного квантования в КЯ удастся оценить лишь очень грубым образом, равно как и сечения межзонной рекомбинации неравновесных электрон-дырочных пар. Впрочем, значения этих величин не являются критическими для построения качественной картины ЭФЛ.

Г) Необходимо рассчитать скорости «лавинообразующих» процессов – кросс-релаксации в локальных центрах или процессов оже-типа в КЯ. Соответствующие вычисления оказываются техническими весьма сложными, однако удастся получить достаточно надежные оценки скоростей этих процессов [10, 11, 13].

Д) Требуется составить и решить систему уравнений нелинейных дифференциальных уравнений баланса населенностей различных электронных состояний. Качественный анализ такой системы удастся провести лишь для некоторых наиболее простых моделей. Решения системы получаются исключительно численными методами. Следует иметь в виду, что применимость системы уравнений баланса для анализа фотоиндуцированной динамики далеко не во всех случаях очевидна, и этот вопрос нуждается в отдельном исследовании.

3. Модель каскадно-лавинойной *ip*-конверсии и переключения в квантовой яме типа I

В данной работе рассмотрим более подробно ЭФЛ в гетероструктуре типа I с глубокими КЯ (в структуре типа I энергетическая щель узкозонного материала ямы лежит полностью внутри щели широкозонного материала барьера). Несмотря на одинаковую природу эффекта в различных системах, для каждой из них имеется ряд особенностей. Эти особенности могут играть существенную роль при практическом использовании ЭФЛ, в том числе для накачки коротковолновых лазеров длинноволновым светом, полностью оптического переключения и управления светом с помощью света.

Итак, рассмотрим гетероструктуру с зонной схемой типа I, состоящую из компонент *A* и *B* (рис. 1). Область *A* с шириной $2a$ является прямоугольной ямой как для электронов (с глубиной U_c), так и для дырок (с глубиной U_v). Считаем, что глубина ямы U_c достаточно велика (~ 1.5 – 2 эВ), в ней имеются три подзоны размерного квантования, нумеруемые в порядке возрастания энергии, а энергетические зазоры между *i*-й *j*-й под-

зонами $\hbar\omega_{ij}$ велики по сравнению с температурой T . Предполагается также, что $\omega_{32} > \omega_{21}$, а частота падающего света $\omega \approx \omega_{32}$ ($\hbar\omega \sim 0.5\text{--}0.8$ эВ). Считается, что $U_v < U_c$ и в яме для дырок имеется лишь одна подзона размерного квантования v_1 . В отличие от [11–13], предполагается, что в равновесных условиях электронные состояния в квантовой яме не заселены.

Развитие процесса (рассматривается также случай, когда имеется небольшое число электронов в подзоне 1) фотонной лавины происходит следующим образом. При малых интенсивностях света j появляется лишь небольшое число неравновесных электронов в подзоне 1 за счет прямых двухфотонных переходов (ДФП) из подзоны v_1 . Поскольку частота света больше частоты прямых переходов ω_{21} между подзонами 1 и 2, между этими подзонами могут идти лишь слабые непрямые (в \mathbf{k}_\perp -пространстве) переходы. При этом заселенности подзоны 2 и более высоких состояний остаются пренебрежимо малыми. Быстрые резонансные фотопереходы могли бы идти между подзонами 2 и 3, но эти подзоны при малых j остаются пустыми. Картина резко меняется при высоких значениях j . За счет двухступенчатого каскада слабых переходов $v \rightarrow 1$ и $1 \rightarrow 2$ некоторое число электронов все же оказывается в состояниях с $\mathbf{k}_\perp \neq 0$ подзоны 2. Эти электроны быстро релаксируют на дно подзоны. Затем они могут либо «свалиться» в подзону 1 благодаря релаксации с участием фононов, либо поглотить фотон $\hbar\omega$ и оказаться в подзоне 3. Для резонансных разрешенных переходов $2 \rightarrow 3$ сила осциллятора очень велика. Из подзоны 3 электроны могут «свалиться» в подзоны 2 и 1 благодаря релаксации с участием фононов, но достаточно эффективным, как показано в [11], оказывается и механизм оже-релаксации: благодаря кулоновскому взаимодействию между электроном в подзоне 3 и электроном в подзоне 1 они оба переходят в подзону 2. Каждый из этих электронов может таким же образом привести к появлению двух электронов в подзоне 2 и т.д. При больших интенсивностях света j в подзону 2 благодаря этому механизму попадает больше электронов, чем их уходит в подзону 1. В этом случае происходит лавинообразное увеличение заселенностей в подзонах 2, 3 и в состояниях непрерывного спектра зоны проводимости c .

В рассматриваемой модели существенную роль играет процесс оже-типа $3 \rightarrow 1v_1$: электрон из подзоны 3 переходит в подзону 1, передавая высвободившуюся энергию на рождение пары, состоящей из электрона в подзоне 1 и дырки в подзоне v_1 . При этом увеличивается общее число неравновесных электрон-дырочных пар в КЯ, что приводит к снижению пороговой интенсивности света j_{th} . Этот процесс отличается от рассмотренного в разд. 4 работ [14, 15] тем, что теперь обе начальные частицы находятся в КЯ (область A).

В данной работе в дополнение к процессам, исследованным в [14, 15], мы также включаем в рассмотрение фотопереходы дырок из подзоны v_1 (область A) в состояния непрерывного спектра валентной зоны (область B) и процесс релаксации дырок из непрерывного спектра в подзону v_1 . Таким образом, выполнен пункт А раздела 2. По пунктам Б (частично), В и Г (частично) мы можем воспользоваться результатами, полученными ранее в работах [11–15]. Фактически по пункту Б нам необходимо лишь рассчитать вероятности рождения электрон-дырочных пар за счет ДФП между подзонами v_1 и 1. Реализация программы вычислений по пунктам Г и Д будет описана ниже.

4. Вероятности двухфотонных переходов $v_1 \rightarrow 1$ и переходов оже-типа $3 \rightarrow 1v_1$

Для оценки вероятности прямых (в координатном пространстве и в пространстве поперечных импульсов) ДФП между подзонами v_1 и 1 была использована модель с однопериодными волновыми функциями (см. [14]) и учитывалась специфика правил отбора для межзонных и межподзонных оптических переходов. Вычисления проводились во

втором порядке теории возмущений по взаимодействию электронной подсистемы со светом. Были получены аналитические выражения для вероятности ДФП через интенсивность света, приведенную массу электрона и дырки, межзонный матричный элемент оператора импульса электрона и превышение энергии двух фотонов над порогом ДФП.

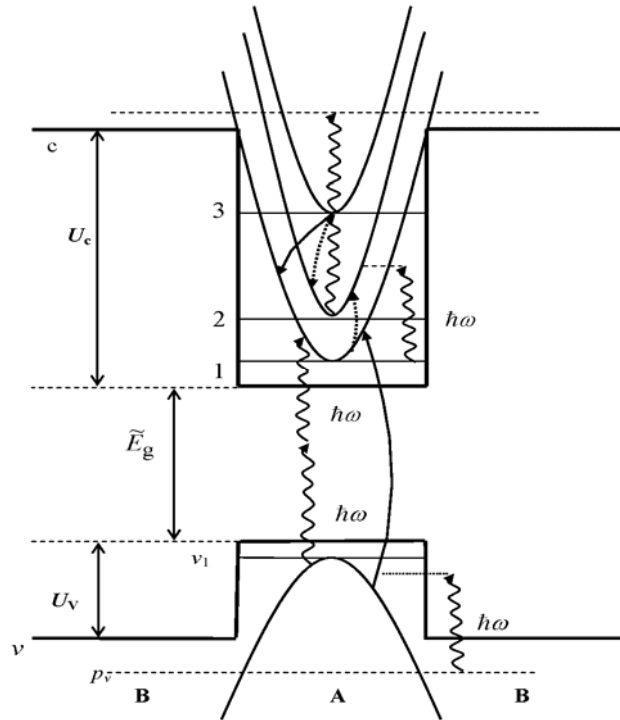


Рис. 1. Схема переходов в системе с квантовыми ямами типа I при каскадно-лавинной ip -конверсии. Вертикальные волнистые линии со стрелками обозначают оптические переходы, сплошные линии со стрелками – оже-переходы $3 \rightarrow 1$ и $3 \rightarrow v_1$, а пунктирные линии со стрелками – оже-переходы $31 \rightarrow 22$

Далее рассмотрим переходы оже-типа $3 \rightarrow 1$ и $3 \rightarrow v_1$: электрон с двумерным волновым вектором $\mathbf{k}_{3\perp}$, находящийся в подзоне 3 КЯ зоны проводимости, переходит в подзону 1 в состояние с волновым вектором $\mathbf{k}_{11\perp}$. Высвободившаяся энергия расходуется на рождение пары, состоящей из электрона в подзоне 1 с волновым вектором $\mathbf{k}_{12\perp}$ и дырки в подзоне v_1 с волновым вектором $-\mathbf{k}_{v\perp}$. Переданный при этом импульс в плоскости, перпендикулярной оси роста наноструктуры, равен:

$$\mathbf{q}_{0\perp} = \mathbf{k}_{3\perp} - \mathbf{k}_{11\perp} = \mathbf{k}_{12\perp} - \mathbf{k}_{v\perp}. \quad (3)$$

Выражение для вероятности данного перехода запишем в виде

$$W(\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}) = \frac{2\pi}{\hbar} \sum_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}, \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}} f(\mathbf{k}_{3\perp}) \left| M_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}}^{(d)} + M_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}}^{(e)} \right|^2 \times \delta(E_{3\mathbf{k}_{3\perp}} + E_{v\mathbf{k}_{v\perp}} - E_{1\mathbf{k}_{11\perp}} - E_{1\mathbf{k}_{12\perp}}), \quad (4)$$

где $f(\mathbf{k}_{3\perp})$ – функция распределения электронов в подзоне 3 зоны проводимости, а $M_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}}^{(d)}$ и $M_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}}^{(e)}$ – прямой и обменный матричные элементы оператора межэлектронного взаимодействия. При написании формулы (4) предполагалось, что электронные состояния $\mathbf{k}_{11\perp}$, $\mathbf{k}_{12\perp}$ и дырочные состояния $-\mathbf{k}_{v\perp}$ не заняты. Это обусловлено, с одной стороны, большой величиной расстройки резонанса $\Delta'_v = (E_{03} - 2E_{01} - \tilde{E}_g - E_{0p_1})$, (E_{0i} – энергия экстремума i -й подзоны размерного кванто-

вания, \tilde{E}_g – см. рис. 1) и короткими временами релаксации электронов в подзоне размерного квантования, а с другой — уходом дырок под действием излучения накачки в состоянии непрерывного спектра валентной зоны).

Подставляя кулоновское взаимодействие в виде разложения в трехмерный ряд Фурье, после стандартных преобразований в результате получим следующее выражение для прямого матричного элемента:

$$M_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}, \mathbf{k}_{3\perp} - \mathbf{q}_{0\perp}, \mathbf{k}_{v\perp} + \mathbf{q}_{0\perp}}^{(d)}(q_{0\perp}) = \frac{2e^2}{\varepsilon_l S} \chi_{\parallel} \left[i\tilde{I}_1(q_{0\perp}) + \sqrt{2} \cos \theta_{\mathbf{q}_{0\perp} \pi_{\perp}} \tilde{I}_2(q_{0\perp}) \right], \quad (5)$$

где $\chi = \frac{\hbar \pi_{cv}}{mE_g}$, $\pi = \mathbf{p} + \frac{\hbar}{4mc^2} [\boldsymbol{\sigma} \nabla V_L]$, $\boldsymbol{\sigma}$ – матрица Паули, S – площадь КЯ, V_L – периодический потенциал решетки, \mathbf{p} – оператор импульса, $\theta_{\mathbf{q}_{0\perp} \pi_{\perp}}$ – угол между векторами $\mathbf{q}_{0\perp}$ и π_{\perp} , а величины $\tilde{I}_i(q_{0\perp})$ представляют собой следующие выражения:

$$\tilde{I}_1(q_{0\perp}) = \iint dz_1 dz_2 e^{-q_{0\perp} |z_1 - z_2|} \text{sign}(z_1 - z_2) \beta_3^*(z_1) \beta_1(z_1) \beta_{v1}^*(z_2) \beta_1(z_2), \quad (6)$$

$$\tilde{I}_2(q_{0\perp}) = \iint dz_1 dz_2 e^{-q_{0\perp} |z_1 - z_2|} \beta_3^*(z_1) \beta_1(z_1) \beta_{v1}^*(z_2) \beta_1(z_2), \quad (7)$$

где $\beta_i(z)$ – огибающие волновые функции подзон размерного квантования.

Выражение для вероятности перехода запишем в виде:

$$W_{3,11v} = \frac{2^3 S^3}{(2\pi)^5 \hbar} \int d^2 \mathbf{k}_{v\perp} \int f(\mathbf{k}_{3\perp}) d^2 \mathbf{k}_{3\perp} \int d^2 \mathbf{q}_{0\perp} \left| M_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}, \mathbf{k}_{3\perp} - \mathbf{q}_{0\perp}, \mathbf{k}_{v\perp} + \mathbf{q}_{0\perp}}^{(d)} + M_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}, \mathbf{k}_{3\perp} - \mathbf{q}_{0\perp}, \mathbf{k}_{v\perp} + \mathbf{q}_{0\perp}}^{(e)} \right|^2 \delta \left[\Delta'_v + \frac{\hbar^2}{m_c} (\mathbf{k}_{3\perp} - \mathbf{k}_{v\perp}) \cdot \mathbf{q}_{0\perp} - \frac{\hbar^2 q_{0\perp}^2}{m_c} - \frac{\hbar^2 k_{v\perp}^2}{2m_r} \right]. \quad (8)$$

Для упрощения вычисления шестикратного интеграла в выражении для вероятности будем считать, что все электроны в подзоне 3 находятся вблизи ее дна. Такое предположение обусловлено тем обстоятельством, что релаксация электронов внутри подзоны 3 происходит значительно быстрее других релаксационных процессов. Тогда в подынтегральной функции в правой части (8) можно положить $\mathbf{k}_{3\perp} = 0$. При этом

$$\int f(\mathbf{k}_{3\perp}) d^2 k_{3\perp} = (2\pi)^2 n_3. \quad (9)$$

Интегрирование по углам между $\mathbf{k}_{3\perp}$ и \mathbf{q}_0 снимается с помощью δ -функции. После некоторых вычислений получим:

$$\frac{1}{\tau_{\text{aug}}} = \frac{W_{3,11v}}{S n_3} = \frac{4 e^4 m_c \chi_{\parallel}^2}{\pi^2 \hbar^3 a^2 \varepsilon_l^2} \int_0^{k_{v\perp}^{(m)}} dk_{v\perp}^2 \int_{g_-}^{g_+} \frac{dq_0^2}{\sqrt{(q_{0\perp}^2 - g_-)(g_+ - q_{0\perp}^2)}} \left| \tilde{M}_{\mathbf{q}_{0\perp}}^{(d)} + \tilde{M}_{\mathbf{q}_{0\perp}}^{(e)} \right|^2, \quad (10)$$

$$g_{\pm} = \frac{1}{2} (k_{v\perp} \pm \rho), \quad \rho = \sqrt{k_{v\perp}^2 + \frac{4m_c a^2}{\hbar^2} \left(\Delta'_v - \frac{\hbar^2 k_{v\perp}^2}{2m_r} \right)}, \quad (11)$$

$$k_{v\perp}^{(m)} = \frac{a^2}{\hbar^2} \frac{4m_v m_c}{2m_c + m_v} \Delta'_v. \quad (12)$$

В приведенных выше формулах фигурируют безразмерные волновые векторы, измеряемые в единицах a^{-1} .

5. Математическое моделирование и оптимизация параметров гетероструктуры типа I с глубокими квантовыми ямами

Математическое моделирование фотонной лавины в квантовых ямах проводится с использованием уравнений баланса для заселенностей электронных состояний. При составлении уравнений баланса следует иметь в виду, что при больших интенсивностях накачки j концентрации электронов в подзонах 2 и 3 становятся сопоставимыми, и следует учитывать фотоиндуцированные переходы как с поглощением, так и с испусканием фотона. Сказанное не относится к переходам $p_v \rightarrow p_1$, $p_1 \rightarrow 1$, $1 \rightarrow 2$ и $3 \rightarrow c$, так как конечные состояния для этих переходов практически не заполнены из-за отдаленности от минимумов соответствующих подзон в КЯ валентной зоны и КЯ зоны проводимости. Обладая относительно большой кинетической энергией, электроны или дырки в этих конечных состояниях быстро теряют ее за счет внутризонной или внутривалентной релаксации и покидают область резонанса.

Система уравнений для концентраций неравновесных дырок непрерывного спектра валентной зоны p_v , дырок p_1 в КЯ валентной зоны, электронов $n_{1,2,3}$ в трех подзонах КЯ зоны проводимости и электронов в непрерывном спектре зоны c n_c имеет вид

$$\begin{aligned}
 \dot{n}_c &= \sigma_{3c} j n_3 - W_{c3} n_c - d_c n_c (p_0 + p_v), \\
 \dot{n}_3 &= W_{c3} n_c - \sigma_{3c} j n_3 - (W_{31} + W_{32} + \tau_{aug}^{-1}) n_3 + \sigma_{23} j (n_2 - n_3) - W_{31,22} (n_1, n_3), \\
 \dot{n}_2 &= -W_{21} n_2 - \sigma_{23} j (n_2 - n_3) + W_{32} n_3 + 2W_{21,22} (n_1, n_3) + \sigma_{12} j n_1, \\
 \dot{n}_1 &= \sigma_{12} j n_1 + W_{21} n_2 + (W_{31} + 2\tau_{aug}^{-1}) n_3 - W_{31,22} (n_1, n_3) \\
 &\quad + \sigma_{p_1} j^2 (1 - f_{p_1}) (1 - f_{c_1}) - d_1 n_1 (p_{10} + p_1), \\
 \dot{p}_1 &= d_1 n_1 (p_{10} + p_1) + \sigma_{p_1} j^2 (1 - f_{p_1}) (1 - f_{n_1}) + \tau_{aug}^{-1} n_3 + \sigma_{p_v p_1} p_v j - W_{p_1 p_v} p_1, \\
 \dot{p}_v &= d_c n_c (p_0 + p_v) - \sigma_{p_v p_1} p_v j + W_{p_1 p_v} p_1
 \end{aligned} \tag{13}$$

с начальными условиями: $n_1, n_2, n_3, n_c = 0, p_1 = p_{01}, p_v = p_0$ при $t = 0$. В формулах (13) W_{c3} , $W_{p_1 p_v}$ – скорости захвата электронов и дырок из непрерывного спектра зоны проводимости в подзону 3 КЯ зоны проводимости и подзону p_1 КЯ валентной зоны. Вероятность оже-переходов типа 31→22 описывается членом $W_{31,22}$ (см. [11, 13]). Оценка сечений σ_{ij} оптических переходов между i -й и j -й подзонами проводится с учетом того, что типичные ширины полос межподзонного поглощения составляют 10–30 мэВ.

В уравнениях (13) фигурируют «двумерные» концентрации p_v, p_0, n_c частиц непрерывного зонного спектра, которые отличаются от обычных концентраций $\tilde{p}_v, \tilde{p}_0, \tilde{n}_c$ множителем $n_w = 5 \cdot 10^5 \text{ см}^{-1}$ (количество ям на единицу длины вдоль оси роста наноструктуры): $\tilde{n}_c = n_c n_w$ и т.п. Члены $-d_c n_c (p_0 + p_v)$ и $-d_1 n_1 (p_{01} + p_v)$ описывают обычную бимолекулярную рекомбинацию электронов в зоне проводимости и в подзоне 1 с дырками в валентной зоне и в подзоне p_1 области A . Величины p_0, p_{10} обозначает равновесную концентрацию дырок в непрерывном спектре валентной зоны и в подзоне p_1 . Равновесную концентрацию электронов в зоне проводимости считаем равной нулю.

При высоких интенсивностях света состояния вблизи дна нижней подзоны размерного квантования КЯ для электронов оказываются заполненными, а состояния вблизи потолка подзоны в КЯ для дырок опустошаются (т.е. возникает высокая концентрация дырок p). Это влияет на скорости многофотонных межзонных переходов и многофотонных переходов оже-типа. Данные эффекты не являются критическими для рассматриваемой задачи. Тем не менее, их желательно учесть хотя бы в самом грубом приближении. С этой целью реальные распределения неравновесных электронов и ды-

рок аппроксимируются фермиевскими функциями распределения, соответствующими мгновенным концентрациям $n_1(t)$ и $p_1(t)$ электронов и дырок. Так, в формулах (13) фигурируют функции распределения f_{n_1} и f_{p_1} электронов и дырок в подзоне n_1 и p_1 соответствующие энергиям $\varepsilon_c(\mathbf{k}_\perp)$ и $\varepsilon_v(\mathbf{k}_\perp)$ в точке \mathbf{k}_\perp двухфотонного резонанса между подзонами.

$$f_{n_1} = \left\{ \exp \left[(\varepsilon_{n_1}(\mathbf{k}_\perp) - \mu_{n_1}(n_1)) / k_B T \right] + 1 \right\}^{-1}, \quad f_{p_1} = \left\{ \exp \left[(\varepsilon_{p_1}(\mathbf{k}_\perp) - \mu_{p_1}(p_1)) / k_B T \right] + 1 \right\}^{-1}. \quad (14)$$

Для вычисления химических потенциалов μ_{n_1} , μ_{p_1} , ($\mu_{p_1} > 0$) для электронов и дырок, фигурирующих в формулах для функций распределения, воспользуемся трансцендентными уравнениями:

$$\frac{n_1(t)}{N_{n_1}(T)} - \Phi(\mu_{n_1}^*) = 0, \quad \frac{p_1(t)}{N_{p_1}(T)} - \Phi(\mu_{p_1}^*) = 0, \quad (15)$$

$$N_{n_1} = \frac{m_c k_B T}{\pi \hbar^2}, \quad N_{p_1} = \frac{m_v k_B T}{\pi \hbar^2}, \quad \Phi_{1/2}(x) = \int_0^\infty \frac{dz}{1 + \exp(z - x)}, \quad (16)$$

где $\mu_{n_1, p_1}^* = \mu_{n_1, p_1}^* / k_B T$, N_{n_1} и N_{p_1} – эффективные плотности состояний соответственно в подзоне 1 КЯ зоны проводимости и подзоны в КЯ валентной зоны.

Таким образом, при математическом моделировании фотонной лавины в гетероструктуре типа I с глубокими КЯ требуется получить самосогласованное решение системы нелинейных дифференциальных уравнений (13) и трансцендентных уравнений (14–17). Сложность этой задачи делает невозможным как качественный анализ динамики системы, подобного [12, 13], так и получение простых формул для пороговых интенсивностей света и времен установления квазиравновесных населенностей в электронной системе. Поэтому будем использовать результаты численных расчетов, которые приводятся ниже.

Оптимизация параметров зонной структуры для уменьшения пороговой интенсивности света, необходимой для запуска фотонной лавины, проводилась в пределах, обусловленных (а) соотношениями между энергетическими параметрами, допускаемыми выбранной моделью эффекта, и (б) значениями параметров, встречающихся в реально существующих структурах. В результате были выбраны следующие значения параметров, фигурирующих в правых частях уравнений (13): $a = 3 \cdot 10^{-7}$ см, $p_{01} = 0$, $p_0 = 3 \cdot 10^{10}$ см $^{-2}$, $W_{31} = 0.02$ пс $^{-1}$, $W_{p_1 p_v} = W_{c3} = 0.01$ пс $^{-1}$, $W_{32} = 0.07$ пс $^{-1}$, $W_{21} = 0.1$ пс $^{-1}$, $\tau_{aug} = 0.06$ пс, $\delta = 0.05$ эВ, $\alpha = 1.2$, $m_v = 0.5m$, $m_c = 0.04m$, $\sigma_{23} = 2$ см 2 /пс $^{-1}$ МВт, $\sigma_{12} = 0.003$ см 2 /пс $^{-1}$ МВт, $\sigma_{p_1 p_1} = \sigma_{3c} = 0.025$ см 2 МВт, $E_g = 1$ эВ, $T = 300$ К, $\sigma_{p_1 n_1} = 4.7 \times 10^4$ см 2 МВт $^{-2}$ пс $^{-1}$, $d_1 = 0.01$ см 2 с $^{-1}$, $d_c = 0.003$ см 2 с $^{-1}$, $U_v = 0.3$ эВ, $U_c = 1.74$ эВ.

Результаты численного решения системы нелинейных уравнений (13) даны на рис. 2 и 3. На рис. 2 приведены зависимости квазиравновесных заселенностей $n_{1,2,3,c}$ и $p_{1,v}$ от интенсивности накачки. Видно, что имеется пороговое значение интенсивности $j = j_{th}$, вблизи которого происходит резкое увеличение заселенностей подзон в КЯ для электронов, подзоны в КЯ для дырок, непрерывного спектра в зоне проводимости и валентной зоне. В данном случае $j_{th} \approx 0.056$ МВт/см 2 .

На рис. 3 изображены зависимости времени установления квазиравновесного распределения электронов τ_{eq} от интенсивности накачки j . Видно, что вблизи $j = j_{th}$ величины τ_{eq} резко возрастают.

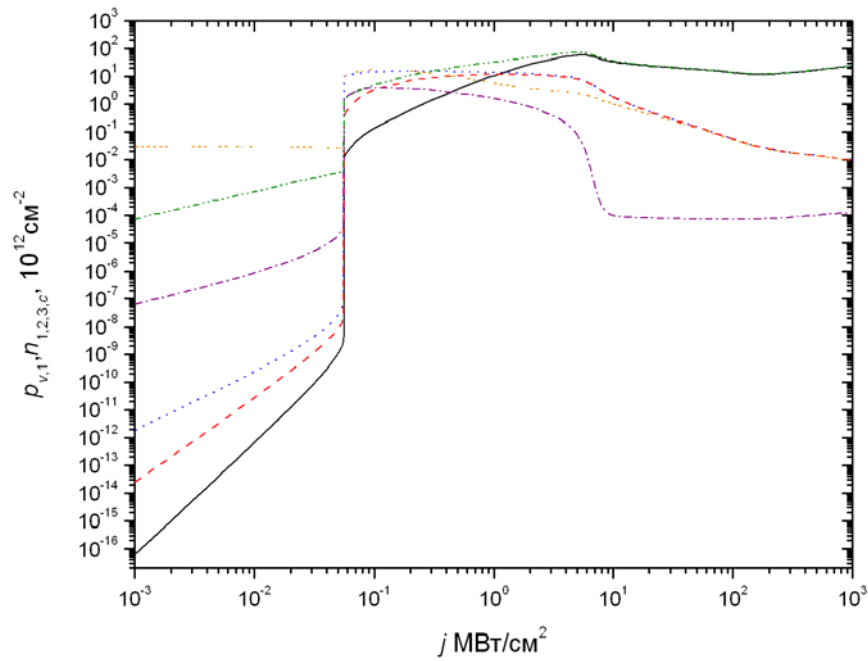


Рис. 2. Зависимость равновесных заселенностей от интенсивности света, где жирная сплошная линия – n_c , штриховая – n_3 , пунктирная – n_2 , штрих-пунктирная – n_1 , штрих два пунктира – ρ_1 , сплошная – ρ_v .

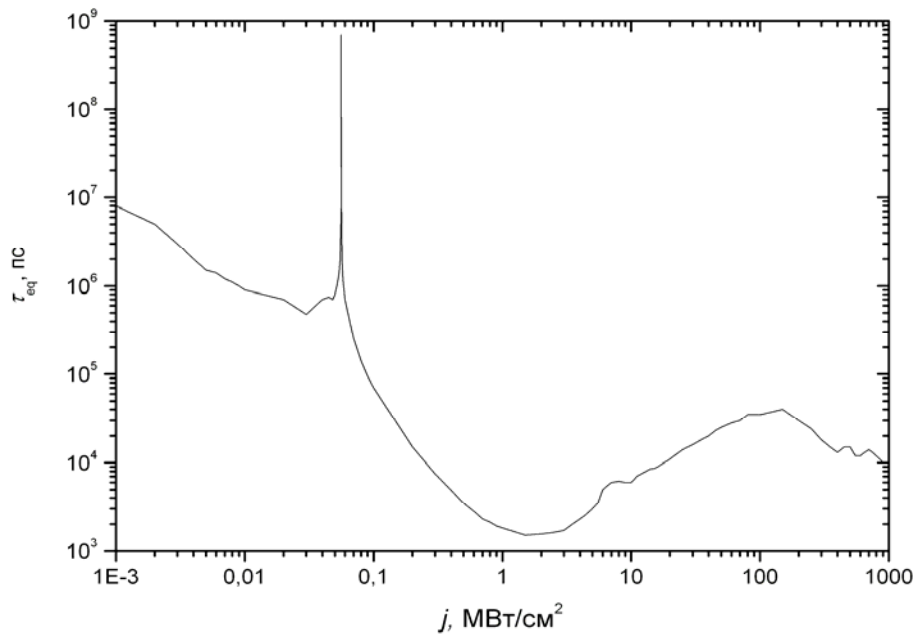


Рис. 3. Зависимость времени установления равновесных заселенностей τ_{eq} от интенсивности света j

Указанные эффекты являются следствием возникновения фотонной лавины. Полученные результаты схожи с результатами для системы с глубокими КЯ типа II [14, 15]. Таким образом, благодаря ЭФЛ можно осуществить генерацию неравновесных ЭДП слабым ИК светом в системе с квантовыми ямами типа I [16]. При этом пороговые интенсивности ИК света, вызывающего лавину, составляют сотни кВт/см².

Следует отметить, что эффект фотонной (точнее, многофотонной) лавины при определенных условиях может играть ключевую роль в оптическом пробое прозрачных широкозонных диэлектриков или полупроводником короткими мощными световыми импульсами (см. [23, 24]).

6. Сравнительный анализ моделей оптического переключения и *up*-конверсии в наноструктурах различных типов

Можно сопоставить параметры переключения на основе эффекта фотонной лавины и переключения на основе оптического эффекта Штарка в наноструктурах ([25–29]). В этом случае переключение можно осуществить за времена 0.1–1 пс, более короткие, чем в случае ЭФЛ, с плотностью энергии в переключающем световом импульсе ~10–100 фДж/мкм². Это сопоставимо с энергией, необходимой для переключения на основе ЭФЛ. Однако в случае штарк-эффекта требуются интенсивности переключающего света, примерно на 3 порядка более высокие, чем в случае ЭФЛ. Таким образом, для работы в пико- и субпикосекундном диапазонах предпочтительнее использовать переключение на основе оптического эффекта Штарка, а для работы, скажем, в наносекундном диапазоне удобнее использовать переключение на основе ЭФЛ в КЯ.

Сопоставим теперь оптическое переключение и *up*-конверсию в квантовых ямах различных типов. В случае легированных КЯ при начальной (двумерной) концентрации свободных электронов в нижней подзоне размерного квантования ~10¹² см⁻² времена переключения в актуальном диапазоне интенсивностей света ($j \sim 1$ МВт/см²) могут составлять от единиц до десятков пикосекунд, а энергии переключения весьма малы (~ 10–100 фДж/мкм²). Весьма низкими являются и пороговые интенсивности света ($j_{th} \sim 10$ –100 кВт/см²). Гетероструктуры с подходящими зонными параметрами достаточно распространены. Существенным недостатком данной схемы является невысокая степень *up*-конверсии: фактически можно получить люминесценцию (и лазерную генерацию) на частоте, менее чем в 2 раза превосходящей частоту накачки.

Этот недостаток преодолевается в моделях, использующих двухфотонную накачку нижней подзоны размерного квантования в яме для электрона в гетероструктурах типов I и II. Так, в структуре с квантовыми ямами типа II можно получить люминесценцию на межзонных переходах на частоте, почти в 5 раз превышающей частоту накачки. При этом пороговые интенсивности света остаются примерно такими же, как в случае легированных квантовых ям. Однако из-за того, что в такой модели требуется время для создания достаточно высокой концентрации носителей в нижней подзоне размерного квантования, времена (а вместе с ними и энергии) переключения оказываются примерно на два порядка большими, чем в случае эффекта фотонной лавины в материале с легированными квантовыми ямами. Это и есть та цена, которую приходится платить за увеличение степени *up*-конверсии. Примером гетероструктуры типа II с глубокими КЯ могут служить системы In_{0.53}Ga_{0.47}As (область A)/AlAs_{0.56}Sb_{0.54} (см. [30, 31]) (зонные параметры этой системы близки к использованным в наших расчетах [14]) либо In_{0.3}Ga_{0.7}As (область A)/AlAs (область B) (см. [32]).

В рассмотренном в настоящей работе эффекте фотонной лавины в гетероструктуре типа I с глубокими квантовыми ямами степень *up*-конверсии может превышать 5, а время (~ 1 нс) и энергия переключения (~ 1 пДж/см²) принимают значения, промежуточные между теми, которые можно реализовать в рассмотренных ранее моделях леги-

рованных КЯ и КЯ типа II. В принципе, такой вариант мог бы считаться оптимальным, однако в настоящее время известна лишь одна гетеросистема CdSe/SrS [33], зонные параметры которой соответствуют требованиям данной модели. Вместе с тем отсутствуют принципиальные ограничения для создания в будущем и других гетеросистем, позволяющих реализовать рассмотренную модель оптического переключения и η -конверсии.

7. Качественный анализ принципиально новых моделей многократной η -конверсии («оптический трамплин») в низкоразмерных системах

В рассмотренных в предыдущих разделах моделях фотонной лавины ключевую роль играли процессы типа Оже, которые обеспечивали «размножение» электронов в возбужденных состояниях. В модели многократной η -конверсии (или оптического трамплина), которая будет рассматриваться в данном разделе, главная роль принадлежит более сложным процессам типа Оже, а именно, процессам второго порядка с участием в элементарном акте фотонов. Поясним модель оптического трамплина на простом примере. Рассмотрим диэлектрик или полупроводник с широкой запрещенной зоной ($E_g \sim 4-6$ эВ). Пусть в кристалле имеется достаточно высокая концентрация глубоких двухуровневых примесных центров. Обозначим энергетический зазор между возбужденным (2) и основным (1) состоянием примесного центра через Δ_{21} . Примем для определенности, что расстояние Δ_{c2} от уровня 2 до дна зоны проводимости c и расстояние Δ_{v1} от потолка валентной зоны v до уровня 1 таковы, что

$$\Delta_{21} < \Delta_{c2}, \Delta_{v1} < 2\Delta_{21}, \quad (17)$$

(см. рис. 4).

Пусть на материал падает свет с частотой $\omega = \Delta_{21}/\hbar$. Каждый из двух соседних примесных центров (A и B), поглощая фотон $\hbar\omega$, переходит из основного состояния 1 в возбужденное состояние 2. Далее может иметь место процесс $A^02+B^02+\hbar\omega \rightarrow A^01+B^+e_c$: электрон на центре A переходит из состояния 2 в состояние 1, передавая высвободившуюся энергию электрону на центре B , который переходит в зону проводимости c , поглощая в том же элементарном акте фотон $\hbar\omega$ (см. рис. 4а). На освобожденное состояние центра B в результате аналогичных процессов может перейти электрон из валентной зоны. Сначала центр A , снова поглощая фотон $\hbar\omega$, переходит из основного состояния 1 в возбужденное состояние 2. Затем происходит процесс $A^02+B^+1+\hbar\omega \rightarrow A^01+B^01+h_\nu$ (см. рис. 4б). В результате всех перечисленных процессов образуется электрон-дырочная пара (электрон в зоне проводимости c и дырка в валентной зоне v), энергия возбуждения которой составляет почти $5\hbar\omega$. При грубой оценке вероятностей рассмотренных процессов считалось, что глубокий примесный центр описывается двухзонной моделью потенциала нулевого радиуса, а взаимодействие между примесными центрами является диполь-дипольным. Было показано, что при величине $\hbar\omega \sim 1$ эВ, концентрации центров $n_1 \sim 10^{20} \text{ см}^{-3}$ предлагаемый здесь механизм генерации электрон-дырочных пар более эффективен, чем прямые 5-фотонные межзонные переходы, по крайней мере, при интенсивностях света $j \sim 10 \text{ ГВт/см}^2$.

Механизм оптического трамплина может действовать и в квантовых наноструктурах. Рассмотрим, например, глубокую легированную квантовую яму с тремя подзонами размерного квантования, нумеруемыми в порядке увеличения энергии как 1, 2 и 3 (рис. 5). Как и в предыдущих разделах, где рассматривался эффект фотонной лавины, считаем, что $\omega_{21} < \omega_{32}$, однако частота света ω теперь выбирается равной частоте перехода ω_{21} между нижней и второй подзонами.

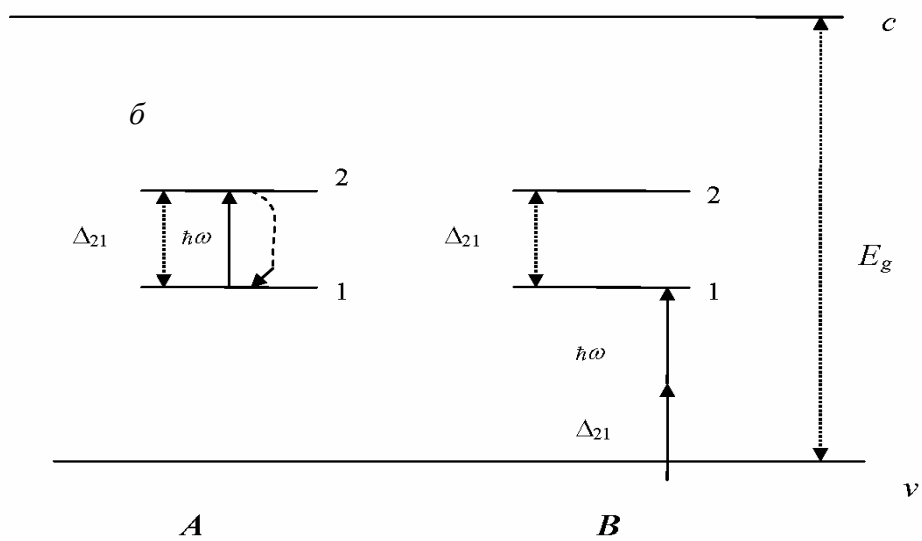
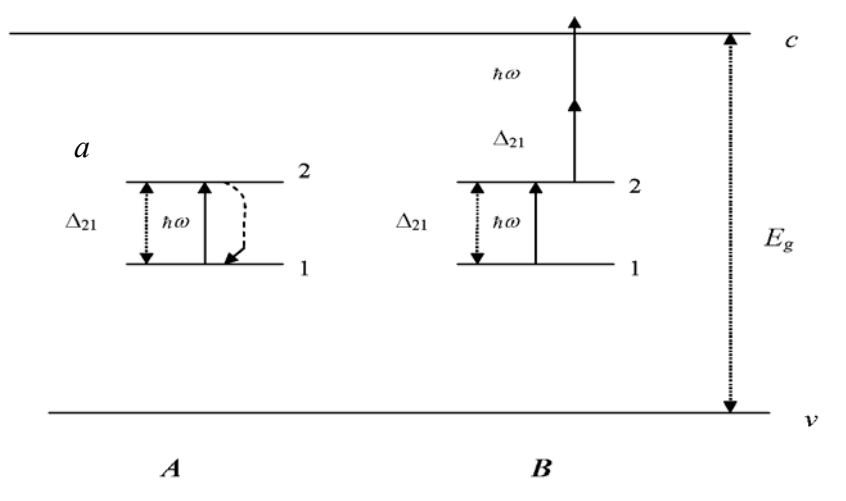


Рис. 4. Схема оптического трамплина в материале с глубокими примесными центрами. Пояснения в тексте

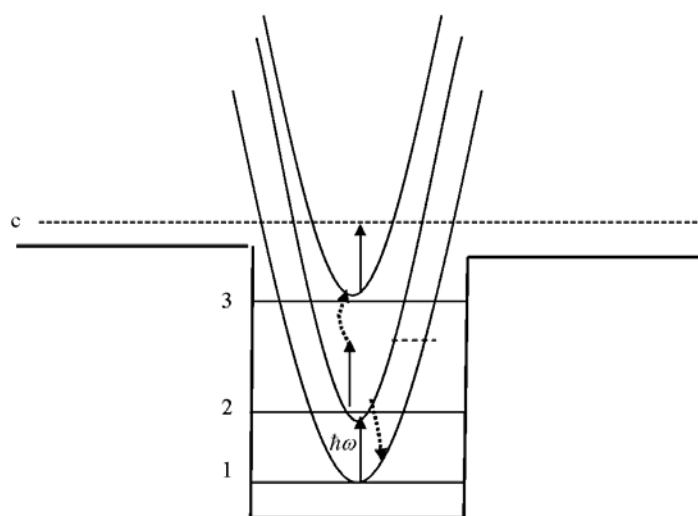


Рис. 5. Схема оптического трамплина в квантовой яме. Пояснения в тексте

При $\hbar\omega_{21} \gg kT$ в отсутствие света практически все электроны находятся в нижней подзоне 1. После включения света электроны могут переходить в подзону 2. Далее начинает работать следующий механизм: электрон из подзоны 2 переходит в нижнюю подзону 1, отдавая высвободившуюся энергию другому электрону из подзоны 2, который переходит в верхнюю подзону 3 с одновременным поглощением фотона $\hbar\omega$.

Электрон, попавший в подзону 3, может поглотить еще один фотон и перейти в состояние непрерывного спектра. При этом возникает фотопроводимость в направлении оси роста наноструктуры. Вероятности переходов $22+\hbar\omega \rightarrow 13$ вычисляются во втором порядке теории возмущений – один порядок по межэлектронному кулоновскому взаимодействию и один порядок по взаимодействию электронов со светом. Оценки показывают, что при концентрации электронов в яме $n_0 \sim 10^{11} \text{ см}^{-2}$, $\hbar\omega \approx 0.4 \text{ эВ}$ переходы указанного типа доминируют над двухфотонными непрямыми переходами между подзонами 2 и 3 при всех интенсивностях света ниже порога разрушения для материалов с легированными квантовыми ямами.

Таким образом, механизм оптического трамплина может быть, в принципе, использован для генерации низкочастотным светом высокоэнергетических возбуждений примесных систем и наноструктур. Очевидно, что этот механизм нуждается в подробном исследовании.

Работа выполнена по Госконтракту № 02.467.11.7006 при частичной поддержке РФФИ, грант 04-02-16175.

Литература

1. J.S. Chivian, W.E. Case, and D.D. Eden. The photon avalanche: a new phenomenon in Pr^{3+} -based infrared quantum counters. // Appl. Phys. Lett. 1979. Vol. 35. № 2. P. 124–126.
2. A.W. Kueny, W.E. Case, and M.E. Koch. Nonlinear-optical absorption through photon avalanche. // JOSA B. 1989. Vol. 6. № 5. P. 639–643.
3. W.E. Case, M.E. Koch, and A.W. Kueny. Photon avalanche in rare-earth crystals. // Journ. of Lumin. 1990. Vol. 45. P. 351–354.
4. H. Ni and S.C. Rand. Avalanche upconversion in $\text{Tm}:\text{YAlO}_3$. // Opt. Lett. 1991. Vol. 16. № 10. P. 14241–14243.
5. A.W. Kueny, W.E. Case, and M.E. Koch. Infrared-to-ultraviolet photon-avalanche pumped upconversion in $\text{Tm}:\text{LiYF}_4$. // JOSA B. 1993. Vol. 10, № 10. P. 1834–1838.
6. S. Guy, M.-F. Joubert, and B. Jacquier. Photon avalanche in the mean field approximation. // Phys. Rev. B. 1997. Vol. 55. № 13. P. 8240–8249.
7. M.-F. Joubert. Photon avalanche upconversion in rare-earth laser materials. // Optical materials. 1999. Vol. 11. P. 181–212.
8. D.B. Gatch, W.M. Dennis, and W.M. Yen. Photon avalanche effect in $\text{LaCl}_3:\text{Pr}^{3+}$ // Phys. Rev. B. 2000. Vol. 62. № 16. P. 1790–1796.
9. M.P. Hehlen, A. Kudicher, et al. Nonradiative dynamics of avalanche upconversion in $\text{Tm}:\text{LiYF}_4$ // Phys. Rev. B. 2000. Vol. 61. № 2. P. 1116–1124.
10. Е.Ю. Перлин, А.М. Ткачук, М.-F. Joubert and R. Moncorge. Каскадно-лавиная up-конверсия в кристаллах $\text{YLF}:\text{Tm}^{3+}$. // Опт. и спектр. 2001. Т. 90. В. 5. С. 772–781.
11. Е.Ю. Перлин. Фотонная лавина в легированной квантовой яме. // Оптика и спектроскопия. 2001. Т. 5. В. 5. С. 777–783.
12. E.Yu. Perlin. Photon Avalanche Effect in Doped Quantum Wells. // Journ. of Luminescence. 2001. Vol. 94–95. P. 249–253.
13. Е.Ю. Перлин, Р.С. Левицкий. Фотонная лавина в легированных квантовых ямах: up-конверсия и эффект переключения. // Оптический журнал. 2006. Т. 73. В. 1. С. 3–11.
14. Е.Ю. Перлин, А.В. Иванов, Р.С. Левицкий. Каскадно-лавиная генерация электрон-дырочных пар в квантовых ямах типа II. // ЖЭТФ. 2003. Т. 123, В. 3. С. 612–624.

15. Е.Ю. Перлин, А.В. Иванов, Р.С. Левицкий. Каскадно-лавинная ап-конверсия и генерация неравновесных электрон-дырочных пар в гетероструктурах типа II с глубокими квантовыми ямами. // *Оптический журнал*. 2006. Т. 73. В. 1. С. 12–21.
16. Р.С. Левицкий, А.В. Иванов, Е.Ю. Перлин. Эффект фотонной лавины в гетероструктурах типа I с глубокими квантовыми ямами. // *Оптический журнал*. 2006. Т. 73. В. 2. С. 3–8.
17. M.H. Degani and O. Hipolito. Electron-phonon interaction effects in a quasi-two-dimensional electron gas in the GaAs-Ga_{1-x}Al_xAs heterostructure. // *Phys. Rev. B*. 1987. Vol. 35. № 14. P. 7717–1722.
18. B.K. Ridley. Electron scattering by confined LO polar phonons in a quantum well. // *Phys. Rev. B*. 1989. Vol. 39. No 8. P. 5282–5289.
19. B.K. Ridley. Electron-hybridon interaction in a quantum well. // *Phys. Rev. B*. 1993. Vol. 47. № 8. P. 4592–4594.
20. K.J. Nash. Electron-phonon interactions and lattice dynamics of optic phonons in semiconductor heterostructures. // *Phys. Rev. B*. 1992. Vol. 46. № 12. P. 7723–7729.
21. C. Trallero-Giner, F. Comas, and F. Garsia-Moliner. Polar optical modes and electron-phonon interaction in semiconductor nanostructures. // *Phys. Rev. B*. 1994. Vol. 50. № 3. P. 1755–1764.
22. J.L. Gondar, F. Comas, and F. Castro. Scattering rates in a semiconductor heterostructure: the effects of intersubband transitions. // *Physica B*. 2000. Vol. 292. P. 354–356.
23. Е.Ю. Перлин, А.В. Иванов, Р.С. Левицкий. Новый механизм предпробойной генерации электрон-дырочных пар в кристаллах: эффект многофотонной лавины. // *Известия РАН*. 2005. Т. 69. В. 8. С. 1133–1135.
24. Е.Ю. Перлин, А.В. Иванов, Р.С. Левицкий. Предпробойная генерация неравновесных электрон-дырочных пар: эффект многофотонной лавины. // *ЖЭТФ. Физ.* 2005. Т. 128, В. 2 (8). С. 411–421.
25. Е.Ю. Перлин. Оптический штарк-эффект при переходном двойном резонансе в полупроводниках. // *ЖЭТФ*. 1994. Т. 105. В. 1. С. 186–197.
26. Е.Ю. Перлин, А.В. Федоров. Двухфотонное поглощение, контролируемое резонансным оптическим штарк-эффектом в кристаллах и квантовых наноструктурах. // *Оптика и спектроскопия*, 1995. Т. 78. В. 3. С. 445–456.
27. Е.Ю. Перлин, А.В. Федоров. Переходный двойной оптический резонанс на экситонах в квантовых наноструктурах. // *Оптика и спектроскопия*, 1995. Т. 78. В. 3. С. 445–456.
28. Е.Ю. Перлин, А.В. Федоров. Квазистационарный оптический штарк-эффект при двойном межзонном резонансе в анизотропных полупроводниках. // *Известия РАН. Сер. физ* 1996. Т. 60. В. 6. С. 164–179.
29. Е.Ю. Перлин, Д.И. Стаселько. Многофотонные переходы и резонансный оптический эффект Штарка в нанокристаллах AgBr. // *Оптика и спектроскопия*. 2005. Т. 98. В. 6. С. 944–950.
30. A. Neogi, T. Mozume, H. Yoshida, O. Wada. Intersubband Transitions at 1.3 and 1.55 μm in a Novel Coupled InGaAs–AlAsSb Double-Quantum-Well Structure. // *IEEE Photon. Technol. Lett.* 1999. Vol. 11. No. 6. P. 634–636.
31. A. Neogi A, H. Yoshida, T. Mozume, N. Georgiev, T. Akiyama, and O. Wada. Absorption saturation of near-infrared intersubband transition in lattice-matched InGaAs/AlAsSb quantum wells // *Physica E*. 2000. Vol. 7. P. 183–187.
32. C.P. Garcia, A. De Nardis, V. Pellegrini, et al. 1.26 μm intersubband transitions in In_{0.3}Ga_{0.7}As/AlAs quantum wells // *Appl. Phys. Lett.* 2000. Vol. 77, № 23. P. 3767–3769.
33. R. Engelmann, J. Ferguson, and R. Solanki. Quantum-well activated phosphors: A new concept for electroluminescent displays. // *Appl. Phys. Lett.* 1997. Vol. 70. № 4. P. 411–413.

ГОЛОГРАФИЧЕСКИЙ ПРЕДСКАЗАТЕЛЬ СЛУЧАЙНЫХ ПРОЦЕССОВ

А.В. Павлов

Показано, что схема Фурье-голографии с обращением волнового фронта в корреляционной плоскости позволяет реализовать модель множественной линейной регрессии при использовании на этапе обратного прохождения света голограммы с инверсной передаточной характеристикой.

Введение

Задача предсказания значения случайного процесса имеет большое прикладное значение. В первую очередь это относится к проблеме управления работой сложных технических и транспортных комплексов – систем водоснабжения, управления уличным движением, управления работой узлов телекоммуникационных сетей [1–3].

Модель

Задача предсказания в теории случайных процессов рассматривается как частный случай задачи наилучшей оценки [4–6]. Модель множественной линейной регрессии [6] для реализации $Im(x)$ стационарного случайного процесса с автокорреляционной функцией $C(\xi)$, наблюдаемого на интервале $[x_{Min}, x_{Max}]$, где x_{Max} – момент или точка наблюдения, определяет значение $Im_{Pred}(x_{Max} + x_k)$ в точке x_k , выражением

$$Im_{Pred}(x_{Max} + x_k) = \int_{x_{Min}}^{x_{Max}} Im(x)a(x)dx, \quad (1)$$

где весовая функция $a(x)$ определяется из решения уравнения

$$\int_{x_{Min}}^{x_{Max}} a(x)C(|x - \xi|)dx = C(x_k + \xi). \quad (2)$$

Для решения задачи дважды применим к (1) фурье-преобразование и теорему Бо-реля о свертке. В результате, учтя инверсию координат в силу нереализуемости в опти-ке обратного Фурье-преобразования, получим

$$Im_{Pred}(-(x_{Max} + x_k)) = F(F(Im(x)) \cdot F(a(x))). \quad (3)$$

Выражение (1.2) в фурье-пространстве примет вид

$$F(a(x)) \cdot F(C(|x - \xi|)) = F(C(x_k + \xi)),$$

откуда получим

$$F(a(x)) = \frac{F(C(x_k + \xi))}{F(C(|x - \xi|))}. \quad (4)$$

Подставив (4) в (3), получим

$$Im_{Pred}(-(x_{Max} + x_k)) = F\left(F(Im(x)) \cdot \frac{F(C(x_k + \xi))}{F(C(|x - \xi|))}\right)$$

и, применив теорему Винера-Хинчина, и учтя инверсию координат в корреляционной плоскости, получим окончательное выражение для значения процесса, даваемого ли-нейным предсказателем, где астериск обозначает комплексное сопряжение:

$$\text{Im}_{Pred}(x_{Max} + x_k) = F^* \left(\frac{F^*(C(x_k + \xi))}{F(\text{Im}(x))} \right). \quad (5)$$

Оптическая схема

Архитектура «4-f схема фурье-голографии с фазосопрягающим зеркалом в корреляционной плоскости» приведена на рис. 1. Во входной плоскости In помещается эталонное изображение Im_A , освещаемое плоским волновым фронтом U . Входная плоскость In есть передняя фокальная плоскость первой фурье-преобразующей линзы L_1 , которая в своей задней фокальной плоскости формирует фурье-образ этого изображения $F(Im_A)$. При записи голограммы в эту плоскость помещается фоточувствительная среда Н, регистрирующая картину интерференции сигнального $F(Im_A)$ и внеосевого плоского опорного пучка R . Передаточная характеристика проявленной голограммы для интересующего нас +1 порядка дифракции $\eta(F(Im_A))$, где η – оператор голографической регистрирующей среды. При помещении во входную плоскость In объектного изображения Im_B за фурье-голограммой Н формируется распределение $F(Im_B)\eta(F(Im_A))$, а в задней фокальной плоскости С второй фурье-преобразующей линзы L_2 формируется функция взаимной корреляции $Im_B \otimes Im_A^\eta$, где \otimes – символ операции корреляции, а Im_A^η – результат фильтрации изображения Im_A оператором η . На этом этапе схема работает как классический коррелятор Ван дер-Люгта.

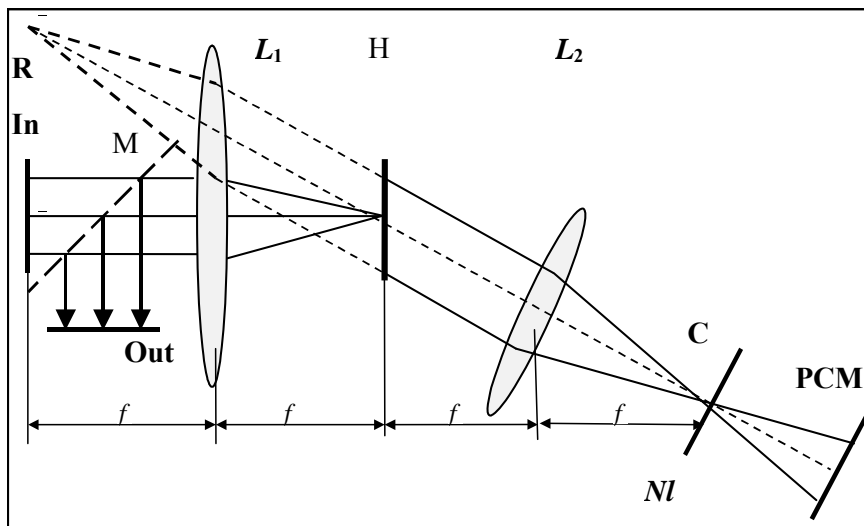


Рис.1. Схема Фурье-голографии с обращением волнового фронта

Для обеспечения прохождения возбуждения в обратном направлении, т.е. $C \rightarrow H \rightarrow \text{Out}$, где Out – выходная плоскость, оптически сопряженная с входной In посредством полупрозрачного зеркала М, в плоскость С (или за ней) помещается устройство обращения волнового фронта РСМ (фазо-сопрягающее зеркало). Нелинейный оператор NI должен быть локализован в плоскости С и подбирается таким образом, чтобы выделить из корреляционного распределения только глобальный максимум автокорреляционной функции (АКФ). Как правило, эта процедура реализуется посредством многоитерационного процесса, для обеспечения которого аналогичное фазосопрягающее зеркало помещается также и в плоскость In. Процесс в плоскости С может быть описан как

$$NI(Im_B \otimes Im_A^\eta) \rightarrow \delta,$$

где δ – дельта-функция, описывающая дифракционно-ограниченный точечный источник. В результате такой процедуры в обратном ходе лучей $C \rightarrow H \rightarrow \text{Out}$ глобальный

максимум АКФ как точечный опорный источник восстанавливает в Out изображение Im_A^{η} .

Таким образом, схема рис. 1. реализует модель автоассоциативной памяти. Эта схема хорошо известна под названием голографической автоассоциативной памяти. Отметим, что само эталонное изображение Im_A , с которого была записана голограмма H, может быть восстановлено только при линейном операторе η , последнее для большинства реальных изображений физически невыполнимо.

Нетрудно видеть, что выражение (5) может быть реализовано в обсуждаемой оптической архитектуре посредством процедуры, состоящей из следующих двух этапов.

1 этап – формирование в слое С распределения $C(x)$. На этом этапе НС работает как классический голографический коррелятор Ван-дер-Люгта и фурье-голограмма H согласована с эталоном, т.е. $H(\nu) = F(Im_A^{\eta})$.

2 этап – прохождение возбуждения от выделенного из АКФ фрагмента $C(x_k + \xi)$ в обратном направлении $C \rightarrow H^{-1} \rightarrow I$ через матрицу связей с инверсной передаточной характеристикой $H^{-1}(\nu) = (F(Im_A^{\eta}))^{-1}$. Методы реализации инверсных голограмм известны и заключаются, в частности, в использовании голографических регистрирующих сред с обратной зависимостью дифракционной эффективности от экспозиции. В результате реализации второго шага амплитуда света в точке x_k плоскости Out будет пропорциональна искомому значению $Im_{Pred}(x_{Max} + x_k)$.

В следующем разделе мы рассмотрим структуру связей, реализуемую этой архитектурой, и увидим, что требование на выделение на втором шаге фрагмента $C(x_k + \xi)$ из автокорреляционной функции выполняется автоматически в случае, если голограмма обладает свойством угловой инвариантности.

Структура связей

Схема рис. 2а реализует архитектуру двуслойной нейронной сети (НС) со связями только между слоями I и С. Слой I объединяет оптически сопряженные плоскости In и Out. Нейрон может быть реализован как в виде дискретного элемента матричной структуры (пиксела), так и в виде дифракционно-ограниченного элемента разрешения для непрерывной среды, размер которого определяется в соответствии с теоремой Котельникова.

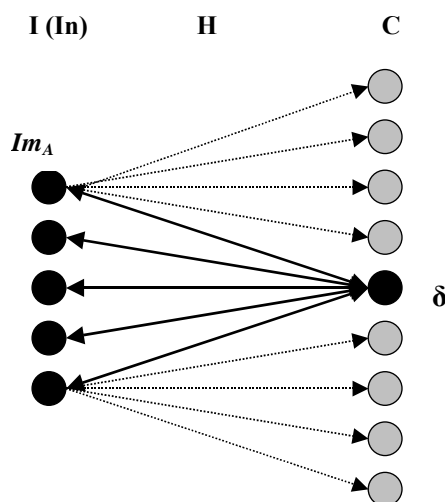


Рис. 2а. НС типа «звезда Гроссберга»

Операция суммирования входов такого нейрона осуществляется в соответствии с принципом суперпозиции, связи в свободном пространстве формируются в соответст-

вии с принципом Гюйгенса-Френеля, а нелинейная активационная функция реализуется обычно за счет нелинейных свойств самой регистрирующей среды. Матрица связей нейронных слоев I и C реализуется фурье-голограммой H. Особенность голографического метода реализации межнейронных связей заключается в том, что, в отличие от НС с локализованными связями, голограмма восстанавливает более сложную структуру связей, нежели структура, сформированная при обучении НС. Использование именно этой особенности и позволило нам реализовать модель предсказателя.

Основные отличия структуры связей, восстанавливаемой Фурье-голограммой, от фиксированных связей состоят в следующем.

Во-первых, если при обучении предъявлялось звезда из N_A лучей, то в результате действия оператора η восстанавливается звезда из N_A^n лучей, где N_A^n – число нейронов в изображении Im_A^n . Это отличие актуально для любого типа голограмм – как «тонких», так и объемных.

Второе существенное для рассматриваемой задачи отличие имеет силу только для «тонкой» голограммы и заключается в том, что проявленной голограммой воспроизводится единственная «звезда Гроссберга» только в том случае, если голограмма объемна, т.е. обладает свойством угловой селективности. Если голограмма тонкая, то, поскольку каждый I-нейрон независимо от других I-нейронов является точечным источником, в силу свойств угловой инвариантности тонкой голограммы и независимости распространения световых волн при предъявлении во входной плоскости изображения Im_B звезда «outstar» $\delta \rightarrow Im_A^n$ восстанавливается независимо для каждого активированного I-нейрона. Размер I-нейрона на этапе предъявления обученной сети изображения Im_B определяется оператором η . В результате формируется N_B независимых звезд Гроссберга «outstar», где N_B – размерность (число нейронов) Im_B , показанных на рис. 2б пунктиром.

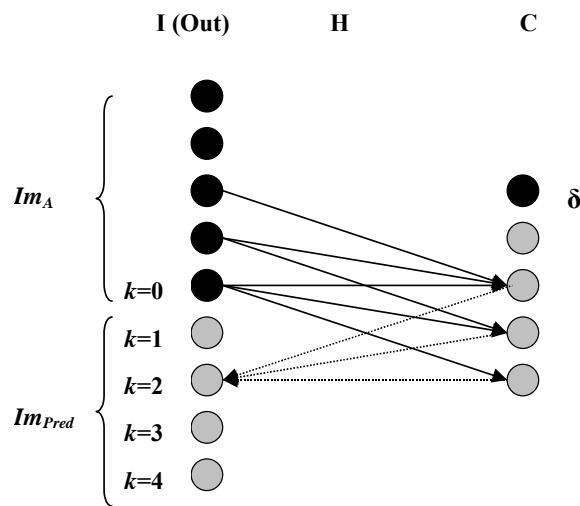


Рис. 2б. Формирование независимых звезд Гроссберга «outstar»

Каждый I-нейрон через свою звезду активирует N_A^n нейронов в слое C (на рис. 2б для простоты пунктиром показаны только связи от крайних I-нейронов). Всего в слое C активируются $(N_A^n + N_B - 1)$ C-нейронов. При этом только активированный глобальным максимумом АКФ δ -нейрон связан со всеми N_A^n I-нейронами, остальные C-нейроны, затухеванные на рис. 2б серым, связаны каждый с $(N_B - n)$ I-нейронами, где n – порядковый номер C-нейрона относительно δ -нейрона, принятого за начало отсчета.

При обращении волнового фронта в слое C для каждого C-нейрона как независимого точечного источника голограмма также формирует звезду «outstar». В результате

каждый С-нейрон активирует N_A^n связей к I-нейронам, а общее количество активированных I-нейронов, которые могут быть активированы этими связями, составляет $(2N_A^n + N_B - 2)$.

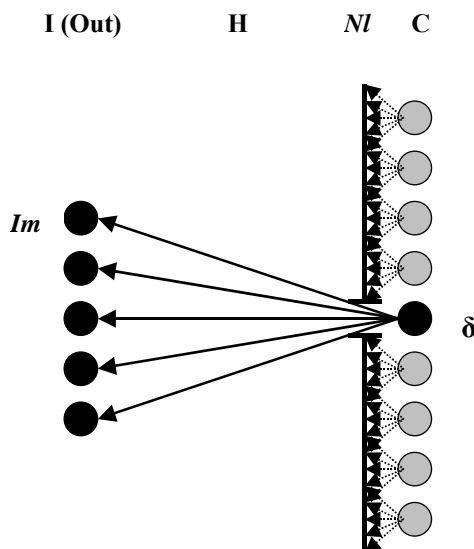


Рис. 2в. Восстановление одной звезды

Для реализации модели ААП из этого массива связей слоев $C \rightarrow I$ посредством нелинейного оператора Nl , действующего в слое C , выделяется единственная звезда $\delta \rightarrow Im_A^n$, восстанавливающая Im_A^n (сплошные стрелки на рис. 2в), а связи от нейронов, активированных боковыми максимумами корреляционной функции (на рис. 2в веера из коротких стрелок), режектируются. Если эти связи не режектируются, то созданный ими ореол вокруг изображения Im_A^n (на рис. 2.в I-нейроны, затушеванные серым) в оптике традиционно трактуется как шум.

Как следует из рассмотрения, проведенного выше, именно эти, режектируемые при реализации ААП, связи от С-нейронов, возбужденных боковыми максимумами корреляционной функции, и необходимы для реализации модели линейного предсказания (5).

На рис. 2в приведена структура связей, обеспечивающих формирования решения (1.5). Для упрощения рисунка приведена только одна (нижняя) половина. Из рис. 2.в видно, что k -тый I-нейрон получает возбуждение от $(N_A^n - k)$ С-нейронов и через них связан с $(N_A^n - k)$ I-нейронами, активированными Im_A^n . Иными словами, предсказание на глубину k вычисляется по базе $(N_A^n - k)$.

Экспериментальная реализация

Модель была экспериментально проверена в схеме рис. 1 на примере аэрофотоизображения леса (рис. 3).

Для записи голограммы использовался центральный фрагмент, ограниченный на рис. 3 вертикальными белыми линиями. Размер эталона составлял 650×650 пикселей.

Голограмма фазосопрягающего зеркала РСМ располагалась не в корреляционной плоскости, а за ней, на расстоянии, на котором распределение амплитуд по сечению пучка было примерно равномерным. Такое решение было обусловлено невозможностью записи линейной голограммы корреляционного поля в плоскости C в силу того, что перепад амплитуд в корреляционном поле существенно превышает динамический

диапазон регистрирующих сред. Принятая схема обеспечила возможность режекции ГМ АКФ не на этапе записи РСМ, а на этапе восстановления голограммы.

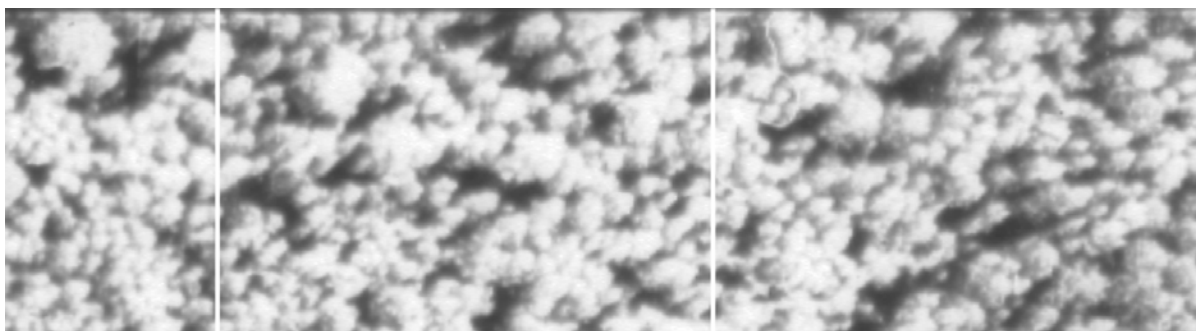


Рис. 3. Аэрофотоизображение леса, использованное в экспериментах

Как следует из (5), в построении предсказания участвуют только боковые максимумы корреляционной функции. В то же время яркость изображения эталона, восстанавливаемого ГМ АКФ, настолько превосходит яркость ореола-предсказания, что «забывает» телевизионный сенсор. Поэтому на этапе восстановления голограммы для обеспечения возможности регистрации изображения ГМ АКФ также режектировался.

Для реализации модели необходима именно тонкая голограмма. Принципиальная ограниченность дифракционной эффективности тонкой голограммы, вкупе с использованием голографической техники для реализации фазо-сопрягающего зеркала, ведет к существенным энергетическим потерям в схеме рис.1. Эти потери усугубляются применением техники инверсной голограммы. Поэтому для обеспечения интенсивности восстановленного изображения (5), достаточной для его регистрации телевизионным сенсором, вместо инверсной голограммы вынужденно была использована схема узкополосной фильтрации с записью фурье-голограммы в режиме переэкспозиции. В результате в восстановленном изображении были утрачены практически все детали, за исключением контуров крон.

На рис. 4 приведен фотометрический профиль фрагмента изображения рис. 3 вне пределов кадрового окна (белая кривая) и фотометрический профиль ореола, обрамляющего восстановленное в схеме рис. 1 изображение эталона (черная кривая), т.е. предсказания. По оси X размер дан в пикселях, эталон занимал участок от 0 до 650 пикселей.

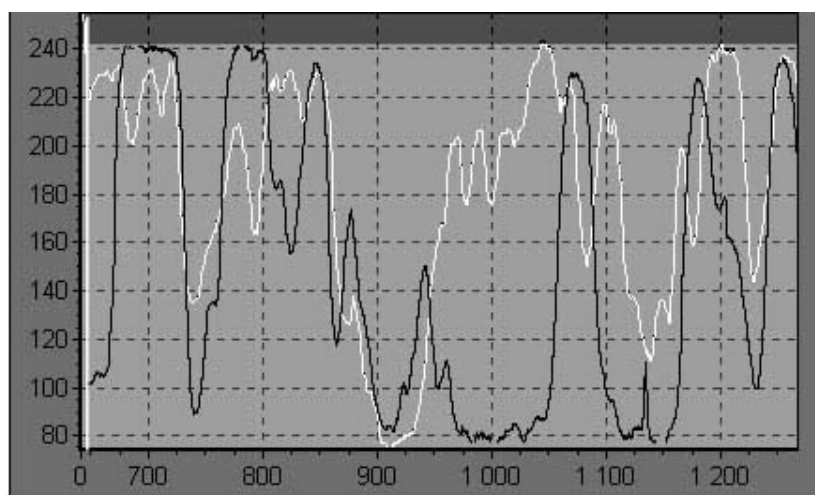


Рис. 4. Фотометрический профиль оригинального аэрофотоизображения рис. 3 (белая кривая) и восстановленного ореола – предсказания (черная кривая)

Заключение

Таким образом, теоретически и экспериментально показано, что схема Фурье-голографии с обращением волнового фронта строит линейное предсказание в случае использования на этапе обратного прохождения голограммы с инверсной передаточной характеристикой.

Литература

1. Pedricz W., Vasilakos A., Linguistic Models and Linguistic Modeling// IEEE Trans. On Systems, Man, and Cybernetics, Part B. 1999. V.29. 6.
2. Grossglauser M., Bolot J.C., On the Relevance of Long-Range Dependence in Network Traffic, //IEEE Trans. on Networking. 1999. V.7. №5.
3. Roughan M., Veitch D., Abry P., Real-Time Estimation of the Parameters of Long-Range Dependence// IEEE Trans. on Networking. 2000. V.8. №4.
4. Вентцель Е.С., Овчаров Л.А. Теория случайных процессов и инженерные приложения, М.: Высшая школа, 2000.
5. Вентцель А.Д., Курс теории случайных процессов, М.: Наука, 1975.
6. Grimmet G.R., Sterzaker D.R. Probability and Random Processes. Oxford. Oxford Sc. Publ., Claredon Press, 1992.

РАЗРАБОТКА ВИДЕОСПЕКТРАЛЬНЫХ ОСНОВ АЭРОКОСМИЧЕСКИХ МЕТОДОВ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ В ОПТИЧЕСКОМ ДИАПАЗОНЕ

К.Н. Чиков, В.В. Гуд, В.М. Красавцев

Рассматриваются особенности построения и методологические возможности бортовых видеоспектрометров, предназначенных для решения экологических и геологических задач. При этом ряд важных экологических проблем эффективно решается с использованием результатов дистанционного видеоспектрального мониторинга биосферы в целях выявления выбросов вредных веществ в атмосферу и обнаружения на ранних стадиях неблагоприятных воздействий на окружающую среду при оперативной оценке последствий стихийных бедствий, техногенных аварий, лесных пожаров. Многие проблемы геологии решаются с привлечением подробного видеоспектрального картирования площадей, на которых выделяются выходы на поверхность горных пород разного возраста и состава, что повышает результативность природно-ресурсных изысканий при значительной экономии сил, средств и времени.

Введение

Современная тенденция к организации глобального спектрального мониторинга, направленного на обнаружение и изучение природных и антропогенных образований, базируется на приборном обеспечении, позволяющем оперативно решать задачи их поиска и ситуационного анализа на возможно большей территории.

Выполнению поставленных задач отвечают в полной мере спектральные и оптико-физические дистанционные методы и средства, применение которых при исследовании естественных и искусственных объектов не искажает поля измеряемых характеристик, в частности, коэффициентов спектральной яркости (КСЯ), что повышает достоверность получаемой информации. Среди таких методов и средств одними из самых перспективных могут считаться видеоспектрометры (imaging spectrometers) и методы, основанные на их применении. Как правило, эти приборы устанавливаются на борту авиационных и космических носителей [1].

Видеоспектрометры отличаются от аналогичных классических приборов тем, что помимо обычной спектральной информации – измерения КСЯ – позволяют получать высококачественные панорамные изображения исследуемого объекта во многих различных спектральных интервалах. При этом видеоспектрометры, в отличие от аналогов, не интегрируют КСЯ со всей поверхности объекта, а обеспечивают поэлементную регистрацию КСЯ. Такая отличительная черта видеоспектрометров делает их наиболее пригодными для ресурсно-сырьевого картирования и выявления объектов повышенной техногенной опасности [2–4].

Проведенные с помощью видеоспектрометров в последние два десятилетия детальные исследования КСЯ природных и антропогенных образований показали, что небольшие различия в регистрируемых спектрах отображают характерные особенности важных объектов и явлений.

Сюда относятся, например, рудоконтролирующие зоны слабых вторичных гидротермальных изменений горных пород; протяженные геохимические аномалии над залежами углеводородов; стресс растительности при неблагоприятном воздействии нитратов и осадков тяжелых металлов из атмосферы; начальные стадии утечек углеводородов из магистральных продуктопроводов; «залповые» выбросы загрязнений и т.п. Было получено подтверждение, что тонкие спектральные различия, являющиеся важнейшими индикаторами ранних и предвестниками значительных изменений состояния биоресурсов, могут быть уверенно зафиксированы только в режиме видеоспектрометрии.

Основные результаты

Существующие бортовые системы дистанционного зондирования, построенные по традиционным схемам и предназначенные для получения панхроматических (черно-белых и цветных) изображений зоны обзора в отдельных широких спектральных интервалах, приближаются к своим предельным информационным возможностям по обнаружению и распознаванию природных и антропогенных образований. Это обусловило в НИГ «Коспектр» СПбГУ ИТМО развитие такого стратегически важного и особо приоритетного направления аэрокосмической техники, как видеоспектральное приборостроение. Данное обстоятельство поясняется тем, что аэрокосмические видеоспектральные средства позволяют не только провести детальный анализ получаемого многоэлементного изображения земной поверхности, но и выполнить регистрацию спектров каждого изображаемого элемента, даже находящегося на пределе пространственного разрешения. Таким образом, отличительной чертой современных видеоспектрометров является совмещение в них высокого пространственного и спектрального разрешений.

Первая характеристика этих видеоспектрометров – пространственное разрешение – определяется минимальными размерами исследуемого участка земной поверхности, который данная аппаратура распознает как отдельный элемент.

Вторая характеристика – спектральное разрешение – определяется способностью аппаратуры различать излучения ближайших соседних спектральных интервалов (например, пользуясь терминологией видимого диапазона – это способность прибора опознавать тонкие цветовые оттенки смежных, очень близких по цвету объектов).

Особенностью аппаратурного построения видеоспектральных средств, содержащих обычно видеоспектрометр и специальный телевизионный модуль, является их объединение в бортовой аэрокосмической информационно-измерительный комплекс. При этом специальный телевизионный модуль позволяет оператору на борту носителя или в камеральных (лабораторных) условиях анализировать в реальном масштабе времени всю наблюдаемую в данный момент картину.

Видеоспектрометр, как правило, включает в себя проецирующий объектив и полихроматор (спектральную часть, снабженную, в частности, дифракционной решеткой). В этих приборах входной проецирующий объектив строит изображение исследуемой местности в неразложенном свете на входе полихроматора в плоскости его входной щели. На выходе полихроматора (рис. 1) формируется континуум монохроматических изображений его входной щели, представляющий собой многополосный спектральный кадр, каждая монохроматическая полоса которого совпадает со строкой применяемого двухкоординатного приемника излучения. Такой приемник в настоящее время чаще всего реализуется на базе прибора с зарядовой связью, т.е. ПЗС-матрицы, содержащей N строк и M столбцов. При этом каждая отдельная строка матричного (двухкоординатного) приемника, располагаемая строго поперек направления полета носителя (т.е. ортогонально маршруту аэросъемки), фактически регистрирует детальное монохроматическое изображение узкой полосы земной поверхности, оптически сопряженной с входной щелью полихроматора. Поэтому сколько таких строк имеет матричный приемник – столько, в принципе, можно зарегистрировать монохроматических узкополосных изображений участка местности, над которой пролетает носитель. Так как входная щель полихроматора оптически сопряжена с земной поверхностью, то длина этой щели определяет протяженность полосы обзора на местности с углом захвата θ (рис. 1), а ширина щели – спектральное и пространственное разрешения.

Направление входной щели полихроматора и направление его дисперсии взаимно перпендикулярны. Поскольку поверхность матричного (двухкоординатного) приемника и входная щель полихроматора расположены в оптически сопряженных плоскостях, то

реальное пространственное разрешение прибора по полосе обзора на местности определяется элементом структуры приемника излучения при максимальном числе таких пространственно разрешаемых элементов, равном M (т.е. количеству столбцов принимаемого матричного приемника).

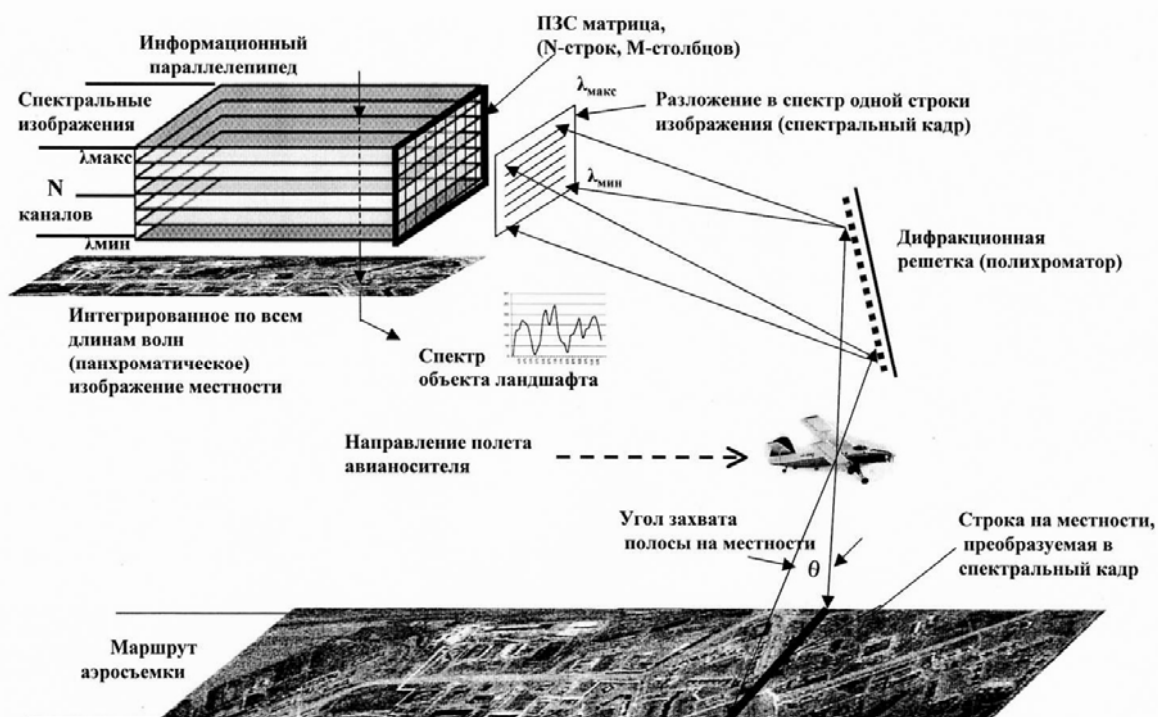


Рис.1. Принцип действия авиационного видеоспектрометра.

В видеоспектрометрах в основу аппаратно-программной обработки получаемой видеoinформации положена модель информационного параллелепипеда (рис. 1). На торце этого параллелепипеда отображается многополосный спектральный кадр как результат разложения в спектр (по направлению спектральной оси: от $\lambda_{\text{мин}}$ до $\lambda_{\text{макс}}$) изображения участка местности в виде узкой полосы, сформированной проецирующим объективом в плоскости входной щели полихроматора.

Это изображение полосы местности (также как и длинная сторона входной щели полихроматора) обычно располагается ортогонально маршруту аэросъемки. И если бы не было разложения в спектр, то изображение данной полосы местности занимало бы лишь одну строку ПЗС-матрицы. Однако спектральное разложение приводит к тому, что количество строк N в ПЗС-матрице (каждая из которых по ширине ориентирована в направлении спектральной оси) равно количеству N разрешаемых спектральных интервалов, называемых спектральными каналами, содержащими детальную информацию в каждой из получаемых монохроматических полос с изображением этой узкой полосы местности.

Вследствие движения носителя шаг за шагом происходит обновление изображения узкого участка местности во входной щели полихроматора. И при каждом таком шаге отображается на ПЗС-матрице новый многополосный спектральный кадр, что вызывает послойный рост информационного параллелепипеда в направлении, перпендикулярном его торцевой плоскости (рис.1).

Результирующие монохроматические видеокadres, получаемые в процессе «пошагового» движения носителя, создаются из слоев, параллельных торцевой плоскости информационного параллелепипеда, путем изъятия из этих слоев одних и тех же вы-

бранных монохроматических строк. После изъятия из этих слоев соответствующих монохроматических строк, с целью получения любого итогового монохроматического видеокадра, адекватные горизонтальные монохроматические строки (адекватные спектральные каналы) последовательно как бы «сшиваются бок о бок». При этом пользователь, исходя из панхроматического изображения местности, может получить спектр любого объекта ландшафта, даже если этот объект находится на пределе пространственного и спектрального разрешений прибора (рис. 1).

Данная модель в виде информационного параллелепипеда позволила упростить разработку алгоритмов для аппаратно-программной обработки видеоинформации.

Таким образом, видеоспектрометры, разрабатываемые в НИГ «Коспектр» СПбГУ ИТМО, обеспечивают разложение в спектр узких протяженных участков земной поверхности, расположенных ортогонально трассе полета носителя и изображаемых проецирующим объективом во входной щели полихроматора. Движение носителя позволяет формировать полноразмерные монохроматические видеокадры в виде последовательности соответствующих монохроматических изображений узких полос на местности, распределенных непрерывно по маршруту аэросъемки поперек трассы полета носителя. При этом реализованные видеоспектрометры совместили в себе высокое пространственное разрешение, свойственное аэрофотоаппаратам, и высокое спектральное разрешение, присущее известным классическим спектральным приборам [2]. Оптическая схема одного из таких видеоспектрометров, успешно прошедших летные полевые испытания [3], представлена на рис. 2.

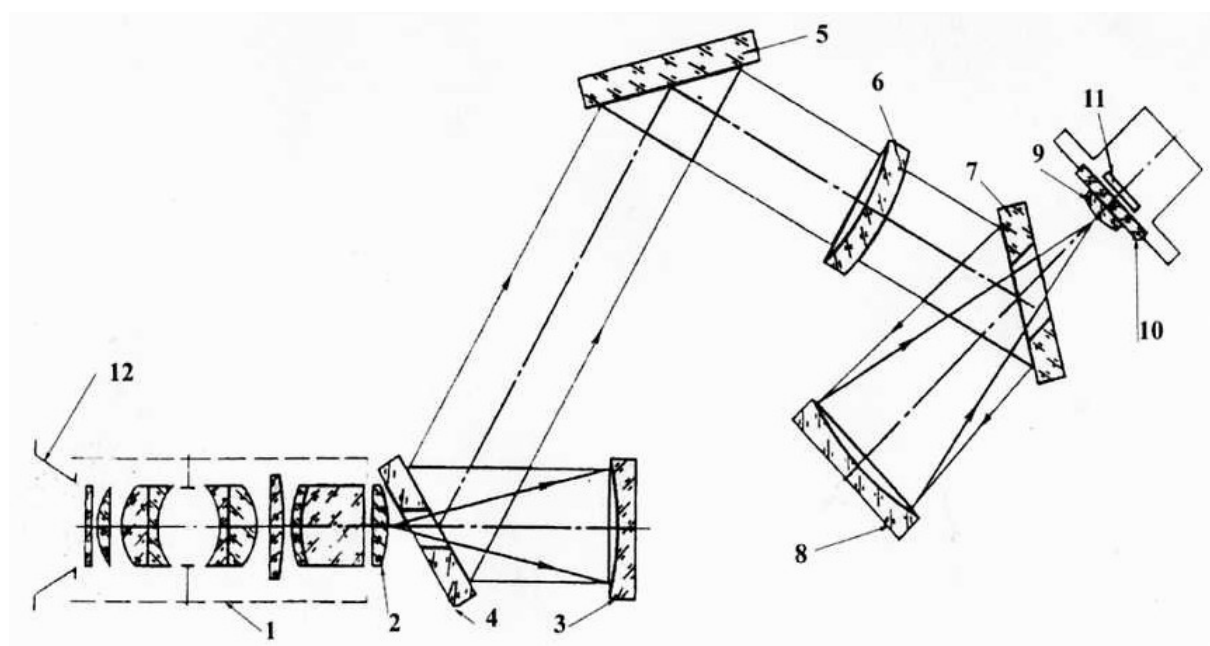


Рис. 2. Оптическая схема видеоспектрометра видимого и ближнего инфракрасного (ВБИК) диапазона

В видеоспектрометре ВБИК диапазона (рис. 2) полихроматор, содержащий элементы 2–9, расположен за входным проецирующим объективом 1. Этот полихроматор состоит из входной щели, совмещенной с плосковыпуклой линзой 2, коллиматорного объектива – вогнутого сферического зеркала 3, плоской отражательной дифракционной решетки 5, вогнуто-выпуклого мениска 6, камерного объектива – вогнутого сферического зеркала 8 и составной выпукло-плоской линзы 9, выполненной из двух соединенных встык полулинз. Плоские зеркала 4, 7 введены в оптическую схему для уменьшения габаритов системы.

Оптическая схема полихроматора (рис. 2) имеет следующие особенности. Плосково-выпуклая линза 2 компенсирует кривизну поля коллиматорного объектива 3 и одновременно выполняет функцию полевой линзы (коллектива). Кроме того, линза 2 обеспечивает совмещение изображения апертурной диафрагмы проецирующего объектива 1 с оправой плоской отражательной дифракционной решетки 5. Плоская поверхность линзы 2 содержит нанесенную на эту поверхность полевую диафрагму в виде узкой входной щели полихроматора и при этом совпадает как с задней фокальной плоскостью проецирующего объектива 1, так и с передней фокальной плоскостью коллиматорного объектива 3. Оправа дифракционной решетки 5 является выходным зрачком коллиматорного объектива 3 и одновременно входным зрачком камерного объектива 8. Мениск 6 исправляет сферохроматическую aberrацию одновременно у трех объективов – проецирующего 1, коллиматорного 3 и камерного 8. Составная выпукло-плоская линза 9, помимо устранения кривизны поля, выполняет фильтрацию высших дифракционных порядков за счет того, что она составлена из двух полулинз из цветных стекол ЖС-10 и ОС-14. Линза 9 крепится на защитное (входное) окно 10 ПЗС-матрицы 11.

При работе видеоспектрометра ВБИК диапазона (рис. 2) излучение от земной поверхности, пройдя после бленды 12 через проецирующий объектив 1, входную щель полихроматора, нанесенную на плоскую поверхность линзы 2, коллиматорный зеркальный объектив 3, плоское зеркало 4, плоскую дифракционную решетку 5, корригирующий мениск 6 и плоское зеркало 7, попадает на камерный зеркальный объектив 8. Этот объектив 8 строит в своей фокальной плоскости многополосный спектральный кадр из развернутых дифракционной решеткой монохроматических изображений, которые соответствуют вырезанной входной щелью узкой полосе на картине поверхности, отображаемой проецирующим объективом 1 в плоскости входной щели полихроматора.

Полученный многополосный спектральный кадр считывается ПЗС-матрицей 11 через закрепленную на окне 10 составную выпукло-плоскую линзу 9. При этом монохроматические строки считываемого многополосного спектрального кадра полностью отвечают видеоспектральному изображению одного и того же узкополосного участка земной поверхности, но в разных спектральных интервалах. Каждая такая монохроматическая строка регистрируется соответствующей строкой ПЗС-матрицы и записывается отдельно в определенном месте запоминающего устройства, отведенном для данной рабочей длины волны.

При полете носителя с видеоспектрометром, т.е. при осуществлении сканирования по пространству, проекция входной щели полихроматора перемещается шаг за шагом по земной поверхности, что приводит к обновлению изображений в монохроматических строках. Это позволяет сформировать в запоминающем устройстве полноразмерные монохроматические видеокадры для каждой рабочей длины волны из складываемых (или, другими словами, «сшиваемых бок о бок») монохроматических строк, сообразно выбранной рабочей длине волны. При этом каждая из таких строк, «сшиваемых» по ее длинной стороне, отвечает очередному шагу при сканировании по пространству.

Из полученных монохроматических видеокадров с исследуемыми удаленными объектами пользователями осуществлялась дискретная выборка только тех, которые были необходимы для обнаружения, определения и изучения определенных свойств интересующих потребителя конкретных объектов.

Полевые летные экспериментальные работы [3] проводились на самолете АН-30. Этот тип самолета является целевым для проведения аэросъемочных работ. В фюзеляже самолета имелись фотолюки с оптическими стеклами, обеспечивающими герметичность кабины. Крепление оптического блока видеоспектрометра не вызвало трудностей. Установка прибора производилась к штатным местам крепления, предназначен-

ным для аэрофотосъемочной аппаратуры, как показано на рис. 3, где приведены также технические характеристики видеоспектрометра ВБИК диапазона.



Рис. 3. Летный эксперимент с видеоспектрометром ВБИК диапазона

Полеты выполнялись в середине дня в условиях хорошей освещенности на высотах 600 м, 1000 м и 1500 м при наличии на отдельных маршрутах верхней кучевой облачности.

Для получения полноразмерных видеокладов реализован программно-аппаратный комплекс, включающий в себя четыре основных функциональных компонента:

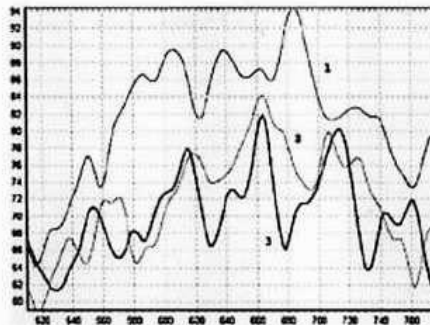
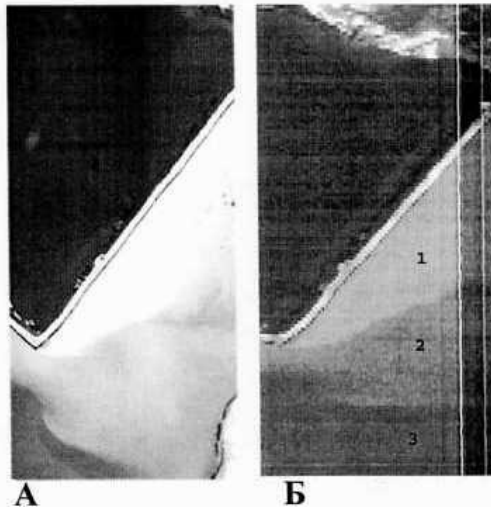
- **анализатор изображения** – ПЗС-матрицу, выполняющую покадровое преобразование свет-сигнал изображения, формируемого на выходе полихроматора;
- **плата видеозахвата**, являющаяся контроллером ввода в оперативную память сигнального кадра, соответствующего анализируемому изображению;
- **компьютер**, в оперативной памяти которого происходит обработка информации;
- **программы управления и обработки информации**, образующие систему отображения и архивации информации видеоспектрометра.

Формирование полноразмерных видеокладов выполнялось в камеральных условиях после аппаратно-программной обработки видеоинформации по отдельным выбранным участкам с наилучшим качеством видеоспектральных изображений различных ландшафтных ситуаций и визуально наблюдаемыми различиями спектральных характеристик ряда элементов ландшафта. Подбор таких участков проводился путем одновременного просмотра на экране монитора двух видеоспектральных изображений с привлечением результатов сопоставления их с данными видеосъемки и визуальных наблюдений.

На рис. 4 представлены некоторые итоги камеральной обработки, которые включают пример видеоспектральной аэросъемки участка прибрежного ландшафта и водной поверхности с тремя выбранными точками спектрометрирования, а также цифровую фотографию той же картины местности.



Черно-белое суммированное (панхроматическое) изображение всех каналов видеоспектрометра



В – Спектры уходящего излучения точек 1, 2, 3.

А – Цветная цифровая фотография участка
Б - Выбор точек спектрометрирования на панхроматическом изображении -1, 2, 3

Рис. 4. Пример видеоспектральной аэросъемки водной поверхности и прибрежной зоны, выполненной с применением видеоспектрометра ВБИК диапазона

На рис. 4 показан участок береговой акватории вблизи промышленного предприятия. Небольшая дамба здесь отделяет сильно загрязненные взвешенными частицами водные пространства от более чистой части акватории. Последняя на изображениях всех спектральных каналов выделяется как очень темная часть картины, благодаря низкому коэффициенту отражения чистой воды. Загрязненные воды выглядят во всех спектральных каналах светлыми, так как взвеси отражают излучение во всех каналах почти одинаково, хотя наиболее четко два вида загрязненных вод разделяются в изображениях зеленых каналов. Из рис. 4 также видно, что цифровая фотография той же местности не способна передать детальное изображение картины со спектральными различиями отдельных ее фрагментов.

Большое значение для решения экологических задач первостепенной важности приобретает разработка и совершенствование видеоспектральной аппаратуры. Это связано с высокой эффективностью ее применения в случае, когда, например, необходим анализ загрязняющих веществ, попадающих в атмосферу с дымовыми газами тепловых электростанций (ТЭС), теплоэлектростанций (ТЭЦ) и других аналогичных источников. Выбрасываемые в атмосферу вредные вещества при этом можно условно классифицировать на две группы. К первой относятся вещества, количество которых определяется в основном составом топлива (оксиды серы, летучая зола, токсичные примеси, содержащиеся в золе, соединения ванадия). Во вторую группу входят вещества, образование которых в значительной мере зависит от технологии, в том числе от режимов сжигания топлива (оксиды азота и углерода, канцерогенные вещества). Выявление источников таких выбросов и внедрение средств, контролирующих вредные вещества, особенно второй группы, позволяет достигнуть оптимальности режимов сжигания топлива, что приводит к его экономии при одновременном снижении вредных выбросов в атмосферу.

Повышенное внимание экологов привлечено в настоящее время к анализу загрязнения атмосферы долгоживущими радиоактивными продуктами, образующимися при

работе атомной промышленности и атомных электростанций (АЭС), что требует глобального видеоспектрального мониторинга концентрации радиоактивных изотопов в атмосфере, например, инертного газа криптона.

Дистанционное выявление выбросов вредных веществ конкретными трубами ТЭС, ТЭЦ и АЭС можно обеспечить только видеоспектральными методами и средствами, так как классические методы и средства всегда интегрируют спектральную картину подстилающей поверхности, сводя к нулю пространственное разрешение на ней. Это делает абсолютно невозможной привязку выполненного спектрального анализа к конкретному источнику загрязнений при использовании классических спектральных приборов и известных методов работы с ними.

Особое значение приобретают видеоспектральные исследования, выполняемые с учетом внутриландшафтных миграционных связей. При таких исследованиях обычно изучают миграцию вредных веществ из атмосферы на подстилающую (т.е. земную) поверхность, в системах почва – растение, почва – водоем.

Высокой результативностью отличается использование аэрокосмических видеоспектральных комплексов для контроля чрезвычайных ситуаций при наблюдении пострадавших районов с целью оценки последствий стихийных бедствий, катастроф, аварий, лесных пожаров и планирования мероприятий по ликвидации этих последствий. При этом во многих случаях информационный вклад аэрокосмических данных оказывается исключительно важным. Это относится, в особенности, к антропогенным бедствиям (аварии на нефте- и газопроводах, «залповые» выбросы загрязнений и т.д.), когда необходимо получение в реальном масштабе времени оперативной информации о процессах, происходящих нередко в труднодоступных районах.

При выполнении природно-ресурсных исследований и наблюдений аэрокосмические изображения подстилающей поверхности, полученные с помощью бортовой видеоспектральной аппаратуры в разных интервалах электромагнитного спектра, служат новым эффективным средством познания строения земной коры.

Видеоспектральная съемка со спутниковых и подспутниковых высот позволяет лучше видеть и понять структуру земной коры, что очень важно для правильной постановки геологами поисковых работ. В этой связи, зная, что месторождение попадает в зону разлома или конкретно – в зону куполовидного поднятия, отчетливо видимого на видеоспектральных изображениях, экономически целесообразно организовать работы по поиску полезных ископаемых не по всему прилегающему району, а строго вдоль простирания этой зоны.

Основой познания строения земной коры того или иного района была и продолжает оставаться геологическая съемка все более крупного масштаба, т.е. подробное картирование площадей, на которых выделяются выходы на поверхность горных пород разного возраста и разного состава. Поэтому становится понятной особая роль, которая отводится результатам съемки (особенно труднодоступных районов) с борта атмосферных или внеатмосферных носителей.

Широкие возможности видеоспектрометрии привели к необходимости разработки новой технологии, объединяющей итоговую видеоспектральную и компьютерную картографию и системы управления базами данных. Применение этой перспективной технологии дает основание прогнозировать создание многослойных электронных видеоспектральных карт, опорный слой которых описывает базовое состояние природных и антропогенных объектов, а каждый из последующих слоев выявляет один из частных аспектов или процессов, интересующих потребителя при получении оперативной экологической, гидрометеорологической и природно-ресурсной информации.

В полной мере позитивные свойства видеоспектральной аппаратуры бортового базирования (на атмосферных или внеатмосферных носителях) раскрываются при выполнении геолого-поисковых исследований в ультрафиолетовом (УФ) диапазоне элек-

ромагнитного спектра с целью прогнозирования геологоразведочных работ и создания на их основе расширенных баз данных. К сожалению, в России в настоящее время это самый малоизученный геологами спектральный диапазон. Вместе с тем, анализ зарубежных публикаций показывает, что в последние годы особое место в открытой научно-технической литературе отводится результатам наблюдений в ультрафиолете как весьма перспективному для геологии спектральном диапазоне. Однако объем таких публикаций по тематике применения в геологии УФ видеоспектрометрии крайне незначителен [4].

Важнейшим из условий грамотного проектирования оптико-электронной системы бортового УФ видеоспектрометра является правильный выбор оптических схем объективов, входящих в состав такой аппаратуры. Критерием правильности выбора оптических схем является сочетание достаточно хорошего качества получаемого спектрального изображения с возможно большей простотой прибора в целом.

Развитие бортовой оптической аппаратуры и в особенности космической видеоспектральной аппаратуры УФ диапазона тесно связано с разработкой зеркально-линзовых систем. Для приборов космического базирования требования минимальных габаритов и веса вынуждают применять компактные зеркально-линзовые системы со светосильными зеркальными и линзовыми элементами, работающими со значительными углами падения лучей, что потребовало разработки специфической методики их расчета. Зеркально-линзовые объективы, рассчитанные по такой методике, обладают улучшенными изобразительными и измерительными свойствами, что является фактором, гарантирующим повышение технико-экономической эффективности бортовой видеоспектральной аппаратуры.

В обобщенном плане следует отметить, что главным требованием, которому должен удовлетворять любой зеркально-линзовый объектив видеоспектрометра, является достаточно совершенная коррекция не только монохроматических аберраций, но и особенно аберраций хроматических, причем в УФ спектральной области, более широкой по сравнению с обычными ахроматами или апохроматами. Все указанные обстоятельства чрезвычайно затрудняют выбор оптических сред, которые должны обладать одновременно высоким пропусканием в широкой УФ области спектра и хорошими физико-химическими свойствами. Наличие сравнительно малого числа таких сред предопределило характер последних исследований в НИГ «Коспектр» СПбГУ ИТМО, направленных на создание оптимальных оптических композиций зеркально-линзовых систем, в которых применяются линзовые компоненты из материалов, прозрачных в ультрафиолетовой области спектра, и весьма ограниченной номенклатуры.

В работе над оптико-электронной системой УФ видеоспектрометра в НИГ «Коспектр» СПбГУ ИТМО привлекались современные методы технического проектирования, основанные на использовании новых знаний: метод машинного моделирования, примененный к входной фокусирующей оптике и к оптике спектральной части прибора (как к объектам автоматизированного проектирования), а также математические методы обработки модели оптической системы в целом.

Эти методы обеспечили:

- высокое качество выходных монохроматических изображений вследствие компенсации большинства имеющихся аберраций посредством специальных корригирующих приемов;
- повышенную надежность аппаратуры из-за внедрения перспективных конструкторских решений и технологий при разработке главных оптических и оптико-механических узлов прибора.

Реализованная в НИГ «Коспектр» СПбГУ ИТМО комплексная видеоспектральная система, состоящая из видеоспектрометра ВБИК диапазона, УФ видеоспектрометра и специального УФ телевизионного модуля, представлена на рис. 5.

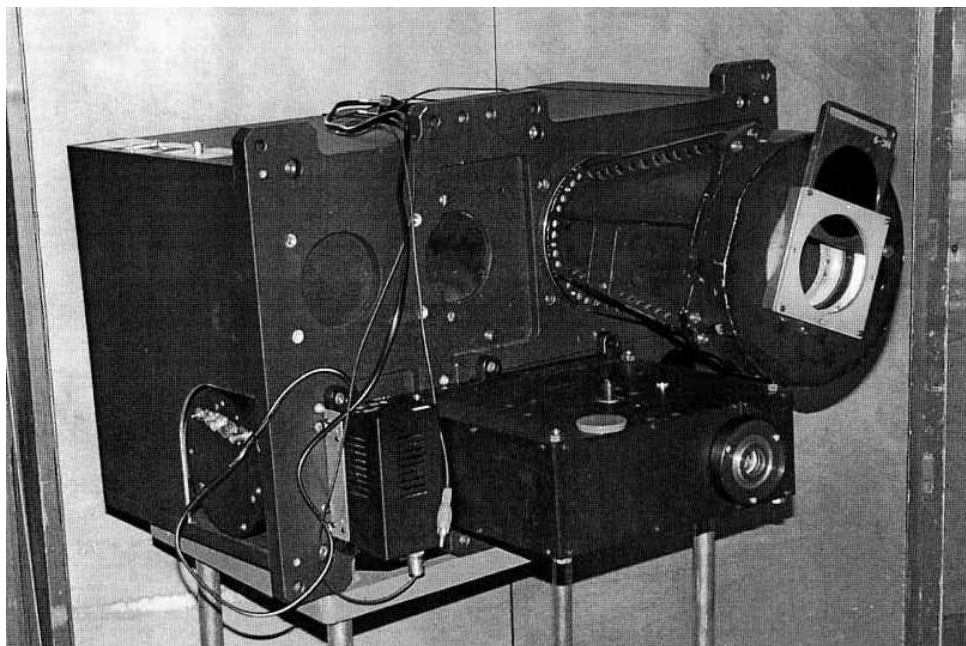


Рис. 5. Внешний вид комплексной видеоспектральной системы, реализованной в НИГ «Коспектр» СПбГУ ИТМО.

Входящий в состав комплексной видеоспектральной системы УФ видеоспектрометр, обладает следующими техническими параметрами и характеристиками:

- рабочий спектральный диапазон: ультрафиолетовая область от 200 нм до 350 нм;
- относительное отверстие прибора, характеризующее его светосилу: 1 : 2,5;
- мгновенное поле зрения, определяющее пространственное разрешение: не более 1 угловой минуты;
- спектральное разрешение: не хуже 2 нм;
- угловое поле, определяющее протяженность полосы обзора: не менее 8 угловых градусов;
- приемники излучения - высокочувствительные гибридные телевизионные приборы ультрафиолетового диапазона (типа УПЗС - 023).

Лабораторные испытания подтвердили значения расчетных параметров разработанной в СПбГИТМО (ТУ) комплексной видеоспектральной системы и высокое качество видеоспектральных изображений, формируемых этим прибором.

Литература

1. Гуд В.В., Красавцев В.М., Сандаков А.Н., Чиков К.Н. Видеоспектрометрические космические комплексы для исследования Земли и планет Солнечной системы. // Оптический журнал. 1995. № 8.
2. Чиков К.Н., Гуд В.В., Красавцев В.М., Сандаков А.Н. Видеоспектрометры для экологического мониторинга. // Известия вузов. Приборостроение. 1998. Т.41. № 3.
3. Батян П.В., Гуд В.В., Коняхин И.А., Красавцев В.М., Чиков К.Н., Груздев В.Н., Шилин Б.В. Опыт использования видеоспектрометра ИТМО в натуральных условиях. // Известия вузов. Приборостроение. 2002. Т. 45. № 2.
4. Чиков К.Н., Панков Э.Д., Порфирьев Л.Ф., Гуд В.В., Красавцев В.М., Тимофеев А.Н. Особенности построения бортовых оптико-электронных систем спектрального мониторинга. // Известия вузов. Приборостроение. 2004. Т. 47. № 9.

10 ГВТ ФЕМТОСЕКУНДНАЯ ЛАЗЕРНАЯ СИСТЕМА ФС-1 – УСТРОЙСТВО И ХАРАКТЕРИСТИКИ

В.Г. Беспалов, В.Н. Васильев, Ю.Н. Ефимов, С.А. Козлов, В.Н. Крылов,
Г.В. Лукомский, С.Э. Путилин, Д.И. Стаселько

Прогресс науки и радикальные изменения информационных технологий XXI века во многом будут связаны с достижениями оптики световых импульсов сверхкороткой длительности (фемтосекундной оптики). Создание качественно новых технологий, основанных на взаимодействии фемтосекундных импульсов света с веществом (фемтотехнологии), явится важнейшим движущим фактором для многих областей науки и техники. Фемтосекундные комплексы – это быстро прогрессирующая область лазерной техники, находящаяся на самом переднем крае науки и технологии. Фемтосекундные комплексы совсем недавно перешли из разряда уникальных лабораторных инструментов в разряд доступного оборудования, но их создание и применение по-прежнему предъявляет самые высокие требования к процессам разработки, конструирования, сборки и обслуживания. Разработка в СПбГУ ИТМО 10 ГВт фемтосекундной лазерной системы ФС-1 – один из первых шагов на пути становления отечественной оптоинформатики с использованием предельно коротких импульсов света.

1. Введение

Начиная с открытия импульсных лазеров [1], исследователи стремились получить все более короткие световые импульсы. В начале 80-х годов несколькими группами был преодолен рубеж 10^{-13} с, и началось быстрое освоение фемтосекундного диапазона длительностей [2]. Как показывает анализ, именно фемтосекундные импульсы были необходимы для исследования многих элементарных процессов по химии, фотобиологии и физике. Фемтосекундная оптическая техника позволила разработать прямые экспериментальные методы изучения молекулярной динамики сложных (в том числе биологически активных) молекул и конденсированных сред.

В 1984 г. была понята важность осторожного баланса между временем удвоенного пробега по резонатору и дисперсией групповых скоростей при формировании сверхкороткого импульса [3], и были получены суб-100 фс оптические импульсы [4]. Дальнейшее развитие техники привело к созданию лазера на красителе со сталкивающимися импульсами и с управляемой призмной системой с отрицательной дисперсией [5], генерирующего световые импульсы длительностью ~ 30 фс.

Для дальнейшего продвижения в сторону более коротких импульсов были применены методы нелинейной оптики. В данном методе [6] 30 фс импульс вводится в одномодовое волокно и посредством объединенного действия фазовой самомодуляции и дисперсии среды становится спектрально уширенным. Этот спектрально и временно уширенный импульс сжимался оптической системой с отрицательной дисперсией на основе либо двух призм, либо двух дифракционных решеток [7], что приводит к более короткому импульсу. В 1987 г. этим методом [8] были получены импульсы длительностью 6 фс в видимом диапазоне

Технология фемтосекундных лазеров коренным образом изменилась с появлением в 1990 г. лазера на кристалле сапфира, активированного ионами трехвалентного титана, с самосинхронизацией мод керровской линзой [9], предложенного английским исследователем В. Сиббетом и его группой. Титан-сапфировый кристалл обладает практически наилучшими параметрами среди твердотельных сред для генерации фемтосекундных импульсов. Так, его полоса оптического усиления (спектра спонтанного излучения) простирается от 700 до 1200 нм, он обладает большой теплопроводностью и запасенной энергией. Уже в первых экспериментах с титан-сапфировыми лазерами были получены импульсы длительностью 10–15 фс [10].

Прорыв в область суб-10 фс для титан-сапфировых лазеров стал возможен при появлении теории влияния дисперсии третьего порядка на длительность сверхкороткого импульса и как итог – теории создания техники так называемых чирпированных зеркал

[11] и полупроводниковых модуляторов с квантовыми ямами [12]. На сегодняшний день рекорд минимальной длительности для данных лазеров поставлен группой Урсулы Келлер [13] – 4,8 фс. Для световых импульсов короче 10 фс применяется новый термин – предельно короткие импульсы, так как их длительность становится сравнимой с периодом колебания электрического поля [14].

Абсолютный рекорд минимальной длительности – 4 фс – поставлен двумя группами исследователей [15–16] при использовании внешней временной компрессии спектрально и временно уширенных 10 фс импульсов титан-сапфирового лазера, сфокусированных либо в кварцевое волокно, либо в полый кварцевый волновод, заполненный аргоном.

Переход к фемтосекундным импульсам – это и очередной скачок по шкале интенсивности. При длительности светового импульса 10 фс сравнительно небольшой энергии в 10 мкДж соответствует мощность $P_0 = 1$ ГВт. Такие параметры в настоящее время можно получить в системах на основе метода усиления chirпированного импульса [17], построенных по схеме: маломощный задающий генератор (титан-сапфировый лазер), стретчер (расширитель импульса), усилитель, компрессор (сжиматель импульса). Таким образом, в современных фемтосекундных системах удается перейти к уровням мощности, которые еще совсем недавно удавалось получать только в мультикилоджоульных установках, предназначенных для управляемого термоядерного синтеза.

Устройство и характеристики фемтосекундной лазерной системы

Оптимальная схема фемтосекундной лазерной системы ФС-1, разработанной в СПбГУ ИТМО и предназначенной для исследований и разработок в области оптики сверхкоротких световых импульсов, была выбрана исходя из наличия компонент и минимума покупных изделий. Блок-схема приведена на рис. 1.



Рис. 1. Блок-схема фемтосекундной лазерной системы

В качестве задающего генератора нами использован фемтосекундный лазер на кристалле титан-сапфира «Фемос-2», выпускаемый отделом лазерной оптики при Новосибирском государственном университете.

Параметры лазера «Фемос-2»: длины волн генерации 750–950 нм, длительность одиночного импульса менее 25 фс, частота следования импульсов 80 МГц, энергия одиночного импульса 1–5 нДж, средняя мощность излучения 400 мВт.

В основу работы лазера (рис. 2) положен метод синхронизации мод наведенной керровской линзой. В качестве излучения накачки используется высокостабилизированное, одномодовое по поперечным индексам излучение непрерывного ионного аргонного лазера SP-2017 фирмы Spectra Physics на всех длинах волн с мощностью не менее 4 Вт. Излучение накачки поступает на зеркало 32 и при помощи линзы Л фокусируется в кристалл титан-сапфира диаметром 3 мм и длиной 6 мм, с торцами, ориентированными под углом Брюстера относительно оптической оси. Для пространственного согласования сфокусированного одномодового пучка накачки и одномодового излучения фемтосекундного лазера кристалл Ti:S помещается в центр системы из двух сферических зеркал С31-С32 с радиусом кривизны 80 мм и коэффициентом отражения $R=99.9\%$. Резонатор лазера образован системой плоских зеркал 31–36, причем выходное зеркало 33 имеет пропускание $T=3-5\%$, а остальные обладают коэффициентом отражения $R=99.9\%$. Для компенсации эффектов, связанных с дисперсионным расплыванием импульсов внутри резонатора лазера, установлен призмный компрессор, образованный двумя кварцевыми брюстеровскими призмами П1-П2. Оптические детали резонатора, кристалл титан-сапфира и система ввода излучения закреплены на трех стержнях из инвара, что обеспечивает температурную стабильность длины резонатора и позволяет реализовать механизм синхронизации мод лазера, основанный на изменении пространственного профиля пучка в результате самофокусировки, вызванной эффектом Керра в кристалле титан-сапфира.

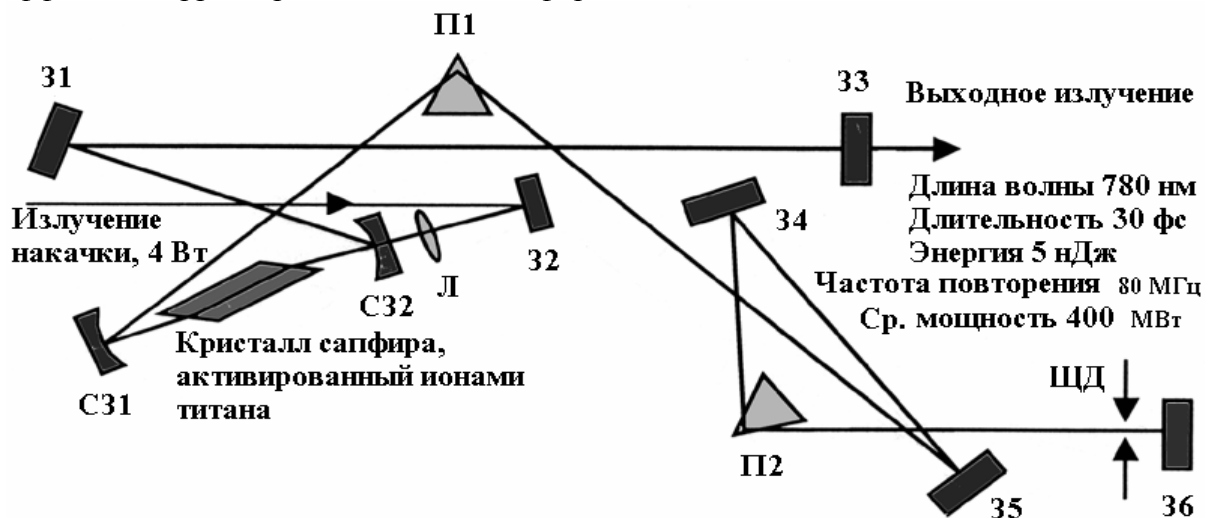


Рис. 2 . Оптическая схема лазера «Фемос-2»

Лазерные импульсы от лазера «Фемос-2» поступают на вход удлинителя фемтосекундных импульсов – стретчера, состоящего из дифракционной решетки и линзы. Растянутые до 100 пс лазерные импульсы с вертикальной поляризацией, отражаясь от входного поляризатора П, поступают на входную (первую) ячейку Погкельса на основе кристалла DKDP с быстродействием менее 5 наносекунд. При отсутствии управляющего напряжения отдельный импульс из цуга проходит ячейку Погкельса, не меняя поляризацию, и не попадает в резонатор регенеративного усилителя, так как, пройдя до зеркала 31 и обратно, он отражается от поляризатора П. Если на ячейку Погкельса подается импульс четвертьволнового напряжения от драйвера длительностью 20 нс, то излучение проходит ячейку с поворотом поляризации на 45° на первом и втором проходах (итого 90°), затем проходит через поляризатор П и поступает в усилитель на кристалле сапфира с титаном. При этом на вторую (выходную) ячейку Погкельса не подается напряжение, она не изменяет поляризацию импульса, и он проходит через поляризатор П. На обратном проходе по резонатору в первой ячейке Погкельса кончается четвертьвол-

новый электрический импульс длительностью 20 нс, поляризация излучения остается неизменной, и оно проходит сквозь первый поляризатор П на второй проход усиления.

Накачка регенеративного усилителя осуществляется импульсным излучением второй гармоники Nd:YAG лазера LQ-129 (фирмы «ЛОТИС-ТИИ», г. Минск, Беларусь) с модуляцией добротности и с энергией в отдельном импульсе до 80 мДж, длительностью 12 нс и частотой следования 50 Гц. Поскольку поляризация излучения в дальнейшем не меняется, то импульс в резонаторе проходит сквозь поляризаторы и начинает циркулировать по резонатору. Через 20–40 проходов по резонатору на вторую (выходную) ячейку Погкельса подается четвертьволновое напряжение, усиленный импульс излучения с вертикальной поляризацией отражается от выходного поляризационного зеркала и выходит из резонатора. В дальнейшем импульс поступает на вход компрессора на основе двух дифракционных решеток и сжимается до фемтосекундной длительности.

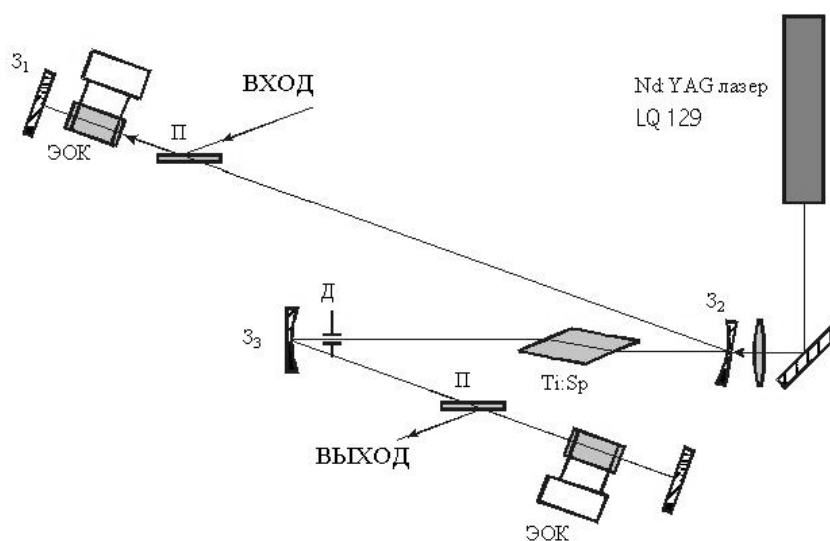


Рис. 3. Оптическая схема регенеративного усилителя: 31–34 – зеркала регенеративного усилителя, П – поляризационные зеркала, ЭОК – ячейки Погкельса, Д – диафрагма, Ti:Sp – кристалл сапфира с титаном

Блок синхронизации предназначен для формирования последовательности импульсов запуска лампы и электрооптического затвора лазера LQ-129, а также запуска блока управления ячейками Погкельса синхронно относительно задающего генератора. Особенностью конструкции блока управления ячейками Погкельса является запуск электрической системы с использованием р-і-п фотодиода и оптоэлектронного высоковольтного тиристорного ключа.

Блок синхронизации и блок управления ячейками Погкельса были реализованы с использованием готовых разработок фирмы «ЛОТИС-ТИИ».

Выходные параметры фемтосекундной лазерной системы: длительность импульсов – менее 150 фс, энергия в отдельном лазерном импульсе – до 1 мДж, частота следования импульсов 50 Гц.

Перспективы применения фемтосекундной лазерной системы

Фемтосекундная лазерная система ФС-1 будет использована для экспериментального изучения широкого круга проблем оптики предельно коротких световых импульсов и сложных оптических волновых образований. Исследования с помощью фемтосе-

кундной лазерной системы позволят определить закономерности протяженной пространственной филаментации, реализовать световые образования из одного колебания электрического и магнитного поля, эффективную генерацию спектрального суперконтинуума и гиперконтинуума в различных средах, в том числе в фотонных микроструктурных волокнах, использовать предельно короткие импульсы и фемтосекундный спектральный суперконтинуум в системах сверхбыстрой передачи информации.

Широкое внимание, связанное с возможностью создания устройств обработки информации с фемтосекундным быстродействием, привлекают сверхбыстрые нелинейные отклики металлических наночастиц, квантовых точек, фуллеренов в твердотельных и жидких матрицах и ряда других объектов. Существенное внимание должно уделяться теоретическим и экспериментальным исследованиям спектроскопической кинетики взаимодействия света с указанными объектами, механизмам их сверхбыстрых откликов, а также возможностям их применения в информационных технологиях.

Основная задача в области фемтосекундной метрологии должна быть связана с разработкой принципов регистрации непосредственно электрического поля однопериодного импульса, а также методам адекватного описания параметров фемтосекундных импульсов.

Одним из перспективных направлений фемтосекундной оптики является получение генерации широкополосного терагерцового излучения фотовозбужденными носителями в полупроводниках и сверхпроводниках. Данная область электромагнитного спектра до последнего времени не была обеспечена источниками когерентного излучения, и по мере их создания предполагается развивать терагерцовую голографию, спектроскопию, а также исследования терагерцовых нелинейных взаимодействий.

Перечисленные фундаментальные исследования в конечном итоге будут направлены на создание конкретных технологий. Например, проблема протяженной филаментации в атмосфере связана с экологическими лидарными технологиями, проблема генерации одинопериодного импульса – с созданием временных стандартов, и т.д. Большую роль в современной науке и ее практических применениях будут играть создание новых технологий в областях, связанных с обработкой материалов фемтосекундными импульсами, с воздействием фемтосекундных электромагнитных импульсов на вещество и, в частности, на биологические объекты и их молекулярные фрагменты, включая молекулы ДНК.

Развитие исследований в указанных направлениях требует оснащения разработанной фемтосекундной системы средствами диагностики и управления параметрами излучения, накопления опыта работы на ней, а также создания специализированных стендов для проведения работ в области фемтосекундной нелинейной оптики, оптоинформатики, фемтосекундной спектроскопии, терагерцовой оптики и фотобиологии.

Выводы

Разработана 10 ГВт фемтосекундная лазерная система ФС-1. Показаны перспективы использования этой системы в различных областях науки и техники, а также при разработке новых технологий.

Работа выполнена при поддержке грантов РФФИ 05-02-20000 и РФФИ 05-02-16556

Литература

1. Maiman T.H. Stimulated Optical Radiation in Ruby. // Nature. 1960. Vol. 187. № 4736. P. 493–494.
2. Ruddock I.S., Bradley D. J. Bandwidth limited subpicosecond pulse generation in mode-locked cw dye lasers. // Appl. Phys. Lett. 1976. Vol. 29. P.296.

3. J.P. Gordon, R.L. Fork. Optical resonator with negative dispersion // *Opt. Lett.* 1984. Vol. 9. P. 153.
4. R.L. Fork, O.E. Martinez. Negative dispersion using pair of prisms // *Opt. Lett.* 1984. Vol. 9. P. 150.
5. J.A. Valdmanis, R.L. Fork, J.P. Gordon. Generation of optical pulses as short as 27 femtoseconds directly from a laser balancing self-phase modulation, group-velocity dispersion, saturable absorption, and saturable gain. // *Opt. Lett.* 1985. Vol. 10. P. 131.
6. W.J. Tomlison, R.H. Stolen, C.V. Shank. Compression of optical pulses chirped by self-phase modulation in fibers // *J. Opt. Soc. Am. B.* 1984. Vol. 1. P. 139.
7. J.D. Kafka, T. Baer. Prism-pair dispersive delay lines in optical pulse compression. // *Opt. Lett.* 1987. Vol. 12. P. 401.
8. R.L. Fork, C.H. Brito Cruz, P.C. Becker, C.V. Shank. Compression of optical pulses to six femtoseconds by using cubic phase compensation. // *Opt. Lett.* 1987. Vol. 12. P. 483.
9. D.E. Spence P.N. Kean, W. Sibbet. 60-fs pulse generation from a self-mode-locked Ti:sapphire laser. // *Opt. Lett.* 1991. Vol. 16. P. 42-44.
10. Ch. Spielmann, P.F. Curley, T. Brabec, E. Wintner, F. Krausz. Generation of sub-20 fs mode-locked pulses from Ti:sapphire laser. // *Electron. Lett.* 1992. Vol. 28. P. 1532.
11. R. Szipocs, K. Ferencz, Ch. Spielmann, F. Krausz. Chirped multilayer coatings for broadband dispersion control in femtosecond lasers. // *Opt. Lett.* 1994. Vol. 19, P. 201.
12. I.D. Jung, F.X. Kartner, N. Matuschek, D.H. Sutter, F. Morier-Genoid, G. Zhang, U. Keller, V. Scheuer, M. Tilsch, T.Tschudi. Self-starting 6.5 fs pulses from a Ti:sapphire laser. // *Optics Lett.* 1997. Vol. 22. P. 1009-1011.
13. D. H. Sutter, G. Steinmeyer, L. Gallmann, N. Matuschek, F. Morier-Genoud, U. Keller, V. Scheuer, G. Angelow, T. Tschudi . Semiconductor saturable-absorbermirrorassisted Kerr-lensmode-locked Tisapphire laser producing pulses inthe two-cycle regime. // *Opt. Lett.* 1999. Vol. 24. P. 631.
14. 21. Bepalov V.G., Kozlov S.A., ShpolyanskyYu.A., Oukrainski A.O., Sazonov S.V. Spectral evolution of propagating extremely short pulses. // *Physics of Vibration*, 1999, Vol. 7. P. 19–27.
15. A. Baltuska, Z. Wei, M.S. Pshenichnikov, D.A. Wiersma, Robert Szipocs. - All-solid-state Cavity-dumped sub-5-fs laser. // *Appl. Phys. B.* 1997. Vol. 65. P. 175–188.
16. M. Nisoli, S. Stagira, S. De Silvestri, A. Svelto, S. Sartania, Z. Cheng, M. Lenzner, Ch. Spielmann, F. Krausz. A novel-high energy pulse compression system: generation of multigigawatt sub-5-fs pulses. // *Appl. Phys. B.* 1997. Vol. 65, P. 189–196.
17. S.Backus, C.G.Durfee III, G.Mourou, H.C.Karteyn, M.M.Murnane 0.2 TW laser system at 1 kHz. // *Opt. Lett.* 1997. Vol. 22. P.1256–1258.

ИСПОЛЬЗОВАНИЕ ПАРАМЕТРОВ ТОНКИХ КОМПОНЕНТОВ ДЛЯ АВТОМАТИЗИРОВАННОЙ КОРРЕКЦИИ ОПТИЧЕСКИХ СИСТЕМ

А.В. Иванов

В работе исследуются проблемы выбора переменных для автоматизированной коррекции оптических систем, доказывается целесообразность применения параметров тонких компонентов в качестве варьируемых на этапах параметрического синтеза.

При использовании в процессе параметрического синтеза оптических систем методики тонких компонентов [1] выбор параметров автоматизированной коррекции aberrаций в области реальных полей и апертур может осуществляться на альтернативной основе. Действительно, в этом случае применимы, по крайней мере, три различных подхода:

а) задавать в качестве параметров конструктивные элементы оптики и углы первого нулевого луча [2–5] (здесь и далее подразумевается, что те или иные переменные, входящие в состав параметров моделей, предварительно соответствующим образом нормируются [5–6]);

б) использовать поправки в вектор свободных членов линейной aberrационной модели, лежащей в основе методики тонких компонентов [1];

в) использовать параметры P , W , C тонких компонентов [7–9].

Первый путь, очевидно, наиболее универсальный, но приводит к существенной нелинейности и полимодальности оценочной функции и других aberrационных характеристик. Именно поэтому конструктивные элементы оптики обычно применяются в качестве параметров на стадиях окончательной доводки (оптимизации) систем [5].

Второй путь прост в реализации, но также имеет ряд существенных недостатков. Дело в том, что, во-первых, подбор поправок в модельные уравнения может быть произведен либо эмпирически, либо нелинейным методом итераций, сходимость которого остается недоказанной; во-вторых, постоянные поправки в модельные уравнения способны привести к компенсации отдельных aberrаций высших порядков всей системы в целом, но отнюдь не обязательно к устранению их по отдельным компонентам, что может отрицательно сказаться на интегральной оценке качества конструкции; в-третьих, каждому набору поправок может соответствовать не одно конструктивное решение, а несколько. В этом случае неясно, какому из вариантов следует отдать предпочтение, чтобы обеспечить автоматически сходимость общего процесса.

Третий путь выбора варьируемых параметров моделей представляет, на наш взгляд, наивысший интерес с позиций параметрического синтеза. Первоначально он использовался при пересчете одиночных тонких компонентов (см., например, работы [7, 10]), а затем был распространен и на случай многокомпонентных систем [9]. В статье [9] выбор параметров P , W , C в качестве варьируемых параметров синтеза обоснован повышением чувствительности метода по сравнению с аналогами. Такое обоснование кажется нам достаточно спорным, и поэтому целесообразно остановиться на аргументах в пользу применения P , W , C более подробно.

Вначале обратимся к линейной aberrационной модели централизованной оптической системы в зейделевой области. В общем виде она может быть представлена следующим образом:

$$\mathbf{S} = \mathbf{S}^{(0)} + \mathbf{A}_s \mathbf{X}, \quad (1)$$

где \mathbf{X} – вектор параметров, состоящий из элементов P , W , C ; \mathbf{A}_s – матрица коэффициентов при них; $\mathbf{S}^{(0)}$ – вектор оставшихся (свободных) членов в правых частях; \mathbf{S} – вектор сумм Зейделя и хроматических сумм [1, 11]. В силу ряда предположений, определяемых непосредственно методикой синтеза, эта модель является вполне адекватной

лишь в зейделева области аббераций. Тем не менее, выражения (1) могут быть использованы и для нахождения параметров оптической системы с учетом реальных полей и апертур, если в них ввести некоторые слагаемые-поправки [1]. Тогда, исправляя абберации при $\mathbf{S} = \mathbf{0}$, имеем:

$$\mathbf{A}_s \mathbf{X} + \mathbf{S}^{(0)} + \Delta \mathbf{S} = \mathbf{0}, \quad (2)$$

где $\Delta \mathbf{S}$ – вектор поправок. Полученная система уравнений может быть использована в качестве базовой для синтеза. Обозначим $\mathbf{B} = -(\mathbf{S}^{(0)} + \Delta \mathbf{S})$; тогда (2) преобразуется к виду

$$\mathbf{A}_s \mathbf{X} = \mathbf{B} \quad (3)$$

В уравнениях (3) вектор правых частей известен лишь с определенной степенью точности, так как значения поправок $\Delta \mathbf{S}$ до расчета конструкции системы и ее аббераций действительных лучей установить невозможно. Однако ограничение на норму вектора вычислить несложно, исходя из допустимых аббераций зейделева области:

$$\|\Delta \mathbf{S}\| < \delta \quad (4)$$

(предполагается, что уравнения (3) предварительно преобразованы и отнормированы; при подходящей нормировке величина δ может иметь смысл максимально допустимого среднеквадратического размера пятна рассеяния или максимально допустимой деформации волнового фронта в области Зейделя [11]). Поскольку вектор правых частей (3) известен с погрешностью, определяемой (4), задача (3) является некорректной по Тихонову. Ее решение, как следует из теории [12], необходимо искать на некотором множестве векторов, сопоставимых по точности с исходными данными (\mathbf{B}), минимизируя при этом стабилизирующий функционал $\Omega(\mathbf{X})$.

Таким образом, корректный подход к решению (3) состоит в вычислении вектора \mathbf{X}^* , доставляющего минимум $\Omega(\mathbf{X})$ при ограничении $\|\mathbf{A}_s \mathbf{X} + \mathbf{S}^{(0)}\| < \delta$.

Вид стабилизирующего функционала обычно связывается с содержанием задачи. В нашем случае уместно положить:

$$\Omega(\mathbf{X}) = \sum_i \left[(f_i(\mathbf{X}) - f_i^{(0)}) / \delta f_i^{(0)} \right]^2 = \mathbf{F}^T(\mathbf{X}) \mathbf{F}(\mathbf{X}), \quad (5)$$

где $f_i(\mathbf{X})$ – отдельные абберации действительных лучей или коэффициенты аббераций высших порядков; $f_i^{(0)}$, $\delta f_i^{(0)}$ – их заданные значения и нормы; $\mathbf{F}(\mathbf{X})$ – вектор так называемых минимизируемых (оптимизируемых) функций, а значок "Т" означает транспонирование.

Из сказанного можно сделать вывод о том, что с точки зрения теории некорректных задач использование линейной модели (1) объективно приводит к необходимости минимизации функционала Ω , связанного с характеристиками оптической системы в реальной области аббераций, по переменным P , W , C . Отметим, что ограничение на переменные (вида $\|\mathbf{A}_s \mathbf{X} + \mathbf{S}^{(0)}\| < \delta$) позволяет в известном смысле контролировать зональные величины аббераций; этот факт весьма благоприятно сказывается на работе методов нелинейного программирования, поскольку данное ограничение является линейным по P , W , C .

Важным аргументом в поддержку использования тех или иных варьируемых параметров при решении задач автоматизированной коррекции аббераций считается «хороший» рельеф оценочной функции, «удобство» которого определяется отсутствием локальных минимумов и близостью корригируемых функций к линейным [6, 13]. С этой точки зрения применение параметров P , W , C по сравнению с конструктивными элементами оптики также часто бывает предпочтительным.

Чтобы продемонстрировать это, сравним зависимости продольной сферической aberrации $\Delta s'$ одиночной тонкой линзы в воздухе от ее прогиба (угла α_2 первого нулевого луча) и от параметра P при заднем фокусном расстоянии $f' = 100$ мм и относительном отверстии 1 : 2. Предмет будем считать расположенным в бесконечности.

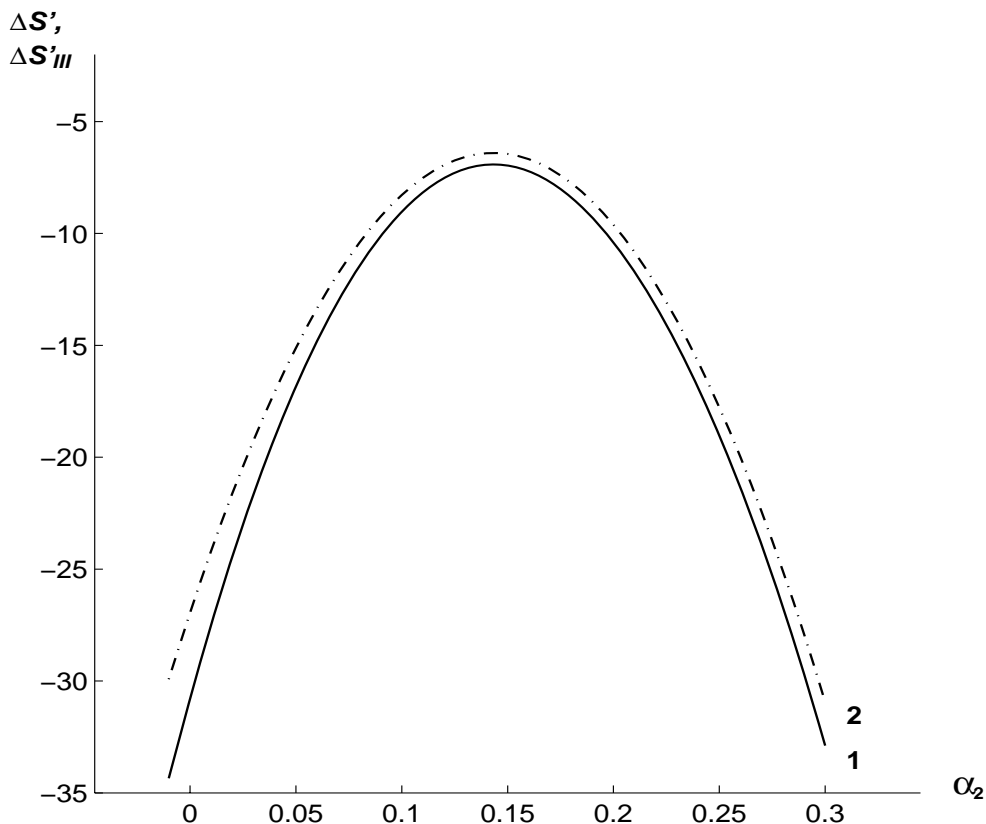


Рис. 1. Исследование зависимости реальной сферической aberrации (1) и сферической aberrации третьего порядка (2) от прогиба линзы

Результаты исследования зависимости $\Delta s'(\alpha_2)$ (см. рис. 1) показывают, что она в первом приближении может быть описана полиномом второй степени:

$$\Delta s'(\alpha_2) \approx a_1 \alpha_2^2 + a_2 \alpha_2 + a_3, \quad (6)$$

где коэффициенты a_1, a_2, a_3 нетрудно вычислить с помощью методов аппроксимации [14]. С другой стороны, для одиночной линзы в воздухе

$$P(\alpha_2) = b_1 \alpha_2^2 + b_2 \alpha_2 + b_3, \quad (7)$$

где b_1, b_2, b_3 – известные постоянные [1]. Пользуясь (6) и (7), имеем:

$$\frac{\partial(\Delta s')}{\partial P} = \frac{2a_1 \alpha_2 + a_2}{2b_1 \alpha_2 + b_2} \quad (8)$$

Если стационарные точки кривых $\Delta s'(\alpha_2)$ и $P(\alpha_2)$ совпадают, то, очевидно,

$$\frac{\partial(\Delta s')}{\partial P} = const,$$

т.е. зависимость продольной сферической aberrации от параметра P в этом случае становится линейной. На практике экстремальные значения функций (6) и (7) достигаются не в одной, а в довольно близко расположенных точках (см. рис. 1). В подобной ситуации $\Delta s'(P)$ является уже нелинейной функцией, однако ее нелинейность существенно сказывается лишь в окрестности точки ветвления $\alpha_2(P)$. По мере удаления угла α_2 от

значения $\alpha_2^{(0)} = -b_2 / (2b_1)$ первая производная $\Delta s'(P)$ приближается к постоянной a_1 / b_1 , а производные высших порядков стремительно уменьшаются по абсолютной величине. Например, вторая производная убывает пропорционально кубу смещения α_2 относительно $\alpha_2^{(0)}$:

$$\frac{\partial^2(\Delta s')}{\partial P^2} = \frac{a_1 b_2 - a_2 b_1}{4b_1^3 (\alpha_2 - \alpha_2^{(0)})^3}, -$$

третья производная – пропорционально пятой степени смещения и т.д. Поэтому на значительной части области определения зависимость $\Delta s'(P)$ с точностью до аппроксимации (6) должна быть близка к линейной. Данный вывод подтверждается результатами прямого построения кривых $\Delta s'(P)$ для обеих ветвей $\alpha_2(P)$ (см. рис. 2).

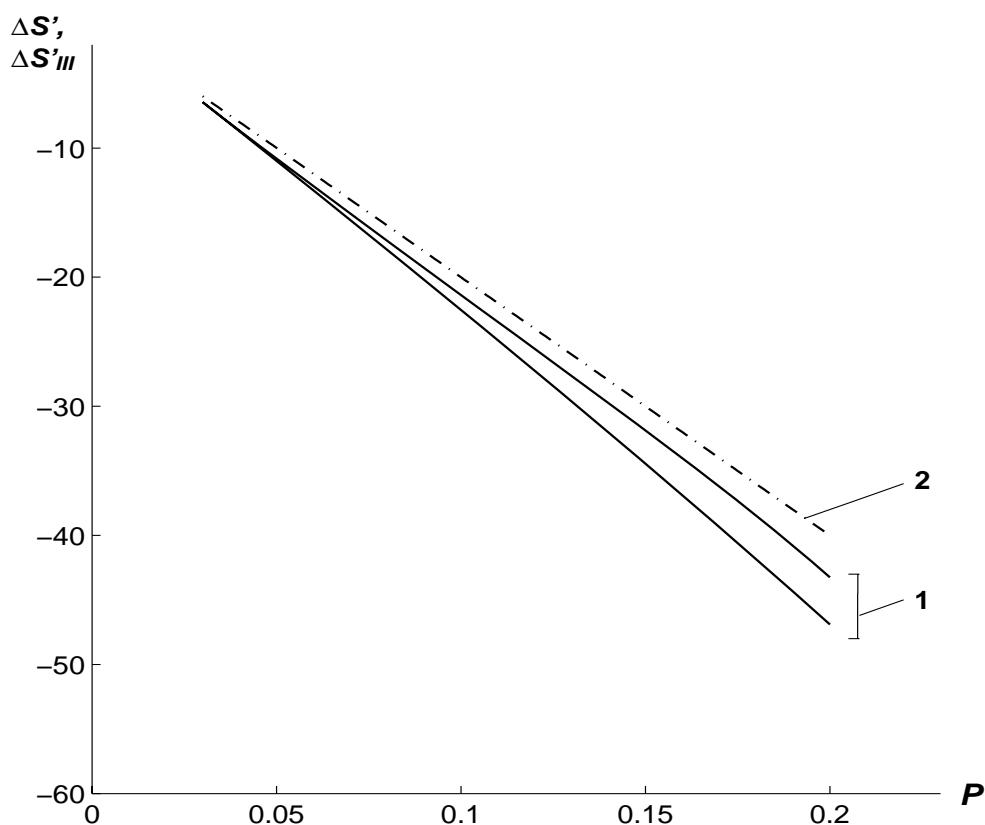


Рис. 2. Исследование зависимости реальной сферической аберрации (1) и сферической аберрации третьего порядка (2) от параметра P для двух вариантов конструктивных решений

Глобальное или локальное снижение нелинейности минимизируемых функций при использовании параметров P, W, C характерно не только для случая сферической аберрации в одиночной линзе. Интересным примером, подтверждающим сказанное, является зависимость астигматизма склеенного тонкого дублета от параметра P при фиксированном параметре W . На рис. 3 эта зависимость построена для сочетания стекол К8/Ф1; основная длина волны – D ; входной зрачок расположен на расстоянии +15 мм от дублета; первый нулевой луч пересекает тонкий компонент на высоте 25 мм и образует углы в пространстве предметов и изображений, равные $\alpha_1 = 0, \alpha_4 = 0.25$; параметр $W = -0.2$. Построение произведено для одного из множества вариантов конструктивных решений.

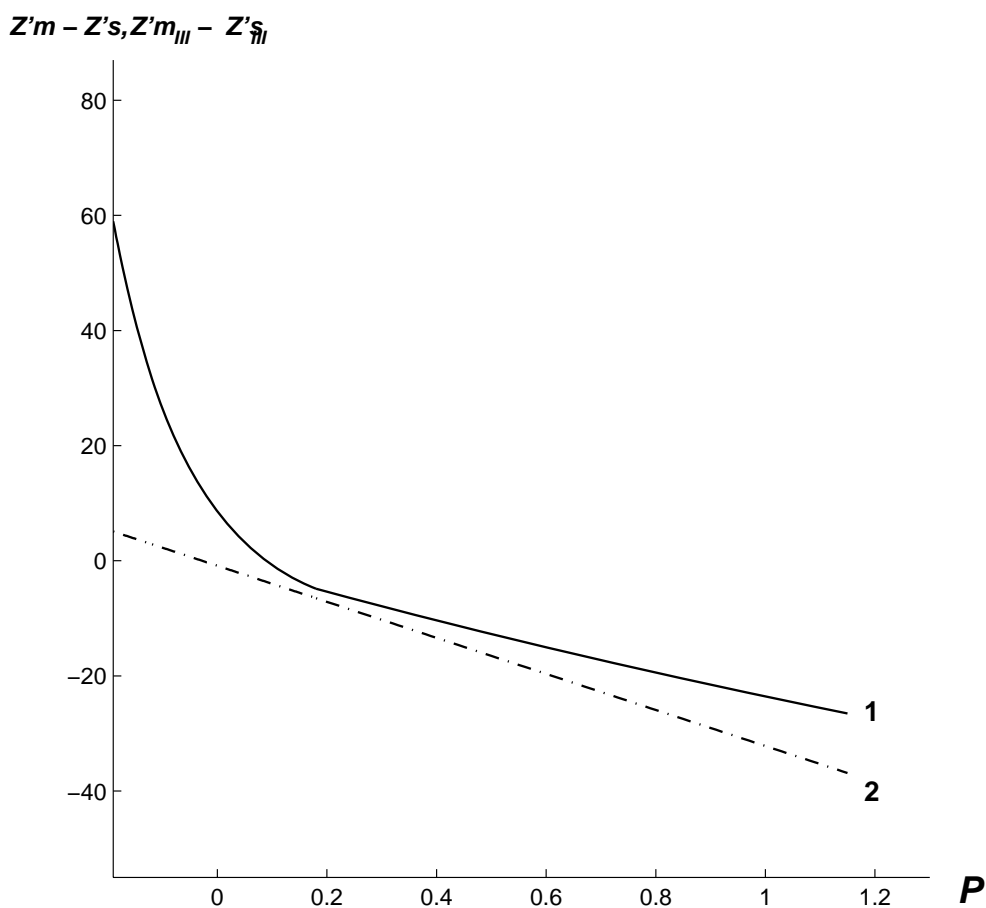


Рис. 3. Исследование зависимости реального астигматизма (1) и астигматизма третьего порядка (2) склеенного дублета от параметра P

Как видно из графика, на значительном диапазоне изменения P функция астигматизма близка к линейной, а нелинейный участок имеет монотонный характер. В то же время зависимость исследуемой aberrации от угла первого нулевого луча α_3 при заданных углах $\alpha_1 = 0$, $\alpha_2 = -0.2$, $\alpha_4 = 0.25$ имеет две стационарные точки (см. рис. 4). Вследствие этого, попытка устранения астигматизма за счет итерационного изменения угла α_3 будет обречена на неудачу при многих начальных значениях параметра: процесс будет сходиться в локальном минимуме, не являющемся решением задачи. Скажем, при старте со значения $\alpha_3 = 0.45$ любой метод нелинейного программирования, обеспечивающий минимизацию квадрата астигматизма, сойдется в точке $\alpha_3 \approx 0.38463$, в которой астигматизм составляет -3.41 мм.

Таким образом, применение параметров P , W , C при решении задачи нелинейного программирования часто является предпочтительным по сравнению с использованием традиционных переменных. Однако следует заметить, что переход от P , W , C к конструкции осложнен неоднозначностью отыскания неизвестных конструктивных переменных. Неправильный отбор корней может привести к потере сходимости итерационного процесса решения задачи автоматизированной коррекции aberrаций. Поэтому параллельно с методикой применения P , W , C должна существовать и методика однозначного определения конструктивных элементов по этим параметрам в процессе их непрерывного изменения.

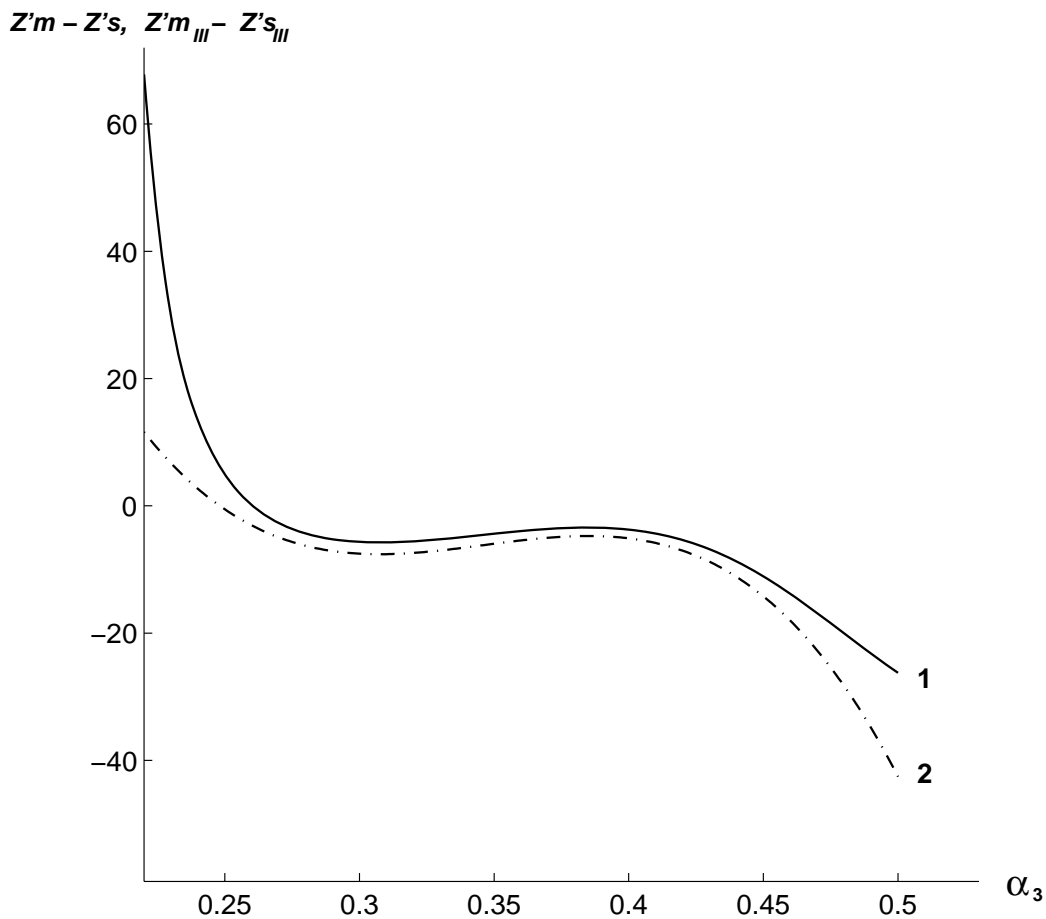


Рис. 4. Исследование зависимости реального астигматизма (1) и астигматизма третьего порядка (2) склеенного дублета от угла α_3

Литература

1. Слюсарев Г.Г. Методы расчета оптических систем. Л.: Машино строение, 1969. 672 с.
2. Бездидько С.Н. Автоматизация проектирования оптических систем: Аналит. обзор за 1945–1980 гг. / ЦНИИ информ. и техн.-экон.исслед. М., 1982. 99 с.
3. Вороненская Т.С. Оптимизация оптических систем: Аналит. обзор за 1957–1981 гг. / ЦНИИ информ. и техн.-экон.исслед. М., 1983. 47 с.
4. Грамматин А.П. Использование углов первого параксиального луча с осью в качестве коррекционных параметров // Тр./ Ленингр. ин-т точн. механики и оптики. Л., 1960. С. 12–24.
5. Родионов С.А. Автоматизация проектирования оптических систем. Л.: Машиностроение, 1982. 270 с.
6. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. М.: Мир, 1985. 509 с.
7. Апенко М.И., Запрыгаева Л.А., Свешникова И.С. Задачник по прикладной оптике. М.: Недра, 1987. 310 с.
8. Волосов Д. С. Методы расчета сложных фотографических систем. Л., М.: ОГИЗ; Гос-техиздат, 1948. 396 с.
9. Зверев В.А., Сакин Л.И. Применение теории основных параметров для расчета оптических систем // Опт.-мех. пром-сть. 1984. № 12. С. 25–29.
10. Вычислительная оптика: Справочник / Русинов М.М., Грамматин А.П., Иванов П.Д. и др.; Под общ. ред. Русинова М.М. Л.: Машиностроение, 1984. 423 с.

11. Иванов А.В. Математический аппарат синтеза оптических систем из тонких компонентов // Изв. вузов СССР. Приборостроение. 1986, № 1. С. 78–82.
12. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1986. 288 с.
13. Родионов С.А. Теория и методы автоматизированного проектирования оптических систем: Дис. на соиск. учен. степ. д-ра техн. наук / Ленингр. ин-т точн. механики и оптики. Л., 1984. 473 с.
14. Бронштейн И.Н., Семендяев К.А. Справочник по математике для инженеров и учащихся втузов. М.: Наука, 1986. 544 с.

ЭКОНОМИЧНЫЙ МЕТОД ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТОВ ПОПЕРЕЧНЫХ АБЕРРАЦИЙ ТРЕТЬЕГО И ВЫСШИХ ПОРЯДКОВ

А.В. Иванов

В статье описан метод расчета абберационных коэффициентов центрированных оптических систем, основанный на интерполяции данных расчета поперечных аббераций различных порядков. Метод отличается высокой экономичностью, простотой, а также высокой точностью, позволяющей использовать степенной базис при аппроксимации.

В работе [1] предложен эффективный способ расчета составляющих поперечных аббераций различных порядков центрированных оптических систем, основанный на прогонке так называемых псевдолучей [2–4]. Однако для оценки качества коррекции часто требуются не сами значения аббераций, а коэффициенты их разложения в степенные ряды по относительным зрачковым и полевой координатам. Поэтому преимущества подхода с использованием псевдолучей в значительной степени теряются.

С целью устранения отмеченного недостатка можно воспользоваться методом интерполяции по результатам расчета нескольких псевдолучей. Метод интерполяции, как известно, сводится к решению системы линейных уравнений [5]:

$$\mathbf{AX} = \mathbf{F} \quad (1)$$

где \mathbf{A} – конструкторная матрица, элементы которой зависят от базиса разложения аббераций и от выбора узловых точек на зрачке и поле (здесь и далее предполагается, что входной и выходной зрачки оптической системы определены в соответствии с [6]); \mathbf{F} – значения аббераций в узловых точках; \mathbf{X} – неизвестные коэффициенты интерполяции. Очевидно, что точность и трудоемкость нахождения \mathbf{X} обусловлены принципами построения \mathbf{A} . Как ни странно, анализ этих принципов не получил должного освещения в литературе. Лишь в работе [4] описан несложный, но трудоемкий прием определения неизвестных в (1). Он состоит в приведении системы (1) к простому, легко разрешаемому виду за счет прогонки избыточного числа псевдолучей со специально подобранными координатами.

Рациональный способ интерполяции, конечно, должен обеспечивать разбиение системы уравнений на несколько элементарных подсистем возможно меньшей размерности: это позволяет получить максимальную точность вычислений и упростить решение уравнений. Но в то же время повышение точности не должно осуществляться посредством увеличения числа рассчитываемых псевдолучей, т.е. за счет повышенной трудоемкости процесса.

Ниже мы приводим в некотором смысле оптимальную схему интерполяции, отвечающую обоим противоречивым требованиям.

Введем декартовы системы координат на зрачке и предмете подобно тому, как это сделано в работе [4]. Обозначим через ρ_x, ρ_y – относительные зрачковые, а через H – относительную предметную координаты. Рассмотрим разложение поперечных аббераций центрированной оптической системы по степенному базису при фиксированной длине волны [4]:

$$\begin{aligned} \varepsilon_y &= \varepsilon_y^{(1)} + \varepsilon_y^{(3)} + \varepsilon_y^{(5)} + \varepsilon_y^{(7)} + \dots; \quad \varepsilon_x = \varepsilon_x^{(1)} + \varepsilon_x^{(3)} + \varepsilon_x^{(5)} + \varepsilon_x^{(7)} + \dots; \\ \varepsilon_y^{(1)} &= G_1 \rho_y + G_2 H; \quad \varepsilon_y^{(3)} = P_1 \rho^2 \rho_y + P_2 H \rho^2 + P_3 H \rho_y^2 + P_4 H^2 \rho_y + P_5 H^2 \rho_y + P_6 H^3; \\ \varepsilon_y^{(5)} &= S_1 \rho^4 \rho_y + S_2 H \rho^4 + S_3 H \rho^2 \rho_y^2 + S_4 H^2 \rho^2 \rho_y + S_5 H^2 \rho^2 \rho_y + S_6 H^2 \rho_y^3 + \\ &+ S_7 H^3 \rho^2 + S_8 H^3 \rho_y^2 + S_9 H^3 \rho_y^2 + S_{10} H^4 \rho_y + S_{11} H^4 \rho_y + S_{12} H^5; \end{aligned}$$

$$\begin{aligned}
\varepsilon_y^{(7)} &= T_1 \rho^6 \rho_y + T_2 H \rho^6 + T_3 H \rho^4 \rho_y^2 + T_4 H^2 \rho^4 \rho_y + T_5 H^2 \rho^4 \rho_y + T_6 H^2 \rho^2 \rho_y^3 + \\
&+ T_7 H^3 \rho^4 + T_8 H^3 \rho^2 \rho_y^2 + T_9 H^3 \rho^2 \rho_y^2 + T_{10} H^3 \rho_y^4 + T_{11} H^4 \rho^2 \rho_y + T_{12} H^4 \rho_y^3 + \\
&+ T_{13} H^4 \rho^2 \rho_y + T_{14} H^4 \rho_y^3 + T_{15} H^5 \rho^2 + T_{16} H^5 \rho_y^2 + T_{17} H^5 \rho_y^2 + T_{18} H^6 \rho_y + \\
&+ T_{19} H^6 \rho_y + T_{20} H^7; \\
\varepsilon_x^{(1)} &= G_1 \rho_x; \quad \varepsilon_x^{(3)} = P_1 \rho^2 \rho_x + P_3 H \rho_x \rho_y + P_5 H^2 \rho_x; \\
\varepsilon_x^{(5)} &= S_1 \rho^4 \rho_x + S_3 H \rho^2 \rho_x \rho_y + S_5 H^2 \rho^2 \rho_x + S_6 H^2 \rho_x \rho_y^2 + S_9 H^3 \rho_x \rho_y + S_{11} H^4 \rho_x; \\
\varepsilon_x^{(7)} &= T_1 \rho^6 \rho_x + T_3 H \rho^4 \rho_x \rho_y + T_5 H^2 \rho^4 \rho_x + T_6 H^2 \rho^2 \rho_x \rho_y^2 + T_9 H^3 \rho^2 \rho_x \rho_y + \\
&+ T_{10} H^3 \rho_x \rho_y^3 + T_{13} H^4 \rho^2 \rho_x + T_{14} H^4 \rho_x \rho_y^2 + T_{17} H^5 \rho_x \rho_y + T_{19} H^6 \rho_x
\end{aligned} \tag{2}$$

где $\rho^2 = \rho_x^2 + \rho_y^2$; $\varepsilon_y, \varepsilon_x$ – меридиональная и сагиттальная составляющие поперечных aberrаций; $G_1, G_2, P_1, \dots, P_6, S_1, \dots, S_{12}, T_1, \dots, T_{20}$ – коэффициенты aberrаций. Из разложения (2) следует, что для интерполяции невыгодно использовать меридиональные лучи ($\rho_x = 0$), поскольку в этом случае $\varepsilon_x = \varepsilon_x^{(1)} = \varepsilon_x^{(3)} = \varepsilon_x^{(5)} \dots \equiv 0$ и теряется часть уравнений для отыскания неизвестных коэффициентов. Напротив, расчет нескольких сагиттальных лучей ($H \neq 0, \rho_y = 0$) оказывается чрезвычайно полезным, так как позволяет определить ряд неизвестных из систем уравнений малой размерности. Например, для коэффициентов 7-го порядка имеем:

$$\begin{aligned}
(\rho_x)_i (T_1 \rho_i^6 + T_5 H_i^2 \rho_i^4 + T_{13} H_i^4 \rho_i^2 + T_{19} H_i^6) &= (\varepsilon_x^{(7)})_i \\
H_i (T_2 \rho_i^6 + T_7 H_i^2 \rho_i^4 + T_{15} H_i^4 \rho_i^2 + T_{19} H_i^6) &= (\varepsilon_y^{(7)})_i
\end{aligned} \tag{3}$$

где i – индекс псевдолуча. Выразив часть переменных из уравнений типа (3), найдем оставшиеся коэффициенты сагиттальной составляющей aberrаций. С этой целью осуществим расчет «косых» псевдолучей ($\rho_y \neq 0, \rho_x \neq 0, H \neq 0$). Соотношения для $\varepsilon_x^{(7)}$ дают:

$$\begin{aligned}
(\rho_x)_i (\rho_y)_i H_i (T_3 \rho_i^4 + T_6 H_i \rho_i^2 (\rho_y)_i + T_9 H_i^2 \rho_i^2 + T_{10} H_i^2 (\rho_y)_i^2 + \\
+ T_{14} H_i^3 (\rho_y)_i + T_{17} H_i^4) &= (\varepsilon_x^{(7)})_i - \delta_i^{(1)}
\end{aligned} \tag{4}$$

где $\delta_i^{(1)} = (\rho_x)_i (T_1 \rho_i^6 + T_5 H_i^2 \rho_i^4 + T_{13} H_i^4 \rho_i^2 + T_{19} H_i^6)$.

Систему типа (4) можно разбить во всех порядках на две подсистемы за счет прогонки симметричных относительно сагиттальной плоскости «косых» псевдолучей. Например, для 7-го порядка имеем:

$$2(\rho_y)_i^2 (\rho_x)_i H_i^2 (T_6 \rho_i^2 + T_{14} H_i^2) = (\varepsilon_x^{(7)})_i + (\varepsilon_x^{(7)})_{i+1} - 2\delta_i^{(1)} \tag{5}$$

причем $(\rho_y)_i = -(\rho_y)_{i+1}, (\rho_x)_i = (\rho_x)_{i+1}, H_i = H_{i+1}$.

С учетом (5) коэффициенты T_3, T_9, T_{10}, T_{17} находятся из (4) путем решения системы 4 уравнений с 4 неизвестными:

$$(\rho_x)_i (\rho_y)_i H_i (T_3 \rho_i^4 + T_9 H_i^2 \rho_i^2 + T_{10} H_i^2 (\rho_y)_i^2 + T_{17} H_i^4) = (\varepsilon_x^{(7)})_i - \delta_i^{(1)} - \delta_i^{(2)} \tag{6}$$

где $\delta_i^{(2)} = (\rho_y)_i^2 (\rho_x)_i H_i^2 (T_6 \rho_i^2 + T_{14} H_i^2)$.

Последний способ разбиения систем уравнений, на который мы обратим внимание, характерен для aberrаций 9-го и более высоких порядков. Здесь возникает ситуация, когда пары неизвестных имеют постоянные коэффициенты, отличающиеся только степенью координаты H . Положив $H = 1$, мы можем сначала найти значения сумм этих неизвестных из систем уравнений уменьшенной размерности, а затем определить и значения самих слагаемых с помощью оставшейся части уравнений с измененной величиной H .

Получив все коэффициенты, описывающие сагиттальные составляющие aberrаций, можно перейти к отысканию оставшейся части \mathbf{X} по выражениям для $\varepsilon_y^{(3)}, \varepsilon_y^{(5)}, \varepsilon_y^{(7)}, \dots$. Набор псевдолучей, выделенный ниже, позволяет и в этом случае произвести значительные упрощения систем уравнений. Максимальный размер обращаемых при интерполяции матриц равен: 3 – при нахождении коэффициентов 5-го порядка, 4 – при нахождении коэффициентов 7-го порядка, 5 – при нахождении коэффициентов 9-го порядка. Это гарантирует хорошую обусловленность матриц [7], а, следовательно, высокую точность метода при минимально допустимом количестве рассчитываемых псевдолучей, которое определяется уже самой задачей.

Представляет интерес выбор ненулевых узловых значений координат псевдолучей. Рассмотрим его на примере отыскания коэффициентов aberrаций до 9-го порядка включительно. Оказывается, что все матрицы, подлежащие обращению при интерполяции, могут быть приведены либо к диагональному, либо к универсальному виду:

$$\mathbf{C} = \{(c_i)^j\}_j \quad j = 0, 1, \dots, n \quad (7)$$

где $n \leq 4$, а под символом c_i понимается координата ρ_i^2 . Матрица \mathbf{C} обладает спецификой [8] и может быть обращена аналитически, что дает возможность добиться еще более высокой точности нахождения коэффициентов.

Набор псевдолучей, позволяющий привести конструкционные матрицы интерполяции к виду (7), описан в таблице 1. В этой таблице $\cos \theta = \sqrt{1 - (\rho_x / \rho)^2}$.

Координата	Порядок aberrаций														
	3, 5, 7, 9			5, 7, 9			7, 9				9				
H	1	1	1	1	1	1	1	1	1	1	1	1	1	c_1	c_2
ρ^2	c_1	c_2	c_1	c_3	c_1	c_2	c_4	c_2	c_3	c_4	c_5	c_3	c_4	c_1	c_2
$\cos \theta$	0	0	c_1	0	$-c_1$	c_2	0	$-c_2$	c_3	c_4	0	$-c_3$	$-c_4$	c_1	c_2

Таблица 1. Входные координаты псевдолучей, используемых при вычислении коэффициентов aberrаций: H – предметная координата; ρ^2 , $\cos \theta$ – зрачковые координаты

Что касается выбора самих значений c_i , то для улучшения свойств матрицы \mathbf{C} целесообразно положить их равными корням ортогональных полиномов [5], скажем, полиномов Лежандра или Чебышева, на отрезке $[0; 1]$.

Описанный способ нахождения коэффициентов поперечных aberrаций обладает низкой трудоемкостью по количеству рассчитываемых псевдолучей, высокой точностью (эта точность позволяет использовать степенной базис интерполяции как наиболее экономичный с точки зрения «восстановления» aberrаций) и удобством использования на компьютере.

Немаловажным фактором является и то, что его очень просто обобщить на случай вычисления хроматических aberrаций. Пусть $(\varepsilon_x)_{\lambda_i}$, $(\varepsilon_y)_{\lambda_i}$, $i = 0, 1, 2$ – составляющие поперечных aberrаций на длине волны λ_i . Тогда для расчета коэффициентов первичного хроматизма нужно при интерполяции вместо $\varepsilon_x, \varepsilon_y$ использовать разности $(\varepsilon_x)_{\lambda_2} - (\varepsilon_x)_{\lambda_1}$, $(\varepsilon_y)_{\lambda_2} - (\varepsilon_y)_{\lambda_1}$, а для расчета коэффициентов вторичного хроматизма – использовать $0,5[(\varepsilon_x)_{\lambda_2} + (\varepsilon_x)_{\lambda_1}] - (\varepsilon_x)_{\lambda_0}$ и $0,5[(\varepsilon_y)_{\lambda_2} + (\varepsilon_y)_{\lambda_1}] - (\varepsilon_y)_{\lambda_0}$. В остальном схема интерполяции полностью сохраняется.

Литература

1. Иванов А.В. Применение метода «псевдолучей» при автоматизированном синтезе оптических систем // Изв. вузов СССР. Приборостроение. 1986. № 3. С. 79--83.
2. Cox A. A system of the optical design. London: Focal, 1964. P. 175–181, 231–233.
3. Hopkins G.W. Proximate ray tracing and optical aberration coefficients // J. Opt. Soc. Am. V.66. №5. P. 405–410.
4. Hopkins G.W. Proximate ray tracing and wave aberration coefficients // J. Opt. Soc. Am. V.66. №9. P. 942–949
5. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. М.: Наука, 1987. 600 с.
6. Вычислительная оптика: Справочник / Русинов М.М., Грамматин А.П., Иванов П.Д. и др. / Под общ. ред. Русинова М.М. Л.: Машиностроение, 1984. 423 с.
7. Воеводин В.В. Вычислительные основы линейной алгебры. М.: Наука, 1977. 304 с.
8. Курош А.Г. Курс высшей алгебры. М.: Наука, 1971. 432 с.

МАТЕМАТИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ПРОЦЕССА ИЗМЕРЕНИЯ ПАРАМЕТРОВ МАССОПЕРЕНОСА В ЖИДКОЙ БИНАРНОЙ СРЕДЕ С ГРАНИЦЕЙ РАЗДЕЛА

З.Г.Симоненко

Работа посвящена вопросам описания и математического представления процесса измерения параметров массопереноса в жидкой бинарной среде с границей раздела для разработки алгоритмов математического обеспечения проектируемых вариантов моделей ИИС измерения параметров массопереноса сплошных сред.

Введение

Разработка математических моделей сложных систем связана с проведением очень трудоемких и разнообразных исследований. Обычно эти исследования начинают на этапе проектирования и заканчивают либо в процессе эксплуатации, либо после экспериментальных работ, организованных на средствах системы. Описания могут принимать различную языковую форму, поэтому математическое моделирование становится неотъемлемым этапом современного проектирования

На этапе проектирования в первую очередь решают вопросы выбора методов и способов реализации математических моделей при помощи вычислительных машин.

При решении этих вопросов определяющим фактором является ожидаемая сложность модели. На практике предполагается, что процессы в элементах исследуемой системы могут быть описаны с помощью решений дифференциальных, разностных или функциональных уравнений. Для определения параметров, входящих в такие уравнения, в процессе разработки системы проводят экспериментальные исследования на элементах и на средствах всей системы в условиях ее нормального функционирования.

Для объединения полученной информации разбивают сложную систему на такую совокупность подсистем, которая бы наилучшим образом отображала работу и функциональное взаимодействие всех ее элементов, участвующих при постановке того или иного вида физического эксперимента.

Разработка алгоритма математической модели системы на этих принципах позволяет создать программу на ЭВМ, состоящую из подпрограмм, которые можно при необходимости заменить или скорректировать по результатам физических экспериментов другими, более точными аналогами. Для сложных систем разрабатываемые модели обычно получаются чрезвычайно громоздкими. В значительной степени сложность создаваемых моделей зависит и определяется математическим обеспечением моделей, разрабатываемым для обработки результатов моделирования.

Основу алгоритмов математического обеспечения моделей составляют расчетные формулы методов, используемых при постановке экспериментов на модели и обработке всех полученных результатов. Поэтому на этапе создания моделей сложных систем стремятся выбрать метод математического моделирования, удовлетворяющий следующим требованиям:

1. расчет оценок выходных показателей должен осуществляться с использованием достаточно простых алгоритмов обработки;
2. определение необходимого объема моделирования из условий достижения заданной точности оценок выходных показателей должно происходить на основании простых и, в то же время, достаточно точных соотношений;
3. методика организации эксперимента на модели должна быть по возможности простой и реализуемой на средствах используемой вычислительной техники.

Целью данной работы является математическое представление процесса измерения параметров массопереноса в жидкой бинарной среде с границей раздела, полученных в результате экспериментальных исследований с помощью модели информационно-измерительной системы (ИИС), для разработки алгоритмов математического обеспечения проектируемых вариантов моделей ИИС измерения параметров массопереноса сплошных сред.[1].

Основная часть

В настоящее время основной тенденцией развития экспериментальной науки является повышение информативности исследований, достигаемое разработкой и применением высокоточных быстродействующих комплексных методов, обеспечивающих получение сведений о совокупности свойств в одном эксперименте, что достигается за счет повышения скорости измерительного процесса и скорости измерения, сокращения времени на подготовку измерения, а также сокращения времени самих измерений и обработки результатов. Сюда же относятся и методы, позволяющие имитировать рассматриваемые процессы с помощью программного обеспечения и последующей обработки информации.

Опτικο-физические методы и средства являются неконтактными, неразрушающими и обеспечивают локальность и экспрессность, наряду с высокой точностью и надежностью измерений. Разработка этих методов при изучении процессов нетрансляционного переноса массы компонентов (или диффузии) в жидких средах с границей раздела позволяет измерить ряд параметров массопереноса: разность показателей преломления, концентрацию диффундирующего раствора, а также коэффициент массопереноса (диффузии) – рабочего параметра для контроля за скоростью массопереноса исследуемых растворов.

В настоящее время разработаны и созданы новые лазерные методы и аппаратура для исследования явления массопереноса в жидких бинарных средах с границей раздела путем измерения параметров массопереноса в реальном масштабе времени, обеспечивающие объективный контроль за измеряемыми величинами в непрерывном и дискретном аппаратных режимах с учетом специфики протекающего динамического процесса переноса [2].

Для оптических методов при условии линейной зависимости показателя преломления раствора от градиента концентрации диффундирующей среды для растворов малой концентрации уравнение диффузии при соблюдении граничных условий приведено к виду

$$\frac{dn}{dx} = \Delta n \sqrt{\frac{k}{\pi}} h^{-kx^2}, \quad (1)$$

где Δn – градиент показателя преломления в слое, отстоящем на расстоянии X от границы раздела, $(n_1 - n_2)$ – первоначально заданная разность показателей преломления двух жидкостей,

$$K = 1 / (4D \tau),$$

D – коэффициент диффузии, τ – время от момента образования границы раздела.

Разность фаз, приобретаемая двумя лучами на пути h в диффундирующем слое, при введении переменной $y = \sqrt{kx}$ приводит соотношение (1) к виду:

$$\delta = \frac{\Delta n h}{2\lambda} L, \quad (2)$$

где

$$L = \operatorname{erf}\left[\frac{2}{\sqrt{4D\tau}}(X_0 + a/2)\right] - \operatorname{erf}\left[\frac{2}{\sqrt{4D\tau}}(X_0 - a/2)\right]. \quad (3)$$

Модель информационно-измерительной системы [3], представленной на рис. 1, содержит следующие элементы: 1 – источник излучения, 2 и 3 – светоделительные элементы (полупрозрачные зеркала), 4 – поляризатор (полуволновая пластина $\lambda/2$), 5 и 9 – двулучепреломляющие пластины из исландского шпата, 6 и 8 – четвертьволновая пластина ($\lambda/4$), 7 – кювета диффузионная, 10 – третья четвертьволновая пластина ($\lambda/4$), 11 – анализатор (полуволновая пластина $\lambda/2$), 12 – компенсатор Сенармона, 13 – фотодиод, 14 – аналого-цифровой преобразователь, 15 – блок управления, 16 – блок оперативной памяти, 17 – таймер, 18 – процессор, 19 – ЦПУ (цифropечатающее устройство), 20 – самописец.

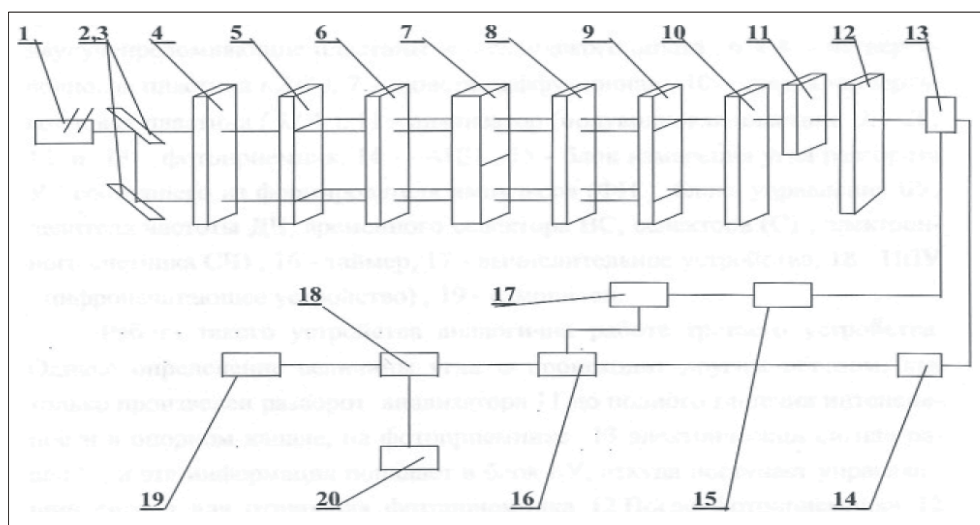


Рис. 1. Схема модели двухканальной информационно-измерительной системы для измерения параметров массопереноса

В данной схеме линейно-поляризованный пучок разделяют на два – рабочий и опорный, причем опорный пучок проходит значительно ниже границы раздела двух исследуемых жидкостей, т.е. вне рабочей зоны. Это условие выполняется в том случае, когда объем, в котором в течение опыта протекает процесс переноса, значительно меньше общего объема кюветы.

После сведения двух пучков в один полученный пучок эллиптически поляризованного света преобразуют в линейный. Далее производится анализ состояния поляризации светового пучка при пропускании его через ряд элементов, а именно, через систему поляризатор – оптические элементы – анализатор.

Анализ такого пучка производится следующим образом. Анализатор 11 установлен так, чтобы его плоскость поляризации была повернута на угол 90° относительно плоскости поляризации поляризатора 4, при этом изменение составляющей интенсивности в зависимости от меняющейся величины разности фаз происходит по синусоидальному закону. Затем анализатор устанавливают так, чтобы его плоскость поляризации была параллельна плоскости поляризации поляризатора, с этим последующим разворотом плоскости поляризации рабочего пучка происходит изменение составляющей интенсивности в зависимости от меняющейся величины разности фаз по косинусоидальному закону.

Для реализации измерений выделяются два значения – синусной (I_1) и косинусной (I_2) составляющих интенсивности поляризованного излучения для рабочего пучка в жидкой системе.

$$I\tau_1 = kI_0 \sin^2((\delta_1 + \delta_{\text{сист}})/2), \quad (4)$$

$$I\tau_2 = kI_0 \cos^2((\delta_2 + \delta_{\text{сист}})/2), \quad (5)$$

где k – коэффициент преобразования, I – интенсивность составляющих синусной (I_1) и косинусной (I_2) интенсивности излучения, δ_1, δ_2 – разность фаз, обусловленная наличием градиента концентраций исследуемой среды в момент, соответствующий мгновенному экстремальному значению величины синусной или косинусной составляющих интенсивности поляризованного излучения, прошедшего через диффундирующий слой, $\delta_{\text{сист}}$ – дополнительная разность фаз, обусловленная наличием систематической погрешности.

Для опорного пучка также выделяются два значения – синусной (I_3) и косинусной (I_4) составляющих интенсивности поляризованного излучения

$$I\tau_3 = kI_0 \sin^2((\delta_{\text{сист}})/2), \quad (6)$$

$$I\tau_4 = kI_0 \cos^2((\delta_{\text{сист}})/2). \quad (7)$$

Далее производится разворот плоскости поляризации излучения в рабочем и опорном каналах до полного гашения излучения в опорном канале. При этом величина интенсивности I_3 становится равной нулю: $I_3=0$. Отсюда и величина $\delta_{\text{дон}}=0$. Тогда разность фаз выражается в следующем виде:

$$\delta = 2 \arctg\left(\sqrt{\frac{I_1}{I_2}}\right) = \varphi. \quad (8)$$

Соотношение (8) позволяет получить зависимость разности фаз от разности углов поворота плоскостей поляризации в рабочем и опорном каналах, выраженной в единицах измеряемых величин синусной и косинусной составляющих величин (I_1 и I_2) интенсивности излучения. При этом за счет произведенного поворота производится исключение дополнительной разности фаз, вносящей погрешность в измерение.

Приравнявая обе части уравнений (3) и (8), получим следующее уравнение:

$$\operatorname{erf}\left[\frac{1}{\sqrt{4D\tau}}(X_0 + a/2)\right] - \operatorname{erf}\left[\frac{1}{\sqrt{4D\tau}}(X_0 - a/2)\right] = \frac{2\lambda\varphi}{\pi\Delta nh}. \quad (9)$$

Для успешного решения задачи массопереноса необходимо рассмотреть его физико-математическую модель (9), представляющую собой совокупность уравнений, формул, констант, логических условий, которые определяют взаимосвязь параметров изучаемого процесса, где параметрами являются измеренные в ходе эксперимента величины: τ – интервалы времени, D – коэффициент массопереноса, a также заданные граничные условия: X_0 – расстояние от границы раздела двух жидких сред до прямой, равноудаленной от двух ортогонально поляризованных пучков, α – расстояние между двумя ортогонально поляризованными пучками.

Численные значения параметров приведены ниже.

$$\alpha = 4 * 10^{-4}, X_0 = 4.5 * 10^{-4}, D = 1,88 * 10^{-9}, \tau_1 = 470, \tau_2 = 1000, \tau_3 = 1436.$$

Использование численных методов для решения задач эллипсометрии [4] и разработанная для рассматриваемой модели программа [5] для проверки корректности полученных экспериментальных данных и выполнения условия в окрестностях значений переменных, реализованная с использованием компилятора Borland C++ Builder под операционной системой Windows, позволяют произвести требуемые расчеты.

Большим удобством программы является наличие в ней подпрограммы для функции erf (при реализации которой использован метод аппроксимации гамма-функции по

6 параметрам [6]. Общая точность аппроксимированной функции составляет $3 \cdot 10^{-7}$ при 100 итерациях с использованием чисел с плавающей запятой двойной точности.

Заключение

В работе рассмотрено математическое представление процесса измерения параметров массопереноса в жидкой бинарной среде с границей раздела, которое необходимо для разработки алгоритмов математического обеспечения проектируемых вариантов моделей ИИС измерения параметров массопереноса сплошных сред.

Разработка алгоритма математической модели рассматриваемого варианта ИИС измерения параметров массопереноса позволяет создать программу, состоящую из подпрограмм, которые можно при необходимости заменить или скорректировать по результатам физических экспериментов другими более точными аналогами

Разработана программа для расчета параметров массопереноса, позволяющая объективно проводить оценку корректности величин, полученных в результате экспериментальных исследований бинарной жидкой среды методами эллипсометрии с помощью лазерных поляризационных интерферометров с элементами нуль-эллипсометрии, при одинаковых заданных начальных и граничных условиях. Это обеспечивается возможностями программы оптимизировать большое количество параметров с заданной степенью точности.

Проведенная работа позволяет решить большой комплекс задач, связанных с информационной и системной интеграцией процесса автоматизированного проектирования для различных задач моделирования ИИС измерения параметров массопереноса сплошных сред.

Литература

1. Симоненко З.Г., Порай-Кошиц А.Б., Шмуylieвич Г.А. Исследование молекулярной диффузии в жидкостях / В сб.: VI Менделеевская дискуссия. Результаты экспериментов и их обсуждение на молекулярном уровне. Тезисы докладов, ч. II, Харьков, 1983.
2. Симоненко З.Г. Оптические методы анализа процессов диффузии жидких сред. // Прогрессивные методы анализа объектов окружающей среды Л.: ЛДНТП, Материалы краткосрочного семинара, 1985, с. 75.
3. Симоненко З.Г., Порай-Кошиц А.Б., Москалев В.А. Способ определения коэффициента молекулярной диффузии в жидкостях и устройство для его реализации. / Авторское свидетельство СССР № 1349452. Бюллетень ОИ ПОТЗ, 1987. № 45, с. 218.
4. Симоненко З.Г., Ткалич В.Л. Использование программ численного решения некоторых задач эллипсометрии в учебном процессе. // Сб. трудов конф. «Оптика и образование» (16–17 октября 2002 г. Санкт-Петербург).
5. Симоненко З.Г. Исследование параметров скорости массопереноса в жидких бинарных системах с границей раздела. // Материалы IV Международной научной конференции «Проблемы пространства, времени и движения», СПбГИТМО (ТУ), Санкт-Петербург, 2000. с. 22.
6. Erf calculation using 6 parameter Lanczos approximation <http://www.rskey.org/gamma.htm>

ПРИМЕНЕНИЕ ФИЛЬТРА КАЛМАНА ВТОРОГО ПОРЯДКА ДЛЯ КОМПЬЮТЕРНОЙ ОБРАБОТКИ ИЗМЕРИТЕЛЬНОЙ ИНФОРМАЦИИ

А.М. Озерский

В статье рассматриваются особенности разработки и проектирования систем компьютерной обработки информации с использованием фильтра Калмана второго порядка на примере различных прикладных систем. Приведен анализ характеристик разработанных систем с учетом типичного жизненного цикла программного обеспечения.

Введение

Широкое применение систем компьютерной обработки данных в различных областях техники требует усовершенствования методик разработки таких систем. Системы компьютерной обработки измерительной информации, как и все компьютерные системы, имеют четко выраженные этапы проектирования. Один из типичных вариантов цикла создания системы обработки состоит из следующих фаз:

1. постановка задачи;
2. разработка технического задания;
3. выбор методов обработки информации;
4. проектирование и разработка программного обеспечения;
5. тестирование системы;
6. установка системы;
7. сопровождение системы.

В рамках приведенного выше цикла разработки предполагается, что свойства обрабатываемой измерительной информации определены до начала проектирования и разработки компьютерной системы в рамках отдельной исследовательской работы.

Первые две фазы цикла (постановка задачи и разработка технического задания) определяют результаты, которые необходимо получить при использовании системы. На последующих фазах выбираются и реализуются методы, с помощью которых возможно решить сформулированные задачи.

При выборе методов обработки информации учитываются, прежде всего, их возможность полностью обеспечить реализацию технического задания. С другой стороны, следует учитывать также необходимость минимизации затрат на разработку и различные риски, к которым могут относиться временные, финансовые и т.п. Таким образом, неправильный выбор подхода к созданию системы обработки информации может привести к невозможности выполнить любой из последующих этапов разработки.

Особенности применения фильтра Калмана второго порядка

В системах компьютерной обработки измерительной информации достаточно часто встречаются две задачи, для решения которых оптимальным выбором является использование фильтра Калмана второго порядка: это выделение заданной спектральной полосы сигнала и определение параметров модели сигнала.

Основными требованиями, которые при выборе метода обработки данных являются определяющими для использования фильтра Калмана второго порядка, являются следующие:

- высокое быстродействие;
- небольшое время разработки;
- небольшой объем программного кода;

- наличие готовых реализаций.

Несмотря на то, что в настоящее время вычислительные возможности аппаратного обеспечения значительно возросли, существует несколько классов задач, для которых требования по быстродействию алгоритмов являются критическими. Типичными видами компьютерных систем, требующих высокого быстродействия, являются системы обработки значительных объемов предварительно измеренных данных, а также системы, работающие в реальном времени в современных многозадачных средах, в которых на обработку данных зачастую отводятся единицы процентов процессорного времени или иных аппаратных ресурсов. Требования к используемому компьютерной системой объему оперативной памяти в настоящее время, по-видимому, не являются определяющими для большинства задач.

Небольшое время разработки программного обеспечения с использованием фильтра Калмана второго порядка определяется относительно небольшим объемом программного кода, а также относительной простотой реализуемых алгоритмов. Кроме этого, небольшой объем программного кода позволяет упростить процедуры тестирования и сопровождения системы.

Дискретный линейный фильтр Калмана определяется в общем случае векторными уравнениями наблюдения и системы

$$\mathbf{s}(k) = \mathbf{C}(k)\boldsymbol{\theta}(k) + \mathbf{n}(k), \quad (1 \text{ а})$$

$$\boldsymbol{\theta}(k+1) = \mathbf{A}(k)\boldsymbol{\theta}(k) + \mathbf{w}(k), \quad (1 \text{ б})$$

где $k = 1, \dots, K$ - номер отсчета дискретного сигнала, $\mathbf{s}(k)$ - наблюдаемый сигнал, $\mathbf{C}(k)$ - матрица измерений, $\mathbf{A}(k)$ - матрица перехода, $\boldsymbol{\theta}(k)$ - вектор состояния системы, $\mathbf{n}(k)$ и $\mathbf{w}(k)$ - шумы наблюдения и системы соответственно.

В одномерном случае алгоритм фильтрации Калмана второго порядка выражается уравнениями

$$\theta_1(k-1) = \theta_1(k-1) - \theta_1(k-2), \quad (2)$$

$$\theta_1(k) = \theta_1(k-1) - 2a\theta_1(k-1) - b^2\theta_1(k-1) + b^2w(k),$$

где a, b - постоянные коэффициенты.

Фильтр, определяемый уравнениями (2), может быть использован как полосовой фильтр, причем полоса пропускания такого фильтра определяется уравнением

$$G_\theta(u) = \frac{b^4 N_\theta}{2[(b - 2\pi u)^2 + a^2][(b + 2\pi u)^2 + a^2]}. \quad (3)$$

и имеет вид, представленный в качестве примера на рис. 1.

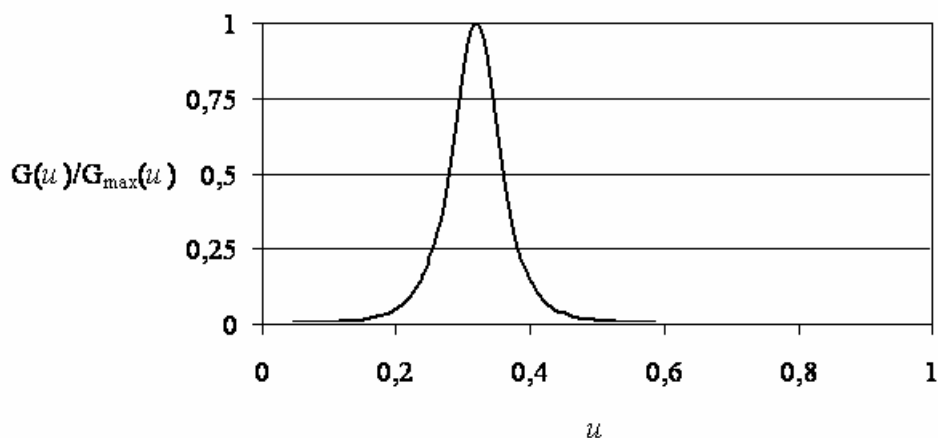


Рис 1. Полоса пропускания фильтра Калмана второго порядка

Таким образом, изменяя соответствующие параметры, фильтр можно настроить на различные полосы пропускания. При этом следует учитывать, что фильтр может терять устойчивость, если значительная часть энергии сигнала сосредоточена в области частот, более высоких, чем половина частоты Найквиста.

Анализ процесса проектирования и разработки прикладных систем

Анализ характеристик фильтра Калмана второго порядка проводился на представительной выборке различных прикладных систем компьютерной обработки измерительной информации:

1. Система анализа электрокардиограмм. Фильтр Калмана использовался для подавления различных помех известной частоты.
2. Система анализа вибрационных и акустических данных. Фильтр Калмана использовался для выделения полезной составляющей сигнала.
3. Подсистема фильтрации выходных воздействий авторулевого устройства. Фильтр использовался для согласования выходных воздействий и исполнительного механизма.
4. Фильтр навигационных данных. Фильтр использовался для определения параметров движения судна.

Был проведен анализ следующих свойств и метрик систем: объем программного кода, суммарное время разработки и быстродействие алгоритмов. В тех случаях, когда реализация систем проводилась с использованием нескольких альтернативных методов обработки данных, возможно провести сравнительный анализ использования различных методов.

Для оценки объема программного кода был использован общепринятый метод, заключающийся в подсчете числа строк кода программного обеспечения на языке высокого уровня, в разработанных системах – на языке C++. Известно, что такой показатель следует рассматривать как приблизительный, так как в зависимости от конкретной реализации алгоритма, в частности, от стиля и способа кодирования и оптимальности программного кода, получаемая оценка может иметь изменчивость до 50%, а в некоторых случаях и более.

В простейшем случае, линейный дискретный одномерный фильтр Калмана второго порядка реализуется примерно 50-ю строками программного кода.

В системах 1, 2 и 3, где описанный выше фильтр используется для выделения частотных полос, для его реализации потребовалось приблизительно 100 строк кода. Аналогичный код, реализующий алгоритм быстрого преобразования Фурье (БПФ) потребовал около 150 строк.

Фильтр, реализованный в системе 4, имел две различные реализации: основанную на гидродинамической модели судна и с использованием фильтра Калмана второго порядка для анализа измерительной информации без учета гидродинамики судна. В первом случае код занял около 5000 строк, во втором случае – около 2000, причем точность полученных результатов различалась незначительно.

Оценка времени проектирования системы обработки данных является, как правило, еще менее точной и субъективной величиной, чем объем программного кода. Получение такой оценки может быть затруднено как природой измеряемой величины, так и объективными сложностями самого процесса измерения.

При оценивании времени разработки подсистемы фильтрации для выделения спектральных полос сигнала время разработки составило около 40 часов, в то время как система на базе БПФ была создана приблизительно за 30 часов. Более быстрая разработка системы с использованием БПФ определяется существованием готовых про-

граммных библиотек, обеспечивающих соответствующую функциональность, а также относительной простотой и более широкой известностью алгоритма БПФ.

При разработке фильтра навигационных данных было затрачено при использовании подхода, основанного на анализе гидродинамических свойств судна, более 450 часов, в то время как при анализе свойств поступающих сигналов без такого анализа около 150 часов, что соответствует более чем трехкратному выигрышу по времени проектирования. Такое значительное преимущество объясняется в основном тем, что создание более компактного кода не только выполняется быстрее, но и позволяет значительно сократить время на тестирование, отладку и верификацию системы.

Таким образом, можно предположить, что использование фильтра Калмана второго порядка может дать значительные преимущества по времени разработки в случае, когда не существует готовых алгоритмов для решения поставленных задач. В этом случае более компактный программный код и относительная простота алгоритмов позволяют уменьшить время разработки.

Еще одной важной характеристикой компьютерной системы, которую следует учитывать на этапе проектирования, является быстродействие программного обеспечения и вычислительная сложность алгоритмов.

При сравнении быстродействия различных методов анализа сигналов следует учитывать различные требования к реализации конкретных алгоритмов, а также конкретные особенности поступления обрабатываемой измерительной информации. Так, некоторые методы анализа требуют вычислений по скользящему окну, что может вызывать «краевые эффекты» в случае, если данные поступают последовательно по одному отсчету.

В предположении, что необходимо обработать n отсчетов данных, результаты оценки вычислительной сложности полосовой фильтрации с использованием алгоритмов типа БПФ дают оценку порядка $n \ln(n)$, в то время как аналогичная фильтрация и использованием фильтра Калмана имеет оценку порядка n . Таким образом, использование фильтра Калмана второго порядка позволяет добиться заметного увеличения быстродействия разрабатываемой системы.

Априорная оценка быстродействия алгоритмов обработки данных, как правило, требует моделирования или предварительной реализации алгоритмов на языке программирования и в среде исполнения, адекватных разрабатываемой системе. В противном случае, использование высокоуровневых средств моделирования, таких как известные программные пакеты MathCad и MatLab, может дать неверные данные по причине различия свойств сред выполнения программ. В описанном выше случае сравнительная оценка производительности алгоритмов БПФ и полосовой калмановской фильтрации может быть существенно искажена наличием в большинстве программных пакетов встроенной реализации БПФ, написанной на языке относительно низкого уровня и оптимизированной под установленный тип процессора. Анализируемый на скорость выполнения алгоритм фильтра Калмана второго порядка, написанный на интерпретируемом языке высокого уровня, покажет неадекватно низкую производительность в силу особенностей работы конкретного программного пакета.

Если ставится задача определения параметров модели сигнала, то вычислительная сложность алгоритма пропорциональна квадрату размерности вектора параметров системы. Таким образом, если на этапе проектирования системы удастся добиться частично или полностью независимой обработки в нескольких каналах измерения, это позволяет существенно увеличить быстродействие. Возможность обеспечить независимую многоканальную обработку можно определить исходя из априорной информации о сигналах, полученной на этапе исследования первичных свойств обрабатываемых данных. В некоторых случаях точность одного из датчиков может существенно превышать

точность остальных, что позволяет понизить размерность модели за счет отбрасывания малозначительных данных, в том числе с использованием различных динамических критериев значимости данных.

При выборе методов для решения задачи определения параметров модели сигнала следует учитывать, что существует множество методов, например, анализ статистических характеристик исследуемых данных, вычислительная сложность которых не превосходит вычислительную сложность фильтра Калмана второго порядка. В этих случаях выбор метода целесообразно проводить исходя из априорно известных свойств сигналов и выбранной модели системы. В данном случае значительную роль при проектировании системы обработки измерительной информации важную роль приобретают субъективные факторы, такие, как полнота априорных знаний и критичность объема программного кода. Как показано выше, зачастую меньший объем программного кода, обеспечиваемый при использовании фильтра Калмана второго порядка, оказывается важнее преимуществ в производительности при применении других упрощенных алгоритмов.

Сводная таблица свойств программного обеспечения на базе различных методов обработки данных приводится ниже.

Таблица 1. Характеристики программного обеспечения на основе различных методов обработки данных

Методы обработки данных	Сложность проектирования	Объем программного кода	Сложность верификации	Быстродействие
Фильтры Калмана 2-го порядка	Умеренная	Малый	Умеренная	Высокое
Спектральные методы на основе БПФ	Невысокая	Малый	Малая	Низкое
Статистические методы	Невысокая	Умеренный	Умеренная	Среднее

Следует отметить, что при проектировании прикладных систем первостепенным фактором при выборе того или иного метода является возможность обеспечить решение задач, для которых проектируется прикладная система. Приведенный выше сравнительный анализ различных методов обработки данных позволяет принять решение о выборе наиболее подходящего метода в случае, когда требуемая функциональность системы может быть обеспечена несколькими различными методами.

Заключение

При проектировании систем компьютерной обработки измерительной информации корректный выбор методов и алгоритмов обработки данных является критически важным для успешного создания таких систем. При выборе методов и алгоритмов обработки следует учитывать не только их принципиальную пригодность для получения необходимых результатов, но и конкретные технические ограничения процесса проектирования создаваемой системы, носящие как объективный, так и субъективный характер, например, объем создаваемого для реализации алгоритма программного кода, время написания такого кода и его верификации.

Использование фильтра Калмана второго порядка позволяет значительно сократить время разработки за счет относительной компактности реализующего соответствующие алгоритмы программного кода. Таким образом, снижаются не только времен-

ные затраты на разработку, но и уменьшается время тестирования и верификации программного обеспечения.

Как показывает сравнительный анализ прикладных систем из различных областей техники, время создания программного обеспечения и объем программного кода могут быть уменьшены в несколько раз при использовании алгоритмов фильтрации Калмана невысоких порядков, за счет отказа от избыточно сложных математических моделей систем.

Благодарности

Автор выражает признательность проф. И.П. Гурову за методическую помощь и полезное обсуждение результатов исследований по проблематике проектирования систем динамической обработки данных методом фильтрации Калмана второго порядка.

Литература

1. Бассвиль М. и др. Обнаружение изменения свойств сигналов и динамических систем. М.: Мир, 1989
2. Бейзер Б. Тестирование черного ящика. Технологии функционального тестирования программного обеспечения и систем. СПб.: Питер, 2004.
3. Гуров И.П., Жиганов П.Г., Озерский А.М. Анализ акустических сигналов на основе метода фильтрации Калмана //Научно-технический вестник СПбГИТМО (ТУ). 2002. Вып. 6. С. 200-206.
4. Дьяконов В., Круглов В. Математические пакеты расширения MATLAB. Специальный справочник. СПб.: Питер, 2001.
5. Колчин А.Ф., Овсянников М.В., Стрекалов А.Ф., Сумароков С.В. Управление жизненным циклом продукции. М.: Анахарсис, 2002.
6. Макс Ж. Методы и техника обработки сигналов при физических измерениях. – М.: Мир, 1983.
7. Kalman R.E. A new approach to linear filtering and prediction problems //Trans. ASME, J. Basic Eng. 1960. V.82. P. 35-45.

ПРИМЕНЕНИЕ ТЕОРИИ КООРДИНАЦИИ ДЛЯ УПРАВЛЕНИЯ КАЧЕСТВОМ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ СО СЛАБО ФОРМАЛИЗУЕМЫМИ КРИТЕРИЯМИ

Н.Ф. Гусарова

Введение

Организация управления в слабо формализуемых (СФ) системах все более активно входит в круг проблем, решаемых современной теорией управления [1]. Понятие слабой формализуемости принято связывать с предложенной Г. Саймоном классификацией типов проблем управления и принятия решений [2]:

- хорошо структурированные (количественно сформулированные);
- неструктурированные (качественно выраженные);
- слабо (плохо) структурированные (смешанные). В литературе имеются различные, хотя и концептуально близкие, трактовки понятия слабой формализуемости применительно к объектам управления. Приведем некоторые примеры.

«СФ временные процессы – процессы, структура и законы, поведения которых не известны либо недостаточно изучены для принятия решений» [3].

«СФ процессом называется динамический процесс, относящийся к классу неструктурированных и слабо структурированных проблем управления и обладающий следующими характеристиками:

- уникальность процесса;
- качественная природа параметров предметной области;
- неоднородность (разнотипность) шкал измерений параметров;
- нелинейный (имплицативный) характер взаимосвязи характеристик;
- многоуровневая иерархическая организация взаимосвязи подпроцессов;
- многообразие возможных форм взаимодействия подпроцессов между собой, порождающее неоднородность информации, циркулирующей в системе;
- многокритериальность, зачастую с противоречивыми критериями, которым должен удовлетворять процесс» [4].

«СФ системы можно охарактеризовать следующим образом:

- их структура и/или функционирование не могут быть описаны формальными моделями в виде некоторой совокупности уравнений;
- они являются активными и обладают «свободой воли»;
- цели их существования и критерии управления ими также не могут быть формализованы и меняются во времени» [5].

Традиционно к СФ, в первую очередь, относят объекты управления высших уровней классификации [6] – социально-экономические и организационно-технические системы. При разработке технических средств управления такими объектами наибольшее внимание уделяется методам коллективной многовариантной экспертизы [7, 8], теории кооперативных игр [9], искусственного интеллекта [10, 11]. В технических приложениях к СФ системам, как правило, относят крупномасштабные производственные объекты оборонной, химической, ракетно-космической и других отраслей промышленности, рассматриваемые как большие системы, моделирование которых затруднено вследствие их размерности [12].

Однако переориентация промышленного производства на индивидуального потребителя заставляет относить к СФ все большее число традиционных технологических процессов (ТП) легкой, пищевой, полиграфической и др. отраслей. Качество продукции здесь определяется не только в формально-количественных, но и в когнитивных категориях. Как правило, эти ТП реализуют выпуск большой номенклатуры продукции ма-

лыми (вплоть до единичных экземпляров) партиями, каждая из которых с точки зрения статистики составляет отдельную генеральную совокупность, что не позволяет использовать при разработке принципы статистической классификации, распознавания образов и т.д. Попытки перевода их в класс больших систем (за счет экстенсивного наращивания количества контролируемых параметров) приводят только к формированию Парето-множества недоминируемых альтернатив. В то же время непосредственно ход ТП вполне удовлетворительно описывается методами классической теории управления.

Поэтому СФ характер таких ТП целесообразно моделировать посредством СФ критериев, а весь ТП – базируясь на принципах самоорганизации [13] и координационного управления [14]. Эти принципы активно развиваются в направлении создания интеллектуальных систем управления динамическими процессами, описание которых выходит за рамки количественных моделей [4, 15–17]. Однако одним из ключевых элементов предложенных решений является база знаний, построение которой в явном виде для рассматриваемых ТП представляет значительные трудности. В настоящей работе рассматриваются возможности организации координационного управления такими ТП с непосредственным включением человека как носителя интегрального критерия в контур управления. Результаты конкретизируются для двух ТП – ТП трикотажного производства [18] и компьютерной допечатной подготовки полиграфической продукции [19].

1. Математическая модель ТП как иерархической системы

Основные идеи координации [14] применительно к задаче работы поясняются рис. 1. ТП моделируется системой S , состоящей из $(n+2)$ основных подсистем: вышестоящей управляющей системы (координатора) C_0 , n нижестоящих управляющих систем C_1, \dots, C_n и управляемого процесса P . Процесс P состоит из отдельных подпроцессов P_i , взаимодействие между которыми осуществляется с помощью связующих сигналов u_i .

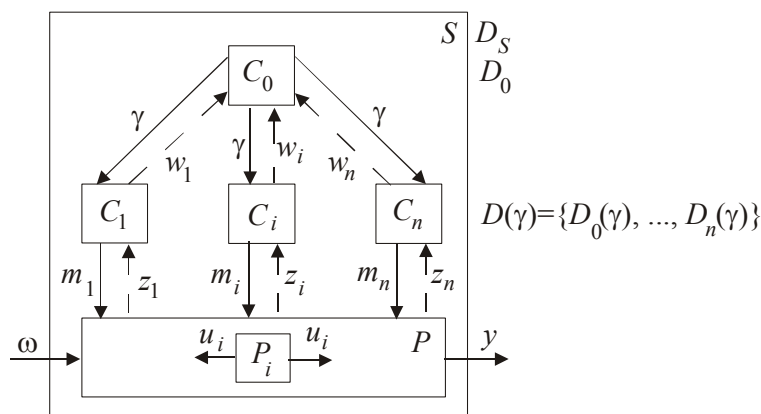


Рис. 1. Двухуровневая иерархическая система

Ни один из элементов системы S не решает ее глобальную задачу D_S , хотя последняя и определена в терминах всего процесса. Каждая из подсистем решает свои задачи: D_0 соответствует цели координатора, $D(\gamma) = \{D_1(\gamma), \dots, D_n(\gamma)\}$ – целям нижестоящих управляющих систем. В общем случае все эти цели различны, однако цели элементов нижестоящего уровня зависят от координирующего параметра γ , получаемого от координатора. Координатор, в свою очередь, выбирает значение γ так, чтобы обеспечить решение своей собственной задачи. Если между D_S , D_0 и $D(\gamma)$ имеются определенные закономерности, то глобальная цель системы может быть достигнута. Для создания этих закономерностей координатор может использовать различные пути параметризации, измеряя либо множество целевых функций подсистем C_1, \dots, C_n , либо

множество связующих входов для управляемых ими подпроцессов, либо то и другое вместе.

ТП с СФ критериями, как правило, имеют вполне определенную матрицу технологических связей. Поэтому множество $U \supset u_i$ оказывается недоступным для управления, кроме элемента u_n , который одновременно является глобальным выходным сигналом ТП в целом, и в этом случае применима параметризация путем прогнозирования взаимодействий [14]: управляющие подсистемы C_i , $1 \leq i \leq n$, получают от координатора C_0 не только значение параметра β своей целевой функции, но и прогнозные значения связующих сигналов, α , и включают их в решение своих задач $D(\gamma)$, т.е. $\gamma = \{\alpha, \beta\}$. Тогда решение задачи координации в содержательном смысле сводится к подбору текущих значений β , которые при соблюдении матрицы технологических связей, т.е. при совпадении прогнозных и реальных значений α , доставляют максимум глобальной функции качества системы в целом.

Математическая модель ТП (рис. 1) строится на уровне теоретико-множественного описания системы S в каузальной

$$S: X \rightarrow Y \quad (1.1)$$

и критериальной

$$G: X \times Y \rightarrow V \quad (1.2)$$

формах. Во входном объекте X процесса P целесообразно выделить множество M управляющих сигналов m , $m \subset M$, и множество Ω внешних возмущений ω , $\omega \subset \Omega$, т.е. считать, что процесс P реализует отображение

$$P: M \times \Omega \rightarrow Y. \quad (1.3)$$

Оценочный объект V системы S можно интерпретировать как множество оценок качества конечного продукта ТП. Предполагается, что множество M представимо в виде декартова произведения n подмножеств локальных управляющих сигналов M_i :

$$M = M_1 \times M_2 \times \dots \times M_n, \quad 1 < i < n. \quad (1.4)$$

Каждый локальный управляющий сигнал $m_i \subset M$ формируется локальной управляющей системой C_i и воздействует на подпроцесс P_i . Тогда управляющий сигнал для процесса в целом может быть записан в виде вектора

$$m = \{m_i\}. \quad (1.5)$$

Координатор C_0 воздействует на локальные управляющие системы C_i с помощью координирующих сигналов

$$\gamma = (\gamma_1, \dots, \gamma_n), \quad (1.7)$$

образующих множество Γ , $\gamma \subset \Gamma$, причем будем полагать, что на вход i -ой локальной управляющей системы поступает только компонента γ_i .

Опишем соотношение между процессом P в целом и его подпроцессами P_i . Полагая, что $U = U_1 \times U_2 \times \dots \times U_n$, определим функции H на множестве $M \times Y$ и \bar{P} на множестве $M \times U \times \Omega$ в виде

$$H(m, y) = (H_1(m, y), \dots, H_n(m, y)); \quad (1.8)$$

$$\bar{P}(m, u, \omega) = (P_1(m, u, \omega), \dots, P_n(m, u, \omega)). \quad (1.9)$$

Подпроцессы связаны между собой, если условие

$$y = \bar{P}(m, H(m, y), \omega) \Leftrightarrow y = \bar{P}(m, \omega) \quad (1.10)$$

выполняется для всех $(m, y, \omega) \subset M \times Y \times \Omega$. Из (1.10) следует, что u является результатом отображения

$$K: M \times \Omega \rightarrow U, \quad (1.11)$$

или, в покомпонентной записи,

$$K_i: M \times \Omega \rightarrow U_i. \quad (1.12)$$

которое, в свою очередь, определяется уравнением

$$K(m, \omega) = H(m, P(m, \omega)). \quad (1.13)$$

Таким образом, функция K – функция взаимодействия подпроцессов – определяется соотношением

$$P(m, \omega) = \bar{P}(m, K(m, \omega), \omega). \quad (1.14)$$

При стационарности системы S и окружающей среды множество Ω может быть сужено до одного элемента. Предполагая, что система в целом работает в условиях определенности, можно записать

$$P: M \times U \rightarrow Y; \quad (1.15)$$

$$G: M \times U \rightarrow V; \quad (1.16)$$

$$P_i: M \times U_i \rightarrow Y_i. \quad (1.17)$$

Целевые свойства системы S характеризуются следующими функциями:

- глобальная целевая функция $G: M \rightarrow V$ или

$$g(M) = G(m, P(m)); \quad (1.18)$$

- межуровневые функции качества $\Psi_\gamma: V^n \rightarrow V$ или

$$\Psi_\gamma = \{(g_{1\gamma}(m_1, K_1(m)), \dots, g_{n\gamma}(m_n, K_n(m))), g(m): m \subseteq M\}; \quad (1.19)$$

- кажущаяся глобальная целевая функция $g_\gamma: \Gamma \times M \times U \rightarrow V$ или

$$g_\gamma(\gamma, m, u) = \Psi_\gamma(g_{1\gamma}(m_1, u_1), \dots, g_{n\gamma}(m_n, u_n)); \quad (1.20)$$

существенно, что для любых $\gamma \subseteq \Gamma$ и $m \subseteq M$

$$g_\gamma(\gamma, m, K(m)) = g(m). \quad (1.21)$$

т.е. $g_\gamma(\gamma, m, u) = g(m)$ при $u = K(m)$.

Условия координируемости системы, кроме общих свойств объектов системы S как подмножеств нормированных линейных пространств, базируются на двух предпосылках: монотонность межуровневых функций качества системы

$$g(m) \leq g(m') \text{ при } g_i(m) \leq g_i(m'), 1 \leq i \leq n, m \subseteq M, m' \subseteq M; \quad (1.22)$$

существование аддитивного представления глобальной целевой функции G через заданные локальные функции качества $G_i, 1 \leq i \leq n$, т.е. функции

$$\tilde{G}(m, y) = \sum_{i=1}^n a_i G_i(m_i, y_i, H_i(m, y)) \quad (1.23)$$

такой, что глобальные управляющие воздействия, оптимальные с точки зрения G , одновременно глобально оптимальны по отношению к \tilde{G} .

Так как функция предпочтений потребителя не имеет глобального экстремума, то при ориентации на индивидуального потребителя задачу ТП в общем случае целесообразно ставить не как задачу оптимизации, а как задачу улучшения. Здесь эффективной оказывается координация путем оценки взаимодействий [14], когда подсистемы $C_i, 1 \leq i \leq n$, получают от координатора C_0 , помимо параметра β , еще и диапазон прогнозных значений связующих сигналов α , и при решении своих задач $D(\gamma)$ рассматривают его как диапазон возмущений. Легко видеть, что прогнозирование взаимодействий можно рассматривать как частный случай оценки взаимодействий, когда множество возмущений является одноэлементным. Таким образом, координацию в ТП с СФ критериями целесообразно строить на базе оценки (в частном случае – прогнозирования) взаимодействий. При прогнозировании взаимодействий для каждого $\gamma \subseteq \Gamma$ выделяемое множество U^γ состоит из одного элемента u^γ . Тогда задача i -го локального управляющего элемента есть задача оптимизации на множестве M_i с заданным значением α^γ связующего входа. Как правило, прогнозирование взаимодействий используется параллельно с модификацией целей. В этом случае значение $\gamma \subseteq \Gamma$ в (1.10) понимается расширительно, а именно:

$$\gamma = (\alpha, \beta); \quad \Gamma = A \times B, \quad (1.24)$$

где α – прогнозное значение связующего входа, $\alpha \subseteq A$, а β – параметр, конкретизирующий локальные функции качества $G_{i\gamma} = G_{i\beta}, \beta \subseteq B$.

Если для двухуровневой системы заданы два множества A и B такие, что $A=K(M)$, где K описывается выражениями (1.11), (1.12), а каждое $\beta \in B$ определяет локальные целевые функции $g_{i\beta}$, $1 \leq i \leq n$, то для каждого координирующего сигнала $\gamma = (\alpha, \beta)$ из $A \times B$ i -ая локальная задача состоит в минимизации $g_{i\beta}(m_i, \alpha)$ на множестве M , т.е. нахождении такого значения m_i^γ , что

$$g_{i\beta}(m_i^\gamma) = \min_{M_i} g_{i\beta}(m_i, \alpha). \quad (1.25)$$

При этом в фиксированной паре $\gamma = (\alpha, \beta)$ из $A \times B$ прогноз считается правильным, если

$$\alpha = K(m^\gamma), \quad (1.26)$$

где $m^\gamma = m_1^\gamma \times \dots \times m_n^\gamma$. Кроме того, нужно исключить те пары $\gamma = (\alpha, \beta)$, для которых прогноз правилен, но управляющее воздействие не является глобальной оптимальным.

2. Задание СФ критериев в модели ТП

Основные проблемы, связанные с организацией управления в ТП с СФ критериями, состоят в поиске инструментальных методов отображения представлений конкретного потребителя о будущем изделии на множество параметров изделия, которые регулируются технологическим процессом.

Система предпочтений заказчика, в соответствии с теорией потребительского выбора [20], описывается отношением предпочтения P между товарами x и y . В [20] постулируется, что P обладает следующими свойствами: иррефлексивность (xPx для любого $x \geq 0$), транзитивность (из xPy и yPz следует xPz), монотонность (из $x \geq y$ следует xPy) и выпуклость (из xPz и zPy следует $[tx+(1-t)y]Pz$ для всех $0 < t < 1$). Здесь также вводится следующее определение: если существует такая числовая функция $U(x)$, $x \geq 0$, что

$$U(x) > U(y) \Leftrightarrow xPy, \quad (2.1)$$

то U называется индексом полезности для P .

К сожалению, для ТП с СФ критериями предположение (2.1) оказывается слишком сильным. В этом случае правомерно считать, что оценочный объект заказчика V_k , порождаемый его отношением предпочтений P , представляется одномерным линейным метрическим пространством

$$V_k = R. \quad (2.2)$$

Множество V_k является открытым и, в соответствии с [20], не имеет экстремума (т.е. не существует изделия, которое потребитель признал бы идеальным, и набора изделий, который потребитель признал бы репрезентативной выборкой). Это означает, что задача управления ТП, по крайней мере, на уровне глобальной целевой функции, в общем случае представляет собою задачу улучшения, а не строгой оптимизации.

Для задания оценочного объекта изделия V (1.2) введем в рассмотрение характеристическое пространство изделия размерностью n ,

$$V = R^n, \quad (2.3)$$

накрывающее прямое произведение множеств, соответствующих всем возможным признакам (параметрам) изделия. Структура этого прямого произведения и его свойства зависят от возможностей описания каждого возможного признака – точечное упорядоченное множество, линейное пространство с метрикой, многообразие. Непосредственная организация управления ТП с такими характеристиками оценочного объекта представляется проблематичной, и целесообразно переходить к его подмножествам.

Выделять подмножества в V можно различными способами. Так, в практике организации технического контроля рассматриваемых ТП распространено представление множества V в виде декартова произведения двух подмножеств V_T и V_S , описывающих технические (нормативные) и когнитивно-эстетические признаки ТП, соответственно:

$$V_k = V_T \times V_3. \quad (2.4)$$

В свою очередь, V_T разделяется на два подмножества – V_T^1 и V_T^2 :

$$V_T = V_T^1 \times V_T^2. \quad (2.5)$$

где V_T^1 включает в себя показатели качества продукции, задаваемые нормативной документацией в абсолютных шкалах, а V_T^2 – показатели качества продукции, задаваемые нормативной документацией в шкалах слабее абсолютной.

Однако в контексте настоящей работы более перспективным оказывается выделение подмножеств в V по обобщенным признакам, соотносимым со структурой семантического пространства человека. Например, такой обобщенный признак, как «артикул полотна» в ТП трикотажного производства или «тип технологического процесса» в ТП допечатной подготовки полиграфической продукции, выделяет в $V = R^n$ несвязные (в общем случае) области размерностью $m < n$, а в каждом из оставшихся m признаков – определенные подобласти. Тогда, выбрав тип ТП, потребитель фиксирует подмножество в характеристическом пространстве изделия V , в котором $V_{\text{реал}}$ представлено сечением V плоскостью размерностью $n-m$. В качестве обобщенных признаков, как показано в [5], могут выступать и признаки, формируемые непосредственно в ментальном пространстве человека – например, «художественный стиль» – если они одинаково понимаются заказчиком и исполнителем, т.е. существует гомоморфизм отображений $V_k \rightarrow V_{\text{реал}}$ исполнителя на отображение $V_k \rightarrow V_{\text{реал}}$ заказчика. Для обеспечения такого гомоморфизма можно, например, использовать выбор по эталонным моделям (образцам изделий, выполненных в соответствующем стиле).

Таким образом, формируется итерационная процедура, состоящая в последовательном решении задачи улучшения с одновременным снижением размерности и связности признакового пространства V . Процедура повторяется до выделения односвязной области признакового пространства, V_n , на котором выполняется завершающее улучшение. Очевидным условием правомерности такого процесса является

$$V_i \supseteq V_{i+1}, 1 \leq i \leq n. \quad (2.6)$$

Внутри этой области решение задачи улучшения организуется уже по формальным признакам. По существу, предложенный механизм находится в русле решения задач выбора [21]. Особенностью процедуры (2.1)–(2.6) является то, что она реализуется в признаковом пространстве более общего вида, позволяет учитывать неформализуемые альтернативы, а также не требует задания в явном виде функции выбора.

Существенно, что даже на уровне V_n решение задачи улучшения производится в терминах СФ критериев, т.е. не требует задания критериальной функции в явном виде. Ее носитель – потребитель или другое лицо, персонифицирующее его требования, – выступает здесь в роли обобщенного наблюдателя, что обеспечивает инструментальные средства перевода его текущего состояния как ЛПР со своими «поведенческими инвариантами» в параметры конкретного ТП. Возможность такого перевода применительно к исполнителю – активному звену ТП – базируется на его модельном представлении, предложенном в [22].

3. Определение задач глобального и локальных уровней управления качеством ТП

Глобальная задача управления качеством ТП на содержательном уровне ставится как «производство кондиционной продукции» с ограничениями типа «приемлемые издержки на управление», «приемлемое время выполнения заказа» и т.п. Если ограничения известны, они могут быть представлены как некоторая («эксплуатационная») область пространства состояний, причем для организации управления, согласно [23], существенно, чтобы эта область была односвязной. Но, как правило, в ТП с СФ критерия-

ми и глобальная задача, и ограничения являются слабо формализуемыми и образуют многосвязные области в признаковом пространстве.

Идеология выделения односвязной области с одновременным снижением размерности признакового пространства, представленная в разделе 2 для глобальной задачи, применима и к СФ ограничениям, хотя, как показывает опыт, здесь не всегда удается обеспечить строгое решение задачи улучшения на уровне предварительных экспертных оценок. Например, при разработке комплекса контрольной аппаратуры для ТП трикотажного производства [24] на этапе эскизного проектирования рассматривались две альтернативы – многоканальный прибор, заведомо более простой в разработке, но, по мнению потребителя, более сложный в эксплуатации, или одноканальный прибор со сложной механической схемой пространственного позиционирования, которая представлялась потребителю более привычной. Другими словами, области признакового пространства V_i (2.6), соответствующие этим альтернативам, образовали пересечение по целому ряду признаков – надежность, расходы на разработку, эксплуатационные расходы и др. В этом случае для следующей итерации была выбрана область, соответствующая пожеланию потребителя.

Определение задач D_i локальных управляющих подсистем в ТП с СФ критериями невозможно произвести путем непосредственной декомпозиции глобальной целевой функции (1.18). Во-первых, последняя не представлена в формальном виде. Во-вторых, конфигурация подсистем, как правило, определяется независимо от координатора: например, в ТП трикотажного производства она задается его топологической декомпозицией, а в ТП компьютерной допечатной подготовки полиграфической продукции – характеристиками команды исполнителей, в том числе их активностью. Поэтому локальные задачи D_i ставятся на содержательном уровне, а их координация обеспечивается за счет параметризации критериальных функций локальных подсистем, осуществляемой координатором процесса. Так, для ТП трикотажного производства задача локальной подсистемы определяется как максимизация вероятности обнаружения дефектов продукции, производимой на соответствующем технологическом переделе (группе переделов). Для ТП компьютерной подготовки изданий локальная задача соответствует изготовлению определенной группы технологических элементов издания в целом (рисунков, текста и т.д.) с качеством, соответствующим требованиям потребителя.

Задачей координатора D_C при выполнении определенных условий может служить кажущаяся глобальная целевая функция $g_\gamma: \Gamma \times M \times U \rightarrow V$ (1.20). Такими условиями, как показано в разделе 2, являются строгая выпуклость признакового пространства каждой локальной подзадачи, монотонная связь между глобальной целевой функцией, локальной целевой функцией и соответствующим признаком, а также постоянство конфигурации ТП, т.е. отображения K , в процессе координации. Все они представляются достаточно тривиальными в ТП, ориентированных на полностью формализуемые критерии. Так, условиям монотонности в той или иной форме удовлетворяют распознающие алгоритмы, входящие в большинство известных параметрических моделей [25], а выпуклость признакового пространства – базовое условие в математическом программировании. Тем не менее, их выполнение в ТП трикотажного производства оказалось далеко не рутинной процедурой и потребовало проведения целого комплекса теоретических и экспериментальных исследований опико-физических свойств полотна, трикотажных игл и нитей [26–30]. Постоянству конфигурации ТП препятствуют слабо формализуемые ограничения на объем подзадач, а также активность исполнительных элементов. Возможные способы разрешения этого противоречия обсуждаются в разделе 6.

4. Разработка алгоритма управления качеством ТП с СФ критериями

Выполнение условий, рассмотренных в разделе 3, позволяет конкретизировать выражения для координирующего сигнала $\gamma = \{\alpha, \beta\}$ (1.24):

$$\alpha = K(m(\alpha, \beta)); \quad (4.1)$$

$$\beta = \eta(m(\alpha, \beta)), \quad (4.2)$$

где η есть отображение

$$\eta: M \rightarrow B, \quad (4.3)$$

а функция K (1.3.61) представима в виде матрицы

$$K = [K_{ij}], \quad (4.4)$$

где элемент K_{ij} может принимать только значения $K_{ij} = 1$ или $K_{ij} = 0$. Тогда, согласно [14], модификации локальных функций качества можно задавать как

$$g_{i\beta}(m_i, u_i) = g_i(m_i, u_i) + \beta_i m_i, \quad (4.5)$$

где $\beta = \{\beta_i\}$, и записать уравнения координации (4.1)–(4.3) в виде

$$\alpha = K(m^\gamma), \quad (4.6)$$

$$\beta = \sum_{i=1}^n \{g'(m^\gamma)m_i - g'_{iM}(m_i^\gamma, K(m^\gamma))m_i\}, \quad (4.7)$$

где $g'(m^\gamma)m_i$ – производная Фреше от функции g по m_i в точке m^γ , а $g'_{iM}(m_i^\gamma, K(m^\gamma))m_i$ – производная Фреше от функции g_i по m_i в точке m^γ . Кажущаяся глобальная целевая функция системы (1.20) здесь записывается в виде

$$g_B(\beta, m, u) = \sum_{i=1}^n g_i(m_i, u_i) + \sum_{i=1}^n \beta_i u_i - K_i(m). \quad (4.8)$$

Это позволяет перейти к организации итеративной процедуры, в которой участвуют как локальные решающие системы, так и координатор. Согласно [14], стратегия координации, основанная на принципе прогнозирования взаимодействий с использованием отображения (4.3), в общем случае может быть представлена как отображение

$$s_0: \Gamma \times B \times U \rightarrow \Gamma, \quad (4.9)$$

такое, что условие

$$[s_0((\alpha, \beta), \beta', u) = (\alpha, \beta)] \Rightarrow [\beta = \beta', u = u^\alpha] \quad (4.10)$$

удовлетворяется на всей области определения этого отображения. Соответственно, преобразование, отображающее Γ само на себя, представляется в виде

$$T_\eta(\alpha, \beta) = s_0((\alpha, \beta), \eta(m(\alpha, \beta)), K(m(\alpha, \beta))) \quad (4.11)$$

на множестве Γ . В ходе итеративной процедуры координатор с помощью обратной связи получает информацию в виде пары $(u^\gamma, K(m^\gamma))$, где u^γ – желаемое (прогнозное) значение связующего сигнала, а $K(m^\gamma)$ – его текущее значение (рис. 1), и на основании предварительно сформированного координирующего сигнала γ и стратегии (4.9) улучшает координирование системы за счет подачи нового координирующего сигнала γ' .

Базируясь на экстремизации кажущейся глобальной целевой функции (1.20) и используя стратегию (4.8), можно применить теоремы 6.20, 6.21 [14], согласно которым сходящийся итерационный процесс может быть определен преобразованиями

$$\begin{aligned} T_{iA}(\gamma) &= \alpha_i^\gamma - \mu_i(\gamma) [\eta_i(m^\gamma) - \beta_i^\gamma] \\ T_{iB}(\gamma) &= \beta_i^\gamma - \lambda_i(\gamma) [K_i(m^\gamma) - \alpha_i^\gamma] \end{aligned} \quad (4.12)$$

где, как и прежде, $\gamma = (\alpha, \beta)$, $1 \leq i \leq n$, а $\lambda_i(\gamma)$ и $\mu_i(\gamma)$ – положительные числа. Для отдельного шага итерации уравнения (4.12) примут вид

$$\alpha' = \alpha + \mu [\beta - \eta(m(\alpha, \beta))], \quad (4.13)$$

$$\beta' = \beta + \lambda [\alpha - K(m(\alpha, \beta))]. \quad (4.14)$$

Результатом применения итерационной процедуры (4.13)–(4.14) должна, очевидно, явиться единственная пара α_0, β_0 , для которой

$$\beta_0 = \eta(m(\alpha_0, \beta_0)), \quad (4.15)$$

$$\alpha_0 = K(m(\alpha_0, \beta_0)). \quad (4.16)$$

При этом для каждого координирующего сигнала $\gamma = (\alpha, \beta)$ i -я локальная задача состоит в минимизации $g_{i\beta}(m_i, \alpha)$ на множестве M_i , в то время как глобальная задача системы в целом – минимизация g на множестве M .

Хотя выражения (4.12) описывают задачу оптимизации, их можно рассматривать как базовые для дальнейшего развития. В частности, при решении задачи улучшения они сохраняют силу, но точкой схождения является уже не точка экстремума, а точка, соответствующая значению глобальной целевой функции, в неявном виде задаваемая координатором. Этот вопрос, а также особенности организации итеративной процедуры в случае, когда локальная задача решается активным звеном, рассматриваются более подробно в разделе 6. В разделе 5 на примере ТП трикотажного производства представлены особенности применения итеративной процедуры (4.12) в условиях возмущений статистического характера.

5. Особенности реализации алгоритма управления для ТП трикотажного производства

Специфика ТП трикотажного производства состоит в том, что его СФ характер связан не только с активностью человека как носителя интегрального критерия качества, но и с переработкой естественного сырья, которое характеризуется «внутренней активностью» [26].

Для технологической реализации рассмотренного в разделе 4 алгоритма конкретизируем физический смысл входящих в него переменных. Переменная α (1.26) представляет собою вектор прогнозных значений связующих входов, т.е.

$$\alpha = u(\beta), \quad (5.1)$$

где $u = (u_1, \dots, u_n)$ – связующие сигналы, поступающие на входы соответствующих подпроцессов. Функция K (1.13) имеет смысл текущей (актуальной) функции взаимодействия подпроцессов, то есть вход u_i i -го подпроцесса связан с выходом y_j j -го подпроцесса уравнением

$$u_i = K_i(y_1, \dots, y_j, \dots, y_n), \quad (5.2)$$

причем непрерывность ТП выражается здесь в требовании полного использования продукта, полученного на предыдущих операциях, в последующих операциях для всех технологически значимых пар (i, j) . Выполнение этого требования достаточно легко контролируется, так как его технологическим показателем является отсутствие прерываний (принудительных остановок оборудования) на всех подпроцессах ТП.

Составляющие m_i вектора m отражают формируемые отдельными контрольными устройствами сигналы управления подпроцессами ТП. В результате проведенных исследований [27–31] удалось унифицировать эти сигналы для всех контрольных устройств и свести их к сигналу остановки ТП при неадекватном выполнении любого из контролируемых подпроцессов. Таким образом, значения m_i можно интерпретировать как текущее количество прерываний ТП за счет i -го контрольного устройства, отнесенное к единице продукции (например, к 1 м полотна). Очевидно, что бездефектный ход ТП соответствует отсутствию прерываний, т.е. условию

$$m = 0. \quad (5.3)$$

С другой стороны, бездефектный ход ТП можно описать и в терминах качества выпускаемой продукции. Так, в общем случае оценочный объект V для системы S (1. 2) имеет смысл слабо формализуемого множества «хорошее/плохое полотно». Однако введение ЛПР в состав системы позволяет трактовать этот же оценочный объект V как дихотомию «есть дефекты полотна / нет дефектов полотна», причем участие ЛПР сводится в данном случае к фиксации кондиционности вырабатываемого полотна в его понимании, что соответствует установлению границы раздела $V_{\text{доп}}$. Далее, опираясь на показанную монотонность признакового пространства относительно глобальной g (1.18) и ло-

кальных g_i (1.19) целевых функций, можно интерпретировать g_i и g как количество пропущенных дефектов после i -го контрольного устройства и на выходе ТП в целом, отнесенные к единице продукции. В частности, если вероятность обнаружения дефекта i -м контрольным устройством и системой контроля в целом равны $p_i = 1$ и $p = 1$, соответственно, то выпуск бездефектной продукции описывается условиями

$$g_i = 0 \text{ для всех } i \quad (5.4)$$

или

$$g = 0, \quad (5.5)$$

эквивалентность которых вытекает из условия (1.4), и при этом

$$V \geq V_{\text{доп}}. \quad (5.6)$$

Компоненты β_i вектора β описывают (в относительных единицах) уровень ограничения i -го контрольного устройства:

$$\beta_i = C_i / \sigma_i, \quad (5.7)$$

где C_i – абсолютное значение уровня ограничения i -го контрольного устройства, а σ_i^2 – дисперсия i -го подпроцесса на его входе. Другими словами, в соответствии с (4.3), вектор β задает локальные целевые функции системы мониторинга, но, в отличие от α , не имеет непосредственного аналога среди параметров, описывающих ТП. В этой связи целесообразно, используя соотношение (4.9), перейти от уравнений (4.3), (4.4) к уравнениям

$$(\sigma_i \alpha_i)' = (\sigma_i \alpha_i) + \mu [C_i - \sigma_i \eta(m(\alpha_i, \beta_i))], \quad (5.8)$$

$$C_i' = C_i + \sigma_i \lambda [\alpha - K_i(m(\alpha_i, \beta_i))], \quad (5.9)$$

где η описывает отображение $\eta: M \rightarrow C$, а K , как и раньше, представляет собою матрицу, содержащую только нулевые и единичные элементы, что позволяет оперировать технологически достижимыми величинами C_i . В этом случае точка схождения (4.15), (4.16) приобретает вид

$$C_0 = \sigma_i \eta[m(\sigma_i \alpha_0, C_0)]; \quad (5.10)$$

$$(\sigma_i \alpha_0) = K[m(\sigma_i \alpha_0, C_0)]. \quad (5.11)$$

Сформулированные условия бездефектного хода ТП (5.3)–(5.6) позволили провести дальнейшее упрощение уравнений итерации (5.8), (5.9), сохранив в них только технологически достижимые переменные. Разработанная итеративная процедура идентична для всех подсистем, входящих в состав системы мониторинга ТП, и проводится над сигналом $\xi(t)$, поступающим на вход порогового устройства подсистемы с уровнем ограничения C . Как показали исследования [26–30], этот сигнал для всех подсистем представляет собою стационарный случайный процесс, имеющий на конечном интервале времени T конечное число k_T максимумов с различными значениями амплитуды H_k , $0 < k < k_T$, причем в момент времени t , соответствующий $\xi(t) > C$, контрольное устройство формирует сигнал для сервопривода трикотажной машины. В качестве первого приближения в итеративной процедуре используется значение уровня ограничения

$$C_1 = 0, \quad (5.12)$$

т.е. точка схождения (5.10), (5.11) достигается снизу, а μ и λ принимаются равными $\mu = \lambda = 1$. Итерации выполняются в дискретном режиме, а в качестве интервала дискретизации используется интервал между двумя соседними максимумами сигнала H_k и H_{k+1} . При этом на протяжении всей процедуры обучения наблюдатель (ЛПР) визуально следит, что выпускаемое полотно является кондиционным; это означает, что выполняется условие (5.6).

Организацию итеративной процедуры можно пояснить следующим образом. Если при определенном значении уровня ограничения C_i (5.6) удовлетворяется, то команды прерывания ТП со стороны координирующего устройства должны отсутствовать, т.е. $m(\sigma_i \alpha_0, C_0) = 0$; это соответствует выполнению условия $\alpha_i = K_i(m(\alpha_i, \beta_i))$ для всех i , и (5.9) удовлетворяется в виде

$$(\sigma_i \alpha)' = (\sigma_i \alpha). \quad (5.13)$$

В противном случае, при $m(\sigma_i \alpha_i, C_i) \neq 0$, необходимо, в соответствии с (5.9), изменять значение C_i . Целесообразно проводить такое изменение после каждого последующего k -го импульса сигнала с амплитудой H_k в соответствии с алгоритмом:

$$C_{i+1} = A_k \text{ при } A_k > C_i; \quad (5.14)$$

$$C_{i+1} = C_i \text{ при } A_k \leq C_i. \quad (5.15)$$

Аппаратная реализация рассмотренной итеративной процедуры может быть осуществлена различными способами, например, на базе схемы пикового детектора.

Определенные сложности на пути технологической реализации процедуры обучения в системе управления создает статистический характер ТП. Действительно, как показывает (4.9), итеративный процесс заканчивается при достижении ситуации $(\alpha, \beta) = \text{const}$ независимо от номера итерации. Однако для ТП трикотажного производства это равенство может выполняться только статистически, т.е. в пределах определенной диссипативной области, которая связана, в первую очередь, с правильной организацией обучающей последовательности, в качестве которой для всех подсистем ТП выступает производимое трикотажное полотно. Анализ траекторных и терминальных потерь [5] в процессе обучения системы позволил принять в качестве оценки времени обучения

$$T_{об} = 1 \text{ мин}, \quad (5.16)$$

что по порядку величины соответствует прохождению вероятности ложной тревоги $\sim 2 \cdot 10^{-5}$ (в зависимости от типа полотна). Кроме того, резко понизить частоту ложных тревог позволяет выбор в качестве порога срабатывания не абсолютного максимума отрезка реализации сигнала за время обучения $T_{об}$, а величины

$$u_n = \sup\{\xi(t); 0 < t \leq T_{об}\} + \Delta p, \quad (5.17)$$

где

$$\frac{\Delta p}{\sup\{\xi(t); 0 < t \leq T_{об}\}} = 0,1. \quad (5.18)$$

Отметим, что обе оценки – (5.16) и (5.17) – являются оценками сверху, т.е. не требуют какой-либо перестройки в процессе работы оборудования.

Выбор и оценка показателей эффективности разработанного алгоритма управления качеством ТП трикотажного производства подробно рассмотрены в [5]. Экспериментально подтверждены преимущества контрольных устройств, входящих в состав разработанной системы управления качеством ТП, по сравнению с зарубежными аналогами по большинству технических показателей, прежде всего при переработке сырья естественного происхождения.

6. Особенности реализация алгоритма управления для ТП компьютерной допечатной подготовки полиграфической продукции

Структурная схема ТП компьютерной допечатной подготовки изданий представлена на рис. 2. Отличием схемы от базового варианта (рис. 1) является введение добавочного звена управления с критериальной функцией G_n , описывающей подмножество нормативных показателей оценочного объекта ТП, которые не выявляются программным или аппаратным путем. Кроме того, структурная схема рис. 2 использует обоснованное в [22] модельное представление функций человека в ТП компьютерной допечатной подготовки изданий. Функции оператора моделируются здесь блоком P_i , когнитивные функции – блоком C_s , а аффективные – блоком C_0 . Первые две функции всегда совмещаются в одном лице – s -м исполнителе, в то же время функции C_0 – функции обобщенного наблюдателя – могут принадлежать и заказчику, и исполнителю, и менеджеру проекта в зависимости от стадии его выполнения.

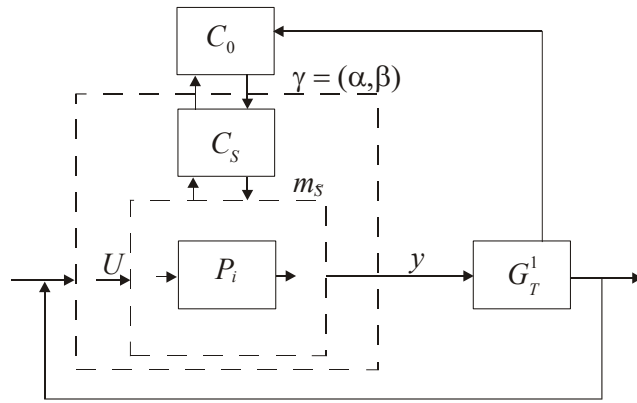


Рис. 2. Структурная схема ТП компьютерной допечатной подготовки изданий

Анализ ТП компьютерной допечатной подготовки изданий как концептуальной среды [5] позволяет считать формализмы (4.12)–(4.14) адекватными для описания ТП как двухуровневой иерархической системы. В частности, здесь остаются справедливыми уравнения (4/13)–(4.14), задающие итеративную организацию процесса координации. Но активность звеньев ТП обуславливает специфику перевода этого формализма в технологически достижимые переменные, которая проявляется, прежде всего, на уровне их смысловой интерпретации. Так, компоненты вектора $m(\alpha, \beta) = \{m_s(\alpha, \beta)\} = \{m_s(\alpha)m_s(\beta)\}$, т.е. отдельные локальные управляющие сигналы, в данном случае целесообразно трактовать как уровень качества ($m_s(\beta)$) и объем ($m_s(\alpha)$) подзадач, выбираемых для исполнения s -м участником ТП.

Смысловое содержание компонентов α и β , координирующего сигнала $\gamma = (\alpha, \beta)$ обусловлено структурой оценочного объекта ТП. В соответствии с моделью [22] в составе β можно выделить компоненты β_s и β_n . Локальная функция качества G_s , соответствующая β_s , рассматривается как неформализуемая, и ее уровень устанавливается в сознании исполнителя в результате процедуры обучения с координатором. Локальная функция качества G_n , задаваемая β_n , отражает подмножество V_n нормативных показателей оценочного объекта ТП, которые не могут быть выявлены программным или аппаратным путем и подлежат выявлению при корректуре. Легко видеть, что

$$V_n \subset V_T^1 \times V_T^2, \quad (6.1)$$

где V_T^1 и V_T^2 определены выражением (2.5). G_n контролируется в отдельном блоке G_n структурной схемы (рис. 2). Локальная функция качества $G_{па}$, связанная с подмножеством $V_{па}$,

$$V_{па} \subset V_T^1 \times V_T^2, \quad (5.2)$$

отражает программно-аппаратные требования к ТП (например, перечень допустимых в конкретном ТП аппаратуры, программных продуктов, их опций и т.д.). Ей соответствует компонент $\alpha = \{\alpha_s\}$ координирующего сигнала $\gamma = (\alpha, \beta)$ ТП. В общем случае α_s имеет смысл объема подзадач, задаваемого координатором каждому s -му исполнителю, а $K = \{K_s\}$ – объема подзадач, фактически им выполненных. Для докомпьютерного издательско-полиграфического ТП характерно задание в явном виде векторов α и K , а также смыслового содержания множества P_s и таблицы соответствий $\{s \rightarrow P_s\}$, т.е. распределения работ по исполнителям, с последующей проверкой равенства $\{\alpha_s\} = \{K_s\}$. Однако, как показано в [5], в ТП компьютерной подготовки изданий объем подзадачи, выполняемой отдельным исполнителем, следует рассматривать как слабо формализуемый.

Таким образом, параметры $\{\alpha_i, \beta_i\}$ координирующего вектора, отражающие объем и качество работы, выполняемой отдельным исполнителем, одновременно оказываются слабо формализуемыми. Такая ситуация позволяет определить лишь область возможных решений, что может привести к разрыву непрерывности ТП в целом. Однако содержа-

тельный анализ причинно-следственных связей, характерных для ТП компьютерной подготовки изданий, позволяет свернуть многомерное множество признаков $V_{па}$ в одномерное, задаваемое в номинальной шкале и характеризуемое следующим интегральным критерием. Для любой пары подзадач $\{i, j\}$, отвечающей условию $K_{ij} = 1$ (1.4.40), выходы которой связаны уравнением

$$y_j = F(y_i), \quad (6.3)$$

должна быть технологически разрешима обратная задача,

$$y_i = F^{-1}(y_j). \quad (6.4)$$

Выполнение условий (6.3), (6.4) означает обратимость ТП на любой его стадии и поддерживается современными программными продуктами для настольно-издательских систем. Таким образом, локальная функция качества $G_{па i}$, задаваемая α_i , формализуется условиями (6.3), (6.4) и легко контролируется инструментально. Важно отметить, что этот контроль может осуществляться непосредственно самим исполнителем и, следовательно, использоваться как средство само- или взаимообучения. В то же время в рамках локальной функции качества G_s , задаваемой параметром β_s , задача исполнителя остается слабо формализованной и рассматривается им как задача улучшения.

Поясним технологическую реализацию итеративной процедуры (4.13)–(4.14) применительно к ТП компьютерной допечатной подготовки издания. В качестве первого приближения здесь используется продуцируемое координатором C_0 значение α , под воздействием которого каждый локальный управляющий элемент C_s задает (т.е. s -тый участник устанавливает для себя) значение $m_s = 1$. В ходе и по результатам выполнения этой подзадачи в паре C_0 – C_s в режиме обучения формируются уровни β_s , β_n и, отчасти, $\beta_{па}$, которые будут воспроизводиться локальным управляющим элементом (то есть s -м исполнителем) в других подзадачах.

По завершении процесса обучения, т.е. достижении приемлемого для C_0 уровня β_s , координатор инициирует процедуру итеративного структурирования ТП. Для пояснения особенностей методики введем в рассмотрение следующие величины: P – множество возможных подзадач в ТП, рассматриваемом как отдельный проект; P_s – множество подзадач, выполняемых s -м участником проекта, где $1 \leq s \leq S$, а S – общее число участников проекта; P_s^{start} – исходное множество подзадач, задаваемое (предлагаемое) s -му участнику координатором проекта; P_s^{fin} – множество подзадач, фактически выполненных s -м участником на момент окончания проекта.

В традиционной методике множества P и P_s являются четкими, при этом

$$\sum_s P_s = P, \quad (6.5)$$

$$\prod_s P_s = 0, \quad (6.6)$$

а смысловое содержание P_s и таблица соответствий $\{s \rightarrow P_s\}$ однозначно задаются руководителем проекта. В предлагаемой методике множества P и P_s рассматриваются как нечеткие, что позволяет каждому участнику формировать свои подзадачи в соответствии со своим пониманием и текущей мотивацией, т.е. достаточно адекватно отражает сочетание в нем функций оператора и ЛПР [22]. Координатор задает (предлагает) только исходное разбиение P_s^{start} , причем

$$\sum_s P_s^{start} \subset P_s. \quad (6.7)$$

В ходе выполнения проекта s -ый участник может (но не обязан) самостоятельно изменить множество P_s в диапазоне

$$0 \subset P_s \subset P, \quad (6.8)$$

т.е. происходит итеративное структурирование множества P . Условие окончания итераций

$$\sum_s P_s^{fin} \supseteq P \quad (6.9)$$

контролируется координатором по отсутствию прерываний в ТП. Ситуация

$$\prod_s P_s^{fin} \neq 0, \quad (6.10)$$

отражающая возможное дублирование подзадач отдельными участниками проекта, создает для них конкурентную среду (по типу аукциона), являясь дополнительным мотиватором. Глобальная задача ТП в целом является задачей улучшения с соответствующими ограничениями (на суммарное время выполнения проекта,

$$T_{\Sigma} \leq T_{\text{доп}}, \quad (6.11)$$

на суммарную стоимость его выполнения и др.). Локальная задача, выполняемая отдельным исполнителем, также является задачей улучшения и решается в ограничениях, которые задаются выбранной стратегией аукциона и формируют структуру мотивации исполнителей.

Результаты выполнения отдельных подзадач в электронном виде заносятся в единую директорию, управляемую координатором (менеджером проекта). Это позволяет отдельным участникам (C_s) вносить текущие изменения в задание m_s , т.е. реализует коммуникацию между локальными управляющими системами через координатора. В то же время контроль таблицы соответствий $\{s \rightarrow P_s\}$ не производится. Случаи (6.10) разрешаются согласно установленной для ТП стратегии аукциона, например, по типу бесприоритетной дисциплины обслуживания FIFO (first in – first out).

Точка схождения итеративного процесса (4.13)–(4.14) диагностируется координатором по отсутствию сигналов прерывания ТП. Готовый продукт (оригинал-макет) поступает в блок G_T^{-1} (рис. 2), где производится его контроль (корректурa) в соответствии с критериальной функцией

$$G_{\square}^1 = \prod_N (G_{\square}^1)_n, \quad 1 \leq n \leq N, \quad (6.12)$$

где $(G_T^{-1})_n$ – критериальные функции, описывающие отдельные требования нормативных документов. Внесение изменений в продукт по результатам этого контроля рассматривается как подзадача (подзадачи) ТП и выполняется в соответствии с уже описанными процедурами (4.12)–(4.14), (6.7)–(6.9). При необходимости цикл может повторяться несколько раз, причем вместо G_T^{-1} могут использоваться другие критериальные функции, персонифицируемые различными участниками процесса (в том числе координатором и/или заказчиком).

Выбор и оценка показателей эффективности разработанного алгоритма управления качеством ТП компьютерной доредакционной подготовки изданий рассмотрены в [5]. Обобщенно их можно представить следующим образом:

- ТП во многом является самоорганизующимся. Координатор (менеджер проекта) не занимается подбором исполнителей, распределением подзадач между ними и контролем их выполнения, регулированием взаимных претензий;
- номенклатура параметров, не охватываемых самоорганизацией, во многом определяется принятой в ТП стратегией конкуренции. Эти параметры – скорость схождения процесса и воспроизводимость качества выполнения подзадач – достаточно легко регулируются координатором;
- экспериментально подтвержденная положительная динамика основных показателей ТП позволяет говорить о робастности разработанной системы управления к вариациям внутренних и внешних информационных потоков;
- итеративная организация ТП с участием заказчика позволяет получить продукт, соответствующий не только исходным требованиям заказчика, но и динамике их изменений в процессе работы;

- большинство контрольных операций в ТП осуществляются обезличенно, что во многом снимает проблемы взаимных претензий участников процесса и обеспечивают самоадаптацию процесса к текущим изменениям их мотивации;
- обеспечивается более быстрое, по сравнению с традиционными методиками, профессиональное обучение и самообучение.

Заключение

В работе предложен алгоритм управления качеством ТП с СФ критериями, построенный на базе теории координации. Полученные теоретические результаты конкретизированы для двух ТП трикотажного производства и компьютерной допечатной подготовки полиграфической продукции, что позволило продемонстрировать специфику применения алгоритма в зависимости от локализации и характера проявления активности элементов ТП.

Литература

1. Методы классической и современной теории автоматического управления. Учебник в 3-х т. Т.3: Методы современной теории автоматического управления / Ред. Н.Д. Егунов. М.: Изд-во МГТУ им. Н.Э. Баумана, 2000. 748 с.
2. Ларичев О.И., Мошкович Е.М. Качественные методы принятия решений. Вербальный анализ решений. М.: Наука. Физматлит, 1996. 208 с.
3. Ковалев С.М., Долгий А.И. Модель представления и обработки нечетко-временной информации о последовательных событиях в слабо формализованных динамических процессах // Перспективные информационные технологии и интеллектуальные системы. 2004. №3. С. 77-81.
4. Соколов А.Ю. Методология алгебраического проектирования интеллектуальных систем управления// Інформаційно-керуючі системи на залізничному транспорті. 2000. №6. С. 28-31.
5. Гусарова Н.Ф. Координация в технологических процессах со слабо формализуемыми критериями. / Монография. СПб: СПбГИТМО(ТУ), 2001. 271 с.
6. Боулдинг К. Общая теория систем – скелет науки // Исследования по общей теории систем. М.: Прогресс, 1969. С. 106–124.
7. Ларичев О.И. Теория и методы принятия решений, а также Хроника событий в Волшебных странах. / Учебник. Изд. 2, пер. и доп. М.: Логос, 2003. 392 с.
8. Покровская И.В., Дорофеев А.А., Чернявский А.Л. Методы коллективной многовариантной экспертизы и их практическое использование. Труды международной конференции «Современные сложные системы управления». Старый Оскол: «Тонкие наукоемкие технологии», 2002. С. 122 – 127.
9. Губко М.В., Новиков Д.А. Теория игр в управлении организационными системами. М.: Синтег, 2002. 148 с.
10. Борисов В.В., Бычков И., Дементьев А. Компьютерная поддержка сложных организационно-технических систем. М.: Горячая Линия–Телеком, 2002. 160 стр.
11. Антамошин А.Н., Близнава О.В., Бобов А.В., Большаков А.А., Лобанов В.В., Кузнецова И.Н. Интеллектуальные системы управления организационно-техническими системами / Под ред. проф. А.А. Большакова. М.: Горячая линия–Телеком, 2006. 160 с.: ил.
12. Перегудов Ф.И., Тарасенко Ф.И. Введение в системный анализ / Уч. пособие для вузов. М.: Высшая школа, 1989. 367 с.
13. Хакен Г. Синергетика. Иерархия неустойчивостей в самоорганизующихся системах и устройствах. М.: Мир, 1985.

14. Месарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем. М.: Мир, 1973. 344 с.
15. Курейчик В.В. Эволюционные, синергетические и гомеостатические методы принятия решений. Таганрог: Изд-во ТРТУ, 2001. 221с.
16. Курейчик В.В. О синергетическом подходе к проблеме принятия решений // Перспективные информационные технологии и интеллектуальные системы №2 / 2001 С. 65–69.
17. Недзельский И.И. Морские навигационные тренажеры: проблемы выбора. СПб: ГНЦ РФ-ЦНИИ "Электроприбор", 2002. 220 с.
18. Цитович И.Г. Технологическое обеспечение качества и эффективности процессов вязания поперечновязанного трикотажа. М.: Легпромбытиздат, 1992. 240 с.
19. Мильчин А.Э. Издательский словарь-справочник. М.: Юристъ, 1998. 472 с.
20. Карлин С. Математические методы в теории игр, программировании и экономике. М.: Мир, 1964. 838 с.
21. Юдин Д.Б. Вычислительные методы теории принятия решений. М.: Наука. Гл. ред. физ.-мат. лит., 1989. 320 с. (Теория и методы системного анализа).
22. Гусарова Н.Ф., Маятин А.В. Моделирование человека при проектировании информационных систем // Научно-технический вестник СПбГУ ИТМО. 2005. № 22.
23. Справочник по теории автоматического управления / Под ред. А.А. Красовского. М.: Наука. Гл. ред. физ.-мат. лит.-ры, 1987. 712 с.
24. Гусарова Н.Ф., Тихомиров А.С., Беляев А.И. Прибор контроля трикотажного полотна // Научно-технические разработки Санкт-Петербургского государственного института точной механики и оптики / Под ред. В.Н. Васильева. СПб: СПбГИТМО(ТУ), 2001. С. 142–143.
25. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики. Вып. 33. М.: Наука, 1978. С. 5–68.
26. Гусарова Н.Ф., Маятин А.В. Человеко-машинные методы обработки изображений // Научно-технический вестник СПбГУ ИТМО. 2005. № 22.
27. Гусарова Н.Ф., Горелик С.Я., Тихомиров А.С. Оптимизация структуры датчика первичного сигнала при оптической дефектоскопии трикотажного полотна // Оптический журнал. 1995. № 1. С. 35–39.
28. Гусарова Н.Ф. Исследование структуры трикотажного полотна как объекта оптического анализа // Оптический журнал. 2001. Т. 68. № 1. С.82–83.
29. Гусарова Н.Ф. Оптический контроль пряжи при ее переработке // Оптический журнал. 2001. Т. 68. № 8. С.88–92.
30. Гусарова Н.Ф. Оптико-электронная дефектоскопия трикотажного полотна в процессе его изготовления. // Дефектоскопия. 2001. № 10. С. 62–71.
31. Гусарова Н.Ф. Дефектоскопия рабочих органов трикотажной машины в процессе изготовления полотна. // Дефектоскопия. 2003. № 2. С. 61–70.

ПОСТРОЕНИЕ ВИЗУАЛИЗАТОРОВ АЛГОРИТМОВ ДИСКРЕТНОЙ МАТЕМАТИКИ

Г.А. Корнеев, А.А. Шалыто

В статье описана технология построения визуализаторов алгоритмов. Технология иллюстрируется на примере построения визуализатора алгоритма поиска максимума в массиве натуральных чисел. Описываемая технология также применима для построения визуализаторов сложных алгоритмов дискретной математики.

Введение

При изучении дискретной математики [1, 2], важную роль играют визуализаторы алгоритмов, позволяющие в наглядной форме динамически отображать детали их работы. Это открывает возможность использования нового подхода к обучению дискретной математике и программированию [3, 4].

Визуализатор – это программа, в процессе работы которой на экране компьютера динамически демонстрируется применение алгоритма к выбранному набору данных. Визуализаторы позволяют изучать работу алгоритмов как в автоматическом, так и в пошаговом режиме, аналогичном режиму трассировки программ. При этом трассировка может осуществляться малыми и большими шагами как в прямом, так и в обратном направлении.

Для построения визуализаторов авторами была предложена технология *Vizi* [5] и программный пакет с тем же названием, позволяющие автоматизировать построение визуализатора за счет генерации логики визуализатора на основе конечных автоматов [6]. Для иллюстрации применения этой технологии достаточно построить визуализатор простого алгоритма, что и сделано в настоящей работе. Однако рассматриваемая технология применима и для построения визуализаторов сложных алгоритмов, примеры которых приведены на сайте <http://is.ifmo.ru> в разделе «Визуализаторы».

Технология построения визуализатора

Опишем порядок разработки визуализатора с применением технологии *Vizi*.

1. Постановка задачи и анализ литературы.
2. Создание визуализируемой программы:
 - реализация алгоритма,
 - отладка построенной программы.
3. Разработка концепции визуализации:
 - выделение «интересных» шагов алгоритма,
 - разработка концепции визуального представления,
 - разработка набора комментариев,
 - разработка элементов управления.
4. Построение XML-описания визуализируемой программы:
 - выделение модели данных,
 - преобразование программы,
 - запись XML-описания отдельных процедур,
 - запись XML-описания визуализируемой программы,
 - отладка XML-описания визуализируемой программы,

- интеграция набора комментариев в XML-описание.
- 5. Реализация визуального представления.
- 6. Реализация элементов управления.
- 7. Интеграция и отладка визуализатора:
 - интеграция визуального представления в XML-описание,
 - генерация кода по XML-описанию визуализатора,
 - отладка визуализатора.

На первом этапе производится постановка задачи и анализ литературы. При этом рассматриваются существующие модификации алгоритма и одна из них выбирается для визуализации.

На следующем этапе создается и отлаживается программа, реализующая выбранную модификацию алгоритма.

На третьем этапе разрабатывается концепция визуализации. При этом сначала выделяются те шаги алгоритма, которые представляют наибольший интерес («интересные» шаги). Затем разрабатывается общая концепция визуального представления, в рамках которой разрабатываются слайды для каждого «интересного» шага. Одновременно с разработкой визуального представления для каждого «интересного» шага пишутся комментарии, поясняющие действия, выполняемые алгоритмом. После этого разрабатываются элементы управления визуализатором, в частности, определяется, какие параметры и в каких пределах сможет регулировать пользователь.

На четвертом этапе производится построение XML-описания визуализируемой программы, созданной на втором этапе. Первоначально из программы выделяется модель данных для того, чтобы визуализатор имел доступ к переменным, используемым при визуализации. Так как в работе [5] было принято, что в XML-описании могут использоваться только операторы присваивания, ветвления, цикла с предусловием и блочные операторы, то предварительно программу необходимо преобразовать к такому виду. После этого записываются XML-описания отдельных процедур, которые затем объединяются в XML-описание визуализируемой программы. Затем созданное XML-описание отлаживается при помощи средств, предоставляемых пакетом *Vizi*. Этап завершается добавлением комментариев, разработанных на третьем этапе к XML-описанию визуализируемой программы.

На пятом и шестом этапах производится реализация визуального представления и элементов управления, в соответствии с концепцией, разработанной на третьем этапе.

На заключительном этапе интегрируются результаты четвертого и пятого этапов, и производится отладка визуализатора.

Проиллюстрируем применение этой технологии на простом примере – построении визуализатора алгоритма поиска максимума в массиве натуральных чисел.

Постановка задачи и анализ литературы

Задан массив из N натуральных чисел. Требуется найти в нем максимальное число. Алгоритм решения этой задачи очень прост, и поэтому сразу перейдем к следующему этапу.

Создание визуализируемой программы

Задача поиска максимума в массиве решается следующей программой:

```
void main() {
    int max = 0;
    for (int i = 0; i < a.length; i++) {
        if (max < a[i]) {
            max = a[i];
        }
    }
}
```

```

    }
  }
}

```

Здесь a – массив, в котором производится поиск максимума, max – значение текущего максимума (после i -ой итерации – среди первых i элементов).

Отметим, что инициализация максимума нулем не приводит к ошибке, так как по условию задачи в массиве содержатся только натуральные числа.

Разработка концепции визуализации

Выделение «интересных» шагов алгоритма. В визуализаторе поиска максимума наиболее интересным являются шаги, осуществляющие ветвление и связанное с ним обновления текущего максимума. Кроме того, должны быть выделены начальное и заключительное состояния. Специально визуализировать переход к следующему элементу массива вряд ли имеет смысл.

В табл. 1 приведен список «интересных» шагов алгоритма.

Таблица 1. «Интересные» шаги алгоритма

Шаг	Пояснение
<i>Начальное состояние</i>	Описывается цель алгоритма
<i>Инициализация максимума</i>	Инициализация текущего максимума нулем и соответствующие пояснения
<i>Проверка на обновление максимума</i>	Проверяется необходимость обновить текущий максимум
<i>Обновление текущего максимума</i>	Присваивание текущему максимуму значения текущего элемента
<i>Заключительное состояние</i>	Отображение результатов вычисления

Разработка концепции визуального представления. Основными данными в визуализаторе являются элементы массива и значение текущего максимума. Их значения постоянно должны быть представлены на экране. Поэтому для визуального представления будем использовать схему визуального представления, изображенную на рис. 1.

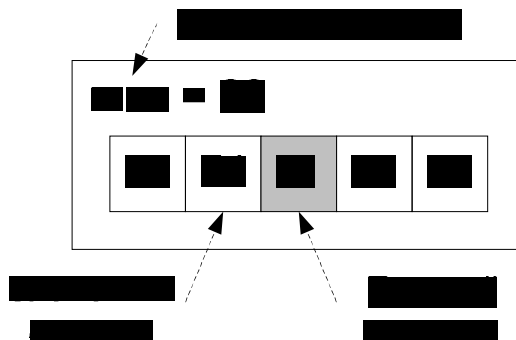


Рис. 1. Схема визуального представления

Отметим, что индекс текущего элемента в массиве явно не отображается. Вместо этого текущий элемент выделяется цветом.

Для визуализации выделенных «интересных» шагов требуются слайды, представленные в табл. 2.

Разработка набора комментариев. Для каждого из «интересных» состояний разрабатываются комментарии. Отметим, что, так как шаг «Проверка на обновление максимума» соответствует оператору ветвления, то для него должны быть разработаны два комментария: для истинного и ложного условий.

Таблица 2. Слайды для “интересных” шагов

Шаг	Пояснение	Слайд
Начальное состояние	Текущий элемент и максимум не подсвечиваются	
Инициализация максимума	Текущий элемент не подсвечивается	
Проверка на обновление максимума	Текущий элемент подсвечивается зеленым цветом	
Обновление текущего максимума	Текущий элемент подсвечивается красным цветом	
Заключительное состояние	Текущий элемент не подсвечивается	

Комментарии могут содержать параметры, места включения которых задаются следующим образом:

{номер параметра}

Значения параметров вычисляются и подставляются в процессе работы визуализатора. Приведем набор комментариев для визуализатора поиска максимума (табл. 3). Выражения, соответствующие параметрам визуализатора, указаны в столбце «Параметры».

Таблица 3. Набор комментариев

Шаг алгоритма	Комментарий	Параметры
Начальное состояние	На экране изображен массив, в котором будет осуществляться поиск максимума	
Инициализация максимума	Инициализируем максимум нулем (так как в массиве только натуральные числа)	
Проверка на обновление максимума (условие истинно)	{0} больше текущего максимума ({1})	$a[i]$, \max
Проверка на обновление максимума (условие ложно)	{0} не больше текущего максимума ({1})	$a[i]$, \max
Обновление текущего максимума	Обновляем текущий максимум	
Заключительное состояние	Максимум найден ({0})	\max

Разработка элементов управления. Список элементов управления визуализатора приведен в табл. 4. Отметим, что только два из них не являются стандартными, а все остальные входят в библиотеку *Vizi*.

Таблица 4. Элементы управления

Элемент управления	Назначение	Стандартный
	Шаг назад и вперед	Да
	Начало визуализации сначала	Да
	Вход в автоматический режим и выход из него	Да
	Регулирование задержки между шагами в автоматическом режиме. Задержка измеряется в миллисекундах	Да
	Вывод информации о визуализаторе	Да
	Генерация случайного набора данных	Нет
	Загрузка/сохранения состояния визуализатора	Да
	Задание количество элементов в массиве	Нет

Элементы управления визуализатором komponуются в области элементов управления, как показано на рис. 2.

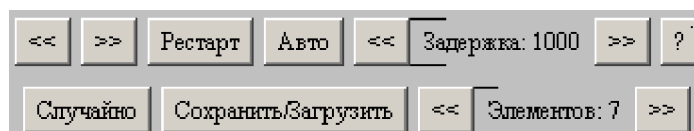


Рис. 2. Область элементов управления

Построение XML-описания визуализируемой программы

Рассмотрим этапы построения XML-описания.

Выделение модели данных. В визуализируемой программе используются три переменные:

- `max` – значение текущего максимума;
- `a` – массив, в котором осуществляется поиск;
- `i` – индекс текущего элемента массива `a`.

Эти переменные должны быть вынесены в модель данных для того, чтобы, как отмечено выше, визуализатор имел доступ к ним. При этом значения переменных `max` и `a` будут визуализироваться, а значения переменной `i` – нет. Поэтому сделаем переменные `max` и `a` глобальными, а переменную `i` – локальной для процедуры поиска максимума. XML-описание переменных имеет следующий вид:

```
<variable
  description = "Массив для поиска"
  name       = "a"
  type      = "int[]"
  value     = "new int[]{1, 2, 3, 1, 3, 5, 6}"
/>
<variable
  description = "Текущий максимум"
  name       = "max"
  type      = "int"
  value     = "0"
/>
<variable
  description = "Переменная цикла"
  name       = "i"
  type      = "int"
/>
```

Отметим, что для глобальных переменных указаны их начальные значения, которые могут быть использованы при отладке.

В модель данных также вынесем переменную, содержащую ссылку на экземпляр визуализатора. Она будет применяться для формирования визуального представления текущего шага. Опишем ее следующим образом:

```
<variable
  description = "Экземпляр апплета"
```

```

    name      = "visualizer"
    type      = "FindMaximumVisualizer"
    value     = "null"
  />

```

Для ссылок на переменные, вынесенные в модель данных, применяется нотация вида

@ИМЯ переменной

Таким образом, после выделения модели данных программа имеет следующий вид:

```

void main() {
    @max = 0;
    for (@i = 0; @i < @a.length; @i++) {
        if (@max < @a[@i]) {
            @max = @a[@i];
        }
    }
}

```

Преобразование программы. Сначала преобразуем цикл со счетчиком в цикл с предусловием. После преобразования программа выглядит следующим образом:

```

void main() {
    @max = 0;
    @i = 0;
    while (@i < @a.length)) {
        if (@max < @a[@i]) {
            @max = @a[@i];
        }
        @i++;
    }
}

```

Теперь преобразуем выражение @i++ к виду простого присваивания. В соответствии с семантикой языка *Java* [7] это приводит к оператору

```
@i = @i + 1
```

Для автоматического построения кода, осуществляющего обратную трассировку алгоритма, требуется заменить операторы присваивания на операторы обратимого присваивания [5], имеющие вид @=. В результате визуализируемая программа будет записана следующим образом:

```

void main() {
    @max @= 0;
    @i @= 0;
    while (@i < @a.length)) {
        if (@max < @a[@i]) {
            @max @= @a[@i];
        }
    }
}

```

```

    }
    @i @= @i + 1;
  }
}

```

Запись XML-описания отдельных процедур. В XML-описании процедуре соответствует элемент `auto`, который содержит описание локальных переменных (элемент `variable`) и шаги алгоритма (элементы `step` – для присваивания, `if` – для ветвления и `while` – для цикла с предусловием) [5].

Отметим, что для каждого шага алгоритма должен быть указан его идентификатор (атрибут `id`), используемый в дальнейшем при генерации кода, и словесное описание (атрибут `description`), применяемое при отладке [5].

Визуализируемая программа содержит одну процедуру `main`, XML-описание которой будет иметь вид:

```

<auto id="Main" description="Ищет максимум в массиве">
  <variable description="Переменная цикла" name="i" type="int"/>
  <step id="Initialization" description="Инициализация">
    <action>@max @= 0;</action>
  </step>
  <step id="LoopInit" description="Начало цикла">
    <action>@i @= 0;</action>
  </step>
  <while id="Loop" description="Цикл" test="@i &lt; @a.length">
    <if id="Cond" description="Условие" test="@max &lt; @a[@i]">
      <then>
        <step id="newMax" description="Обновление максимума">
          <action>@max @= @a[@i];</action>
        </step>
      </then>
    </if>
    <step id="inc" description="Инкремент">
      <action>@i @= @i + 1;</action>
    </step>
  </while>
</auto>

```

Отметим, что `i` как локальная переменная процедуры `main` описана в теле этой процедуры.

Запись XML-описания визуализируемой программы. Для построения полного описания визуализируемой программы добавим к описанию процедуры `main` описания переменных модели данных и метода `toString`, который применяется при отладке.

```

<algorithm>
  <... описание переменных ...>

  <toString>

```

```

        StringBuffer s = new StringBuffer();
        s.append("max = ").append(@max).append("\n");
        s.append("i = ").append(@Main@i).append("\n");
        return s.toString();
    </toString>

    <... XML-описание процедуры main ...>
</algorithm>

```

Для получения XML-описания визуализатора в целом требуется включить описание программы в описание визуализатора:

```

<!DOCTYPE visualizer PUBLIC
    "-//IFMO Vizi//Visualizer description"
    "http://ips.ifmo.ru/vizi/dtd/visualizer.dtd"
>
<visualizer
    id="FindMaximum"
    package="ru.ifmo.vizi.find_max"
    main-class="FindMaximumVisualizer"

    name-ru="Поиск максимума"
    author-ru="Георгий Корнеев"
    author-email="kgeorgiy@rain.ifmo.ru"

    supervisor-ru="Георгий Корнеев"
    supervisor-email="kgeorgiy@rain.ifmo.ru"

    copyright-ru="Кафедра КТ, СПбГУ ИТМО, 2005"

    preferred-width="400"
    preferred-height="250"
>
    <... XML-описание алгоритма ...>
    <configuration>
    </configuration>
</visualizer>

```

В атрибутах элемента `visualizer` описания визуализатора указываются пакет и имена классов визуализатора, а также информация об авторе [5].

Отладка XML-описания визуализируемой программы. Отладка XML-описания визуализатора производится при помощи средств, входящих в пакет *Vizi*.

Для этого вызовом *build*-скрипта с параметром `debug-source` [5] генерируется файл с реализацией алгоритма, заданного XML-описанием:

```
/**
```

```

    * Ищет максимум в массиве.
    */
private final void Main() {
    // Инициализация
    d.max = 0;
    // Начало цикла
    d.Main_i = 0;
    // Цикл
    while (d.Main_i < d.a.length) {
        // Условие
        if (d.max < d.a[d.Main_i]) {
            // Обновление максимума
            d.max = d.a[d.Main_i];
        }
        // Инкремент
        d.Main_i = d.Main_i + 1;
    }
}

```

Данный файл позволяет убедиться в том, что XML-описание корректно. Отметим, что сгенерированный код содержит комментарии с информацией, перенесенной из атрибутов `description` XML-описания.

Проверим также корректность автоматического обращения визуализируемой программы. Для этого вызовом `build`-скрипта с параметром `debug-check` создадим тест и запустим его командой `CheckFindMaximum` из каталога `deploy` [5]. Удостоверимся, что тест прошел успешно:

```

Check 1 steps... OK ()
Check 2 steps... OK ()
...
Check 36 steps... OK ()
Check 37 steps... OK ()
OK

```

Интеграция набора комментариев в XML-описание. Добавим разработанные комментарии к XML-описанию.

Для отображения комментариев в начальном и конечном состояниях к описанию процедуры `main` требуется добавить шаги `start` и `finish`:

```

<auto id="Main" description="Ищет максимум в массиве">
  <variable description="Переменная цикла" name="i" type="int"/>
  <start comment-ru="На экране изображен массив, в котором будет
    осуществляться поиск максимума"/>
  <step id="Initialization" description="Инициализация"
    comment-ru="Инициализируем максимум нулем (так как в
    массиве только натуральные числа).">
    <action>@max @= 0;</action>
  </step>
  <step id="LoopInit" description="..." level="-1">...</step>
  <while id="Loop" description="Цикл" test="@i < @a.length"
    level="-1">
    <if id="Cond" description="Условие" test="@max < @a[@i]"
      true-comment-ru="{0} больше текущего максимума ({1})"

```

```

        false-comment-ru="{0} не больше текущего максимума ({1})"
        comment-args="new Integer(@a[@i]), new Integer(@max)"
    >
    <then>
        <step id="newMax" description="Обновление максимума"
            comment-ru="Обновляем текущий максимум">
            <action>@max @= @a[@i];</action>
        </step>
    </then>
</if>
<step id="inc" description="..." level="-1">...</step>
    <action>@i @= @i + 1;</action>
</step>
</while>
<finish comment-ru="Максимум найден ({0})"
    comment-args="new Integer(@max)"/>
</auto>

```

Здесь и далее полужирным шрифтом выделены внесенные дополнения.

Отметим, что шагам присвоены уровни (атрибут `level`), которые определяют, какие шаги будут отображаться пользователю в различных режимах работы:

- -1 – не отображается пользователю
- 0 – отображается при выполнении обычных шагов и пропускается при больших шагах (значение по умолчанию).
- 1 – отображается при выполнении как обычных, так и больших шагов.

Реализация визуального представления

Разработанное визуальное представление весьма просто (табл. 2), и его отображение может производиться вызовом одного метода

```
updateArray(<индекс элемента>, <номер цвета>)
```

При вызове этого метода первым параметром передается индекс выделенного элемента, а вторым – способ его отображения: 0 – без выделения, 1 – зеленый, 2 – красный.

Реализация элементов управления

Визуализатор содержит два нестандартных элемента управления: кнопку генерации случайного набора данных и панель задания количество элементов. Первый из них может быть реализован на основе кнопки с подсказкой (`HintedButton`), а второй – на основе панели выбора (`AdjustablePanel`), входящих в пакет *Vizi*.

Отметим, что в результате воздействия на эти элементы управления изменяются входные данные. Поэтому при их использовании визуализатор должен переходить в начальное состояние.

Интеграция и отладка визуализатора

Интеграция визуального представления в XML-описание. Добавим вызовы, осуществляющие обновление визуального представления, к шагам, имеющим неотрицательный уровень, путем указания их в элементах `draw` [5]:

```

<auto id="Main" description="Ищет максимум в массиве">
    <variable description="Переменная цикла" name="i" type="int"/>
    <start ...> <draw>@visualizer.updateArray(0, 0);</draw></start>
    <step ...>
        <draw>@visualizer.updateArray(0, 0);</draw>

```

```

    <action>@max @= 0;</action>
  </step>
</step ...>...</step>
<while ...>
  <if ...>
    <draw>@visualizer.updateArray(@i, 1);</draw>
    <then>
      <step ...>
        <draw>@visualizer.updateArray(@i, 2);</draw>
        <action>@max @= @a[@i];</action>
      </step>
    </then>
  </if>
  <step ...>...</step>
</while>
<finish ...><draw>@visualizer.updateArray(0, 0);</draw></finish>
</auto>

```

Генерация кода по XML-описанию визуализатора. Код визуализатора генерируется вызовом *build*-скрипта с параметром *all* [5].

max = 59

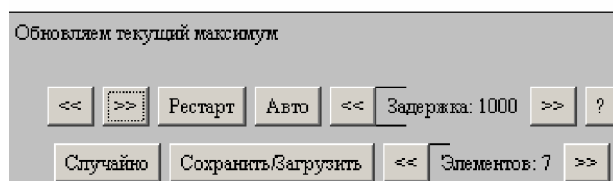


Рис. 3. Визуализатор в одном из состояний

В результате этого по XML-описанию будет сгенерирован код, реализующий, в том числе, и логику визуализатора [8]. Для данного визуализатора будут автоматически построены два взаимодействующих автомата с девятью состояниями в каждом.

На рис. 3 приведен скриншот построенного визуализатора, запущенного на массиве значений 56 15 59 10 87.

Отладка визуализатора. После построения визуализатора запустим его командой *FindMaximum_ru* из каталога *deploy* [5] и проверим его работоспособность. В случае возникновения ошибок они устраняются обычным способом.

Заключение

Из рассмотренного примера следует, что построение визуализаторов с применением технологии и пакета *Vizi* автоматизировано, в отличие от подходов, изложенных в работе [9].

В настоящей работе приведен пример визуализатора простого алгоритма, однако технология *Vizi* позволяет визуализировать и сложные алгоритмы. Например, в работе [10] описывается реализация визуализатора алгоритма нахождения максимального потока в сети методом Диница и Малхотры – Кумара-Махешвари. Реализация этого алгоритма содержит девять процедур, по которым генерируются 18 автоматов, имеющих в сумме более ста состояний.

Визуализатор алгоритма поиска максимума в массиве натуральных чисел и его исходные коды будут опубликованы на сайте <http://is.ifmo.ru> в разделе «Статьи».

Литература

1. Кнут Д. Искусство программирования. Том 1. Основные алгоритмы. М.: Вильямс, 2000.
2. Кормен Т., Лейзерсон Ч., Ривест Р. Алгоритмы. Построение и анализ. М.: МЦНМО, 1999.
3. Казаков М.А., Столяр С.Е. Визуализаторы алгоритмов как элемент технологии преподавания дискретной математики и программирования // Международная научно-методическая конференция «Телематика-2000». СПб: СПбГИТМО (ТУ), 2000. С. 189–191.
4. Корнеев Г.А., Парфенов В.Г., Столяр С.Е., Васильев В.Н. Визуализаторы алгоритмов как основной инструмент технологии преподавания дискретной математики и программирования // Труды международной научно-методической конференции «Телематика-2001». СПб: СПбГИТМО (ТУ), 2001. С. 119–120.
5. Vizi home page // <http://ctddev.ifmo.ru/vizi>
6. Корнеев Г.А., Шалыто А.А. Преобразование программ в систему взаимодействующих конечных автоматов. // Труды Второй Всероссийской Научной конференции «Методы и средства обработки информации». М.: МГУ, 2005. С. 385–387.
7. Joy B., Steele G., Gosling J., Bracha G. Java Language Specification (Second Edition). Addison-Wesley. 2000. <http://java.sun.com/docs/books/jls/>
8. Корнеев Г.А., Казаков М.А., Шалыто А.А. Разработка логики визуализаторов алгоритмов на основе конечных автоматов // Телекоммуникации и информатизация образования. 2003. № 6. С. 27–58. <http://is.ifmo.ru/works/vis/>
9. Казаков М.А., Шалыто А.А. Использование автоматного программирования для реализации визуализаторов // Компьютерные инструменты в образовании. 2004. № 2. С. 19–33. http://is.ifmo.ru/works/art_vis/
10. Бедный Д.Ю. Нахождения максимального потока в сети методом Диница и Малхотры – Кумара–Махешвари // Мир ПК-Диск. 2005. № 8. <http://is.ifmo.ru/vis/dmkm/>

VIZI — ЯЗЫК ОПИСАНИЯ ЛОГИКИ ВИЗУАЛИЗАТОРОВ АЛГОРИТМОВ

Г.А. Корнеев, А.А. Шалыто

В статье предлагается язык описания логики визуализаторов алгоритмов Vizi, основанный на XML. Этот язык позволяет автоматизировать процесс построения логики визуализаторов алгоритмов дискретной математики.

Введение

Визуализаторы алгоритмов широко применяются для обучения дискретной математике и доказали свою эффективность [1, 2]. При этом часто используются визуализаторы, создаваемые студентами [3].

Основной проблемой при разработке визуализаторов алгоритмов является обеспечение возможности просматривать алгоритм не только в прямом, но и в обратном направлениях [4]. При этом много ошибок допускается именно при реализации «обратного хода». Таким образом, автоматизация создания логики визуализаторов алгоритмов, обеспечивающей трассировку алгоритма как в прямом, так и в обратном направлении весьма полезна.

В настоящей работе предлагается язык описания логики визуализаторов алгоритмов Vizi, построенный на основе расширяемого языка разметки XML (*Extensible Markup Language* [5]). Применение этого языка позволяет автоматизировать разработку логики визуализаторов алгоритмов, что реализовано в проекте Vizi [6].

Концепции языка

В рамках предлагаемого языка алгоритм рассматривается как набор процедур и глобальных переменных. Среди процедур выделяется главная процедура, запускаемая при выполнении алгоритма.

Процедура рассматривается как последовательность операторов следующих типов:

- оператор присваивания;
- оператор ветвления;
- оператор цикла с предусловием;
- оператор вызова процедуры;
- блочный оператор.

Каждому типу оператора, кроме блочного, соответствует XML-элемент. Блочные операторы кодируются неявно.

Отметим, что операторы вызова процедуры позволяют задавать рекурсивные алгоритмы как с явной, так и с косвенной рекурсией. Оба этих случая обрабатываются корректно. Операторы могут использовать как глобальные переменные, определенные на уровне алгоритма, так и локальные переменные, определенные в рамках процедуры.

Для шага может быть указан его уровень, шаблон комментария и связь с визуальным представлением. Уровень шага определяет, в каком случае визуализатор останавливается при исполнении данного шага. Уровень задается целым числом. При этом значение -1 обозначает пропуск шага, 0 – остановку при просмотре алгоритма маленькими шагами, 1 – остановку при просмотре алгоритма большими шагами.

Шаблон комментария определяет вид и параметры комментария, отображаемого на данном шаге. Связь с визуальным представлением позволяет обновлять визуальное представление при отображении шага.

На рис. 1 приведена диаграмма XML-элементов, применяемых при описании логики визуализаторов. Диаграмма основана на диаграмме классов *UML* [7]. При этом классы соответствуют XML-элементам, а их атрибуты – атрибутам элементов. Атрибуты, имеющие стереотип «text», соответствует вложенному тексту элемента.

Абстрактные элементы введены для отображения общности элементов. При этом конкретные XML-элементы им не соответствуют.

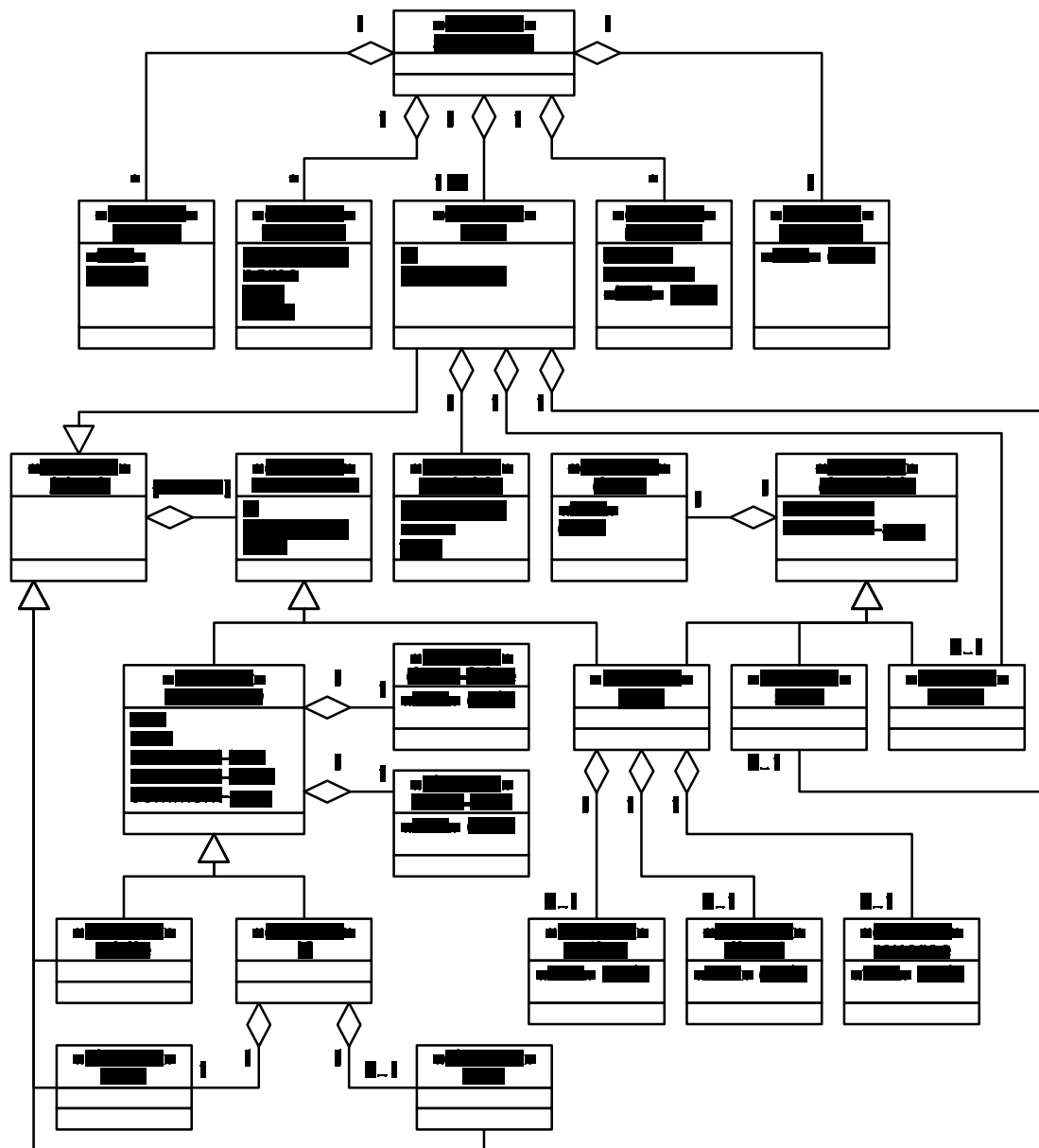


Рис. 1. Диаграмма элементов

Описание алгоритма

Как следует из рис. 1, описание алгоритма задается элементом `algorithm`, содержащим описания процедур (элемент `auto`), глобальных переменных (элемент `variable`), а также вспомогательных конструкций (элементы `import`, `toString` и `method`).

Описание процедур будет рассмотрено ниже.

Описание глобальных переменных. Для каждой глобальной переменной указывается ее описание (атрибут `description`), имя (`name`), тип (`type`) и значение по умолчанию (`value`).

Описание переменной применяется при отладке алгоритма и не используется при генерации логики визуализатора [6].

Имена переменных должны быть корректными *Java*-идентификаторами и не содержать символов подчеркивания.

Значение по умолчанию применяется при автоматизированной проверке корректности построения логики визуализатора [6].

Описание вспомогательных конструкций. Элемент `import` позволяет описывать связь визуализатора с другими типами аналогично конструкции `import` языка *Java* [8]. Имя импортируемого класса (пакета) задается обычным образом и записывается внутри элемента `import`.

Элемент `toString` задает метод, представляющий значения основных переменных визуализатора в виде строки. Этот метод используется при автоматизированной проверке корректности построения логики визуализатора [6]. При этом значения, выдаваемые этим методом при прямом и обратном проходах, сравниваются, и при их неравенстве выдается сообщение об ошибке.

Элемент `method` позволяет задавать не визуализируемые методы, которые могут, например, применяться для построения визуального представления.

Для метода указывается сигнатура метода (атрибут `header`), комментарий (`comment`) и тело метода (текст элемента). Сигнатура и тело метода записываются также как в языке *Java*.

Описание процедур

Описание процедуры задается элементом `auto`. При этом указываются имя (атрибут `name`) и описание (`description`).

В рамках процедуры описываются локальные переменные, начальное и конечное состояния, а также шаги алгоритма.

Описание локальных переменных. Локальные переменные описываются элементом `auto`. Для каждой локальной переменной указываются ее описание (атрибут `description`), имя (`name`) и тип (`type`).

Отметим, что для локальных переменных значения по умолчанию не указываются. Таким образом, исходное значение локальной переменной не определено.

Описание начального и конечного состояний. Описание начального (элемент `start`) и конечного (элемент `finish`) состояний позволяет задать комментарий и визуальное представление, отображаемые пользователю при входе и выходе из процедуры. Они обычно применяются в главной процедуре алгоритма.

Комментарий задается шаблоном комментария (атрибут `comment`) и параметрами комментария (`comment-args`).

Шаблон комментария содержит ссылки на параметры, записываемые в виде

{номер аргумента}

Параметры комментария перечисляются через запятую. Каждый параметр представляет собой выражение, вычисляемое в момент отображения комментария. Таким образом, комментарии могут включать в себя значения переменных алгоритма.

Связь с визуальным представлением указывается во вложенном элементе `draw`, содержащем код, исполняемый при отображении визуального представления пользователю. Таким образом, визуальное представление может зависеть не только от текущего состояния, но и от значений переменных.

Описание операторов

С точки зрения удобства визуализации, операторы (абстрактный элемент `statement` на рис. 1) разделены следующим образом:

- неотображаемые операторы:
 - оператор вызова процедуры;
 - блочный оператор (абстрактный элемент `block` на рис. 1).
- отображаемые операторы:
 - оператор присваивания;
 - условные операторы (абстрактный элемент `condition` на рис. 1):
 - оператор ветвления;
 - оператор цикла с предусловием.

Оператор вызова процедуры. Оператор вызова процедуры описывается элементом `call-auto`. При этом указывается имя вызываемой процедуры (атрибут `id`).

Блочный оператор. Предлагаемый язык не содержит выделенной конструкции для описания блочных операторов. Вместо этого тело процедуры, цикла с предусловием и ветвей оператора ветвления являются блочными операторами.

Отображаемые операторы. Для отображаемых операторов указываются идентификаторы (атрибут `id`), описание (`description`) и уровень оператора (`level`), указывающий остановится ли визуализатор на данном операторе.

Значение уровня по умолчанию равно нулю.

Оператор присваивания. В операторе присваивания могут быть изменены значения одной или нескольких переменных.

Для трассировки в обратном направлении требуется осуществить обращение операторов – построить код, исполняемый при обратном проходе [4].

Оператор присваивания может обращаться либо автоматически, либо вручную. Ручное обращение иногда позволяет сэкономить память и/или время по сравнению с автоматическим выполнением этой операции [4].

При автоматическом обращении выполняемые действия записываются в элементе `action`. При этом для изменения переменных применяются операторы обратимого присваивания, имеющие вид `@=`. При обратном проходе все изменения, выполненные операторами обратимого присваивания, автоматически «откатываются» [4].

В случае ручного обращения отдельно указываются действия, выполняемые при прямом (элемент `direct`) и обратном (элемент `reverse`) проходах. В этом случае проверка верности обращения может быть автоматизирована [6].

В случае отображения пользователю для оператора присваивания указываются комментарий и связь с визуальным представлением, как было описано выше для начального и конечного состояний.

Условные операторы. Для условного оператора указывается условие (атрибут `test`), позволяющее выбрать следующий оператор при прямом проходе.

Выбор следующего оператора при обратном проходе может осуществляться как автоматически, так и вручную. Во втором случае соответствующее условие указывается в атрибуте `rtest`.

Для условных операторов также указываются комментарии и связь с визуальным представлением. При этом шаблоны комментариев и действия по обновлению визуального представления указываются отдельно для истинного (атрибут `comment-true` и элемент `draw-true`) и ложного (`comment-false`, `draw-false`) значений условия.

Оператор ветвления задается элементом `if`, в который вложены описания ветвей, исполняемых при истинности (элемент `then`) и ложности (`else`) условия. Для каждой ветви указывается набор операторов, составляющих эту ветвь.

Отметим, что в случае укороченного оператора ветвления ветвь `else` опускается.

Оператор цикла с предусловием задается элементом `while`, в который вложены описания операторов, составляющих тело цикла.

Переменные

При записи описания логики визуализатора переменные делятся на глобальные и локальные. Глобальные переменные доступны во всех процедурах, а локальные – только в той процедуре, в которой они объявлены.

Для доступа к переменным используется `@`-нотация. При этом настоящее имя переменной не используется в явном виде, а подставляется автоматически.

Доступ к переменной осуществляется при помощи выражения вида:

```
@<имя переменной>
```

Например, оператор

```
@max = @a[@i]
```

выполняет присваивание переменной `max` значения `i`-го элемента массива `a`. Заметим, что при наличии в области видимости глобальной и локальной переменных с одним и тем же именем используется локальная переменная.

Для создания строкового представления автомата в процедуре `toString` введен синтаксис, позволяющий производить доступ к локальным переменным других процедур. Доступ к переменной, объявленной в другой процедуре, осуществляется следующим образом:

```
@<имя процедуры>@<имя переменной>
```

Например, оператор

```
buffer.append(@Main@i);
```

осуществляет добавление к буферу значения локальной переменной `i` процедуры `Main`.

Для доступа к переменным модели из кода визуализатора применяются для глобальных переменных выражения вида

```
<переменная модели>.<имя переменной>
```

а для локальных переменных – выражение вида

```
<переменная модели>.<имя процедуры>_<имя переменной>
```

Например:

```
data.max = 0;
System.out.println(data.Main_i);
```

Пример

Рассмотрим описания логики визуализатора алгоритма поиска максимума в массиве натуральных чисел.

Данная задача может быть решена следующей программой:

```
void main() {
    int max = 0;
    for (int i = 0; i < a.length; i++) {
        if (max < a[i]) {
            max = a[i];
        }
    }
}
```

```

    }
  }
}

```

Здесь `a` – массив, в котором производится поиск максимума, а `max` – значение текущего максимума (после i -ой итерации – среди первых i элементов).

Отметим, что инициализация максимума нулем не приводит к ошибке, так как по условию задачи в массиве содержатся только натуральные числа.

На предложенном языке данный алгоритм может быть записан следующим образом:

```

1: <algorithm>
2:   <variable name="a" type="int[]" value="new int[]{1, 2, 3, 1,
      6}" description="Массив для поиска"/>
3:   <variable name="max" type="int" value="0"
      description="Текущий максимум"/>
4:   <auto id="Main" description="Ищет максимум в массиве">
5:     <variable description="Переменная цикла" name="i"
      type="int"/>
6:     <start comment-ru="На экране изображен массив, в котором
      будет осуществляться поиск максимума">
7:       <draw>@visualizer.updateArray(0, 0);</draw>
8:     </start>
9:     <step id="Initialization" description="Инициализация"
      comment-ru="Инициализируем максимум нулем (так как в
      массиве только натуральные числа).">
10:      <draw>@visualizer.updateArray(0, 0);</draw>
11:      <action>@max @= 0;</action>
12:    </step>
13:    <step id="LoopInit" description=":" level="-1"></step>
14:    <while id="Loop" description="Цикл" test="@i < @a.length"
      rtest="@i >= 0" level="-1">
15:      <if id="Cond" description="Условие" test="@max < @a[@i]"
      true-comment-ru="{0} больше текущего максимума ({1})"
      false-comment-ru="{0} не больше текущего максимума
      ({1})"
      comment-args="new Integer(@a[@i]), new Integer(@max)">
16:        <draw>@visualizer.updateArray(@i, 1);</draw>
17:        <then>
18:          <step id="newMax" description="Обновление максимума"
      comment-ru="Обновляем текущий максимум">
19:            <draw>@visualizer.updateArray(@i, 2);</draw>
20:            <action>@max @= @a[@i];</action>
21:          </step>
22:        </then>
23:      </if>
24:      <step id="inc" description=":" level="-1"></step>
25:      <forward>@i = @i + 1;</forward>

```

```

26:         <reverse>@i = @i - 1;</reverse>
27:     </step>
28: </while>
29:     <finish comment-ru="Максимум найден ({0})"
        comment-args="new Integer(@max)">
30:         <draw>@visualizer.updateArray(0, 0);</draw>
31:     </finish>
32: </auto>
33: </algorithm>

```

Рассмотрим некоторые части этого описания подробнее.

Данный алгоритм содержит две глобальные переменные: `a` – массив, в котором осуществляется поиск (объявлена во второй строке) и `max` – текущее значение максимума (объявлена в третьей строке).

В четвертой строке начинается описание основной процедуры `main`, которое заканчивается в строке 31. В процедуре объявлена локальная переменная – индекс текущего элемента массива (строка 4).

За описанием локальной переменной следует описание начального состояния (строки 6–8), включающее комментарий (строка 6) и связь с визуальным представлением (строка 7).

Далее приводится описание шагов алгоритма, начинающееся операторами присваивания начальных значений переменным `max` (строки 9–12) и `i` (строка 13). Отметим, что первое из этих присваиваний визуализируется, а второе – нет, так как для него указан уровень `-1`.

В строках с 14 по 28 описан оператор цикла с предусловием. Отметим, что для него применяется обращение вручную. Соответствующее условие указано в атрибуте `rtest`.

Тело оператора цикла включает два оператора: ветвления (строки 15–23) и присваивания (строки 24–27). При этом оператор присваивания использует косвенное обращение: в строке 25 указано действие, осуществляемое при прямом проходе, а в строке 26 – действие, выполняемое при обратном проходе.

Заключение

Предложенный язык описания логики визуализаторов позволяет записывать логику визуализатора в привычной форме, дополняя ее описанием комментариев и связью с визуальным представлением. По такому описанию логика визуализатора может быть сгенерирована автоматически [6].

Автоматизация процесса построения логики визуализатора позволяет реализовывать визуализаторы таких сложных алгоритмов, как, например, алгоритм Укконена построения суффиксного дерева [9] или алгоритм построения кратчайшего дерева путей в графе [10], что практически невозможно при ручном подходе.

Работа выполнена в совместной лаборатории СПбГУ ИТМО и центра разработки Borland «Технологии программирования».

Литература

1. Byrne M., Catrambone R., Stasko J. Evaluating Animations as Student Aids in Learning Computer Algorithms // *Computers & Education*. Vol. 33. № 4. 1999. P. 253–278.

2. Казаков М.А., Столяр С.Е. Визуализаторы алгоритмов как элемент технологии преподавания дискретной математики и программирования / Международная научно-методическая конференция “Телематика-2000”. СПб: СПбИТМО (ТУ), 2000. С. 189–191.
3. Stasko J. Using Student-Built Algorithm Animations as Learning Aids / Proceedings of the ACM Technical Symposium on Computer Science Education (SIGCSE '97), CA, 1997. P. 25–29.
4. Корнеев Г.А., Казаков М.А., Шалыто А.А. Метод построения логики работы визуализатора алгоритмов на основе конечных автоматов // Телекоммуникации и информатизация образования. 2003. № 6., С. 27–58. <http://is.ifmo.ru/works/vis/>
5. Extensible Markup Language (XML) 1.1. <http://www.w3.org/TR/xml11/>.
6. Сайт проекта Vizi // <http://ctddev.ifmo.ru/vizi>
7. Буч Г., Якобсон А., Рамбо Дж. UML. 2-е издание. СПб: Питер, 2006, 736 с.
8. Joy B., Steele G., Gosling J., Bracha G. Java Language Specification, Second Edition. NJ.: Addison-Wesley. 2000. <http://java.sun.com/docs/books/jls/>
9. Ахметов И.Р. Разработка визуализатора алгоритма Укконена построения суффиксных деревьев на основе технологии Vizi. <http://is.ifmo.ru/vis/ukkonen/>
10. Пименов С.Ю., Корнеев Г.А., Шалыто А.А. Алгоритм Чу Йонджина и Лю Цзенхонга построения кратчайшего корневого дерева в ориентированном графе. <http://is.ifmo.ru/vis/ctree/>

РАЗРАБОТКА ТЕОРЕТИЧЕСКИХ ОСНОВ И РЕАЛИЗАЦИЯ МЕТОДОВ ПОСТРОЕНИЯ КОМПЬЮТЕРНЫХ СЕТЕЙ НА БАЗЕ IP-ПРОТОКОЛОВ НОВОГО ПОКОЛЕНИЯ

В.Н. Васильев, Ю.В. Гугель, Ю.Л. Ижванов

В статье приведены результаты исследования методов построения компьютерных сетей на базе IPv6 протокола, предлагается, используя транспортную инфраструктуру сети RUNNet (Федеральная университетская компьютерная сеть России), разработать и ввести в опытную эксплуатацию типовые сегменты, использующие новые технологии построения IP сетей, позволяющие поднять эффективность доступа до ресурсов научно-образовательных сетей России. Это позволит повысить эффективность уже развернутой транспортной инфраструктуры и интегрировать ее в другие современные как российские, так и международные высокоскоростные сети.

Введение

Вопрос развития телекоммуникационной инфраструктуры науки и образования в России стоит очень остро, так как без этого невозможно оставаться в ряду научно-технологических развитых стран, сохранять статус научной державы, обеспечить технологическое развитие, реализовать государственное планирование стратегически важных фундаментальных и прикладных исследований.

На данный момент одним из самых важных вопросов эффективного использования этой инфраструктуры является организация связности по высокоскоростным каналам связи с применением новых технологий построения базовых транспортных сетей и IP-сетей. Наиболее интересными и развивающимися в данной области являются технологии IPv6, групповое вещание (multicast), а также VPN (виртуальные частные сети) с заданными параметрами качества обслуживания (QoS), позволяющие существенно повысить эффективность использования транспортной инфраструктуры и уровень сервиса, организацию защиты сетей, а также осуществлять мониторинг для выработки рекомендаций по оптимизации сетей.

Появление любой новой технологии вызывает большой отклик в мире ученых, операторов сетей связи и пользователей. В случае настолько широкомасштабного явления, как новый протокол сети Интернет – IPv6, который затронет абсолютно все аспекты работы глобальной сети, распространение информации о новой технологии, призванной привести глобальную сеть к качественно новому уровню, расширить ее возможности, должно стать особенно актуальным. Объективные оценки показывают существенное превосходство IPv6 по сравнению с тем протоколом сети, который используется сейчас, а также его существенно лучшую приспособленность для дальнейшего роста глобальной сети по всем направлениям развития. Однако недостаток информации, особенно актуальной среди сетевой общественности в нашей стране, может замедлить или отложить работы по внедрению и исследованиям нового протокола на некоторое время, что неблагоприятно скажется на развитии сетевой инфраструктуры страны в целом.

На настоящее время существует ряд обстоятельств, препятствующих быстрому внедрению протокола IPv6:

- практически полное отсутствие поддержки протокола в аппаратных сетевых экранах. Корпоративный пользователь не может сейчас полностью отказаться от протокола IPv4, а значит, какое-то время в сети будут использоваться обе версии IP;
- недостаточная устойчивость тех версий программного обеспечения маршрутизаторов, в которых реализована поддержка протокола IPv6. Недостаточный набор реализованных в этом программном обеспечении внутренних протоколов маршрутизации IPv6;

- недостаточное количество программных продуктов, использующих преимущества нового протокола;
- удовлетворенность подавляющего большинства пользователей услугами, предоставляемыми на основе протокола IPv4. Недостаточный спрос на IPv6 услуги делает пока внедрение нового протокола экономически рискованным.

Особенностью российского Интернета является высокая степень концентрации пользователей этой сети в крупных городах. Как видно из рис. 1, большая часть российской аудитории Интернета сосредоточена в четырех крупных городах – Москве, Санкт-Петербурге, Новосибирске и Екатеринбурге.

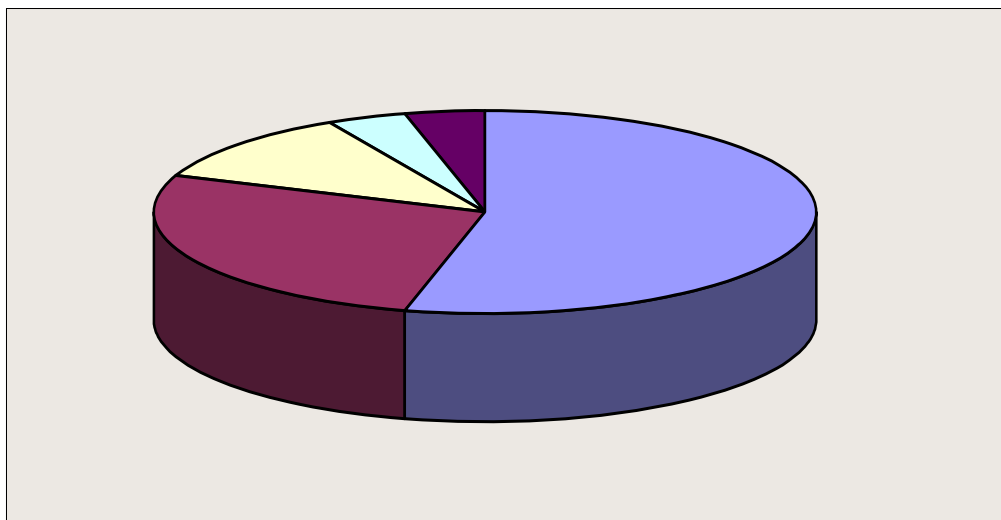


Рис. 1. Распределение пользователей сети Интернет в России

В настоящее время основные работы по внедрению протокола IPv6 в России осуществляются только научно-образовательными сетями. RUNNet, RBNet, FREEnet в настоящее время уже поддерживают IPv6 протоколы. Для этих сетей RIPE-ом [1] выделено соответствующее адресное пространство, удовлетворяющее потребности всего российского научно-образовательного сообщества.

```
inet6num: 2001:0640::/32
netname: RU-FREENET-19991115
descr: IPv6 FREEnet
country: RU
inet6num: 2001:06D0::/32
netname: RU-ROSNIIROS-20010219
descr: Rbnet transition project
country: RU
inet6num: 2001:B08::/32
netname: RU-RUNNET-20030214
descr: RUNNet
descr: Russian Federal University Network
country: RU
```

Критерии эффективности сетей передачи данных на базе IPv6 протокола и сравнение сетей с сетями на базе IPv4

Изменения IPv6 по отношению к IPv4 можно поделить на следующие группы.

- Расширение адресации: в IPv6 длина адреса расширена до 128 бит (против 32 в IPv4), что позволяет обеспечить больше уровней иерархии адресации, увеличить число адресуемых узлов, упростить авто-конфигурацию. Для расширения возмож-

ности мультикастинг-маршрутизации в адресное поле введено субполе «score» (группа адресов). Определен новый тип адреса «anycast address» (эникастный), который используется для посылки запросов клиента любой группе серверов. Эникастная адресация предназначена для использования с набором взаимодействующих серверов, чьи адреса не известны клиенту заранее.

- Спецификация формата заголовков: некоторые поля заголовка IPv4 отбрасываются или делаются опциональными, уменьшая издержки, связанные с обработкой заголовков пакетов с тем, чтобы уменьшить влияние расширения длины адресов в IPv6.
- Улучшенная поддержка расширений и опций: изменение кодирования опций IP-заголовков позволяет облегчить переадресацию пакетов, ослабляет ограничения на длину опций и делает более доступным введение дополнительных опций в будущем.
- Возможность пометки потоков данных: введена возможность пометить пакеты, принадлежащие определенным транспортным потокам, для которых отправитель запросил определенную процедуру обработки, например, нестандартный тип TOS (вид услуг) или обработка данных в реальном масштабе времени.
- Идентификация и защита частных обменов: в IPv6 введена спецификация идентификации сетевых объектов или субъектов, для обеспечения целостности данных и при желании защиты частной информации.

Основные характеристики IPv6 перечислены ниже.

- Адресное пространство увеличено до 128 разрядов. Чтобы наглядно представить себе эту огромную величину, достаточно сказать, что этого количества адресов хватит, чтобы каждый житель планеты владел сетью, соизмеримой с нынешней Internet.
- Гибкий формат заголовков дейтаграмм IP позволяет маршрутизаторам обрабатывать их намного быстрее.
- Встроенная поддержка резервирования ресурсов, в частности, гарантированного обеспечения качества услуг QoS.
- Встроенная поддержка протокола шифрования IPsec.
- Устранение избыточной информации из заголовков дейтаграмм IP.
- Новые опции в заголовках дейтаграмм IP.
- Расширяемость протокола без пересмотра стандарта.
- Гибкое распределение адресного пространства на классы, включая поддержку записи MAC-адресов в IP-адреса. Один из классов выделен для унаследованных адресов IPv4. Адресное пространство IPv6 позволяет создавать многоуровневые схемы адресации.
- Наличие трех типов адресации: unicast, multicast и cluster (anycast). Первые два типа имеются и в IPv4, новым является лишь тип cluster. В этом случае сообщения посылаются не всем членам группы, а только одному (ближайшему).

Несмотря на то, что большинство производителей телекоммуникационного оборудования и программных средств уже встроили поддержку IPv6 в свои продукты, миграция к IPv6 будет представлять собой длительный процесс, причем в большинстве реализаций IPv4 и IPv6 какое-то время будут сосуществовать друг с другом. Наличие подходящих процедур миграции должно стать определяющим фактором для внедрения IPv6. Обеспечение совместимости между компонентами IPv6 и оставшимися элементами IPv4, такими как хосты и маршрутизаторы, представляет собой основную проблему процесса миграции.

Двумя основными стратегиями миграции являются двойные стеки и туннелирование. Первый подход предполагает выполнение одного стека для IPv4 и другого для IPv6 на каждом узле с поддержкой IP. Данная конфигурация предусматривает наличие

интерфейса для идентификации типа трафика и для направления этого трафика адресату. Таким интерфейсом мог бы быть IPv6-совместимый API – скорее всего, расширение IPv4 API. Узлы с двумя стеками поддерживали бы как 32-разрядные, так и 128-разрядные адреса.

Для реализации двойных стеков администраторам скорее необходимо изменить информацию о конфигурации маршрутизаторов и модернизировать DNS для поддержки более длинных адресов IPv6.

При туннелировании датаграммы IPv6 инкапсулируются в заголовок IPv4. На принимающем конце туннелирующий маршрутизатор удаляет заголовок IPv4 из датаграммы IPv6. Два туннеля конфигурируются таким образом, что точка входа в один туннель служит точкой выхода из другого туннеля.

При двунаправленном туннелировании два однонаправленных туннеля объединяются для организации двустороннего потока пакетов. Туннелирование позволяет создать мост для передачи пакетов IPv6 через сеть IPv4.

Одно из преимуществ туннелирования состоит в простоте организации взаимодействия между хостами и узлами IPv6 по унаследованной сети. К тому же, в отличие от альтернативных подходов типа NAT, туннелирование прозрачно для большинства конечных систем и приложений.

Другое достоинство туннелирования заключается в том, что оно обеспечивает эффективную инфраструктуру для тестирования. Данный подход позволяет любой компании создать свою частную тестовую сеть, не нарушая функционирования имеющейся сети на базе IPv4, причем она может быть подключена к таким же тестовым сетям других компаний.

Недостаток туннелирования – в увеличении накладных расходов вследствие необходимости передачи дополнительных заголовков, из-за чего функционирование сети может ухудшиться. Поэтому администратору сети следует заранее убедиться, что глобальная сеть сможет справиться с возросшей нагрузкой.

Основные способы сопряжения участков сетей IPv4 и IPv6

Как уже отмечалось выше, протоколу IPv6 придется длительное время взаимодействовать с IPv4 и на первом этапе даже существовать в его среде. IPv6, к счастью, не создает ограничений, связанных с жестким порядком действий в переходный период: администраторы сетей могут модернизировать свои хост-машины, а затем маршрутизаторы, либо, наоборот, сначала – маршрутизаторы, а потом – хост-машины. Они даже могут модернизировать лишь часть хост-машин и часть маршрутизаторов, отложив остальное на будущее.

Каждый маршрутизатор, поддерживающий оба протокола – IPv4 и IPv6, – не нарушает связность между IPv4-узлами, которые он обслуживает. Хост-машина, которая поддерживает оба стека протокола IP, также не утрачивает связность с другими узлами сети. Создание в сети «островков», поддерживающих только IPv6, также не ухудшает связность, если на границах этих «островков» установлены устройства, поддерживающие обе версии протокола IP.

Порядок модернизации узлов влияет только на способ возможного взаимодействия этих узлов друг с другом. Существуют три метода организации такого взаимодействия: трансляторы протоколов, двухпротокольные стеки (dual stacks) и туннелирование (tunnels).

Метод трансляции протоколов состоит в преобразовании по определенным правилам пакетов IPv6 в пакеты IPv4, и обратно. В настоящее время применяются три механизма трансляции протоколов. Один из них – преобразование заголовков, в котором IP-заголовки транслируются из одной версии в другую протокол-шлюзом (protocol

gateway device). Все шлюзы для трансляции протоколов размещены на границах между сетями IPv4 и IPv6.

Другой механизм трансляции протокола предполагает использование транспортного ретранслятора: протокол-шлюз принимает сегменты TCP и дейтаграммы UDP от исходного узла и воссоздает TCP- и UDP-трафик, передаваемый узлу назначения. Другими словами, такой шлюз не просто исследует и переводит IP-заголовки, но также обрабатывает заголовки протокола транспортного уровня, чтобы заставить их «работать» на соответствующей версии IP-протокола.

Третий механизм представляет собой проху-приложение. В этом случае транслятор является шлюзом между сетями IPv4 и IPv6 на прикладном уровне.

Строение пакета IPv6, однако, слишком сильно отличается от устройства пакета IPv4, чтобы можно было сделать преобразование одного в другой, без утраты некоторых важных свойств протокола IPv6. Одна из проблем связана с преобразованием адресов IPv6 в адреса IPv4. Другая сложность состоит в том, что протокол IPv4 уничтожает или изменяет половину IPv6 заголовков (например, в IPv4 отсутствует аналог идентификатора потока – «flow label»). В результате при трансляции утрачиваются многие из тех функций протокола IPv6, которые выгодно отличают его от IPv4. Кроме того, IPv6 снимает поддержку фрагментации IPv4-пакетов при их передаче.

Второй подход состоит в том, что IPv6-узлы должны поддерживать оба стека протокола IP – и IPv4, и IPv6 – в полном объеме. Узел сам выбирает, какой протокол использовать в каждом конкретном случае: IPv6 может использоваться только в том случае, если удаленный узел также поддерживает IPv6. Признаком того, что удаленный узел поддерживает новую версию протокола IP, служит наличие в DNS специальной записи (AAAA), ассоциированной с данным узлом. Двухпротокольный подход может реализовываться и сам по себе, и в комбинации с третьим из перечисленных подходов – туннелированием.

На первом этапе, пока большая часть Интернета работает по протоколу IPv4, реализация связности по протоколу IPv6 будет осуществляться методом туннелирования. Такое туннелирование достигается путем инкапсуляции IPv6-пакета внутри пакета IPv4. Передача информации при этом осуществляется следующим образом: IPv6-узел создает IPv6-пакеты, которые инкапсулируются в пакеты IPv4 и передаются через IPv4-сеть. Узел на другом конце туннеля разворачивает IPv4-пакет, извлекает IPv6-пакет и затем передает его в узел назначения.

Механизм туннелирования давно используется в IPv4 для транспортировки не IP-пакетов. Соответственно, на другом конце туннеля выполняется обратное преобразование, а в промежутке имеет место обычная доставка пакета IPv4. С точки зрения IPv6, IPv4 играет здесь роль протокола канального уровня.

Можно выделить четыре вида туннелей:

- хост-хост. Два хоста с двойным стеком протоколов, имеющие доступ только к инфраструктуре IPv4, строят туннель «из конца в конец»;
- маршрутизатор-хост. Здесь имеет место туннель «из середины в конец»;
- хост-маршрутизатор. Туннель «из начала в середину»;
- маршрутизатор-маршрутизатор. Здесь туннель соединяет две промежуточные точки на маршруте.

В двух первых случаях конечная точка туннеля совпадает с конечной точкой маршрута пакета IPv6. Следовательно, адрес конца туннеля должен автоматически вычисляться как функция адреса целевого хоста. Говорят, что при этом производится автоматическое туннелирование. Чтобы автоматическое туннелирование было возможным, IPv6-адреса концов туннеля должны быть IPv4-совместимыми и синтаксически, т.е. они должны получаться из адресов IPv4 приписыванием слева 96 нулевых бит.

Если конечная точка туннеля (маршрутизатор) не вычисляется по целевому адресу, приходится прибегать к заранее сконфигурированному туннелированию, когда параметры туннеля задаются маршрутной таблицей в инкапсулирующем узле (например, второй конец может задаваться как подразумеваемый маршрутизатор для IPv6). Подобный подход необходим, когда целевой адрес не является IPv4-совместимым. В этом случае отправитель должен знать IPv4-адрес маршрутизатора с двойным стеком, способного организовать доставку пакета IPv6.

Оба конца любого туннеля (и автоматического, и сконфигурированного) должны обладать IPv4-совместимыми адресами.

Во взаимодействии систем IPv6/IPv4 есть ряд тонкостей, связанных с фрагментацией пакетов и функционированием протоколов маршрутизации. Со временем может возникнуть ситуация, когда уже узлы IPv4 будут составлять островки в море IPv6. В таком случае применимо туннелирование, обратное по сравнению с рассмотренным: пакеты IPv4 инкапсулируются средствами IPv6.

Описанные выше средства являются весьма гибкими, они позволяют каждой организации выбрать свою стратегию перехода на IPv6. Можно начать с хостов, постепенно добавляя к ним периферийные маршрутизаторы. Это позволяет создать начальную массу приложений над IPv6. Можно начать с магистральных маршрутизаторов, готовя транспортную инфраструктуру и инструменты централизованного администрирования. Можно начинать перестройку в рамках передовых рабочих групп, создавая островки IPv6, содержащие хосты и маршрутизаторы с двойным стеком и, может быть, новые хосты только с IPv6.

Переход от IPv4 к IPv6 можно сравнить с увеличением разрядности компьютеров. Процесс этот, безусловно, необходим, но он требует переделки или, по крайней мере, пересмотра реализации всего стека TCP/IP.

У IPv6 есть не только сторонники, но и противники. Противники отмечают, что в рамках IPv4 можно реализовать практически все нововведения, предлагаемые для IPv6. Действительно, разработаны спецификации IPsec, в равной степени ориентированные на IPv4 и IPv6. Протокол DHCP позволяет осуществлять автоконфигурирование адресов. Ведутся работы по поддержке классов обслуживания. Созданы мощные коммутирующие маршрутизаторы, справляющиеся с обработкой заголовков IPv4 на гигабитных скоростях. Наконец, трансляция сетевых адресов позволяет снять или существенно смягчить проблему исчерпания 32-битного адресного пространства.

Эффект от механизма трансляции сетевых адресов зависит от того, какого рода сетевая связность нужна. Если требуются кратковременные выходы «наружу», можно достичь существенной экономии адресов, но, к сожалению, в условиях длительных сеансов взаимодействия, типичных для современных и, тем более, будущих приложений, пользы от трансляции адресов будет немного.

Далее, возможность реализовать в рамках IPv4 перспективные нововведения, как ни странно, скорее является аргументом в пользу перехода на IPv6. Дело в том, что стек TCP/IP меняется, и объем изменений (и количество ошибок, которые при этом могут быть сделаны, а также количество новой информации, которую предстоит усвоить сетевым администраторам) нельзя недооценивать. Есть ли смысл страдать, не избавляясь от неустраимых пороков IPv4?

Сообщество Интернет, накопив огромный опыт и в области стандартизации, и в области эксплуатации колоссальных по масштабам сетей, очень ответственно относится к переходу на IPv6. Предлагаемые и разрабатываемые спецификации охватывают по существу все аспекты функционирования сетевых конфигураций. Для практического опробования реализаций IPv6 создана всемирная экспериментальная сеть 6Bone. Все основные производители сетевого оборудования (3Com, Bay Networks, Cisco Systems и др.) [2] и операционных систем (IBM, Microsoft, Sun Microsystems и многие другие)

предлагают взаимно совместимые (правда, по большей части экспериментальные) реализации IPv6.

Большинство пользователей воспримут переход на фоне обновления версии используемой операционной системы, и сам по себе он не создаст для них особых проблем. Тем же, кто определяет стратегию развития корпоративных информационных систем, стоит уже сейчас начать присматриваться к IPv6 и, выбирая новые сетевые решения, учитывать долговременные перспективы.

Протокол IPv6 в сети RUNNet

Работы по внедрению новых протоколов в сети RUNNet [3, 4] ведутся постоянно. Создана инфраструктура высокоскоростной телекоммуникационной сети, а также реализован полигон на базе протокола IPv6, который является основным полигоном для отработки современных сетевых технологий и можно использовать для выполнения работ по данному проекту. Постоянно проводится анализ мирового опыта реализации телекоммуникационных инфраструктур на базе стека протоколов IPv6. Особое внимание уделяется изучению процесса стандартизации протокола и результатов его практической реализации в общественных сетях. Полигон IPv6 действует в сети RUNNet с начала 2001 года. На сегодняшний момент топология магистральной части полигона имеет вид, представленный на рис. 2.

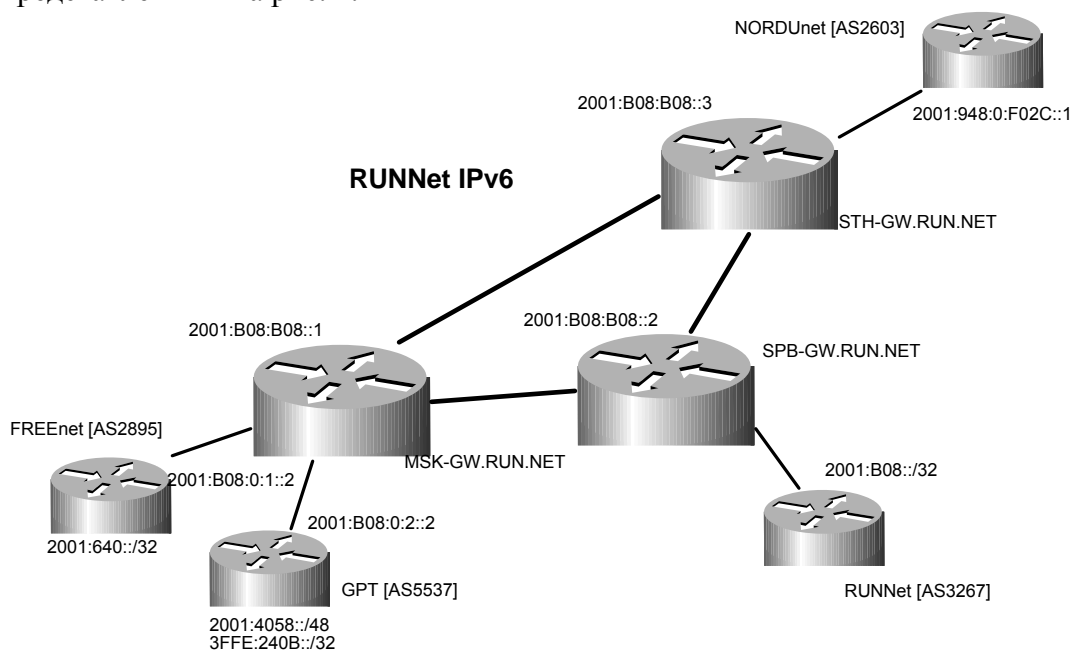


Рис. 2. Опорная IPv6 сеть RUNNet

Полигон организован на основе существующей IPv4 сетевой инфраструктуры. При создании полигона использовалось исключительно имевшееся в наличии и функционирующее на сети оборудование. В качестве маршрутизаторов использовались Cisco 7606, Cisco 7206VXR, хорошо зарекомендовавшие себя в работе в IPv4 сетях. На нескольких компьютерах сети, работающих под управлением ОС FreeBSD и Microsoft Windows XP в параллель существующему IPv4 TCP/IP стеку, было установлено и сконфигурировано программное обеспечение для работы с IPv6. Также было проведено обновление внутреннего программного обеспечения для маршрутизаторов Cisco 7606, Cisco 7206 на новое экспериментальное – с поддержкой протокола IPv6.

Одним из основных исследований протокола IPv6 на полигоне сети RUNNet стало развертывание различных настроек маршрутизации IP-трафика между серверами и

клиентами. Таким образом был получен набор необходимых сетевых конфигураций и политик администрирования. RUNNet имеет свой собственный sTIA, поэтому здесь требуется лишь согласовать, кто из участников подключения выделяет префикс для соединения (линк).

При переходе к новой версии протокола IP на магистральных участках научно-образовательных сетей и сети RUNNet реализована архитектура с двойным стеком IPv4/IPv6 для обеспечения обратной совместимости с доминирующим сейчас в Интернете протоколом IPv4. Для подключения пользователей используется метод тунелирования IPv6 трафика в IPv4 и метод двойного стека, когда на канале до университета используются и IPv4-, и IPv6-протокольные стеки. Такие решения позволят удовлетворять спрос и сегодня, и в будущем, когда резко возрастет количество пользователей и устройств, подключенных к IPv6 Интернет. В настоящее время заключено пиринговое соглашения с сетью GEANT, а также соглашение о сотрудничестве с NORDUNet, что обеспечивает взаимодействие Российских и международных научно-образовательных сетей с использованием протокола IPv6. Инициаторами и организаторами этих работ были Государственный научно-исследовательский институт информационных технологий и телекоммуникаций (ФГУ ГНИИ ИТТ «Информика») и Санкт-Петербургский государственный университет информационных технологий, механики и оптики (СПбГУ ИТМО). В настоящий момент ФГУ ГНИИ ИТТ «Информика» и СПбГУ ИТМО совместно выполняют работы по проекту «Разработка теоретических основ и реализация методов построения компьютерных сетей на базе IP протоколов нового поколения» по приоритетному направлению «Информационно-телекоммуникационные системы» в рамках федеральной целевой научно-технической программы «Исследования и разработки по приоритетным направлениям развития науки и техники» (заказчик – Федеральное агентство по науке и инновациям) [5].

В результате выполняемой работы, с использованием транспортной инфраструктуры федеральной университетской компьютерной сети России RUNNet, разрабатываются и вводятся в опытную эксплуатацию типовые сегменты, использующие новые технологии построения IP-сетей, позволяющие поднять эффективность доступа до ресурсов научно-образовательных сетей России. Это позволит повысить эффективность уже развернутой транспортной инфраструктуры и интегрировать ее в другие современные как российские, так и международные высокоскоростные сети.

Работа включает в себя следующие взаимосвязанные направления.

- Разработка математических моделей маршрутизации в IPv6 сетях. Разработка методов оптимизации маршрутизации трафика. Методы обеспечения QoS в IPv6. Разработка принципов маршрутизации IPv6 в масштабах крупной региональной сети. Существенными вопросами здесь являются, прежде всего, стабильность реализации стека протокола IPv6 на различных клиентских и сетевых устройствах, возможности и эффективность работы соответствующих реализаций протоколов маршрутизации на всех уровнях сети, начиная от магистральных маршрутизаторов сети (устройства серии Cisco 72XX, Cisco76XX) до маршрутизаторов отдельных рабочих групп на базе устройств серии Cisco 1700, 2600, и вплоть до маршрутизаторов, реализованных на устройствах под управлением ОС Unix.
- Эксперименты с мобильными устройствами. Мобильность – одна из важных функций, реализованных в протоколе IPv6 (работа над ее механизмами еще не закончена, но ранние реализации существуют), а вместе с тем – один из признаков современной культуры. Работы над изучением свойств мобильности, заложенных в протокол IPv6, запланированы на более поздние периоды развития проекта, но, тем не менее, являются важным звеном запланированных исследований.
- Организация инфраструктуры, обеспечивающей постепенный переход на новый протокол с сохранением полной совместимости с текущей версией Интернет-

протокола. Известно, что внедрение любого нового протокола будет происходить гораздо сложнее, если он не обеспечивает пользователям возможность беспрепятственно работать, в том числе, с ресурсами, которые доступны только по старой инфраструктуре. В IPv6 такая совместимость может быть достигнута целым рядом внешних механизмов, образующих достаточно сложную систему. Это и методы туннелирования IPv6 трафика поверх IPv4, и всевозможные протокольные трансляторы, разработку которых, в частности, ведут российские ученые, а натурное тестирование которых может быть произведено в RUNNet, например, кэширующий WWW-прокси, работающий одновременно с протоколами IP разных версий, и прокси-серверы, работающие на уровне TCP/IP API (SOCKS).

- Создание защищенного сегмента IPv6 сети с использованием интегрированных средств обеспечения безопасности. В данном контексте представляет существенный интерес не только создание защищенных посредством шифрования на сетевом уровне виртуальных частных сетей, но и создание средств межсетевого разграничения IPv6 трафика по признаку безопасности, в том числе в условиях планового изменения адресного блока IPv6 (т.е. отработка сценариев смены поставщика IPv6 коннективности, что с учетом требования топологичности иерархии IPv6 адресов, ведет за собой изменение всего адресного пространства).
- Создание системы, обеспечивающей возможность организации автоматических туннелей поверх IPv4 для реализации коннективности IPv6 для тех абонентов, которые не имеют выделенного IPv6 соединения.
- Создание системы информационных ресурсов на новой инфраструктуре, обеспечение функционирования серверов WWW, передачи файлов, электронной почты, службы сетевых имен, чтобы пользователи IPv6 сети не только обращались к внешним информационным ресурсам, но и активно использовали и развивали свои.

В первую очередь предполагается решить следующие задачи: обновление центральных DNS серверов RUNNet; проведение модернизации аппаратной части основных маршрутизаторов опорной сети (увеличение памяти, установка дополнительных сетевых модулей и пр.), а также внесение изменений в топологию сети.

Необходимо обновить программное обеспечение DNS серверов с тем, чтобы обеспечить поддержку тех типов записей, которые используются при адресации IPv6 устройств. Модернизацию DNS серверов целесообразно совместить с плановой заменой аппаратной платформы серверов. Это позволит сначала полностью сконфигурировать и отладить новые серверы, а потом поочередно переключить нагрузку с прежних серверов на новые. Поскольку центральные DNS серверы RUNNet поддерживают много важных доменов (например, домен общего пользования «EDU.RU»), необходимо сделать все возможное, чтобы обеспечить бесперебойную работу системы в переходный период. В частности, старые DNS сервера будут отключены только тогда, когда мы убедимся, что на них перестали поступать DNS запросы.

Модернизация маршрутизаторов необходима в связи с тем, что версии программного обеспечения, включающие поддержку протокола IPv6, предъявляют повышенные требования к размеру как оперативной, так и флэш-памяти.

С целью снижения риска возникновения нештатных ситуаций на всей инфраструктуре опорной сети предлагается отказаться от одномоментного включения поддержки нового протокола на всех ее маршрутизаторах. Поэтому важной задачей подготовительного этапа является осуществление выбора группы маршрутизаторов для включения на них поддержки IPv6, с последующим поэтапным расширением этой группы по мере устранения ошибок. В то же время, для целей проекта необходимо по возможности избежать применения туннелей для передачи IPv6 трафика: по опорной сети IPv6 пакеты должны передаваться в своем исходном виде, как «native IPv6». Чтобы обеспечить возможность непосредственного соединения между собой IPv6 маршру-

тизаторов первой очереди, следует частично изменить топологию опорной сети: добавить новые каналы связи и изменить точки подключения некоторых из работающих каналов. После выполнения перечисленных выше мероприятий предполагается осуществить переход к следующему этапу.

На основе сетей RUNNet, FREEnet и RBNet было рассмотрено решение таких вопросов, как работа системы имен в Интернет «DNS lookups over an IPv6 transport», создание туннелей «Manual IPv6 tunnels using GRE», удаленный мониторинг сети с «Static routes IPv6». Принимая во внимание тот факт, что IPv6 во многом повторяет IPv4, был сделан вывод, что при наличии у сети подключения к IPv6 в большинстве случаев следующие подключения к IPv6 могут осуществляться через то же самое соединение. Если маршрутизатор подключаемой сети отделен от точки подключения транзитными сетями, то соединение с одним из маршрутизаторов может быть произведено с помощью туннеля через так называемый «туннельный» маршрутизатор.

Проведена подготовка экспериментальных межсетевых соединений по протоколу IPv6 между научно-образовательными сетями и определен порядок практической реализации подключения. Осуществлен подбор принципов и методов организации доступа пользователей на базе протокола IPv6. Рассмотрены варианты конфигурации: помимо автоконфигурации, IPv6-узел также может быть сконфигурирован вручную, с помощью DHCP сервера или в результате комбинации всех трех способов.

Общим шагом при любом варианте подключения является назначение клиенту IPv6 префикса. В зависимости от того, что представляет собой клиент, выбирается длина префикса (части адреса, начиная со старших разрядов, которая не может быть изменена клиентом). Хотя провайдер может назначать префиксы любой длины в пределах той части, которую он имеет право менять (согласно префиксу, полученному им от Интернет-регистратора), предлагается придерживаться следующих рекомендаций:

- если префикс назначается сети, то его длина – /48;
- если префикс назначается линку, то его длина – /64;
- если префикс назначается линку типа «точка-точка» (point-to-point), то его длина – /127.

В ходе выполнения проекта были отработаны принципы маршрутизации IPv6 в опорной инфраструктуре крупной межрегиональной сети. Особое внимание было уделено исследованию стабильности работы специализированных программных и аппаратных средств, поддерживающих стек протокола IPv6, возможностям и эффективности работы соответствующих реализаций протоколов маршрутизации на уровне магистральных маршрутизаторов сети.

Пример таблицы маршрутизации на роутере в Москве на опорной сети:

```
msk-gw#sh ipv6 route connected
```

```
IPv6 Routing Table - 624 entries
```

```
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
```

```
U - Per-user Static route
```

```
I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
```

```
O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
```

```
C 2001:B08:0:1::/64 [0/0]
```

```
via ::, Vlan199
```

```
C 2001:B08:0:2::/64 [0/0]
```

```
via ::, Vlan203
```

```
LC 2001:B08:B08::1/128 [0/0]
```

```
via ::, Loopback0
```

```
C 2001:B08:B08:1::/64 [0/0]
```

```
via ::, Vlan13
```

```
msk-gw#
```

Организованы экспериментальные межсетевые соединения по протоколу IPv6 между научно-образовательными сетями RUNNet, FREENet и RBNet. Предлагаемые технические решения учитывают организационно-технические особенности научных и образовательных сетей. Особое внимание уделялось мерам, направленным на снижение затрат на реализацию и сопровождение инфраструктурных элементов, поддерживающих стек протоколов IPv6.

Трафик в NORDUnet по IPv6 протоколу приведен на рис. 3. Трафик очень незначителен. Это объясняется тем, что IPv6 протокол в научно-образовательных сетях используется в экспериментальном режиме [6].

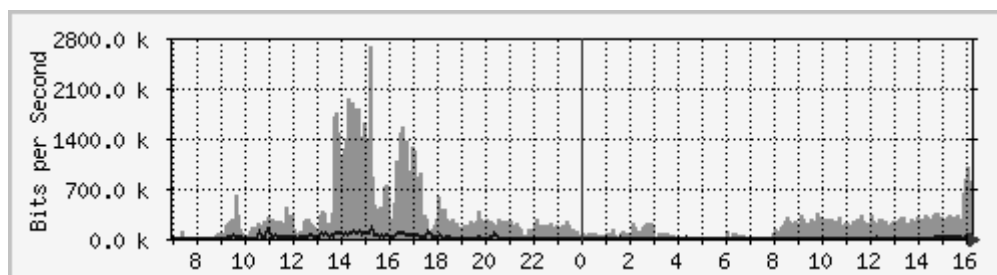


Рис. 3. График объема IPv6 трафика RUNNet-NORDUnet

Сформирован контент сетевого информационно-методического ресурса (<http://ipv6.runnet.ru>), содержащий отечественные и зарубежные материалы по теории и практике использования протоколов нового поколения, на базе «Российского IPv6 форума» [7] для научно-методического обеспечения работ по проекту и популяризации полученных результатов.

Литература

1. RIPE (Réseaux IP Européens) Network Coordination Centre (RIPE NCC) / <http://www.ripe.net> / Technical documentation, 2006.
2. Cisco Systems, IP Version 6 (IPv6) / http://www.cisco.com/en/US/tech/tk872/tsd_technology_support_protocol_home.html / Technical documentation, 2006.
3. Васильев В.Н., Гугель Ю.В., Ижванов Ю.Л., Тихонов А.Н., Хоружников С.Э. Федеральная научно-образовательная сеть RUNNet. Состояние и перспективы развития. // Тезисы докладов XI Всероссийской научно-методической конференции «Телематика'2004». СПб, 2004.
4. Васильев В.Н., Гугель Ю.В., Иванников А.Д., Ижванов Ю.Л., Тихонов А.Н., Хоружников С.Э. Сеть RUNNet: Состояние и перспективы развития. // Тезисы докладов Десятой юбилейной конференции представителей региональных научно-образовательных сетей RELARN-2003. М., 2003.
5. Васильев В.Н., Гугель Ю.В., Ижванов Ю.Л., Куракин Д.В. Разработка теоретических основ и реализация методов построения компьютерных сетей на базе IP-протоколов нового поколения. // Тезисы докладов XII Всероссийской научно-методической конференции «Телематика'2005». СПб, 2005.
6. Герасимов В.В., Гугель Ю.В., Курмышев Н.В., Сигалов А.В. Система образовательных порталов России: анализ телекоммуникационной инфраструктуры, общие требования к аппаратным платформам, технические аспекты размещения // Сб. статей «Образовательные порталы России». Вып. 1. / Научн. ред. В.В. Радаев. М.: Техно-печать, 2004. С. 25–129.
7. Российский IPv6 форум / <http://www.ipv6.ru/>.

КОНЦЕПЦИЯ СЕТЕВОЙ ИНФОРМАЦИОННО-ОБРАЗОВАТЕЛЬНОЙ ТЕХНОЛОГИИ ДЛЯ РАЗРАБОТКИ РЕЗУЛЬТАТОВ ОБРАЗОВАНИЯ

В.Н. Васильев, Л.С. Лисицына

В статье изложен сетевой подход к решению основной проблемы Болонских реформ – разработке результатов образования профессиональным сообществом. Изложена концепция сетевой информационно-образовательной технологии разработки результатов образования, в основе которой лежит направленность на совершенствование системы образования и повышение качества подготовки специалистов. Технология реализована в сетевой среде многопользовательской информационной системы – в методическом Интернет-центре (МИЦ). Рассмотрены особенности предлагаемой технологии и связанные с ними задачи моделирования результатов образования, синтеза ожидаемого результата образования с минимальными затратами на подготовку специалистов, разработки (отбора) содержания образования. Рассмотрена реализация сетевого взаимодействия в МИЦ.

Введение

Сегодня Россия становится все более активной участницей Болонского процесса, о чем свидетельствует большое количество конференций, круглых столов, научных исследований и публикаций, посвященных проблемам Болонских реформ. Основная цель Болонских реформ заключается в формировании «систем сравнительных и сопоставимых квалификаций высшего образования, в которых квалификации описывались бы в терминах *учебной нагрузки, уровня, результатов образования, компетенций и профиля*», а также в создании единого образовательного и исследовательского пространства Европы [1]. Профессор В.И. Байденко делает весьма важный вывод о том, что компетентностный подход к разработке результатов образования призван, в первую очередь, сделать высшее образование более профессиональным за счет перехода от квалификационных моделей специалистов к компетентностным [2].

Наши исследования подтверждают вывод о том, что компетенции и результаты образования, а, следовательно, и компетентностная модель специалиста не зависят от содержания образования. Результаты образования сами определяют содержание образования, являясь языком требований к его разработке. Однако многие уважаемые и авторитетные специалисты российского образования и сегодня пытаются разрабатывать компетентностные модели специалистов, опираясь не на компетенции, а на содержание образования, по сути дела, на ЗУНы (знания–умения–навыки). И сегодня еще существует противоречие между стереотипами понимания «как строить модель специалиста» и требованиями компетентностного подхода к образованию, между желанием сделать более привлекательными образовательные услуги высшей школы и ограниченностью возможностей квалификационного (содержательного) подхода. Это объясняет упорное неприятие компетентностного подхода в профессорско-преподавательской среде вузов. Тем не менее, нам придется научиться общаться и работать, оперируя понятиями компетенций и результатов образования. И это не только понимание объективности реформ Болонского процесса. Это – и требование к российским вузам общеевропейской переводной и накопительной кредитной системы ECTS.

Конечно, с точки зрения рядового преподавателя работа по дисциплине, программа обучения по которой не меняется годами и дается для всех студентов независимо от их профиля, не выглядит очень творческой, но при этом не надо ломать голову над тем, что из того, что Вы преподаете, понадобится в дальнейшей профессиональной деятельности Вашего студента. Безусловно, творческая составляющая в работе любого преподавателя всегда есть; она касается совершенствования учебно-методической работы, разработки новых методик и форм обучения, модернизации содержания образования в связи с новыми достижениями и открытиями в изучаемой предметной области.

Переход от квалификационных моделей специалистов к компетентным заставит каждого преподавателя менять содержание образования «на своем участке» в зависимости от ожидаемого результата образования, т.е. от того, что хочет знать, что хочет уметь делать и кем хочет стать Ваш студент после обучения. Именно поэтому одной из десяти линий действия Болонского процесса является «роль студенчества в осуществлении Болонских реформ». В этой связи «переход к студентоцентрированной парадигме» является одним из важнейших философских аспектов Болонского процесса [2] и требует всестороннего осмысления.

Идентификаторы результатов образования

Ключевыми понятиями Болонских реформ являются понятия компетенций и результата образования. Хотя четкого определения результаты образования до сих пор не имеют, ясно, что именно с ними связаны ожидания обучаемого – что он будет знать, понимать и уметь делать в конце обучения. Эти ожидания невозможно сформулировать ЗУНами. ЗУНы, которые были основным «строительным материалом» для квалификационных моделей специалистов, являются языком преподавателя, языком дисциплин образовательных программ.

Ожидания обучаемого можно выразить только через освоенные им компетенции. При этом результат образования не является простой суммой освоенных компетенций. Понятие результата образования гораздо шире: в нем должен быть итог синтеза (соединения) на практике освоенных ранее компетенций – различных по характеру, но взаимосвязанных целями и задачами ожидаемого результата образования. Вместе с тем, это и подход «и для интеграции академического и профессионального образования, и для оценки предшествующего опыта, и для непрерывного образования и развития структур квалификации в течение всей жизни, и для системы переноса и накопления кредитов» [2].

Поэтому **компетенции** являются важнейшими идентификаторами результатов образования, а результаты образования для них являются языком описания. В этом и состоит диалектическая связь результатов образования и их компетенций [2]. Компетенции обладают способностью к детализации. Детализация компетенций устанавливает в образовательном пространстве элементарные компетенции [3], у которых существуют сопоставимые состояния компетентности, отличающиеся требованиями к уровням знаний и умений обучаемого. Поэтому следующую группу идентификаторов результатов образования представляют уровни **состояний компетентности** обучаемого, например: вводный, базовый и углубленный. Расслоение результатов образования по **уровням и профилям образования** определяет еще две соответствующие группы идентификаторов результатов образования.

Связь результатов и содержания образования

Компетентностный подход к образованию принципиальным образом меняет структурирование образовательного пространства. Теперь не содержание, а результаты образования и их идентификаторы должны определять структуру и состав модульного учебно-методического комплекса (УМК), а идентификаторы ожидаемого результата образования, в свою очередь, должны обеспечивать отбор модулей УМК для развертывания целостного образовательного процесса. Среди идентификаторов результатов образования компетенции имеют определяющее значение. Компетенции не зависят от содержания образования, но не стоит примитивно понимать эту независимость. Компетенции являются итогом отображения содержания образования в пространство компетентностей обучаемых. Каждое состояние компетентности описывается требованиями к

знаниям и умениям обучаемого, что может служить для подбора такого содержания, которое может обеспечить ожидаемый результат образования. Элементарные компетенции и их состояния компетентности находятся ближе других идентификаторов образования к содержанию образования; они описывают результаты образования, дальнейшая детализация которых влечет за собой уже конкретику содержания образования: модели, методы, методики, технологии и т.п. Суть компетентностного подхода и состоит в отделении избыточного содержания образования от того образовательного эффекта, который выражается в компетентности обучаемого и результатах образования. Это отделение должно привести к построению такой надстройки, которая позволит проектировать различные результаты образования и описывать их на языке, понятном всему профессиональному сообществу, а не только узкому кругу преподавателей, занимающихся преподаванием близких по содержанию предметов.

Декомпозиция образовательного пространства на *элементарные результаты образования* с последующим установлением причинно-следственных связей между их состояниями компетентности описывают гиперграфовую модель результатов образования в виде план-графа [3]. План-граф является не только моделью для разработки различных результатов образования, но и основой разработки содержания образования: каждое гиперребро план-графа определяет входную и выходную спецификацию для отдельной учебной программы. Модульная коллекция учебных программ – самая подвижная и изменчивая часть УМК, так как ее состав и содержание ближе всего находится к результатам образования и их компетенциям. Учебные программы могут быть вариативными, т.е. содержать альтернативные способы достижения данного состояния компетентности, например, различные методы, методики, модели и т.п. Остальные компоненты УМК менее подвижны и изменчивы, поэтому модульный подход к разработке тестовых систем, виртуальных лабораторий и т.п. способствует скорейшему внедрению компетентностного подхода к образованию.

Таким образом, при компетентностном подходе необходимо отделить избыточное содержание образования от того образовательного результата, который может быть достигнут с его помощью. Результатом такого отделения и станет надстройка над содержанием в виде модели результатов образования, которая позволит не только планировать различные результаты образования и подбирать для них необходимое и достаточное содержание, но и исследовать образовательное пространство на предмет доработки и совершенствования содержания образования.

Первый опыт практического применения концепции

В СПбГУ ИТМО научные и практические результаты разработки и применения компетентностного подхода впервые были получены в сфере дополнительного профессионального образования работников образования в области информационно-коммуникационных технологий (ИКТ). Они стали итогом наблюдений и обобщений многолетней практической работы (за период 2000–2005 гг. обучено 8614 человек), а также ряда научно-исследовательских работ по Федеральной целевой программе «Электронная Россия (2002–2010 годы)» по направлению «Развитие системы подготовки специалистов по ИКТ и квалифицированных пользователей»). Разнообразие и повторяемость учебно-методических планов, краткосрочность и интенсивность обучения позволили выявить ряд закономерностей, характерных в целом для образовательной деятельности. Важным подтверждением эффективности предложенного компетентностного подхода является создание и внедрение сетевой многопользовательской информационной системы (ИС) – методического Интернет-центра (МИЦ) <http://itocenter.ifmo.ru> [3] в практику разработки результатов дополнительного образования в сетевом сообществе отечественных образовательных учреждений. Главное направле-

ние работ в сообществе – планирование и методическая подготовка компетентностного повышения квалификации в области ИКТ для различных категорий работников образования: учителей, директоров школ и их заместителей, методистов, школьных IT-специалистов и т.д.

Сетевое профессиональное сообщество

Нами разработана *концепция сетевой информационно-образовательной технологии*, в которой главными инструментами разработки результатов образования являются инструменты сетевой среды МИЦ. В основе концепции – совместная работа и общение профессионального сообщества, объединенного средой сетевой многопользовательской ИС, для разработки результатов образования. Сетевой характер образовательной технологии обусловлен тем, что в разработке результата образования принимает участие большой коллектив преподавателей самых различных кафедр, а подвижность и изменчивость компетенций требует постоянного взаимодействия всех участников разработки результата образования.

Под каждый конкретный результат образования МИЦ создает два типа виртуальных центров (рис. 1): центр разработки (ЦР) и центры поддержки (ЦП). Роль ЦР выполняет выпускающая кафедра – модератор разработки, своеобразный агент рынка труда в вузе. В роли ЦП в сообществе МИЦ выступает любая другая кафедра вуза, прямо или опосредованно (через другие кафедры) участвующая в разработке данного результата образования. Заметим, что среда МИЦ позволяет одной и той же кафедре в зависимости от ожидаемого результата образования выступать как в роли центра разработки, так и в роли центра поддержки.

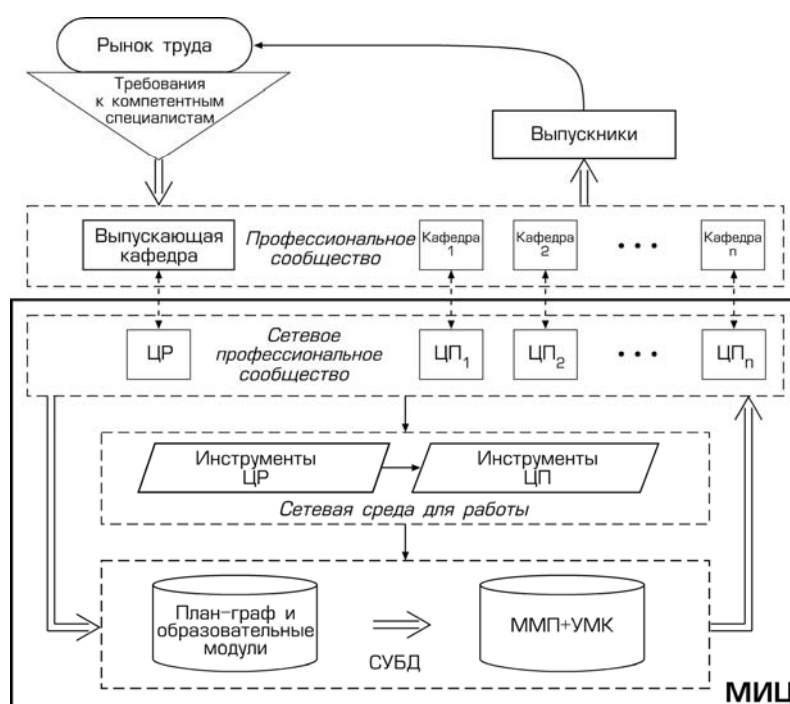


Рис. 1. Взаимодействие в сетевом сообществе МИЦ

Сетевое решение проблемы разработки результата образования имеет еще одно неоспоримое преимущество – удобство тиражирования опыта подготовки компетентных специалистов среди других вузов. Организация доступа к сетевой среде МИЦ для других вузов не только делает общим достоянием системы образования опыт разработки результатов образования, но и позволит кафедрам различных вузов стать актив-

ными участниками сетевого профессионального сообщества в МИЦ как в роли центров разработки, так и в роли центров поддержки.

Реализация взаимодействия в сетевой среде МИЦ

Сетевая среда МИЦ выполнена на основе базовых интернет-технологий. На рис. 2 показана структурная схема среды МИЦ, которая включает в себя Web-обозреватель (клиент), Web-сервер, Java-сервлет, Java-приложение и систему управления базами данных (СУБД).

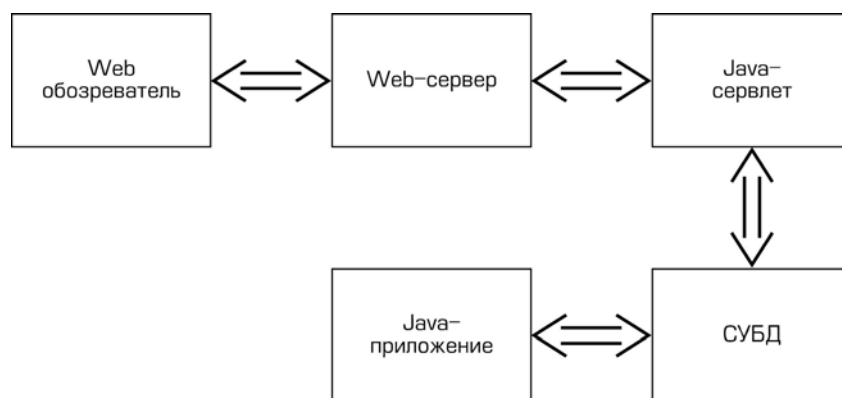


Рис. 2. Структурная схема сетевой среды МИЦ

Доступ к МИЦ осуществляется через Интернет посредством стандартного Web-обозревателя, основными функциями которого являются:

- визуализация пользовательских интерфейсов, описанных в соответствии со спецификацией языка HTML;
- формирование запросов и обработка ответов Web-сервера.

Web-сервер обеспечивает взаимодействие по протоколу HTTP и интерфейс с Java-сервлетом. В качестве Web-сервера может выступать любой программный продукт данного класса с поддержкой Java-сервлетов. Java-сервлет является оригинальной программой, специально разработанной для МИЦ. Java-сервлет выполняет роль вентили между клиентом и БД. В задачи Java-сервлета входят:

- обработка HTTP-запросов пользователя и формирование соответствующих ответов;
- ведение пользовательских сессий;
- взаимодействие с интегрированной БД;
- кэширование элементов системы;
- выполнение команд БД.

В хранилище БД (рис. 1) находятся следующие данные:

- данные о пользователях, их правах;
- модель иерархии результатов образования образовательного пространства профессионального сообщества [3];
- модель результатов образования в виде план-графа [3];
- содержание образования в виде образовательных модули;
- разработанные и апробированные результаты образования в виде минимальных модульных планах (ММП) [4], сведения о центре разработки;

- учебно-методические комплексы (УМК) разработанных результатов образования для повторного развертывания целостного образовательного процесса по подготовке специалистов.

Заметим, что отдельные модули УМК и их компоненты могут находиться за пределами МИЦ, например, тесты – в системе дистанционного обучения, практикум – в виртуальной лаборатории и т.п. Кроме того, на стороне БД хранятся процедуры, реализующие логику работы системы, XSL-шаблоны, определяющие пользовательский интерфейс системы. Взаимодействие Java-сервлета с БД осуществляется посредством хранимых процедур. Каждый запрос клиента обязательно должен содержать параметр с названием хранимой процедуры.

Сетевая информационно-образовательная технология

На рис. 3 схематично представлена предлагаемая нами технология: процессы разработки результатов образования выделены прямоугольниками, а ее состояния – кругами. Важными задачами разработки результатов образования в ней являются задачи моделирования результатов образования, синтеза минимального модульного плана (ММП) и разработки (отбора) содержания образования. Технология направлена на совершенствование системы образования и повышение качества подготовки специалистов. Поясним в этой связи наиболее важные моменты данной технологии.

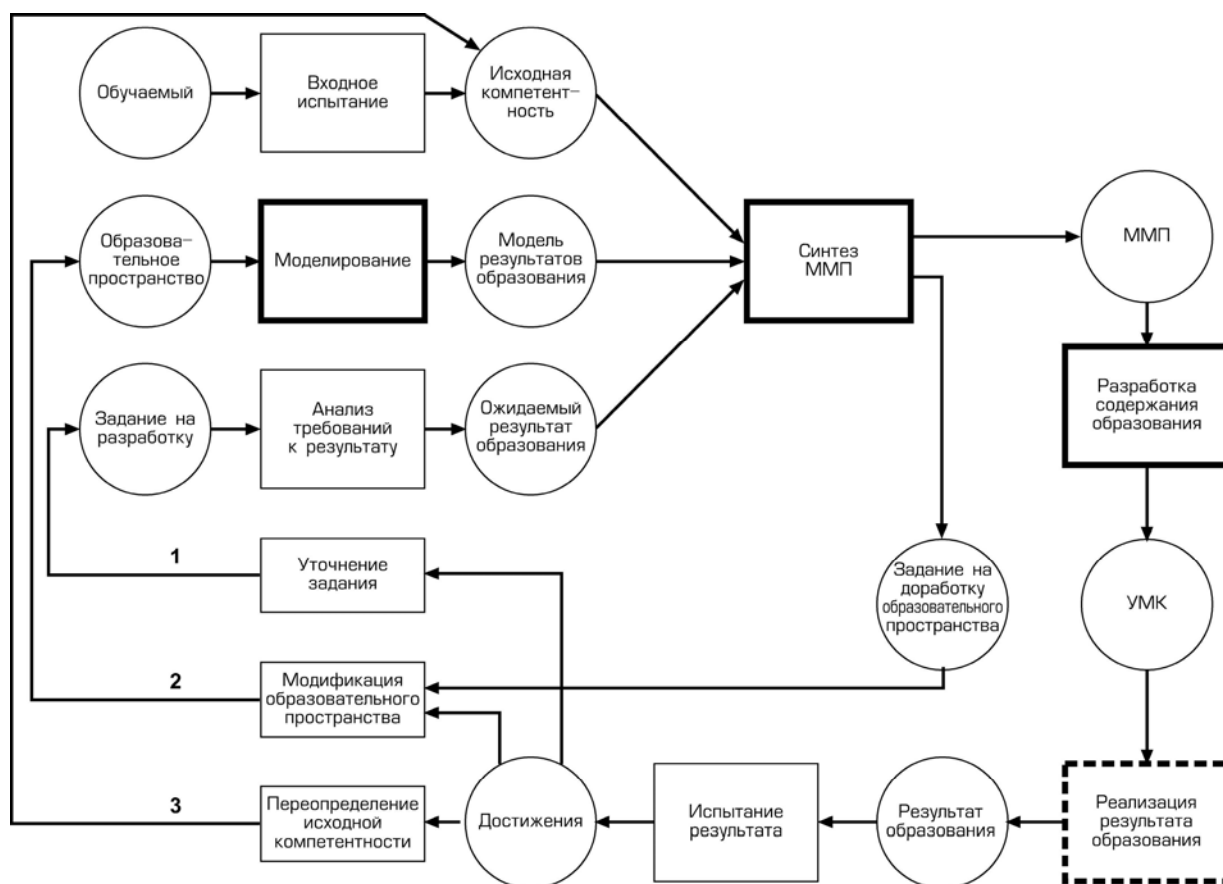


Рис. 3. Технология разработки результата образования

Ожидаемый результат образования [4] описывается с помощью идентификаторов, которые устанавливаются на основе анализа требований рынка труда к компетентности специалистов, а также на основе ожиданий заказчика, в роли которого может и должен выступать сам обучаемый – студент. Построение и систематическая актуализация (об-

ратная связь 2) модели результатов образования [3] описывает весь спектр образовательных возможностей профессионального сообщества с целью:

- автоматизации синтеза ожидаемых результатов образования с минимизацией перечня состояний компетентности, необходимых и достаточных для достижения целевой компетентности обучаемого из исходной;
- структурирования содержания образования;
- планирования целостного образовательного процесса для подготовки компетентных специалистов.

Ядром технологии является синтез ММП. ММП представляет собой минимизированный упорядоченный перечень состояний компетентности для получения ожидаемого результата образования. Исходными данными для этой задачи являются: план-граф, исходная компетентность обучаемого и идентификаторы ожидаемого результата образования. Отсутствие в план-графе необходимых для ожидаемого результата образования идентификаторов указывает на необходимость доработки образовательного пространства, см. обратную связь 2. Если была получена компетентностная модель специалиста в виде ММП, то на следующем этапе разработки ожидаемого результата образования производится отбор модульного УМК. Каждый элемент перечня в ММП неразрывно связан с соответствующей учебной программой модульного УМК – ключом для отбора соответствующего содержания образования.

Реализация в целостном образовательном процессе отобранного УМК формирует целевую компетентность у обучаемого, которая подтверждается испытанием ожидаемого результата образования. В ходе испытаний могут быть выявлены новые достижения образования (новые идентификаторы результатов образования и связанные с ними новые модели, методы, технологии, материалы и т.п., а также новые методики и формы обучения и аттестаций), которые возвращаются в систему образования и расширяют образовательное пространство всего сообщества МИЦ. Испытания результатов образования должны быть связаны с практической работой в профессиональной области специалиста (работа или проект по тематике ожидаемого результата, производственная практика и т.п.).

Контроль качества образования

Компетентностный подход к образованию делает прозрачным и понятным всему профессиональному сообществу процесс контроля качества подготовки специалистов. Теперь контролировать надо не содержание образования, а результаты образования. Требования к качеству подготовки специалистов, выраженные на языке результатов образования, должны подтверждаться итогами испытаний ожидаемого результата образования. Несомненным фактом, подтверждающим качество подготовки специалистов, является наличие новых достижений образования, установленных в ходе испытаний ожидаемого результата образования. Промежуточный контроль знаний и умений также должен быть спланирован на основании компетентностной модели специалиста в виде ММП. Перечень состояний компетентности в ММП, описанных на языке результатов образования, легко поддается формализации для промежуточных аттестаций знаний и умений. При этом такой контроль не зависит, в известном смысле, от содержания образования и может быть унифицирован для определенной выборки контролируемых состояний компетентности. Последнее обстоятельство создает предпосылки для разработки эффективных методов анализа качества подготовки специалистов с целью тиражирования передового педагогического опыта.

Основы функционирования профессионального сообщества МИЦ

Профессиональное сообщество МИЦ работает и взаимодействует в сетевой среде МИЦ для того, чтобы:

1. описывать свои образовательные возможности на языке результатов образования и заявлять потенциальным партнерам о своем образовательном потенциале;
2. определять стратегию доработки содержания образования под планируемые (перспективные) результаты образования;
3. опираться на накопленный потенциал всего сообщества при разработке различных результатов и содержания образования.

Описанием своих образовательных возможностей кафедры сообщества занимаются инициативно, регистрируя у администратора МИЦ новые компетенции, детализируя их и описывая соответствующие состояния компетентности для различных уровней и профилей образования. Сетевая среда МИЦ при этом отслеживает уникальность идентификаторов результатов образования, сверяя с уже имеющимися описаниями. Основная цель таких инициатив – сделать прозрачным и понятным образовательный потенциал своей кафедры для любого потенциального партнера и потребителя образовательных услуг. Другой, не менее важной инициативой кафедры является создание «методической копилки» с содержанием образования, структурированного под состояния компетентности план-графа. Разрабатывая для каждого состояния компетентности результатов образования свою модульную учебную программу и устанавливая ее связь с различными компонентами модульного УМК (слайд-фильмы для лекций, электронные пособия для лабораторных и практических занятий, модули систем дистанционного обучения и виртуальных лабораторий и т. п.), кафедра создает основу для *автоматизации синтеза и тиражирования модульного УМК*. Выступая в сообществе в роли центра разработки, кафедра, опираясь на потенциал всего сообщества, примеряет ожидаемый результат образования к образовательным возможностям профессионального сообщества. Если при синтезе ММП будут выявлены недостающие идентификаторы ожидаемого результата образования, то центр разработки публикует в МИЦ сообщение с приглашением другим участникам сообщества разработать недостающие компетенции.

Литература

1. Болонский процесс: середина пути/ Под ред. д-ра пед. наук, профессора В.И. Байденко. М.: Исследовательский центр проблем качества подготовки специалистов. Российский Новый Университет, 2005. 379с.
2. Васильев В.Н., Лисицына Л.С., Лямин А.В. Методический интернет-центр. СПб: Питер, 2005. 96 с.: ил. (опубликована на сайте www.ito-center.ifmo.ru)
3. Лисицына Л.С. Учимся строить модель результатов образования. // Научно-технический вестник СПбГУ ИТМО. 2005. № 22.
4. Лисицына Л.С. Планирование ожидаемого результата образования // Научно-технический вестник СПбГУ ИТМО. 2005. № 22.

МЕТОДЫ И СРЕДСТВА ЗАЩИТЫ ИНФОРМАЦИИ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ ЭЛЕКТРОННОГО ОБМЕНА СТАНДАРТИЗОВАННЫМИ МЕДИЦИНСКИМИ ДОКУМЕНТАМИ*

В.В. Кириллов, Г.Ю. Громов

В статье приведены результаты исследования средств организации защиты информации в автоматизированной системе электронного обмена стандартизованными медицинскими документами. В качестве инструментальной среды системы используются Microsoft Office 2003 и средство криптографической защиты информации компании КриптоПро. Рассматриваются методы их взаимодействия и результаты использования медицинскими работниками.

Введение

Технические требования к автоматизированной системе электронного обмена стандартизованными медицинскими документами (АС ОСМД) предлагают использовать конкретный инструментарий для визуализации и доступа авторизованных пользователей системы к стандартизованным медицинским документам (СМД) – это проводник Интернета Microsoft Internet Explorer 6.0 SP1 и пакет офисной автоматизации Microsoft Office 2003 Pro Rus SP2, включая InfoPath 2003. Там же указывается, что должна быть предусмотрена возможность электронной подписи СМД (одним или несколькими лицами) и их шифрование при передаче по незащищенным каналам связи. При этом следует применять сертифицированные в России программные средства криптографической защиты информации.

Исследование предлагаемых средств подготовки, пересылки и просмотра СМД показало, что в них предусмотрены механизмы, позволяющие пользователям подписать СМД или его части, а также зашифровать пересылаемый СМД (рис. 1). Естественно, что они позволяют расшифровать зашифрованный СМД и проверить наличие цифровой подписи на СВД, а также определить автора подписи и время подписания.

В связи с этим основное внимание было уделено выбору сертифицированных в России программных средств криптографической защиты информации и их использования совместно с Microsoft Office 2003.

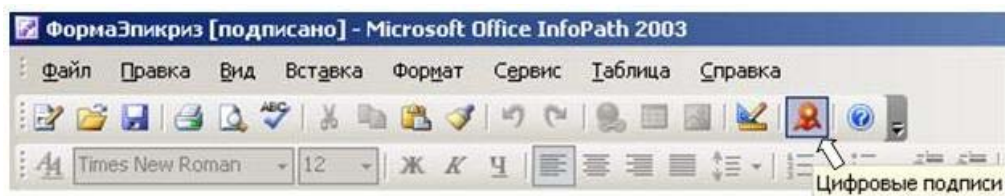
Электронная подпись

Основное назначение электронной подписи документа – обеспечение аутентичности как самого документа, так его автора, т.е. доказательство, что документ действительно заверен отправителем и что содержимое этого документа не было изменено в процессе пересылки по каналам связи. Технология электронных подписей является основой для безопасного функционирования систем документооборота. Эта подпись отличается от обычной подписи на бумажном документе тем, что она не только подтверждает личность отправителя документа, но также и гарантирует, что его содержание не модифицировалось по пути от отправителя к адресату.

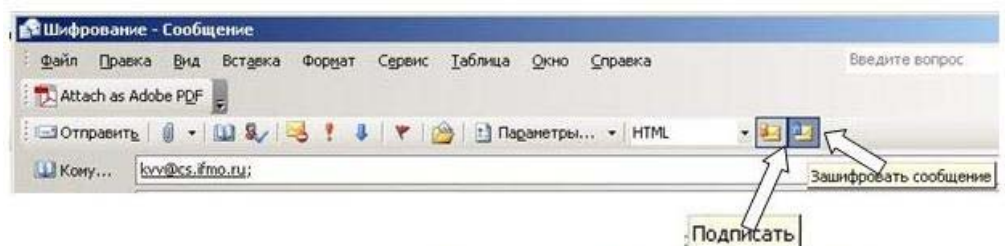
Электронная подпись – это данные в электронной форме, которые содержатся в документе, приложены к нему или логически связаны с ним и которые могут быть использованы для идентификации подписавшего и указания на то, что подписавший согласен с информацией, содержащейся в сообщении [1]. Отправитель формирует ее, обрабатывая исходный документ соответствующей процедурой, называемой функцией

* Работа выполнена в рамках ГК №02.467.11.1007 от "10" октября 2005 года "N2005-ИТ-00.2/002 Разработка системы электронного обмена стандартизованными документами на основе Интернет- и Интранет-порталов для ускорения документооборота в медицинских учреждениях страны"

«Электронной подписи». Внешне электронная подпись выглядит как случайные данные и имеет смысл только при использовании совместно с исходным документом. Получатель этого документа проверяет электронную подпись с помощью другого набора вычислений – так называемой функции «проверки электронной подписи». Результат этой проверки должен подтвердить аутентичность как самого отправителя, так и полученного от него документа.



Кнопка для инициирования подписания СМД или его части в InfoPath 2003



Кнопки для подписания и шифрования СМД в Microsoft Outlook 2003

Рис. 1. Средства шифрования и/или подписания СМД

Надо отметить, что существует два почти синонимических понятия: «электронная подпись» и «электронная цифровая подпись», которые на самом деле имеют существенное различие [2]. Большинство мировых нормативных актов в этой области, например, Типовой закон Комиссии ООН по праву в международной торговле ЮНСИТРАЛ оперируют понятием «электронная подпись». Оно объединяет целый спектр электронных технологий: это не только математические алгоритмы, реализованные в цифре на компьютерах, но и такие биометрические, как сканирование отпечатков пальцев, сетчатки глаза, спектральных характеристик голоса, а в перспективе и молекул ДНК. В подлинности электронного сообщения можно убедиться, используя также специальные ПИН-коды, смарт-карты, маркеры, принадлежащие подписывающему лицу. Такое расширенное толкование понятно, оправдано и дальновидно, ведь сейчас невозможно надежно предсказать, какая именно технология электронной подписи будет доминирующей. Современные электронные и программные средства развиваются так быстро, что делать ставку на какой-то один вариант электронной подписи крайне рискованно. Быстрый прогресс может внезапно свести на нет длительные законодательные усилия и большие затраты на внедрение, тем более, если внедрение будет тотально-принудительным. Однако в Федеральном законе Российской Федерации от 10 января 2002 г. N 1-ФЗ «Об электронной цифровой подписи» она жестко привязана к единственной криптографической технологии, создаваемой методом шифрования с помощью открытых ключей.

Сама по себе эта методика сейчас является одной из самых развитых, и у нее есть серьезные достоинства. Суть шифрования с открытым ключом, частью которого является электронная цифровая подпись, такова. Участники переписки должны иметь одинаковый инструмент шифрования-дешифрования. Обычно это компьютер, снабженный соответствующей программой. В начале программа автоматически создает два ключа, применяющиеся исключительно в паре. Открытый ключ можно послать по почте по-

тенциальным партнерам, разместить в Интернете, отправить в местное отделение ФАПСИ.

Вы подписываете послание с помощью закрытого секретного ключа, а проверить авторство документа может любой, кому вы пошлете или кто не поленится скачать из базы данных в Интернете ваш открытый ключ.

Используя свой закрытый ключ и открытый ключ партнера, вы можете создать ключ шифрования и зашифровывать послание на адрес этого партнера. Для прочтения такого сообщения партнер должен создать ключ расшифровки из своего закрытого и вашего открытого ключей.

Степень защиты цифровой подписи от подделки так же высока, как степень стойкости при шифровании (стойкость оценивается объемами вычислений, которые требуются для ее вскрытия – считается, что ключ шифрования достаточно стоек, если все известные способы его отыскать настолько сложны, что требуют больше времени, чем простой перебор всех возможных ключей). Важно, что можно подписывать сообщение как шифруя, так и не шифруя его – оставляя открытым. Математические алгоритмы шифрования с открытым ключом при длине ключа в несколько килобит делают при нынешних знаниях задачу взлома практически неразрешимой. Такие алгоритмы реализованы практически всеми современными системами.

Итак, схема цифровой подписи или электронно-цифровой подписи – это набор алгоритмов и протоколов, позволяющих построить информационное взаимодействие между двумя и более участниками таким образом, чтобы факт авторства переданного массива данных, «подписанного» одним из участников, мог быть надежно подтвержден или опровергнут третьей стороной – «независимым арбитражем». Любая схема цифровой подписи предполагает добавление к «подписываемому» массиву данных дополнительного кода – собственно «цифровой подписи», выработать который может только автор сообщения, обладающий секретным (закрытым) ключом подписи, а все остальные могут лишь проверить соответствие этой «подписи» подписанным данным. Поэтому каждая схема должна предусмотреть, как минимум, определение трех следующих алгоритмов:

- выработки пары ключей – подписи и проверки подписи с использованием вектора случайных параметров;
- подписи сообщения с использованием ключа подписи;
- проверки подписи с использованием ключа проверки подписи, выдающего в качестве результата булево значение – подтверждается или не подтверждается авторство сообщения.

Цифровые сертификаты

В соответствии с Федеральным законом об электронной цифровой подписи организации, участвующие в документообороте с Федеральными органами государственной власти, органами государственной власти субъектов Российской Федерации и органами местного самоуправления, должны использовать для организации процедуры подписывания только сертифицированное программное обеспечение и иметь лицензию ФСБ России и Государственной Технической комиссии при президенте Российской Федерации.

К такому программному обеспечению относится криптопровайдер «Крипто-Про CSP» и удостоверяющий центр «КриптоПро УЦ». С его помощью учреждение может организовать в своем составе локальный центр сертификации, позволяющий обеспечить пользователей средствами и спецификациями для использования сертификатов открытых ключей в целях:

- контроля целостности электронных документов, передаваемых в автоматизированных информационных системах;
- контроля целостности публичных информационных ресурсов;
- проверки подлинности взаимодействующих программных компонентов и конфиденциальности передаваемых данных при информационном взаимодействии;
- создания системы юридически значимой электронной цифровой подписи в системах электронного документооборота;
- обеспечения безопасности и разграничения доступа при взаимодействии субъектов автоматизированных информационных систем;
- создания иерархической системы управления ключами подписи субъектов автоматизированных информационных систем.

Подчеркнем, что самостоятельно созданный в Microsoft InfoPath цифровой сертификат не является сертификатом, выданным официальным центром сертификации, поэтому все формы, подписанные с использованием такого сертификата, рассматриваются как формы с собственной подписью. Такие сертификаты считаются непроверенными и могут быть причиной появления предупреждения системы безопасности, если для формы задан высокий или средний уровень безопасности. Microsoft InfoPath доверяет сертификатам с собственной подписью только в том случае, если сертификат располагается на компьютере, имеющем доступ к закрытому ключу этого сертификата. Как правило, это означает, что Microsoft InfoPath доверяет сертификатам с собственной подписью только на компьютере, на котором создавался сертификат, если только закрытый ключ не предоставлен для общего доступа другим компьютерам.

В лечебном учреждении должна существовать служба криптографической защиты информации, которая, используя возможности собственного локального или другого Центра Сертификации, снабжает пользователей учреждения личными закрытыми и открытыми ключами ЭЦП и шифрования, а также устанавливает на компьютеры пользователей программное обеспечение, необходимое для подписания подготавливаемых ими документов, а также проверки цифровых подписей получаемых документов.

Подписание шаблонов и форм СМД

В автоматизированной системе электронного обмена СМД необходимо иметь в ее составе набор шаблонов для заполнения следующих документов:

- направление на лабораторное исследование;
- направление на диагностическое исследование;
- направление на консультацию;
- результат лабораторного исследования;
- протокол диагностического исследования;
- заключение консультанта;
- рецепт в аптеку;
- выписка из истории болезни (амбулаторной карты) пациента.

Для обеспечения возможности подписания шаблона формы, созданного в InfoPath, необходимо иметь:

- установленное программное обеспечение криптопровайдера КриптоПро CSP, который имеет сертификат ФАПСИ и может использоваться для формирования ключей шифрования и ключей электронной цифровой подписи, шифрования и имитозащиты данных, обеспечения целостности и подлинности информации, не содержащей сведений, составляющих государственную тайну;
- цифровые сертификаты, полученные у сетевого администратора, занимающегося в учреждении вопросами безопасности.

При разработке шаблона можно разрешить цифровые подписи для всей формы или для ее отдельных частей. Если разрешены подписи для части формы, необходимо определить, какие данные в форме должны быть подписаны. Можно дополнительно связать эти данные с разделом формы. После того, как пользователь подписал форму, данные блокируются, так что их изменение невозможно, и сохраняется их представление с цифровой подписью.

Помимо этого можно указать, разрешается ли использование нескольких подписей для одной формы и, если разрешается, то должны ли это быть совместные подписи (когда каждая подпись не зависит от остальных) или же удостоверяющие подписи (когда каждая следующая подпись удостоверяет как саму форму, так и все предшествующие ей подписи).

Кроме разрешения использования цифровой подписи в формах, благодаря чему пользователи получают возможность подписывать форму, можно подписать цифровой подписью и шаблон формы, что удостоверяет подлинность автора шаблона формы точно таким же образом, как цифровая подпись формы удостоверяет подлинность заполнившего форму.

При заполнении формы можно добавить цифровую подпись ко всей форме или к ее части, в зависимости от настроек в шаблоне формы. Снабжать цифровой подписью можно только те формы и их части, при разработке которых была предусмотрена такая возможность. Можно также удалять или проверять цифровые подписи.

Заключение

Исследование применения Microsoft Office 2003 и средств криптографической защиты информации компании КриптоПро для шифрования и подписания СМД показало, что сотрудники медицинских учреждений, имеющие навыки работы со средствами офисной автоматизации (текстовым редактором, редактором электронных таблиц и т.п.) и стандартными проводниками Интернет, успешно справляются с работой в среде АС ОСМД.

Литература

1. Романец Ю., Тимофеев П., Шаньгин В. Защита информации в компьютерных системах и сетях. М.: Радио и связь, 2001. 376 с.
2. Баричев С., Серов С. Основы современной криптографии. М.: Радио и связь, 2001. 121с.

ПРИМЕНЕНИЕ НЕЙРОННОЙ СЕТИ ДЛЯ УПРАВЛЕНИЯ МОТОЦИКЛОМ

И.М. Аничкин

В статье описана методика применения нейронной сети для построения системы управления равновесием мотоцикла. Данная система, работая в составе более развитой системы управления, включающей в себя также системы получения и анализа данных, позволит мотоциклу беспилотно осуществлять движение по поверхности вдоль заданной траектории. Построенная система обучается на экспериментальных данных, получаемых из опытов по управлению моделью.

Введение

В последнее время весьма актуальна задача компьютерного управления поведением механических или иных объектов. Если рассматривать эти объекты и управляющий ими компьютер как единое целое, то эту задачу можно сформулировать как попытку создания объекта, способного самостоятельно управлять своими действиями. Речь может идти об управлении разной степени сложности – от исполнения простейших команд типа «при нажатии кнопки произвести процедуру автоматического запуска двигателя» до интеллектуального управления, при котором объект может беспилотно (без участия человека) совершать действия, направленные на достижение определенных целей.

Говоря об управлении транспортными средствами, также можно представить себе объекты различной сложности – от системы активной подвески колес автомобиля до полностью автономного транспортного средства, беспилотно передвигающегося, например, на колесах по поверхности земли.

В данной работе предпринята попытка построения одной из частей системы для управления мотоциклом. Эта система позволит ему двигаться по прямой и выполнять повороты, удерживая при этом мотоцикл в равновесии.

Управляемая системой более высокого уровня, данная система может быть использована для реализации движения мотоцикла по произвольной траектории, а в дальнейшем – и для построения автономно управляемого мотоцикла.

Создание системы управления

Порядок разработки системы управления равновесием мотоцикла с применением нейронной сети следующий.

1. Создание среды для физического моделирования поведения мотоцикла (разработка симулятора и создание модели мотоцикла).
2. Исследование поведения мотоцикла с целью выяснения принципов управления им при удержании равновесия.
3. Создание пробной системы управления равновесием мотоцикла, позволяющей ему удерживать равновесие при движении в широком диапазоне скоростей:
 - используется линейное управление;
 - параметры системы управления подбираются эвристически.
4. Анализ поведения построенной системы управления, выявление недостатков такого управления.
5. Построение системы управления моделью мотоцикла на основе нейронной сети:
 - реализация быстрого алгоритма обучения нейронной сети по методу обратного распространения ошибки;
 - разработка методики получения исходных данных для обучения нейронной сети управлению мотоциклом;
 - получение данных для обучения нейронной сети;

- обучение нейронной сети;
- проверка работы системы управления мотоциклом.

Далее в работе приводится более детальное описание приведенной последовательности выполняемых действий

Физическое моделирование

Среда моделирования. Вычислительная модель мотоцикла построена при помощи библиотеки для моделирования механики твердых тел *ODE v. 0.5 (Open Dynamics Engine)* [1]. Данная библиотека имеет открытый исходный код и распространяется свободно. Движок *ODE* разработан для использования в интерактивных симуляторах, а также симуляторах реального времени и хорошо подходит для моделирования шарнирно связанных твердых тел. Встроенная система обнаружения столкновений движка позволяет моделировать взаимодействие тел при их столкновении. В *ODE* также реализовано сухое трение между телами.

Создание модели мотоцикла. В целях упрощения анализа поведения мотоцикла было принято решение сделать модель максимально простой, но так, чтобы сохранялись все основные свойства ее поведения – модель оставалась бы похожей на мотоцикл.

Минимально сложная модель, удовлетворяющая этим требованиям, такова: рама, передняя вилка, поворачивающаяся вокруг наклонной оси, заднее колесо, вращающееся вокруг оси относительно рамы, и переднее колесо, вращающееся вокруг оси относительно передней вилки.

Анализ движения модели

Рассмотрение движения мотоцикла по поверхности при условии нахождения колес мотоцикла на этой поверхности и отсутствии проскальзывания колес позволило сократить число переменных, описывающих движение мотоцикла, до восьми – пяти координат и трех скоростей.

Применительно к задаче удержания равновесия движение мотоцикла описывается всего лишь пятью переменными – двумя координатами и тремя скоростями. Перечислим эти переменные: скорость движения S , угол наклона мотоцикла L , угол поворота руля T , скорость изменения угла наклона $LS = dL/dt$, и скорость изменения угла поворота руля $TS = dT/dt$.

Среди всех вариантов движения мотоцикла можно выделить некоторые особые. Такими будут движения, при которых наклон мотоцикла и угол поворота его руля не изменяются во времени. Такое движение мотоцикла назовем движением по стабильной траектории, или установившимся движением. Этот вариант движения мотоцикла является основным, а все другие движения могут рассматриваться как переходы между различными стабильными траекториями.

Отметим, что стабильная траектория задается двумя параметрами – своей кривизной σ и скоростью движения мотоцикла по траектории S . Другие параметры движения мотоцикла могут быть однозначно определены по указанным двум: $L = L(S, \sigma)$ и $T = T(S, \sigma)$.

Управление мотоциклом

Основная задача данной работы – обеспечение управления равновесием мотоцикла. Решением этой задачи является такой алгоритм управления рулем, который, в зависимости от скорости движения мотоцикла, задаваемой кривизны траектории его движения, текущего наклона и скорости наклона мотоцикла, будет производить управле-

ние рулем («руление») таким образом, чтобы привести мотоцикл на траекторию требуемой кривизны как можно быстрее.

Контрруление. Для удержания мотоцикла в равновесии могут применяться два основных элемента управления: руль и смещение центра масс.

Совместное использование этих элементов при создании системы беспилотного управления (например, если реализовать смещение центра масс посредством перемещения груза), может сделать управление более эффективным по сравнению с управлением только рулем. Однако при этом решаемая задача значительно усложнится.

Поэтому для удержания равновесия будем использовать только повороты руля без перемещения центра масс мотоцикла. Такое управление часто применяется мотоциклистами в случае необходимости выполнить поворот быстро и без приложения дополнительных усилий. Речь идет о приеме, называемом «контрруление» [2], при котором поворот руля в одну сторону, например, вправо, вызывает наклон мотоцикла в противоположную сторону (влево). После этого может быть выполнен левый поворот. Именно этот прием будет использован в настоящей работе при беспилотном управлении мотоциклом.

Управление равновесием. Управление равновесием мотоцикла заключается в том, чтобы в зависимости от переменных, описывающих движение мотоцикла (S, L, T, LS), и заданных параметров движения (кривизны σ) обеспечить переход мотоцикла на траекторию требуемой кривизны. При этом, как уже отмечалось выше, должны выполняться соотношения $L = L(S, \sigma)$ и $T = T(S, \sigma)$. Управляемой переменной в данном случае является скорость поворота руля TS .

Цель настоящей работы – построение оптимального алгоритма управления равновесием мотоцикла. При этом должен быть построен такой алгоритм, который бы позволил мотоциклу осуществлять переход от одной стабильной траектории к другой наиболее быстрым образом (управление, оптимальное по быстродействию).

Аналогия с маятником

Мотоцикл, который едет по поверхности на двух колесах, представляет собой систему, находящуюся в состоянии неустойчивого равновесия. Управление такой системой во многом аналогично управлению простейшей системой, находящейся в состоянии неустойчивого равновесия – перевернутым маятником на тележке (рис. 1). Данная система часто рассматривается (например, в работах [3, 4]) как типичная задача об управлении системами, находящимися в состоянии неустойчивого равновесия – такими системами, которые при отклонении от равновесного состояния начинают уходить от него с течением времени все дальше.

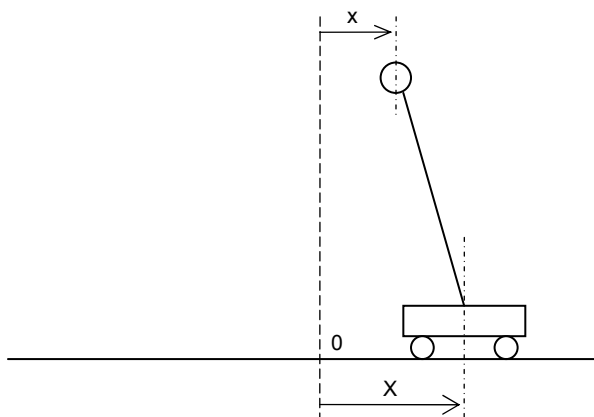


Рис. 1. Перевернутый маятник

В линейном приближении движение маятника описывается формулой

$$x'' = \omega^2 \cdot (x - X) \quad (1)$$

Цель управления маятником – удержание его в равновесии над точкой x_F . Управляющее воздействие – изменение координаты тележки X . Ограничение: $X'_t < X S_{max} = const$.

Предположим, что в момент времени $t = 0$ маятник находится в равновесии над точкой x_0 : $X(0) = x(0) = x_0 < x_F$, и при этом $x'(0) = 0$. Пусть теперь требуется сместить маятник вправо (увеличить значение x). Для этого смещаем тележку влево, так чтобы ускорение было положительным, и скорость x' начала возрастать. Когда скорость возрастет достаточно, необходимо начинать двигать тележку вправо – «догонять» маятник. При этом значения x и X должны сравняться на середине отрезка $[x_0, x_F]$: $x = X = (x_0 + x_F) / 2$. После этого тележку необходимо завести на некоторое расстояние вправо за точку x_F , так чтобы затем переместить ее к точке $X = x_F$. При этом одновременно должны выполняться соотношения $x = x_F$ и $x' = 0$. Графики зависимостей координат и скоростей системы от времени для оптимального управления показаны на рис. 2.

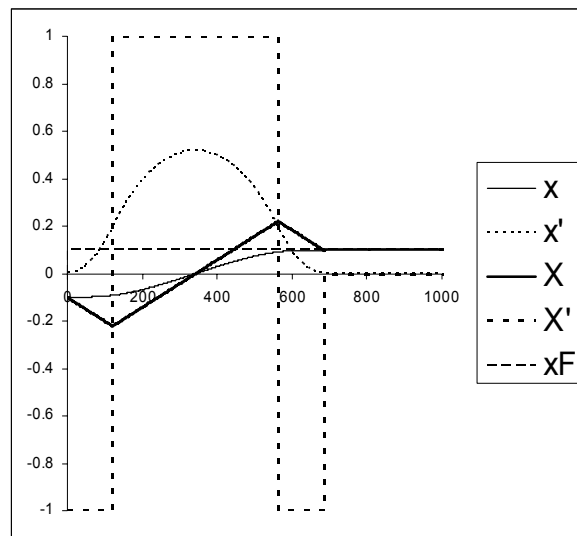


Рис. 2. Оптимальный алгоритм управления маятником

Подобно тому, как маятником можно управлять путем перемещения тележки, на которую он опирается, движущийся мотоцикл управляется посредством поворотов руля. Таким образом, имеет место некоторая аналогия между задачей о перевернутом маятнике и о движущемся мотоцикле, что позволяет рассмотреть задачу о маятнике, как явно более простую, а затем вернуться к рассмотрению основной задачи, применив полученные результаты к управлению мотоциклом.

Стоит отметить, что наличие сходства между двумя задачами не означает их полной тождественности – между этими системами есть заметные различия, которые будут видны из дальнейшего рассмотрения.

Однако даже для такой простой системы, как перевернутый маятник, не удастся построить алгоритм оптимального управления, который мог бы точно вычислять управляющее воздействие непосредственно по значениям переменных, описывающих движение системы – уравнения для этого воздействия не решаются в символьном виде. Итерационное же решение этих уравнений не может быть использовано для управления в реальном времени. В связи с этим требуется какая-то аппроксимация решения уравнений, например, при помощи нейронной сети.

Более простыми способами, такими как линейное управление (ПИД регулирование) или нечеткая логика [4], оптимальное управление маятником обеспечить не удастся.

Нейросеть для управления маятником

Для реализации управления перевернутым маятником при помощи нейронной сети потребуется, во-первых, определить конфигурацию нейронной сети (число слоев, количество нейронов в каждом слое), и, во-вторых, тот набор данных, на котором будет выполняться ее обучение.

Для определения конфигурации нейронной сети, которая, будучи обученной, окажется в состоянии достаточно хорошо управлять маятником, потребовалось провести несколько вычислительных экспериментов. В ходе этих экспериментов были протестированы несколько конфигураций нейронной сети, отличающихся числом слоев и нейронов в слоях. Все нейронные сети, применяемые для управления мотоциклом, имеют следующую структуру: они содержат три или более слоев, из которых первый является входным, последний – выходным, а все остальные – скрытыми. При этом обозначение конфигурации сети «7» понимается как наличие у нейронной сети одного скрытого слоя, состоящего из семи нейронов, обозначение «91» – это нейронная сеть, имеющая 91 нейрон в одном скрытом слое. Обозначение «7–5» понимается как семь нейронов в первом скрытом слое и пять нейронов во втором (всего два скрытых слоя). И, наконец, обозначение 6×3 – это обозначение сети из шести скрытых слоев по три нейрона в каждом.

Рассмотрение графиков зависимостей ошибки обучения от числа итераций для каждой конфигурации сети показало, что увеличивать число нейронов в каждом слое имеет смысл только до определенного предела. Дальнейшее увеличение их числа приводит к уменьшению ошибки аппроксимации ценой значительного увеличения продолжительности каждой итерации обучения. Наоборот, увеличение числа слоев без увеличения общего количества нейронов и весов в сети приводит к уменьшению результирующей ошибки при одновременном уменьшении времени обучения. Однако уменьшение числа нейронов в слое до очень малых значений отрицательно сказывается на точности аппроксимации. Сети с очень малым числом нейронов в каждом слое показывают наихудшие результаты обучения.

Ниже приведена таблица результатов обучения сетей. В этой таблице представлены окончательные ошибки обучения (максимальная и средняя), а также время, затрачиваемое на обучение каждой сети.

Таблица

Конфигурация сети	Ошибка								Время итерации, мс
	7000 итераций				50000 итераций				
	Обучение		Тестирование		Обучение		Тестирование		
	max	сред	max	сред	max	сред	max	сред	
7	0.0628	0.0215	0.0649	0.0217	0.0528	0.0189	0.0631	0.0191	32
37	0.0438	0.0129	0.0478	0.0134	0.0441	0.0101	0.0463	0.0107	130
91	0.0445	0.0122	0.0472	0.0127	0.0433	0.0090	0.0454	0.0096	300
7–5	0.0335	0.0080	0.0356	0.0086	0.0215	0.0045	0.0277	0.0055	50
13–7	0.0230	0.0053	0.0347	0.0061	0.0166	0.0030	0.0300	0.0042	86

Таблица (продолжение)

21–13	0.0214	0.0043	0.0317	0.0055	0.0100	0.0018	0.0320	0.0034	150
13–7–5	0.0158	0.0036	0.0333	0.0050	0.0089	0.0020	0.0295	0.0036	105
21–13–7	0.0128	0.0022	0.0315	0.0038	0.0075	0.0011	0.0364	0.0031	195

7-7-5-5	0.0178	0.0041	0.0379	0.0051	0.0081	0.0018	0.0297	0.0031	105
6 × 3	0.0486	0.0168	0.0524	0.0169	0.0470	0.0103	0.0503	0.0107	80

Столбцы «Тестирование» в этой таблице содержат результаты тестирования обученных нейронных сетей. Тестирование сетей проводилось на большем наборе векторов, чем их обучение – тестовый набор векторов содержит в 16 раз больше точек, чем в обучающем множестве.

Как следует из результатов тестирования, ошибка на тесте может превышать ошибку обучения довольно значительно. Так, например, для сети 21-13-7: максимальная ошибка при тестировании равна 0.0364, а при обучении – 0.0075.

Для реализации управления маятником была выбрана конфигурация нейронной сети, которая обеспечила наилучшие результаты в ходе проведенного тестирования – сеть с конфигурацией 13-7-5. Эта сеть обучалась следующим образом. Начальные веса инициализировались случайными величинами из диапазона $(-1, 1)$. Рассматривалось 10 вариантов случайного задания начальных весов (десять нейронных сетей). Каждая из этих сетей обучалась на протяжении фиксированного числа итераций на обучающем множестве (неравномерной четырехугольной сетке) из $21 \cdot 41 = 861$ вектора. Затем для трех лучших сетей, показавших наименьшую ошибку обучения на этом множестве, обучение продолжалось на большем множестве из $41 \cdot 81 = 3321$ вектора. В ходе этого этапа обучения периодически контролировалась ошибка тестирования на множестве из $81 \cdot 161 = 13041$ вектора. Всякий раз, когда ошибка тестирования становилась меньше ранее достигнутой, значения весов сети сохранялись. Таким способом удалось достигнуть ошибки тестирования $errMax = 0.0209$ (максимальной) и $errSS = 0.0022$ (средне-квадратичной).

Эта сеть, обученная на множестве векторов, позволила добиться достаточно хорошего качества управления, что и явилось решающим фактором в пользу того, чтобы использовать нейронную сеть данной конфигурации для управления маятником.

Исследование нейросети для управления мотоциклом

Посмотрим, каким образом по аналогии с оптимальным управлением маятником можно построить систему оптимального управления равновесием мотоцикла.

Аналогом координаты x маятника будет угол наклона мотоцикла L . Координате X тележки, на которой установлен маятник, соответствует угол поворота руля T . Данная аналогия не полна, и системы «маятник» и «мотоцикл» не тождественны, между ними есть заметные различия.

Во-первых, если равновесие маятника достигается при $x = X$, $x' = 0$, то для равновесия мотоцикла между наклоном мотоцикла L и углом поворота руля T существует более сложная связь, причем ее достаточно сложно записать аналитически.

Второе заметное отличие мотоцикла от маятника – это отсутствие инвариантности относительно сдвига координат. Если для маятника можно было записать зависимость управления от разностей $(X - x_F)$ и $(x - x_F)$ вместо зависимости отдельно от переменных X , x и x_F , то для мотоцикла подобную замену произвести не удастся в силу нелинейности системы.

Тем не менее, оказывается возможным построение системы управления равновесием мотоцикла по аналогии с системой управления маятником. Аналогом координаты x маятника является угол наклона мотоцикла L . Координате X тележки, на которой установлен маятник, соответствует угол поворота руля T .

Пример оптимального управления равновесием мотоцикла (в его идеальной реализации) показан на рис. 3.

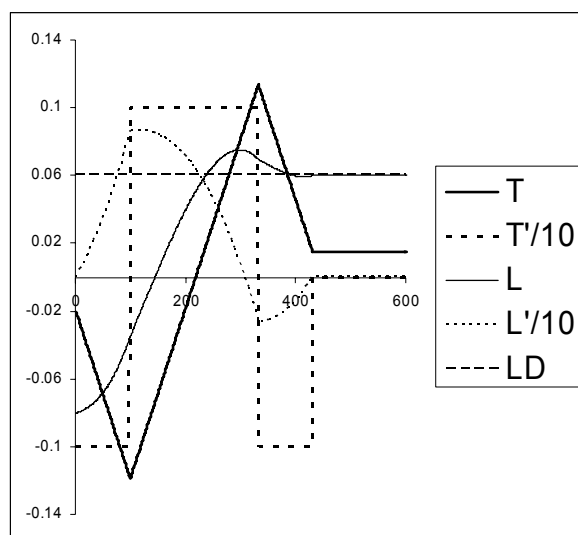


Рис. 3. Пример оптимального (идеального) управления мотоциклом

Создаваемая система управления мотоциклом строилась на основе нейронной сети, которая была обучена на данных, полученных из экспериментов по управлению моделью мотоцикла. Однако, по сравнению с маятником, для уменьшения времени обучения была выбрана одна конфигурация нейронной сети, которая имела четыре входа (четыре нейрона во входном слое), два скрытых слоя по семь нейронов каждый и выходной слой, состоящий из одного нейрона.

Пример графиков зависимостей координат и скоростей системы от времени при нейросетевом управлении приведен на рис. 4.

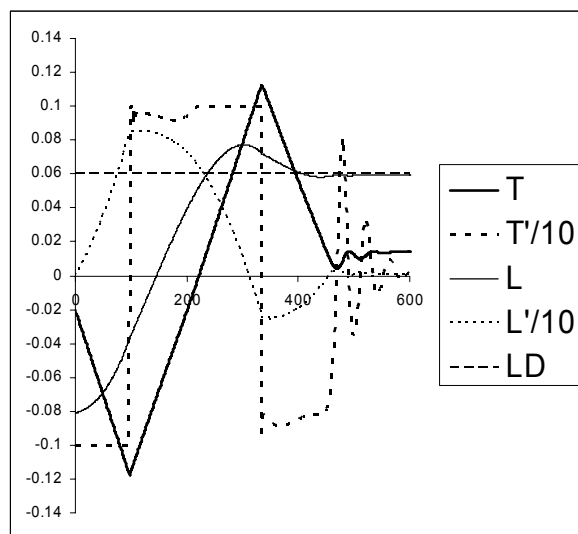


Рис. 4. Пример нейросетевого управления мотоциклом

Здесь управляемая переменная (угол поворота руля) обозначена буквой T .

Сравнивая эти графики с графиками на рис. 3, отметим неидеальность выполняемого нейронной сетью управления равновесием мотоцикла по сравнению с тем, как это было для маятника. Это связано с тем, что в задаче по управлению мотоциклом проводится аппроксимация функции четырех переменных, и размерность задачи равна четырем, в то время как в задаче с маятником было всего две переменных, что позволяло достигать большей точности управления.

При использовании большего числа нейронов и большего числа точек данных, скорее всего, результаты удалось бы улучшить. Однако, даже это управление оказалось

удовлетворительным, что и продемонстрировали эксперименты с виртуальным мотоциклом (рис. 5).

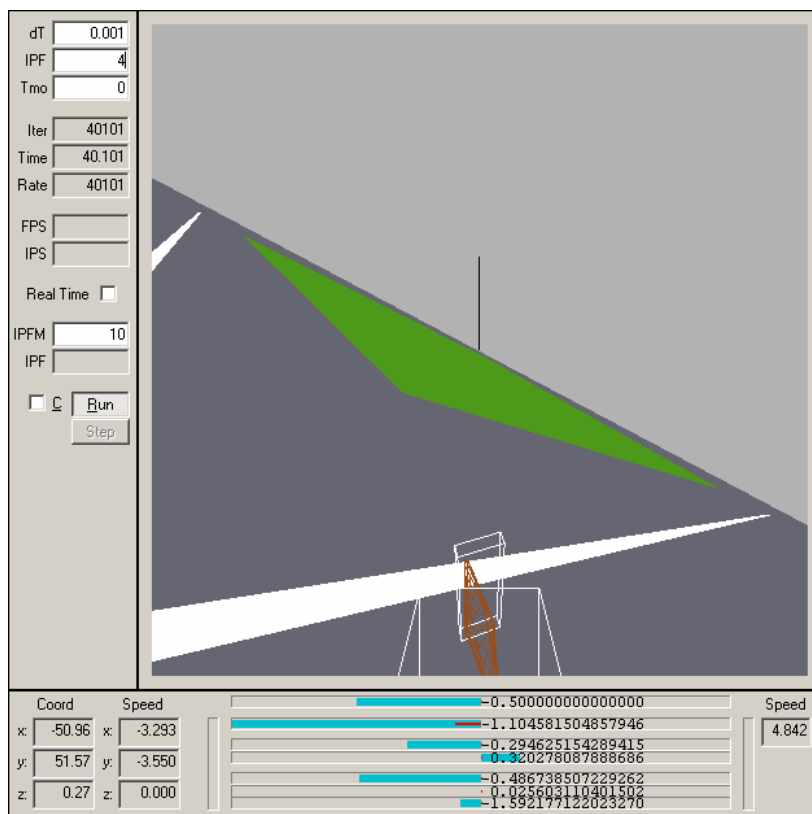


Рис. 5. Эксперимент с виртуальным мотоциклом

Заключение

Поставленная в работе цель (построение системы управления мотоциклом, обеспечивающей его равновесие и оптимальное выполнение поворотов) была достигнута. Построенная система управления обеспечивает равновесие мотоцикла, удерживая его на траектории постоянной кривизны, а также позволяет производить переход между такими траекториями для выполнения поворотов (пример такого перехода показан на рис. 4).

Применение нейронной сети при построении данной системы позволяет мотоциклу выполнять повороты оптимальным образом, когда цель управления – переход к движению по траектории заданной кривизны – достигается максимально быстро при заданном ограничении на скорость поворота руля.

Система управления, построенная на основе нейронной сети, обучалась на экспериментальных данных, получаемых в ходе исследования движения модели. Качество управления, которое обеспечивает обученная нейронная сеть, зависит от числа нейронов и конфигурации этой сети, а также от количества экспериментальных точек, на которых обучается сеть.

Следует понимать, что реализованная здесь модель мотоцикла, как, впрочем, и система управления – это лишь приближения к реальности. При реализации реальной системы управления реальным мотоциклом необходимо будет учесть еще ряд различных особенностей, условий и ограничений. Так, например, здесь предполагалось, что скорость поворота руля можно изменять мгновенно, устанавливая ее равной заданной, но так, чтобы она не превышала максимальную. Это лишь приближение, на самом же деле, для любого реального привода, каков бы он ни был, существует более сложная зависимость его ускорения от скорости, внешнего сопротивления и внутреннего воз-

действия, например, для электромотора угловое ускорение зависит от скорости вращения, момента сил на оси и напряжения на входе. Другое предположение – большой коэффициент трения колес, взятый как у очень хорошей резины на очень хорошем асфальте, например, как в кольцевых мотогонках. В реальности следует предполагать, что коэффициент трения может быть разным в зависимости от состояния дороги, и учитывать этот факт при реализации управления. Исключение из рассмотрения особенностей подвески колес, неровностей дороги и некоторых других эффектов также упростило задачу, при этом несколько отдалив ее от реальности.

Однако, несмотря на неизбежные отличия модели от реального объекта, данная работа продемонстрировала возможность создания оптимальной при заданных ограничениях системы управления и показала путь к созданию такой системы – исследование поведения модели, накопление опыта управления, и обучение системы управления на основе полученного опыта.

Мотоцикл, управляемый оптимальным образом при помощи нейронной сети – это шаг на пути к созданию систем, обладающих способностью действовать в той или иной степени самостоятельно.

Источники

1. Russell Smith "Open Dynamics Engine v0.5 User Guide" <http://ode.org> , "Documentation"
2. Кейт Код "Техника вождения мотоцикла" <http://www.drivemc.ru/articles/40.files/00.html>
3. Гришин А.А., Ленский А.В., Охоцимский Д.Е., Панин Д.А., Формальский А.М. О синтезе управления неустойчивым объектом. Перевернутый маятник //Известия РАН. Теория и системы управления. 2002, № 5.
4. MathWorks Inc. "Fuzzy Logic Toolbox User's Guide" http://www.mathworks.com/access/helpdesk/help/pdf_doc/fuzzy/fuzzy.pdf

НАШИ АВТОРЫ

Алексеева Ирина Петровна – кандидат химических наук, старший научный сотрудник Научно-исследовательского и технологического института оптического материаловедения ВНЦ «ГОИ имени С.И. Вавилова»»

Аничкин Илья Михайлович - аспирант кафедры информационных систем СПбГУ ИТМО

Беспалов Виктор Георгиевич – доктор физико-математических наук, профессор, заместитель декана факультета фотоники и оптоинформатики СПбГУ ИТМО

Васильев Владимир Николаевич – доктор технических наук, профессор, ректор СПбГУ ИТМО

Громов Геннадий Юрьевич – старший преподаватель кафедры вычислительной техники СПбГУ ИТМО

Гугель Юрий Викторович – кандидат технических наук, директор филиала Государственного научно-исследовательского института информационных технологий и телекоммуникаций (ФГУ ГНИИ ИТТ «Информика»)

Гуд Владимир Владимирович – кандидат технических наук, начальник первого отдела СПбГУ ИТМО

Гусарова Наталия Федоровна – кандидат технических наук, доцент кафедры технологий профессионального обучения СПбГУ ИТМО

Дымшиц Ольга Сергеевна – кандидат химических наук, старший научный сотрудник Научно-исследовательского и технологического института оптического материаловедения ГОИ имени С.И. Вавилова»

Егоров Валерий Юрьевич – ведущий инженер ВНЦ «ГОИ имени С.И. Вавилова»»

Ермаков Виктор Анатольевич – студент СПбГУ ИТМО

Ефимов Юрий Николаевич – кандидат физико-математических наук, доцент СПбГУ ИТМО

Жилин Александр Александрович – кандидат химических наук, начальник отдела Научно-исследовательского и технологического института оптического материаловедения ГОИ имени С.И. Вавилова»

Иванов Андрей Викторович – кандидат технических наук, доцент, проректор по экономике и финансам СПбГУ ИТМО

Иванова Светлана Эдуардовна – младший научный сотрудник ЦИОТ, аспирант ГОИ ГОИ имени С.И. Вавилова»

Ижванов Юрий Львович – кандидат технических наук, заместитель директора филиала Государственного научно-исследовательского института информационных технологий и телекоммуникаций (ФГУ ГНИИ ИТТ «Информика»)

Кириллов Владимир Васильевич – кандидат технических наук, профессор кафедры вычислительной техники СПбГУ ИТМО

Козлов Сергей Аркадьевич – доктор физико-математических наук, профессор, декан факультета фотоники и оптоинформатики СПбГУ ИТМО

Коншина Елена Анатольевна – кандидат технических наук, старший научный сотрудник Центра «Информационные оптические технологии» СПбГУ ИТМО

Корешев Сергей Николаевич – доктор технических наук, профессор кафедры прикладной и компьютерной оптики СПбГУ ИТМО

Корнеев Георгий Александрович – аспирант кафедры технологии программирования СПбГУ ИТМО

Красавцев Валерий Михайлович – кандидат технических наук, доцент СПбГУ ИТМО

Крылов Виктор Николаевич – кандидат физико-математических наук, старший научный сотрудник СПбГУ ИТМО

Левицкий Руслан Сергеевич – аспирант, младший научный сотрудник Центра «Информационные оптические технологии» СПбГУ ИТМО

Лисицына Любовь Сергеевна – кандидат технических наук, доцент, заведующая кафедрой информационных образовательных технологий СПбГУ ИТМО

Лукомский Георгий Владимирович – кандидат физико-математических наук, старший научный сотрудник ВНЦ «ГОИ им. С.И. Вавилова»

Никоноров Николай Валентинович – доктор физико-математических наук, профессор, заведующий кафедрой оптоинформационных технологий и материалов СПбГУ ИТМО

Озерский Александр Маркович – старший инженер-программист ЗАО «Транзас Технологии»

Павлов Александр Владимирович – кандидат технических наук, доцент кафедры фотоники и оптоинформатики СПбГУ ИТМО

Перлин Евгений Юрьевич – доктор физико-математических наук, начальник отдела Центра «Информационные оптические технологии» СПбГУ ИТМО

Петров Владимир Иванович – доктор физико-математических наук, ведущий научный сотрудник СПбГУ ИТМО

Пономарева Елена Анатольевна – младший научный сотрудник ВНЦ «ГОИ имени С.И. Вавилова»

Путилин Сергей Эдуардович – кандидат физико-математических наук, старший научный сотрудник СПбГУ ИТМО

Разумова Ирина Константиновна – кандидат физико-математических наук, научный сотрудник ВНЦ «ГОИ им. С.И. Вавилова»

Ратушный Владислав Петрович – инженер ЗАО «ХолоГрэйт»

Рейтеров Владимир Михайлович – кандидат физико-математических наук, ведущий научный сотрудник ВНЦ «ГОИ им. С.И. Вавилова»

Симоненко Зинаида Григорьевна – кандидат технических наук, доцент кафедры информатики и прикладной математики СПбГУ ИТМО

Стаселько Дмитрий Иванович – доктор физико-математических наук, главный научный сотрудник «Центра информационных и оптических технологий»

Ткачук Александра Михайловна – кандидат физико-математических наук, ведущий научный сотрудник ЦОИТ СПбГУ ИТМО

Центер Марина Яковлевна – кандидат физико-математических наук, старший научный сотрудник Научно-исследовательского и технологического института оптического материаловедения ГОИ имени С.И. Вавилова»

Чиков Константин Никитич – кандидат технических наук, начальник отдела технической защиты информации СПбГУ ИТМО

Шалыто Анатолий Абрамович – доктор технических наук, заведующий кафедрой технологий программирования СПбГУ ИТМО

СОДЕРЖАНИЕ

1. ФИЗИКО-ХИМИЧЕСКИЕ ТЕХНОЛОГИИ	3
Коншина Е.А. Развитие физико-химической концепции нанотехнологии ориентации жидких кристаллов.....	3
Алексеева И.П., Дымшиц О.С., Ермаков В.А., Жилин А.А., Петров В.И., Центер М.Я. Фазовый распад и кристаллизация литиевоалюмосиликатных стекол с добавкой ZrO_2	14
Ткачук А.М., Иванова С.Э., Разумова И.К., Рейтеров В.М., Пономарева Е.А., Егоров В.Ю., Никоноров Н.В. Разработка и исследование новых активных сред для твердотельных ир-конверсионных и ик лазеров	22
Корешев С.Н., Ратушный В.П. Основы технологии голографической субмикронной фотолитографии.....	34
Перлин Е.Ю., Иванов А.В., Левицкий Р.С. Оптическое переключение и ир-конверсия в материалах с квантовыми ямами: эффект фотонной лавины	42
2. ТЕХНОЛОГИИ ОПТОИНФОРМАТИКИ.....	57
Павлов А.В. Голографический предсказатель случайных процессов	57
Чиков К.Н., Гуд В.В., Красавцев В.М. Разработка видеоспектральных основ аэрокосмических методов дистанционного зондирования в оптическом диапазоне.....	64
Беспалов В.Г., Васильев В.Н., Ефимов Ю.Н., Козлов С.А., Крылов В.Н., Лукомский Г.В., Путилин С.Э., Стаселько Д.И. 10 ГВт фемтосекундная лазерная система ФС-1 – устройство и характеристики	74
Иванов А.В. Использование параметров тонких компонентов для автоматизированной коррекции оптических систем	80
Иванов А.В. Экономичный метод вычисления коэффициентов поперечных аберраций третьего и высших порядков	87
Симоненко З.Г. Математическое представление процесса измерения параметров массопереноса в жидкой бинарной среде с границей раздела	91
Озерский А.М. Применение фильтра Калмана второго порядка для компьютерной обработки измерительной информации.....	96
Гусарова Н.Ф. Применение теории координации для управления качеством технологических процессов со слабо формализуемыми критериями	102
3. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ	118
Корнеев Г.А., Шалыто А.А. Построение визуализаторов алгоритмов дискретной математики	118
Корнеев Г.А., Шалыто А.А. VIZI – язык описания логики визуализаторов алгоритмов.....	130
Васильев В.Н., Гугель Ю.В., Ижванов Ю.Л. Разработка теоретических основ и реализация методов построения компьютерных сетей на базе ир-протоколов нового поколения	138
Васильев В.Н., Лисицына Л.С. Концепция сетевой информационно-образовательной технологии для разработки результатов образования	149

Кириллов В.В., Громов Г.Ю. Методы и средства защиты информации в автоматизированной системе электронного обмена стандартизованными медицинскими документами.....	157
Аничкин И.М. Нейросеть для управления мотоциклом	162
НАШИ АВТОРЫ.....	171

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК СПбГУ ИТМО
Выпуск 23

**ВЫСОКИЕ ТЕХНОЛОГИИ В ОПТИЧЕСКИХ
И ИНФОРМАЦИОННЫХ СИСТЕМАХ**

Главный редактор
доктор технических наук, профессор
В.Н. Васильев
Дизайн обложки В.А. Петров
Редакционно-издательский отдел СПбГУ ИТМО
Зав. РИО Н.Ф. Гусарова
Лицензия ИД № 00408 от 05.11.99.
Подписано в печать 24.12.05.
Заказ 906. Тираж 100 экз.