

УДК 541.135

ВЛИЯНИЕ СЕРЕБРА НА РОСТ КВАНТОВЫХ ТОЧЕК
ВО ФТОРОФОСФАТНЫХ СТЕКЛАХ

Е.В. Колобкова, Н.В. Никоноров, В.А. Асеев

Изучено влияние образования Ag-кластеров на рост PbSe-квантовых точек во фторофосфатном стекле системы $\text{Na}_2\text{O-P}_2\text{O}_5\text{-Ga}_2\text{O}_3\text{-AlF}_3\text{-ZnO(Se)-PbF}_2$. Обнаружено усиление поглощения первого экситонного перехода при введении Ag при формировании квантовых точек одинакового размера. Зафиксировано увеличение размеров нанокристаллов при одинаковом температурно-временном режиме роста при формировании Ag-кластеров по сравнению с размерами квантовых точек в стеклах, не содержащих серебра.

Ключевые слова: квантовые точки, поверхностный плазмонный резонанс, Ag^0 нанокластеры, спектры поглощения.

Введение

Оптическим свойствам молекулярных кластеров и нанокристаллов благородных металлов (Ag, Au, Pt) сегодня уделяется особое внимание. Они обладают собственной люминесценцией, а также могут усиливать люминесценцию и поглощение других ионов, молекул и нанообъектов. Так, перенос энергии от молекулярных кластеров металлов на редкоземельные ионы позволяет усилить люминесценцию последних [1]. Локальное усиление поля электромагнитной волны при плазмонном резонансе в металлических наночастицах также приводит к усилению люминесценции и поглощению редкоземельных ионов [2] и биологических молекул [3]. По этой причине оптические свойства кластеров благородных металлов, содержащих разное число атомов, в первую очередь серебра, привлекают пристальное внимание исследователей.

В последние десятилетия стекла, содержащие квантовые точки, привлекали заметное внимание как материалы для создания элементов оптоэлектроники. Основой для подобного рассмотрения является радикальное изменение оптических свойств, возникающее вследствие квантового размерного эффекта, проявляющегося в диапазоне размеров, когда нанокристалл становится меньше радиуса экситона Бора. Такой нанокристалл называется квантовой точкой и имеет дискретный оптический спектр, причем дырка и электрон квантуются независимо друг от друга.

Квантовые точки селенида свинца (PbSe) характеризуются широким диапазоном размеров, соответствующих режиму сильного квантования, так как электрон, дырка и экситон имеют относительно большие радиусы экситона Бора. Так, в PbSe радиусы электрона, дырки и экситона Бора равны 23, 23 и 46 нм соответственно. Такие значения радиусов позволяют достигать режима сильного размерного квантования на сравнительно больших частицах.

В настоящее время практически отсутствуют сведения, за исключением работ [4–10], о формировании в стеклах полупроводниковых PbSe-нанокристаллов, характеризующихся сдвигом края оптического поглощения в широком диапазоне длин волн и демонстрирующих квантовые переходы в спектре оптического поглощения.

Недавно были опубликованы интересные результаты о влиянии серебряных кластеров на формирование квантовых точек PbS в силикатных стеклах [11]. Было показано, что при увеличении концентрации вводимого серебра происходит усиление поглощения и люминесценции PbS-нанокристаллов. В [12] проведено сравнение оптических свойств PbS-квантовых точек в слоях стекол до и после проведения ионного Ag⁺-обмена. Показано, что в объеме стекла, подвергнувшегося ионному обмену, происходит сдвиг спектров поглощения и люминесценции PbS-квантовых точек в длинноволновую область по сравнению с областями стекла, в которых ионный обмен не проводился, что соответствует росту более крупных нанокристаллов PbS.

Целью представляемой работы является расширение сведений о взаимодействии PbSe-квантовых точек с введенным в состав исходной матрицы серебром во фторофосфатном стекле системы $\text{Na}_2\text{O-P}_2\text{O}_5\text{-Ga}_2\text{O}_3\text{-AlF}_3\text{-ZnO}$. Это дает возможность синтеза новых фторофосфатных стекол, позволяющих формировать в них квантовые точки PbSe в широком интервале размеров, характеризующихся узким распределением плотности электронных состояний для кристаллов всех размеров в энергетическом диапазоне 1,0–2,5 мкм.

Методика эксперимента и результаты исследований

Для синтеза стекол системы $\text{Na}_2\text{O-P}_2\text{O}_5\text{-Ga}_2\text{O}_3\text{-AlF}_3\text{-ZnO(Se)-PbF}_2$ применялись материалы марки «ХЧ» и «ОСЧ». В качестве образцов для исследования были синтезированы стекла следующих составов: стекло 1, содержащее 10 ppm Ag (AgBr), не содержащее селен; стекло 2, содержащее AgBr и ZnSe + PbF₂;

стекло 3, содержащее $\text{ZnSe} + \text{PbF}_2$. Навеска составляла 50 г, синтез проводился в течение 40 мин при $T = 950\text{--}1000^\circ\text{C}$ в закрытых стеклоглеродных тиглях в атмосфере аргона. Такие условия синтеза позволяли получать стекла с высоким уровнем пропускания в диапазоне от 0,3–5 мкм. Стекломасса вырабатывалась между двумя холодными стеклоглеродными пластинами, закаленные стекла имели толщину не более 2 мм и отжигались при температуре несколько ниже T_g для снятия остаточных напряжений. Характеристические температуры при выборе температурно-временного режима формирования стеклокерамики были определены на основании данных, полученных при анализе кривой дифференциальной сканирующей калориметрии (ДСК). Измерения и математическая обработка данных проводились на дифференциальном сканирующем калориметре STA 449F1 Jupiter фирмы Netzsche. Исходные стекла и подвергнутые вторичной термообработке образцы стеклокерамики для измерения спектрально-люминесцентных характеристик были отшлифованы и отполированы и имели толщину 1 мм.

Спектры поглощения образцов измерялись на спектрофотометре Varian Cary 500 в диапазоне 300–3300 нм (оптическая плотность $D = 0\text{--}10$; спектральный диапазон регистрации 200–3300 нм; разрешение 0,1 нм; время интеграции 0,5 с).

Теплофизические характеристики

Термограмма стекол, перспективных для создания объемных стеклокристаллических материалов (прозрачных стеклокерамик), должна иметь два неперекрывающихся экзопика: первый пик обусловлен объемной кристаллизацией, а второй – поверхностной кристаллизацией. Выбор температуры термообработки в начале первого пика позволяет полностью исключить поверхностную кристаллизацию, которая приводит к неконтролируемому росту больших кристаллов. Характерные термограммы стекол 1 и 2 представлены на рис. 1, а, б, соответственно. Термограмма стекла 1 демонстрирует одну широкую полосу с максимумом 628°C , $T_{\text{нк}} = 570^\circ\text{C}$, соответствующую процессу поверхностной кристаллизации стекламатрицы и формированию микрокристаллического NaPO_3 . Пиков, соответствующих формированию Ag-кластеров, не наблюдается, по-видимому, вследствие их малой концентрации.

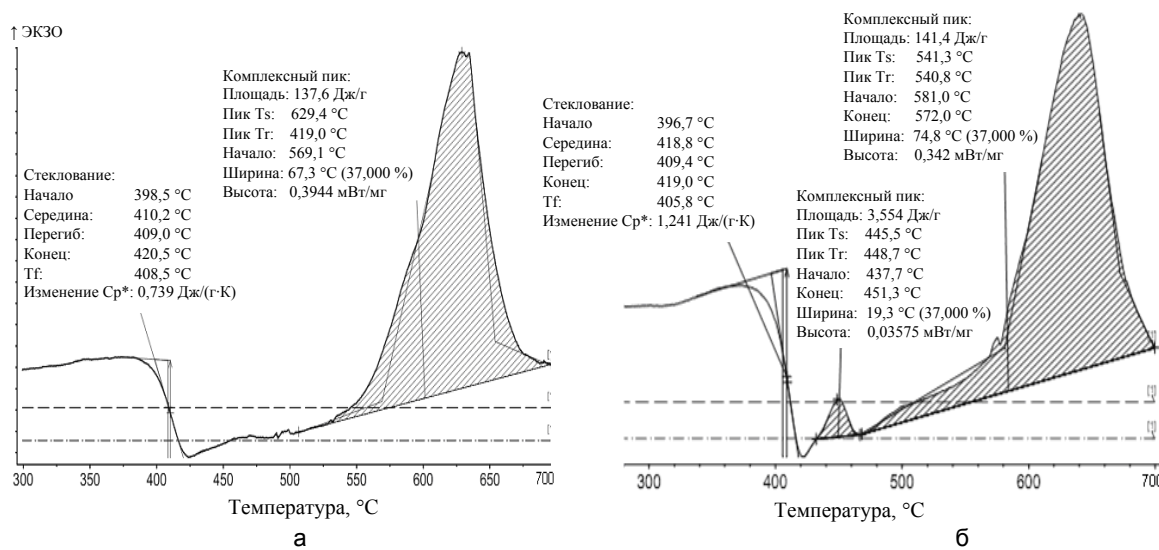


Рис. 1. ДСК стекла 1 (а) и стекла 2 (б)

Стекло 1 было подвергнуто термообработке при $T = 410^\circ\text{C}$ и $T = 430^\circ\text{C}$. При этом в нем наблюдалось возникновение желтой окраски, что свидетельствовало о росте металлических кластеров с $n \geq 100$.

Термограмма стекла 2 демонстрирует два разделенных пика. Первый пик соответствует выделению кубической модификации PbSe [4]. Эти предположения были подтверждены экспериментально: термообработки выше $T_{\text{нк}}$ позволили получить объемную кристаллизацию в стекле 2. На основании ДСК был определен температурный режим для проведения направленной контролируемой кристаллизации. Диапазон температур роста нанокристаллов составил при $T = 430\text{--}470^\circ\text{C}$, площадь пика – $3,6 \text{ Дж/г}$. Однако для контроля за ростом кристаллов необходимо выбирать минимальные из указанного интервала температуры. Известно, что для оптимизации распределения по размерам и приближении этого распределения к монодисперсному необходим двухстадийный режим термообработки. Первая, низкотемпературная стадия (Т1) должна обеспечить рост определенного числа зародышей. Вторая стадия (Т2) должна приводить к росту нанокристаллов преимущественно определенного размера. В соответствии с этой процедурой были выбраны две температуры: $T_1 = 400^\circ\text{C}$ и $T_2 = 430^\circ\text{C}$. Стекла были термообработаны в течение 0,5–1,5 ч.

Формирование Ag_n^0 нанокластеров в стеклах

Отличительной чертой металлических частиц является наличие в них коллективных возбуждений электронов проводимости, называемых также поверхностными плазмонами, которые проявляют себя в определенной области спектра в виде интенсивной полосы поглощения. Спектры поглощения стекла 1, термообработанного при 410°C (кривая 1) и 430°C (кривая 2) в течение 40 мин, представлены на рис. 2. Очевидно, что возникшие полосы поглощения связаны с проявлением поверхностного плазмонного резонанса, характерного для Ag_n^0 . В задачи настоящей работы не входило детальное обсуждение механизмов возникновения и роста металлических кластеров во фторофосфатных стеклах данного состава, поэтому можно только зафиксировать факт возникновения крупных (не молекулярных) кластеров Ag_n^0 в том же интервале температур, что и рост квантовых точек.

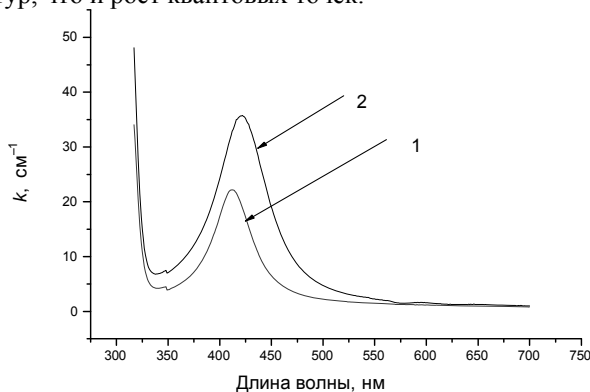


Рис. 2. Спектры поглощения стекла 1, термообработанного при 410°C (кривая 1) и 430°C (кривая 2) в течение 40 мин

Оптические спектры поглощения и рост нанокристаллов

Оптические спектры поглощения стекол, содержащих PbSe и Ag_n^0 , были измерены в спектральном диапазоне 300–3500 нм при комнатной температуре. Полученные спектры четко отражает эффект размерного квантования. Край оптического поглощения смещается от 4 эВ для исходного стекла до 0,56 эВ при увеличении времени термообработки. На рис. 3 показаны спектры поглощения стекол, содержащих нанокристаллы PbSe в присутствии серебряных кластеров, полученные при различной длительности термообработки.

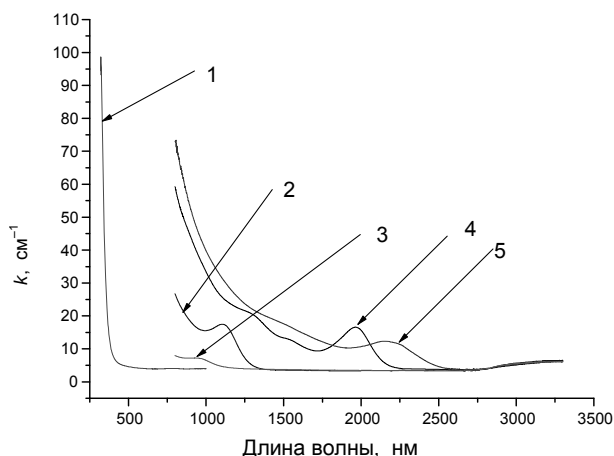


Рис. 3. Спектры поглощения исходного стекла 2 (кривая 1) и стекол, полученных в результате термообработки стекла 2 и содержащих квантовые точки PbSe различных размеров: $D = 2,0$ нм (кривая 2); $D = 3,5$ нм (кривая 3); $D = 8,5$ нм (кривая 4) и $D = 11$ нм (кривая 5)

Спектры поглощения квантовых точек демонстрируют «голубой сдвиг» по отношению к собственному краю поглощения кристалла PbSe ($E_g = 0,28$ эВ) приблизительно на 0,9 эВ для стекла с нанокристаллами наименьших размеров $R \approx 2,0$ нм. На рис. 4 приведена зависимость энергии первого экситонного уровня от размера квантовой точки. Зависимость получена на основании обработки данных электронной микроскопии сверхвысокого разрешения [9]. Сопоставление рис. 3 и рис. 4 позволяет определить, что в нашем эксперименте были сформированы нанокристаллы с диаметрами 2; 3,5; 8,5 и 11 нм.

Влияние введения серебра в состав исходной шихты может быть проанализировано на основе сопоставления спектров поглощения стекол, содержащих и не содержащих в качестве активной добавки AgBr (стекла 2 и 3) и синтезированных в идентичных условиях. На рис. 5 проведено сопоставление двух

стекол, имеющих одинаковую энергию первого экситонного пика. Видно, что коэффициент поглощения выше у стекла, содержащего Ag_n^0 на 2 см^{-1} ($17,5$ и $15,6 \text{ см}^{-1}$ соответственно). Полуширина полосы становится несколько меньше, что означает уменьшение неоднородного уширения, возникающего из-за разброса размеров квантовых точек по размерам.

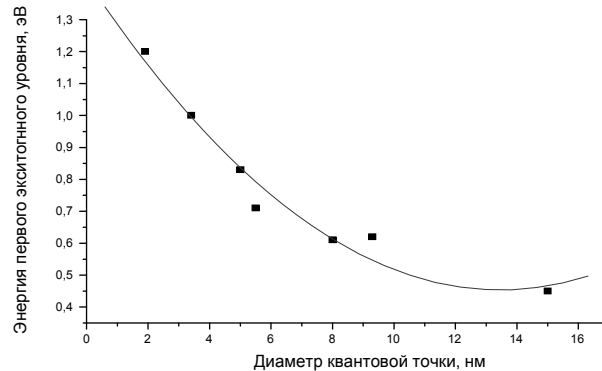


Рис. 4. Зависимость энергии первого экситонного уровня от размера квантовой точки [9].
Выращены нанокристаллы диаметрами 2; 3,5; 8,5 и 11 нм

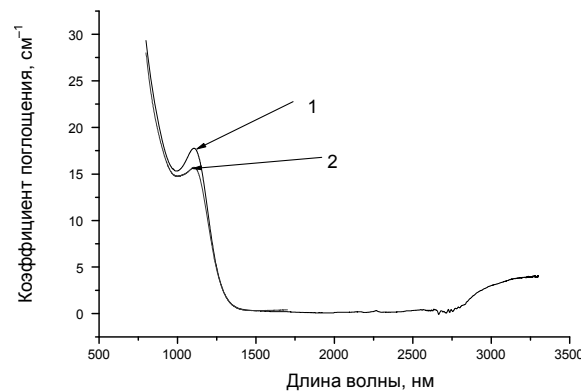


Рис. 5. Спектры поглощения стекла 2 (кривая 1) и стекла 3 (кривая 2) с квантовыми точками одинакового размера ($D = 3,5 \text{ нм}$)

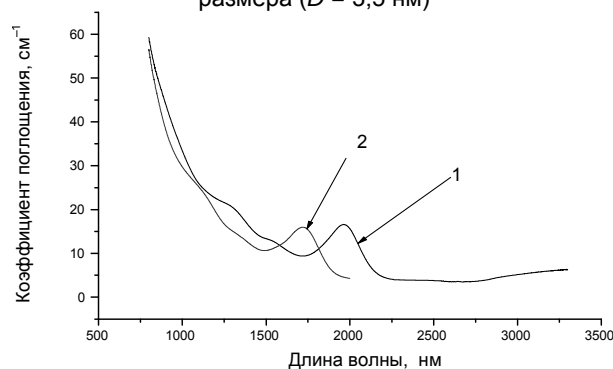


Рис. 6. Спектры поглощения стекла 2 (кривая 1) и стекла 3 (кривая 2) после термообработки в одинаковом температурно-временном режиме

Особенно интересно влияние серебра в области размеров, соответствующих диапазону энергии первого экситона $1,6\text{--}2,3 \text{ мкм}$. На рис. 6 представлены спектры стекол 2 и 3, термообработанных при одинаковых условиях. Сопоставление спектров свидетельствует о формировании нанокристаллов больших размеров при введении AgBr . Так, в случае активной добавки $\text{ZnSe} + \text{PbF}_2$ формируется нанокристалл с диаметром 8 нм , а при введении AgBr диаметр увеличивается до $8,4 \text{ нм}$.

Заключение

Изучено влияние введения Ag на рост PbSe -квантовых точек во фторофосфатном стекле. Обнаружено заметное увеличение интенсивности поглощения первого экситонного перехода при введении Ag -квантовых точек одинакового размера. Зафиксирован рост нанокристаллов больших размеров при одинаковом температурно-временном режиме роста при введении AgBr .

Работа выполнена при поддержке гранта РФФИ №11-08-01226-а «Стекла с квантовыми точками полупроводников IV-VI для модуляторов добротности инфракрасных лазеров».

Литература

1. Hayakawa T., Selvan S.T., Nogami M. Field enhancement effect of small Ag particles on the fluorescence from Eu³⁺-doped SiO₂ glass // Appl. Phys. Lett. – 1999. – V. 74. – P. 1513–1515.
2. Wu Z.K., Lanni E., Chen W.Q., Bier M.E., Ly D., Jin R. High yield, large scale synthesis of thiolate-protected Ag₇ clusters // J. Am. Chem. Soc. – 2009. – V. 131. – P. 16672–16673.
3. Shang L., Dong S., Nienhaus G.U. Ultra-small fluorescent metal nanoclusters: Synthesis and biological applications // Nano Today. – 2011. – V. 6. – № 4. – P. 401–418.
4. Kolobkova E.V., Petrikov V.D., Lipovskii A.A. PbSe quantum dot doped phosphate glass // Electronics Letters. – 1997. – V. 33. – № 1. – P. 101–102.
5. Lipovskii A.A., Kolobkova E.V., Petrikov V.D. Wise Synthesis and characterization of PdSe quantum dots in phosphate glasses // Appl. Phys. Letters. – 1997. – V. 71. – № 23. – P. 3406–3408.
6. Andreev A.O., Lipovskii A.A., Kolobkova E.V. Optical absorption in PbSe spherical QD embedded in glass matrices // J. Appl. Phys. – 2000. – V. 88. – № 2. – P. 750–757.
7. Lipovskii A.A., Kolobkova E.V., Oktovets A., Petrikov V.D., Wise F. Formation of narrowly distributed PbS quantum dots in phosphate glass // J. of Phys. E. – 2000. – V. 5. – № 3. – P. 157–160.
8. Липовский А.А., Петриков В.Д., Колобкова Е.В. Фторфосфатные стекла, содержащие квантовые точки PbSe // Физика и химия стекла. – 2002. – Т. 28. – № 4. – С. 327–332.
9. Silver R.S., Morais P.S., Alcalde A.M., Monte A.F.G., Qu F., Dantas N.O. Optical properties of PbSe quantum dots embedded in oxide glass // J. of Non-Crystalline Solids. – 2006. – V. 352. – P. 3522–3524.
10. Dantas N.O., Qu F., Monte A.F.G., Silver R.S., Morais P.S. Optical properties of IV-VI quantum dots embedded in glass: Size effects // J. of Non-Crystalline Solids. – 2006. – V. 352. – P. 3525–3528.
11. Xu Kai, Heo J. Lead sulfide quantum dots in glasses controlled by silver diffusion // J. Non-Cryst. Solids. – 2012. – V. 358. – № 5. – P. 921–924.
12. Xu Kai, Li C., Dai S., Shen X., Wang X., Heo J. Influence of silver on formation PbS quantum dots glasses // J. Non-Cryst. Solids. – 2011. – V. 357. – № 11–13. – P. 2428–2430.

Колобкова Елена Вячеславовна

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор химических наук, профессор, kolobok106@rambler.ru

Никоноров Николай Валентинович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, зав. кафедрой, Nikonov@oi.ifmo.ru

Асеев Владимир Анатольевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, Aseev@oi.ifmo.ru

УДК 535.015

АНАЛИЗ ПРОХОЖДЕНИЯ ЛУЧЕЙ ЧЕРЕЗ ОПТИЧЕСКИЕ СИСТЕМЫ, ВКЛЮЧАЮЩИЕ ЭЛЕМЕНТЫ ИЗ МЕТАМАТЕРИАЛОВ

А.О. Вознесенская, Д.С. Кабанова

Получено обобщенное выражение закона преломления для моделирования прохождения лучей через оптические среды как с положительным, так и с отрицательным показателем преломления. В результате компьютерного моделирования установлено, что оптические системы, содержащие элементы из метаматериалов, обладают малыми аберрациями.

Ключевые слова: расчет оптических систем, наноструктурированные оптические метаматериалы.

Введение

Одним из перспективных направлений развития современных оптических технологий является синтез и изучение сред с отрицательным показателем преломления – наноструктурированных оптических метаматериалов [1–5]. В целом метаматериалы представляют собой искусственные нанокombинации проводников и диэлектриков. При взаимодействии с электромагнитным полем метаматериалы проявляют уникальные свойства, например, невидимость, зеркальность, «необыкновенное распространение света» (extraordinary light transmission) [6–11] и др.

Среди метаматериалов выделяют фотонные кристаллы, рассмотренные в обзоре [11], и «среды Веселаго» [1, 4, 6, 11]. Последние представляют собой изотропные среды с отрицательными диэлектрической и магнитной проницаемостями. В таких средах волновой вектор и вектор Пойнтинга направлены в

противоположные стороны; на границе раздела сред наблюдается эффект отрицательного преломления – свет отклоняется в противоположную сторону, в отличие от обычных диэлектриков. Для обычной оптической среды изображение не тождественно предмету, поскольку оно формируется без участия затухающих волн. В то же время в метаматериале амплитуда затухающих волн увеличивается при удалении волны от предмета, таким образом, изображение формируется с участием затухающих волн. В силу этого становится возможным получить изображения с разрешением, превышающим дифракционный предел. Таким образом, линза из метаматериала позволяет передавать детали изображений, много меньшие, чем длина волны, а значит, преодолеть классический дифракционный предел разрешения обыкновенных линз [6, 7].

По оценке экспертов [11], несмотря на затруднения с синтезом наноструктурированных оптических метаматериалов, разработка линз с отрицательным показателем преломления, работающих в видимом диапазоне, рассматривается как крайне перспективная. Первые исследования линз из метаматериалов начались в 2004 г. [12]. С тех пор, как в 2005 г. впервые были реализованы и представлены метаматериалы для оптических частот [13, 14], происходит бурное развитие исследований в направлении проектирования оптических систем из метаматериалов. Абберационная теория оптических метаматериалов представлена в [3, 14]. Установлено, что линзы из метаматериалов обладают малыми сферическими абберациями [15, 16], при этом, комбинируя линзы с положительным и отрицательным показателем преломления, можно добиться уменьшения абберации и улучшить качество оптической системы [17].

Несмотря на современные достижения в части изучения свойств оптических систем из метаматериалов, их абберационные характеристики в видимом диапазоне требуют дальнейшего изучения. Настоящая работа посвящена анализу аббераций оптических систем, включающих элементы из метаматериалов, на основе трассировки лучей в оптических средах с положительным и отрицательным преломлением.

Математическая модель трассировки лучей в оптических средах с положительным и отрицательным преломлением

Классическая теория прохождения света не позволяет производить расчет оптических систем, включающих элементы с отрицательным показателем преломления. Сформулируем математическую модель прохождения излучения в оптических средах с положительным и отрицательным показателями преломления. Диэлектрическая проницаемость ϵ и магнитная проницаемость μ являются основными характеристиками, которые определяют распространение электромагнитных волн в веществе; они входят в дисперсионное уравнение, задающее связь между частотой ω монохроматической волны и ее волновым числом k :

$$|(\omega^2/c^2)\epsilon_{ij}\mu_{ij} - k^2\delta_{ij} + k_i k_j| = 0. \quad (1)$$

Если вещество изотропно, уравнение (1) упрощается:

$$k^2 = (\omega^2/c^2) - n^2, \quad (2)$$

где n^2 – квадрат коэффициента преломления вещества, равный

$$n^2 = \epsilon\mu. \quad (3)$$

Если не учитывать потери и считать n , ϵ и μ действительными числами, то из (2) и (3) видно, что одновременная смена знаков ϵ и μ никак не отражается на этих соотношениях.

Для того чтобы выявить электродинамические закономерности, связанные со знаками ϵ и μ , обратимся к уравнениям Максвелла и материальным уравнениям [5]:

$$\text{grad } \mathbf{H} = \mathbf{j} + \partial \mathbf{D} / \partial t, \quad (4)$$

$$\text{grad } \mathbf{E} = -\partial \mathbf{B} / \partial t, \quad (5)$$

$$\text{grad } \mathbf{D} = \rho, \text{ grad } \mathbf{B} = 0,$$

$$\mathbf{D} = \epsilon \mathbf{E}, \mathbf{B} = \mu \mathbf{H}, \quad (6)$$

где $\mathbf{H}$ – напряженность магнитного поля; $\mathbf{E}$ – напряженность электрического поля; $\mathbf{B}$ – магнитная индукция; $\mathbf{D}$ – электрическая индукция; $\mathbf{j}$ – плотность тока; ρ – плотность заряда. Электрическое и магнитное поле можно представить в виде

$$\mathbf{E} = -\text{grad } \phi - \partial \mathbf{A} / \partial t, \mathbf{H} = \text{grad } \mathbf{A},$$

где $\mathbf{A}$ – вектор потенциала; ϕ – скаляр потенциала. Выражения (4) и (5) могут быть представлены как

$$\text{rot } \mathbf{E} = -(1/c) \partial \mathbf{B} / \partial t, \quad (7)$$

$$\text{rot } \mathbf{H} = (1/c) \partial \mathbf{D} / \partial t. \quad (8)$$

Для плоской монохроматической волны выражения (6)–(8) принимают вид

$$[\mathbf{kE}] = -(\omega/c)\epsilon \mathbf{E}, \quad (9)$$

$$[\mathbf{kH}] = (\omega/c)\mu \mathbf{H}. \quad (10)$$

Из выражений (9) и (10) видно, что в случае $\epsilon > 0$ и $\mu > 0$ вектора $\mathbf{E}$, $\mathbf{H}$ и $\mathbf{k}$ образуют правую тройку векторов (рис. 1, а), а в случае $\epsilon < 0$ и $\mu < 0$ – левую тройку векторов (рис. 1, б).

Если ввести для векторов $\mathbf{E}$, $\mathbf{H}$ и $\mathbf{k}$ направляющие косинусы и обозначить их через α_i , β_i , γ_i соответственно, то волна, распространяющаяся в данной среде, будет характеризоваться матрицей

$$G = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix}. \quad (11)$$

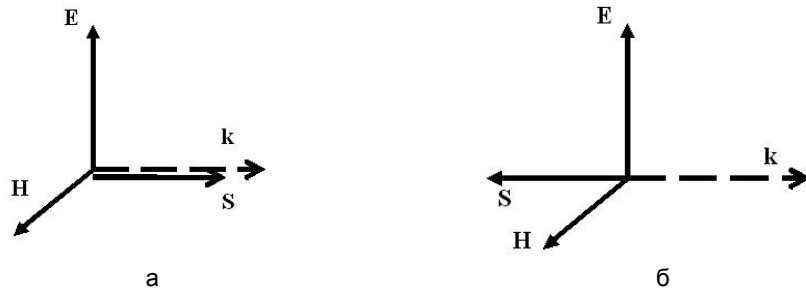


Рис. 1. Оптические среды: обычные («правые») среды, у которых $\varepsilon > 0$ и $\mu > 0$ (а); необычные («левые») среды, у которых $\varepsilon < 0$ и $\mu < 0$ (б)

Обозначим определитель матрицы (11) через параметр «правизны» $p = \det G$. Если тройка векторов $\mathbf{E}$, $\mathbf{H}$ и $\mathbf{k}$ – правая, то $p = +1$, если тройка векторов $\mathbf{E}$, $\mathbf{H}$ и $\mathbf{k}$ – левая, то $p = -1$.

Поток энергии, переносимый электромагнитной волной, определяется вектором Пойнтинга $\mathbf{S}$, который равен

$$\mathbf{S} = (c / 4\pi) [\mathbf{E}\mathbf{H}]. \quad (12)$$

Вектор $\mathbf{S}$ в соответствии с (12) всегда образует с векторами $\mathbf{E}$ и $\mathbf{H}$ правую тройку векторов. Таким образом, для «правых» сред вектора $\mathbf{S}$ и $\mathbf{k}$ направлены в одну сторону, а для «левых» – в противоположные стороны. Поскольку вектор $\mathbf{k}$ совпадает по направлению с фазовой скоростью, то становится ясным, что «левые» среды обладают отрицательной групповой скоростью [1].

При переходе луча света из одной оптической среды в другую граничные условия

$$\begin{cases} E_{t1} = E_{t2}, H_{t1} = H_{t2} \\ \varepsilon_1 E_{n1} = \varepsilon_2 E_{n2}, \mu_1 H_{n1} = \mu_2 H_{n2} \end{cases} \quad (13)$$

должны выполняться вне зависимости от того, имеют эти среды одинаковую или разную «правизну». Из (13) следует, что x - и y -компоненты полей $\mathbf{E}$ и $\mathbf{H}$ в преломленном луче сохраняют свое направление независимо от «правизны» обеих сред. Что касается z -компонент, то они сохраняют свое направление только тогда, когда «правизна» обеих сред одинакова. Если же «правизна» различна, то z -компоненты меняют знак. Таким образом, при переходе в среду с противоположной «правизной» векторы $\mathbf{E}$ и $\mathbf{H}$ не только изменяются по величине из-за различия значений ε и μ , но еще и испытывают зеркальное отражение относительно границы раздела двух сред. То же самое происходит с вектором $\mathbf{k}$. Одновременное зеркальное отражение всей тройки векторов соответствует изменению знака определителя G в (11). В этой связи общепринятая формулировка закона преломления

$$\frac{\sin \sigma}{\sin \psi} = n_{12} = \sqrt{\frac{\varepsilon_2 \mu_2}{\varepsilon_1 \mu_1}} \quad (14)$$

нуждается в уточнении для случаев, когда среды 1 и 2 обладают различной «правизной». Согласно [1] корректная запись формулы (14) – следующая:

$$\frac{\sin \sigma}{\sin \psi} = n_{12} = \frac{p_2}{p_1} \sqrt{\frac{\varepsilon_2 \mu_2}{\varepsilon_1 \mu_1}}. \quad (15)$$

Представив (15) в векторной форме, можно получить выражение

$$\varepsilon_1 \mu_1 p_1 [\mathbf{q}_1 \times \mathbf{g}] = \varepsilon_2 \mu_2 p_2 [\mathbf{q}_2 \times \mathbf{g}], \quad (16)$$

где $\mathbf{q}_1$ – вектор падающего луча; $\mathbf{q}_2$ – вектор преломленного луча; $\mathbf{g}$ – вектор нормали.

Преобразовав (16), имеем

$$[(\varepsilon_2 \mu_2 p_2 \mathbf{q}_2 - \varepsilon_1 \mu_1 p_1 \mathbf{q}_1) \times \mathbf{g}] = 0$$

или

$$[(\mathbf{q}_2 - v \mathbf{q}_1) \times \mathbf{g}] = 0,$$

где $v = \varepsilon_1 \mu_1 p_1 / \varepsilon_2 \mu_2 p_2$.

Векторное произведение двух векторов $\mathbf{g}$ и $(\mathbf{q}_2 - v \mathbf{q}_1)$ равно нулю в случае их коллинеарности, следовательно, вектора отличаются некоторым скалярным множителем $\tilde{A}$:

$$(\mathbf{q}_2 - v \mathbf{q}_1) = \tilde{A} \mathbf{g},$$

откуда $\mathbf{q}_2 = v \mathbf{q}_1 + \tilde{A} \mathbf{g}$. Для нахождения $\tilde{A}$ умножим предыдущее равенство скалярно на $\mathbf{g}$:

$$\mathbf{g}^T \mathbf{q}_2 = \mathbf{g}^T v \mathbf{q}_1 + \tilde{A} \mathbf{g}^T \mathbf{g} \text{ или } u' = v u + \tilde{A} \|\mathbf{g}\|^2,$$

откуда $\tilde{A} = (u' - v u) / \|g\|^2$, где $u = \|g\| \cos \sigma$ и $u' = \|g\| \cos \psi$ равны соответственно косинусам углов падения и преломления, умноженным на длину нормали.

Таким образом, с использованием корректного выражения закона преломления получено соотношение, позволяющее выполнять трассировку лучей через оптические среды как с положительной, так и с отрицательной «правизной»:

$$\begin{aligned} u'^2 &= \|g\|^2 \cos^2 \psi = \|g\|^2 (1 - \sin^2 \psi) = \|g\|^2 (1 - v^2 \sin^2 \sigma) = \\ &= \|g\|^2 [1 - v^2 (1 - \cos^2 \sigma)] = \|g\|^2 - v^2 (\|g\|^2 - \|g\|^2 \cos^2 \sigma). \end{aligned} \quad (17)$$

Результаты компьютерного моделирования

Обобщенное выражение закона преломления (17) положено в основу компьютерной модели трассировки лучей через оптические системы, включающие элементы из метаматериалов, реализованной в визуальной среде разработки C++.

Результаты компьютерного моделирования представлены на рис. 2–5. В качестве прототипа рассмотрена оптическая система трехлинзового объектива «Таир-12» (фокусное расстояние $f' = 150$ мм, относительное отверстие 1:4, поле предмета $2w = 11^\circ$). В первом случае все компоненты объектива выполнены из обычных оптических материалов (рис. 2, а). Во втором случае третий компонент объектива выполнен из метаматериала. Таким образом, двояковогнутая линза исполняет роль собирающей линзы (рис. 2, б).

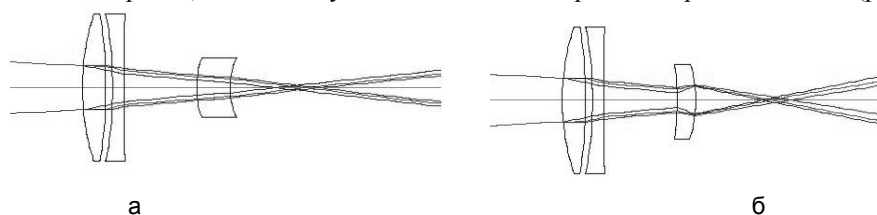


Рис. 2. Прохождение лучей: объектив выполнен из «правых» материалов (а); третий компонент объектива выполнен из метаматериала (б)

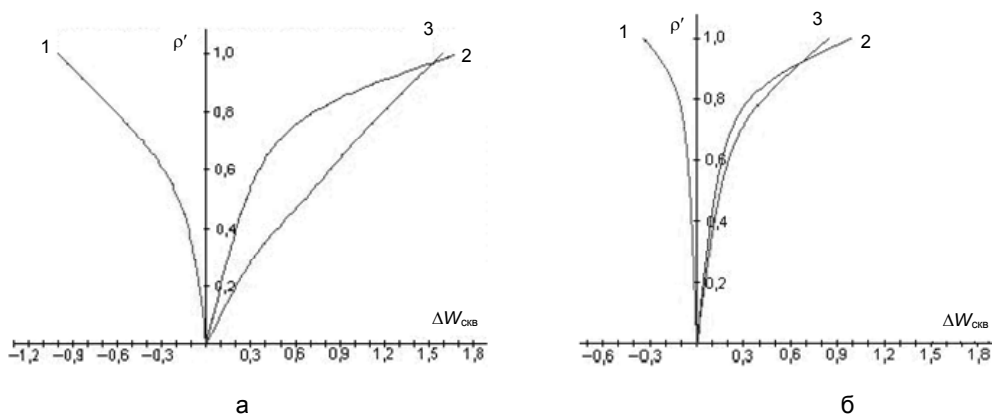


Рис. 3. Волновые aberrации (кривая 1 – $\lambda = 546$ нм; кривая 2 – $\lambda = 480$ нм; кривая 3 – $\lambda = 643$ нм): объектив выполнен из «правых» материалов (а); третий компонент объектива выполнен из метаматериала (б)

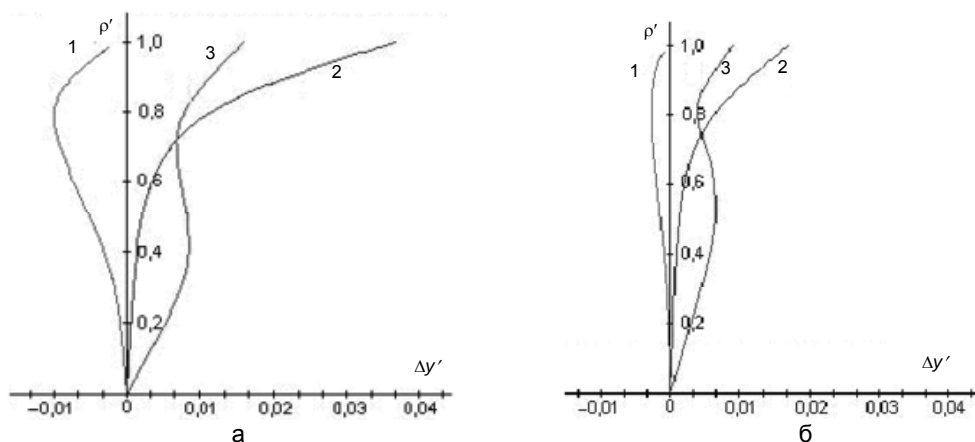


Рис. 4. Поперечные aberrации (кривая 1 – $\lambda = 546$ нм; кривая 2 – $\lambda = 480$ нм; кривая 3 – $\lambda = 643$ нм): объектив выполнен из «правых» материалов (а); третий компонент объектива выполнен из метаматериала (б)

Введение в состав оптической системы элемента из метаматериала способствует компенсации разности хода лучей, что, в свою очередь, приводит к уменьшению сферических aberrаций. Графики волновых, продольных и поперечных aberrаций рассмотренных оптических систем в видимом диапазоне спектра представлены на рис. 3–5. Из графиков aberrаций видно, что после замены третьего элемента объектива на элемент из метаматериала на длине волны $\lambda = 546$ нм волновая aberrация уменьшилась от значения 1,0 до 0,3 длин волн (рис. 3, а, б); поперечная aberrация уменьшилась от значения $-0,010$ до $-0,003$ мм (рис. 4, а, б); продольная aberrация уменьшилась от значения $-0,10$ до $-0,05$ мм (рис. 5, а, б). Аналогичным образом снизились aberrации на границах видимого диапазона (кривые 2 и 3).

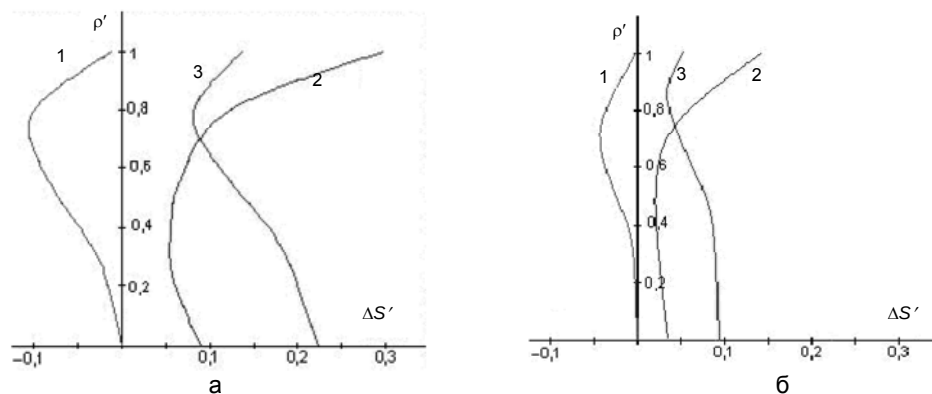


Рис. 5. Продольные aberrации (кривая 1 – $\lambda = 546$ нм; кривая 2 – $\lambda = 480$ нм; кривая 3 – $\lambda = 643$ нм): объектив выполнен из «правых» материалов (а); третий компонент объектива выполнен из метаматериала (б)

Заключение

На основе электромагнитной теории распространения света в оптических средах с различными значениями «правизны» получено обобщенное выражение закона преломления для расчета оптических систем, включающих элементы из метаматериалов. Разработано компьютерное приложение, позволяющее осуществлять трассировку лучей в таких системах, рассчитывать параксиальные характеристики и aberrации в видимом диапазоне. Показано, что введение в состав оптической системы элемента из метаматериала обеспечивает дополнительную возможность компенсации aberrаций и улучшения качества системы в целом.

Литература

1. Веселаго В.Г. Электродинамика веществ с одновременно отрицательными значениями ϵ и μ // Успехи физических наук. – 1967. – № 92. – С. 517–526.
2. Мандельштам Л.И. Полное собрание трудов. – М.: Изд-во АН СССР, 1950. – Т. 5. – 419 с.
3. Eleftheriades G.V., Balmain K.G. Negative-Refractive Metamaterials. Fundamental Principles and Applications. – NJ: John Wiley & Sons Inc, 2005. – 400 p.
4. Cai W., Shalaev V. Optical Metamaterials: Fundamentals and Applications. – Springer, 2010. – 200 p.
5. Кившарь Ю.С., Орлов А.А. Перестраиваемые и нелинейные метаматериалы // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 3 (79). – С. 1–11.
6. Pendry J.B. Negative Refraction Makes a Perfect Lens // Physical Review Letters. – 2000. – V. 85. – № 18. P. 3966–3969.
7. Zang X., Liu Z. Superlenses to Overcome the Diffraction Limit // Nature Materials. – 2008. – № 7. – P. 435–441.
8. Schurig D., Pendry J.B., Smith D.R. Transformation-Designed Optical Elements // Optics Express. – 2007. – V. 15. – № 22. – P. 14772–14782.
9. Li C., Meng X., Liu X. Experimental Realization of a Circuit-Based Broadband Illusion-Optics Analogue // Physical Review Letters. – 2010. – V. 105. – № 23. – P. 233906–233909.
10. Ebbesen T.W., Lezec H.J., Ghaemi H.F., Thio T., Wolff P.A. Extraordinary Optical Transmission Through Sub-wavelength Hole Arrays // Nature. – 1998. – V. 391. – P. 667–669.
11. Белов П.А., Беспалов В.Г., Васильев В.Н., Козлов С.А., Павлов А.В., Симовский К.Р., Шполянский Ю.А. Проблемы когерентной и нелинейной оптики / Под ред. И.П. Гурова и С.А. Козлова. – СПб: СПбГУ ИТМО, 2006. – С. 6–36.
12. Schurig D., Smith D.R. Negative-index Lens Aberrations // Physical Review E. – 2004. – V. 70. – № 6. – P. 065601–065605.

13. Shalaev V., Cai W., Chettiar U., Yuan H., Sarychev A., Drachev V., Kildishev A. Negative Index of Refraction in Optical Metamaterials // Opt. Lett. – 2005. – V. 30. – P. 3356–3358.
14. Zhang S., Fan W., Panoiu N., Malloy K., Osgood R., Brueck S. Experimental Demonstration of Near-Infrared Negative-Index Metamaterials // Phys. Rev. Lett. – 2005. – V. 95. – № 13. – P. 137404–137408.
15. Chen J., Radu C., Puri A. Aberration-Free Negative-Refractive-Index Lens // App. Phys. Lett. – 2006. – V. 88. – P. 071119.
16. Lin Z., Ding J., Zhang P. Aberration-Free Two-Thin-Lens Systems Based on Negative-Index Materials // Chin. Phys. B. – 2008. – V. 17. – P. 954.
17. Sheppard R.H. Metamaterial Lens Design. PhD Dissertation. – The University of Arizona, 2009. – 245 p.

Вознесенская Анна Олеговна

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, voznensenskaya@mail.ifmo.ru

Кабанова Дарья Сергеевна

– ООО «Эксперт-Система», магистр, инженер, kabanova@devexperts.com

УДК 535.317, 628.931

МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ ОПТИЧЕСКОЙ СИСТЕМЫ ДЛЯ ФОРМИРОВАНИЯ ПРЯМОУГОЛЬНОГО СВЕТОВОГО ПЯТНА В НАКЛОННОМ ПУЧКЕ

А.А. Трофимук

Предложено математическое описание и рассчитана оптическая система, формирующая прямоугольное пятно света в широком наклонном пучке при отсутствии наклона источника света и оптической системы. Описание пригодно для использования в процессе автоматической оптимизации и позволяет достичь удовлетворительной равномерности освещенности в световом пятне.

Ключевые слова: светодиод, кривая Безье, прямоугольное световое пятно, неизображающая оптика, уличное освещение.

Введение

В последние годы в связи с возросшими требованиями к энергопотреблению возникает необходимость использования новых источников света, таких как светодиоды. Требования к оптике осветителей также ужесточаются как со стороны производителей оптики, так и потребителя. Перспективным для расчета является использование преломляющих поверхностей свободной формы и оптических элементов на их основе. Требуется также новые методы для расчета неизображающей оптики.

В ряде светотехнических задач (уличное освещение [1], подсветка экранов и т.д.) возникает необходимость получения светового пятна, не лежащего на оси оптической системы и источника света, а также имеющего размеры больше расстояния до источника света. Использование короткофокусных проекторов, вырезающих часть поля формируемого светового пятна, имеет ряд недостатков – большие потери света, сложные оптические схемы, формирование требуемого пятна только в плоскости изображения. Использование наклона оптической системы вместе с источником ведет к искажению формы пятна и перераспределению светового потока по пятну. В связи с этим была поставлена задача предложить математическую модель и рассчитать неизображающую оптическую систему, устраняющую данные недостатки, а также имеющую минимальное количество оптических элементов. Предполагается использование модели в дальнейшем при расчете подобных систем.

Математическое описание начальной формы поверхности

Получим для начала систему, формирующую квадратное пятно, центр которого лежит на оси источника света. Возьмем точечный источник, излучающий по закону Ламберта $I = I_0 \cos(\theta)$, который рассматривается как грубое приближение светодиода. Для создания равномерной освещенности должно выполняться равенство $E = \text{const}$, где E – освещенность в пределах светового пятна. При достаточной удаленности светового пятна от источника света будем иметь $E = \frac{I}{R^2}$. Тогда $E = E(\theta) = I_0 \cos(\theta)/R^2$. Нужно поместить между источником и пятном оптическую систему, например, из двух преломляющих поверхностей, которая реализует условие постоянности освещенности.

Используем следующий алгоритм. Разобьем световой пучок на элементарные световые пучки $I_1 = I_0 \cos(\partial\theta), I_2 = I_0 \cos(2\partial\theta), I_n = I_0 \cos(n\partial\theta)$, где $\partial\theta = \frac{\pi}{2n}$. Пусть каждый пучок (луч) попадает на маленький отрезок прямой. Подберем такие углы наклона этих отрезков, чтобы освещенность,

создаваемая лучами, была приблизительно постоянной на заданном детекторе, а отрезки соприкасались концами друг с другом.

Для трехмерного случая преломления необходимо иметь квадратную форму светового пятна. В данном случае алгоритм усложняется, так как добавляется условие квадратности формы пятна, а отрезки заменяются плоскими поверхностями. В общем случае из-за симметрии можно взять четверть телесного угла источника света и четверть поверхности. В итоге получаем поверхности, формирующие пятно света с приблизительно равномерной освещенностью от точечного источника. На рис. 1, а, показана полученная освещенность для приемника, имеющего разрешение 9×9 пикселей, а на рис. 1, б – освещенность для приемника размером 25×25 пикселей. Провалы на графиках обусловлены фасетчатой структурой поверхностей и грубостью расчета.

Для протяженного источника света, каковым является светодиод, этот метод весьма сложен [2], поэтому необходимо использовать автоматическую оптимизацию, в результате которой будет статистически найдено компромиссное решение для протяженного источника, а поверхность будет иметь гладкую форму, без провалов и ступенек в распределении освещенности. На основе полученного приближения точечного источника было предложено описать поверхность в виде комбинации кривых Безье, которые получают распространение при описании поверхностей неизображающих оптических систем [3, 4].

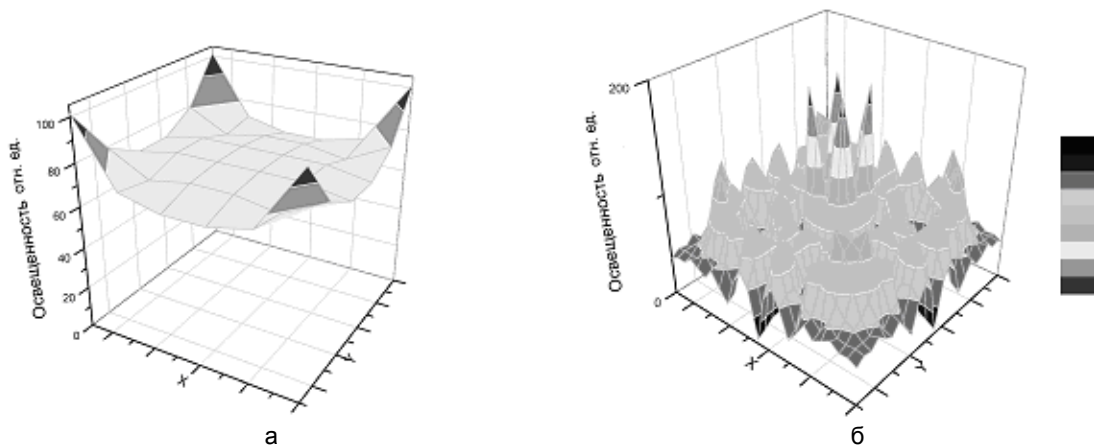


Рис. 1. Освещенность, рассчитанная с помощью алгоритма для точечного источника, на приемнике размером 9×9 пикселей (а); 25×25 пикселей (б)

Как известно [5], параметрическая кривая Безье в декартовых координатах $\mathbf{B}(t) = \{Bx(t), By(t)\}$, $t \in [0, 1]$ определяется всего лишь набором n контрольных точек $\mathbf{P}_i = \{(Px_i, Py_i), i = 1..n\}$ и имеет следующее математическое описание:

$$\mathbf{B}(t) = \sum_{i=0}^n \mathbf{b}_{i,n}(t) \mathbf{P}_i, t \in [0, 1], \quad (1)$$

где $\mathbf{b}_{i,n}(t) = \frac{n!}{i!(n-i)!} t^i (1-t)^{n-i}$, $i = 0..n$ – базисный полином Бернштейна. В полярных координатах кривая

Безье (иногда ее называют п-Безье) $C(t)$ имеет следующее выражение [6]:

$$C(t) = \begin{cases} \rho(t) = 1 / \sum_{i=0}^n (1/c_i) A_{i,n}(t), t \in [-\Delta, \Delta], 2n\Delta < \pi, \\ \theta(t) = nt \end{cases} \quad (2)$$

где $A_{i,n} = \frac{1}{\sin(2\Delta)^n} \frac{n!}{i!(n-i)!} \sin^{n-i}(\Delta-t) \sin^i(\Delta+t)$ – тригонометрический полином Бернштейна; c_i –

контрольные точки кривой п-Безье. В качестве уравнения преломляющей поверхности была предложена поверхность вращения кривой Безье, заданной выражением (1) вдоль направляющей, определяемой уравнением (2). Результирующая поверхность вследствие симметрии состоит из четырех одинаковых поверхностей $\mathbf{F}(u, v)$, имеющих следующее параметрическое уравнение в декартовых координатах:

$$\mathbf{F}(u, v) = \begin{cases} F_x(u, v) = \sqrt{B_y(u) \rho(v)} \cos(\theta(v)) \\ F_y(u, v) = \sqrt{B_y(u) \rho(v)} \sin(\theta(v)), u \in [0, 1], v \in [-\pi/4, \pi/4]. \\ F_z(u, v) = B_x(u) \end{cases} \quad (3)$$

Рассчитанная оптическая система, формирующая квадратное пятно

Для расчета были взяты следующие данные: светодиод Luxeon Rebel White, размер освещаемой площадки 40×40 м, расстояние до площадки 10 м. Поверхность была описана в программе Zemax при помощи пользовательской dll, оптимизация проводилась также в программе Zemax методом «orthogonal descent» [3], более эффективным для такого рода систем, чем метод наименьших квадратов. В качестве переменных использовались контрольные точки кривых Безье. Так как значения точек имеют одинаковый порядок малости, то их вклад в функцию качества примерно одинаков при одинаковых весах, что позволяет избежать использования весовых коэффициентов разных порядков при оптимизации. Габариты рассчитанной системы $15 \times 15 \times 5$ мм. Назовем ее линзой А, она изображена на рис. 2. Материал линзы – органическое стекло. На рис. 3 показано полученное распределение освещенности.

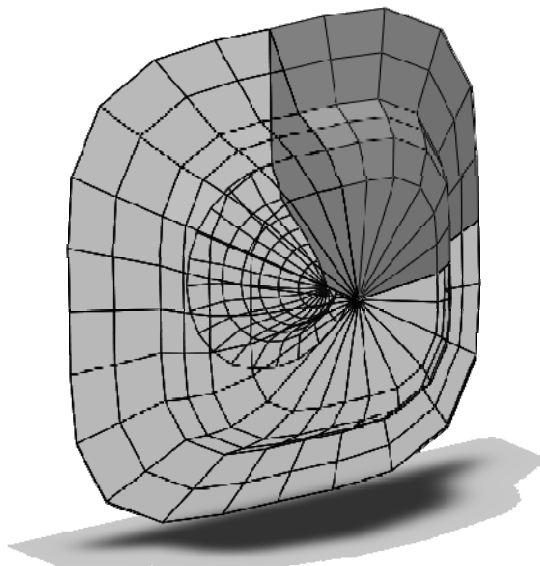


Рис. 2. Форма линзы А, рассчитанная четверть выделена более темным цветом. Масштаб 4:1

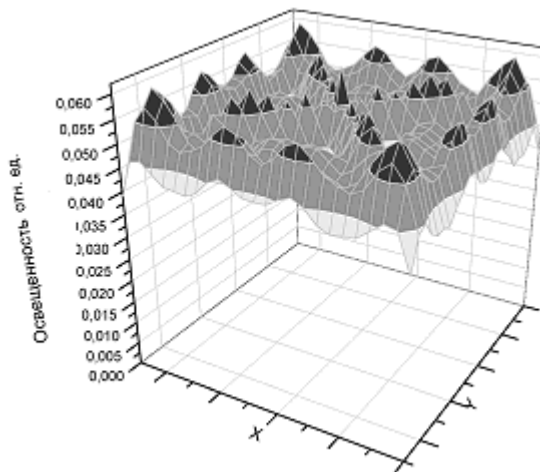


Рис. 3. Освещенность пятна, формируемого линзой А, в относительных единицах

Таким образом, при помощи описанной методики можно рассчитать линзу А, формирующую квадратное пятно с удовлетворительной равномерностью освещенности. Немаловажно, что форма и структура пятна сохраняются на всем протяжении пучка света. Эффективность данной системы составляет 90%, потери возникают за счет френелевского отражения.

Вариант оптической системы для наклонных пучков

Как было показано выше, освещенность для линзы А зависит практически только от углов распространения света, так как мы формируем распределение освещенности, меняя распределение силы света. Значит, ограничивая углы лучей от источника света, можно получать различные формы световых пятен, причем световые пучки будут наклонными. Рассчитаем линзу А с условиями, указанными выше, но с размером пятна 20×40 м. Для этого будем использовать половину поверхности на рис. 3, т.е. она будет

формировать освещенность для светового потока в телесном угле π . Часть света в оставшемся угле π направим вверх при помощи зеркала Б, расположенного ниже линзы А (рис. 4). Форма зеркала описывается поверхностью вращения кривой Безье [7].

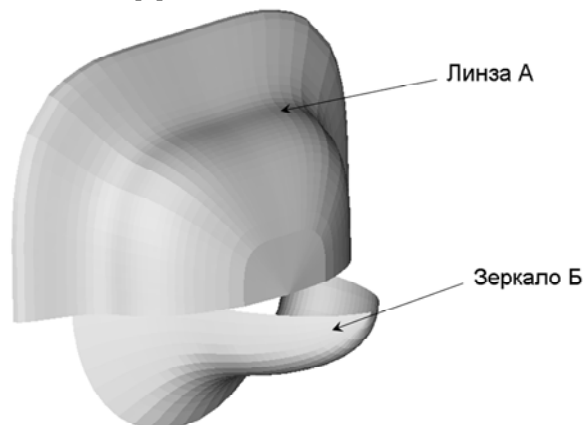


Рис. 4. Оптическая схема для получения прямоугольного пятна. Масштаб 4:1 с сохранением пропорций

В связи с тем, что зеркало полностью не отобразит половину источника света в линзу А, снова требуется искать компромисс при помощи оптимизации и менять как зеркало Б, так и линзу А. Оба элемента в принципе могут быть объединены в один и изготовлены методом литья под давлением. Полученное в результате моделирования распределение освещенности показано на рис. 5.

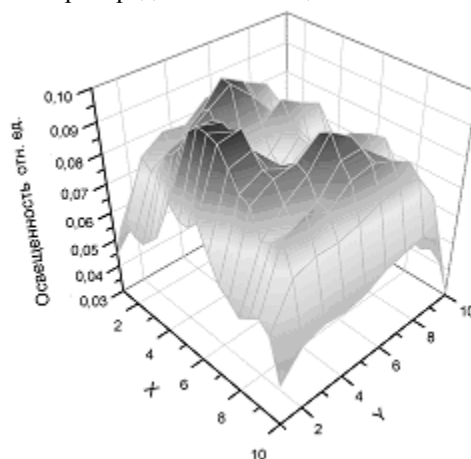


Рис. 5. Освещенность в относительных единицах для светового пятна, формируемого половиной линзы А и зеркалом Б

Заключение

Предложено математическое описание систем для формирования световых пятен прямоугольной и квадратной формы. Рассчитанная оптическая система позволяет формировать пятно с соотношением сторон 2:1. Для получения других соотношений необходимо использовать более сложные преломляющие поверхности, но пригодные для автоматизированного расчета. На основе этой схемы были реализованы светильники уличного освещения типа I и типа III, согласно требованиям [1]. Линза, изображенная на рис. 2 и имеющая поверхности, описываемые формулой (3), позволяет реализовать светильник типа V [1]. Предложенное математическое описание может быть также использовано для автоматизированного расчета осветителей иного назначения.

Литература

1. ANSI IESNA RP-8-00. Practice for Roadway Lighting. – Approved 27.06.2000. – American National Standard Institute Inc. – 70 p.
2. Bortz J., Shatz N. and D. Pitou. Optimal design of a nonimaging projection lens for use with an LED source and a rectangular target // Proceeding of SPIE. – 2000. – V. 4092. – P. 130–138.
3. Kudaev S., Schreiber P. Automated optimization of non-imaging optics for luminaries // Proceedings of SPIE. – 2006. – V. 5942. – P. 87–95.
4. Trofimuk A. Ray tracing illuminates custom design of LED light sources // Laser Focus World. – 2008. – № 11. – P. 51–53.

5. Piegls L., Tiller W. The NURBS Handbook. – Springer, Berlin, 1997. – 639 p.
6. Hill J.M.D., Skillicorn D.B., Casciola G., Morigi S., Sanchez-Reyes J. Degree elevation for p-Bezier curves // Computer Aided Geometric Design. – April 1998. – V. 15. – № 4. – P. 313–322.
7. Trafimuk A. How to Create Surfaces of Revolution via User Defined Objects // Zemax knowledge base [Электронный ресурс]. – Режим доступа <http://www.radiantzemax.com/kb-en/Knowledgebase/How-to-Create-Surfaces-of-Revolution-via-User-Defined-Objects>, свободный. Яз. англ. (дата обращения 13.06.2012).

Трофимук Анатолий Андреевич

– Ellis Amalgamated LLC (Беларусь), инженер-конструктор, аспирант, anatoli@opticsforhire.com

УДК 549.651.24; 535.4

ПЕРИОД ПЕРИСТЕРИТОВОЙ РЕШЕТКИ И ИНТЕРФЕРЕНЦИОННАЯ МОДЕЛЬ ИРИЗАЦИИ

А.П. Симаков, А.В. Смирнов, Б.А. Федоров, М.А. Иванов

Разработана численная интерференционная модель явления иризации плагиоклазов. На основе интерференционной модели рассчитаны спектры иризации для группы эталонных образцов лабрадора с разными цветами иризации. Модельные спектры эталонных образцов соответствуют экспериментальным в пределах погрешности 10–15%. По результатам моделирования обоснована возможность иризации плагиоклазов вне видимого диапазона (в ультрафиолетовой и ближней инфракрасной областях). Экспериментально подтверждено наличие интерференционных максимумов в ближней инфракрасной области.

Ключевые слова: иризация, плагиоклаз, перистеритовая решетка, интерференционная модель.

Введение

Сведения об иризации плагиоклазов встречаются в литературе, начиная с конца XIX в. Характерно, что до 1954 г. исследователи основное внимание уделяли оптическим особенностям иризации, в результате чего были выявлены главные ее закономерности, и было выдвинуто несколько гипотез о физической сущности явления. Вопрос же о его *минералогической причине*, т.е. об особенностях минеральных фаз, на которых происходит иризация, занимает основное место в работах, начиная с 50-х годов прошлого века.

В рентгеноструктурных и электронно-микроскопических исследованиях этого периода [1, 2] наличие иризации рассматривалось как подтверждение явления распада твердого раствора в плагиоклазах. Иризация сама по себе не рассматривалась, а известные ее закономерности не учитывались. В то же время электронно-микроскопическими исследованиями было установлено, что иризирующие кристаллы представляют собой совокупность тонких плоскопараллельных пластинок, размеры которых практически бесконечны по сравнению с их толщиной. Отсюда следует, что явление иризации может быть результатом интерференции лучей, отраженных от поверхностей раздела между пластинками.

Одним из наиболее полных аналитических обзоров в этой области до сих пор остается работа Н.К. Крамаренко [3]. В ней предполагается, что состав пластинок каждого типа остается постоянным для близких по валовому составу образцов. В этом случае интенсивность пучков, отраженных от единичных поверхностей раздела пластинок, будет оставаться (пренебрегая поглощением света в пластинке) постоянной. В той же работе автором представлены уравнения, описывающие явление иризации как результата интерференции света на стопке перистеритовых пластин.

Однако модель Крамаренко, основанная на ряде допущений и упрощений, недостаточна для создания на ее основе строгой интерференционной модели явления иризации:

- в ней не используются реальные (полученные экспериментально) толщины перистеритовых пластин и порядок идентичности;
- модель не предусматривает алгоритм расчета интерференционной картины для бесконечного количества пластин;
- все выкладки ограничиваются частными случаями для пар перистеритовых пластин разного типа;
- типизация пластин основывается на их химическом составе (основности), а не на оптических характеристиках, таких как показатели преломления и поглощения;
- недостаточно полно учтены поляризационные эффекты при отражении света стопкой перистеритовых пластин.

Представленная в настоящей работе интерференционная модель иризации, напротив, использует параметры перистеритов (толщины пластин и период идентичности), полученные экспериментальным путем, и перистеритовые пластины типизируются по показателю преломления. Такой подход к анализу явления иризации в лабрадорах ранее был предложен в диссертации одного из авторов [4]. Расчет интерференционной картины может быть произведен для любого количества слоев. Слои разных типов при

этом могут иметь разную толщину. Такая модель не только позволяет оценить вклад в природу иризации различных конституционных особенностей кристаллов лабрадора, но и дает возможность решить обратную задачу: по известному интерференционному спектру определять толщину перистеритовых пластин. В перспективе такие данные могут быть использованы для оценки физических условий посткристаллизационных изменений плагиоклазов. Подобные задачи ранее решались для просветляющих покрытий и многослойных фильтров [5, 6], но, как правило, для случая нормального падения лучей на многослойную систему. Используемый в настоящей работе метод позволяет производить расчет интерференционной картины при произвольном угле падения луча.

Методика расчета

Для предлагаемой интерференционной модели иризации был использован рекуррентный метод расчета оптических характеристик многослойных систем [7], основанный на формулах Френеля.

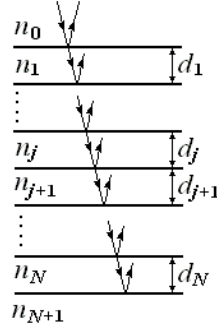


Рис. 1. Взаимное расположение границ

Рассмотрим систему, состоящую из N плоских прозрачных слоев с показателями преломления n_j и толщинами d_j ($j = 1, 2, \dots, N$). Система ограничена сверху средой с показателем преломления n_0 , снизу – средой с показателем преломления n_{N+1} (рис. 1). Свет, падающий из среды n_0 , частично отражается и частично преломляется на каждой из $N+1$ границ.

Амплитудные коэффициенты отражения и пропускания на границах различаются для компонент, поляризованных в плоскости падения (p -компонента) и перпендикулярно плоскости падения (s -компонента). Значения этих коэффициентов $r_{j,j+1}$ и $t_{j,j+1}$ на границе раздела между слоями n_j и n_{j+1} (рис. 2) можно найти по формулам Френеля в следующем виде [7]:

$$r_{j,j+1} = \frac{q_j - q_{j+1}}{q_j + q_{j+1}}, \quad t_{j,j+1} = \frac{2q_j}{q_j + q_{j+1}},$$

где $q_j = \begin{cases} n_j / \cos(\alpha_j), & \text{для } p\text{-компоненты;} \\ n_j \cos(\alpha_j), & \text{для } s\text{-компоненты.} \end{cases}$

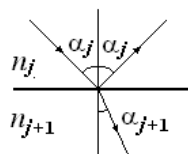


Рис. 2. Отражение и преломление на одной из промежуточных границ

Коэффициент отражения для всей структуры может быть найден с помощью рекуррентного метода, описанного в монографии [7]. На первом шаге вычисляется коэффициент $r_{N-1, N+1}$ отражения от N -го слоя в предположении, что сверху от него находится среда с показателем преломления n_{N-1} , при этом учитывается интерференция волн, отраженных от обеих границ слоя. Далее вычисляется коэффициент отражения $r_{N-2, N+1}$ от системы из $(N-1)$ -го и N -го слоев как результат интерференции волн, отраженных от N -го слоя и от верхней границы $(N-1)$ -го слоя. Коэффициент отражения $r_{j, N+1}$ для системы слоев с номерами $j, j+1, \dots, N$ выражается через коэффициент отражения $r_{j, N+1}$ для системы слоев с номерами $j+1, j+2, \dots, N$ и коэффициент отражения $r_{j-1, j}$ от верхней границы j -го слоя. Коэффициент отражения $r_{j-1, N+1}$ рассчитывается по рекуррентной формуле:

$$r_{j-1, N+1} = \frac{r_{j-1, j} + r_{j, N+1} \exp(-2i\varphi_j)}{1 + r_{j-1, j} r_{j, N+1} \exp(-2i\varphi_j)} \quad (j = N-1, N-2, \dots, 1),$$

где $\varphi_j = \frac{2\pi}{\lambda} n_j d_j \cos(\alpha_j)$.

Амплитудный коэффициент отражения $r_{0,N+1}$ для всей структуры вычисляется на последнем шаге указанной рекуррентной процедуры.

Коэффициент отражения R по интенсивности определяется соотношением

$$R = |r_{0,N+1}|^2.$$

Для расчета оптических характеристик перистеритовых структур в лабрадорах необходимо учесть особенности ориентировки перистеритов в кристалле. Стопка перистеритовых пластин ориентирована по плоскости (081) и образует угол δ с плоскостью второго пинакоида (010). Величина угла δ находится в интервале 5° – 13° и зависит от валового соотношения Ca и Na в конкретном кристалле. Точное значение δ может быть определено экспериментально с помощью гониометра. Таким образом, луч света, падая на кристалл, сначала проходит через клин, частично поляризуется и только после этого попадает на стопку перистеритовых пластин (рис. 3), которые представляют собой последовательно чередующиеся слои натриевой (альбит Na $[\text{Al Si}_3\text{O}_8]$) и калиевой (анортит Ca $[\text{Al}_2\text{Si}_2\text{O}_8]$) фаз распада твердого раствора. Пластины разного состава близки по толщине и имеют приблизительно одинаковую протяженность. Доминирующую толщину пластин d называют периодом идентичности перистеритовой структуры (решетки).

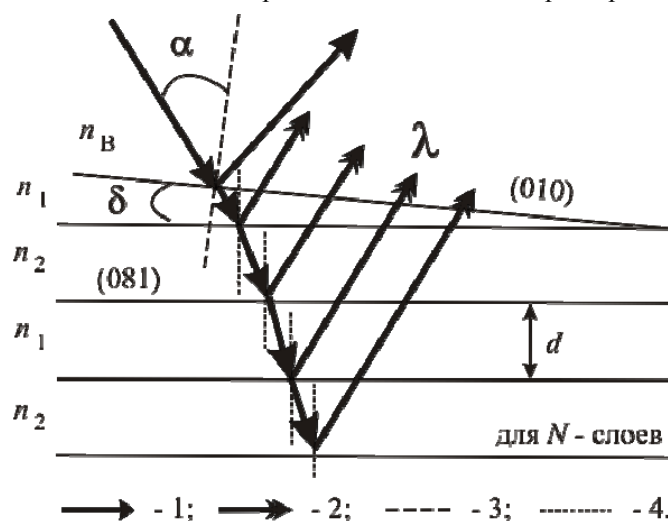


Рис. 3. Схема интерференции света в кристалле лабрадора: 1 – ход лучей сложного света; 2 – ход интерференционно усиленных лучей иризационного спектра; 3 – перпендикуляр к плоскости раздела сред n_B и n_1 ; 4 – перпендикуляры к плоскостям раздела сред; α – угол падения первичного луча сложного света на плоскость (010) кристалла; n_1 и n_2 – показатели преломления кальциевой и натриевой фаз распада; n_B – коэффициент преломления воздуха; δ – угол клина; d – период идентичности перистеритовой структуры; N – число слоев

Присутствие клина ограничивает диапазон углов падения света непосредственно на стопку пластин. Максимальный угол падения $\alpha_{\max}$ может быть рассчитан по формуле

$$\alpha_{\max} = \arcsin\left(\frac{n_B}{n_1}\right) - \delta.$$

Результаты и их обсуждение

Входными параметрами для расчета единичного интерференционного спектра плагиоклаза (единичным считаем спектр, полученный при монохроматическом освещении многослойной системы под углом от 0° до угла полного внутреннего отражения) являются следующие величины: λ – длина волны; α – угол падения луча на плоскость спайности кристалла (010) (рис. 3); d – период перистеритовой решетки; n_1 , n_2 – показатели преломления для соответствующих фаз твердого раствора; N – количество слоев в системе.

Описанная процедура была использована для расчета единичного интерференционного спектра плагиоклаза. Полученная зависимость суммарной интенсивности интерферирующих волн от угла падения первичного луча $I = f(\alpha)$ может быть визуализирована в виде двумерного графика в системе координат $(I_i; \alpha_i)$, где i – порядковый номер дискретного значения угла падения при выбранном шаге изменения этого угла $\Delta\alpha$.

Иризация, однако, возникает при падении на образец не монохроматического, а полихроматического, белого света. По этой причине для более адекватного восприятия результаты расчета лучше пред-

ставлять в виде матрицы значений интенсивности $I_{i,j}$ в координатах $(\alpha_i; \lambda_j)$, где j – порядковый номер дискретного значения длины волны при выбранном шаге ее изменения $\Delta\lambda$. Визуализировать такие матрицы удобно в виде трехмерного графика в координатах $(\alpha_i; \lambda_j; I_{i,j})$.

В качестве эталонных образцов для проведения расчета были взяты четыре образца лабрадора с финского месторождения Юламаа (Ylämaa): неиризирующий, с синей иризацией, с зеленой иризацией и с красной иризацией.

Спаянные сколы по (001), полученные от эталонных образцов, протравливались в растворе соляной кислоты по специальной методике [3] до проявления рельефа перистеритов. Для изготовленных таким образом препаратов в лаборатории кафедры физики Российского государственного педагогического университета им. Герцена (г. Санкт-Петербург) на сканирующем зондовом микроскопе «СОЛВЕР Р47-PRO» (Россия) полуконтактным методом были получены снимки поверхности рельефа травления, по профилям которых были определены значения d .

Определению периода d идентичности перистеритовой решетки предшествовало решение вопроса о способе соотнесения рельефа травления с перистеритовой неоднородностью кристалла. Принято считать, что перистеритовые пластинки разной основности растворяются при травлении с разной скоростью. Таким образом, положению пластинок одного типа на профиле рельефа должны соответствовать «провалы» (минимумы), а положению пластинок другого типа – «возвышения» (максимумы) [3]. Однако при травлении подобной слоистой системы на субмикронном уровне большое значение приобретает кривизна поверхности раздела растворяемой фазы и растворителя [8]. Соотношение растворимостей (или скоростей растворения) участка фазы с плоской поверхностью и участка той же фазы с криволинейной поверхностью описывается уравнением Кельвина [9]. Иными словами, значение d , определенное по описанной Крамаренко модели, будет в два раза меньше значения d , полученного исходя из модели травления Кельвина.

Для выбора способа интерпретации рельефа травления был проведен расчет модельных спектров эталонного образца с синей иризацией по периодам идентичности перистеритовой решетки, соответствующим разным способам интерпретации: $d = 75$ нм (интерпретация по Крамаренко) и $d = 150$ нм (интерпретация по Кельвину). Моделирование спектров проводилось в диапазоне $N = 2$ –1000 слоев.

Известно, что при интерференции света на стопке пластин положение пиков определяется периодом идентичности стопки, а форма пиков и их интенсивность – количеством пластин в стопке [10]. Иначе говоря, при увеличении расчетного количества слоев на модельном спектре максимум интерференции должен увеличиваться по высоте с одновременным уменьшением по ширине.

На спектрах, рассчитанных для $d = 150$ нм, с увеличением N происходит уменьшение ширины пиков с резким увеличением интенсивности в ультрафиолетовой и инфракрасной областях. В видимой же части спектра не наблюдается заметного повышения интенсивности пиков. Большинство максимумов, расположенных в этой области, с увеличением числа слоев, напротив, теряет интенсивность, и только при большом количестве слоев становятся заметны несколько отдельно стоящих максимумов. Положение максимумов в видимой части спектра соответствует синей и зелено-синей областям.

Спектры, рассчитанные для $d = 75$ нм, демонстрируют иную картину. При $N = 2$ наблюдается один широкий максимум с низкой интенсивностью. С увеличением расчетного числа слоев он резко увеличивается по интенсивности и становится заметно уже. При большом количестве слоев ($N > 1000$) наблюдается только непрерывная серия узких максимумов с высокой интенсивностью. Разрывы между пиками в данном случае объясняются дискретностью рекуррентной функции, лежащей в основе представленной модели [7]. По своему положению линия максимумов также соответствует синей и зелено-синей области спектра.

Таким образом, оба варианта расчета дают спектры с характеристическими пиками в области спектра, близкой к визуально наблюдаемой иризационной картине. Если говорить об общих закономерностях, свойственных интерференционным спектрам, то во втором случае ($d = 75$ нм) они проявляются заметно лучше. Исходя из этого, можно заключить, что рельеф травления эталонных образцов является результатом различной скорости растворения перистеритовых пластин разной основности. Полученные таким образом данные свидетельствуют об увеличении периода d перистеритовой решетки с увеличением длины волны иризации в эталонных образцах лабрадора: неиризирующий – $d = 54 \pm 5$ нм, иризирующий синим – $d = 75 \pm 5$ нм, иризирующий зеленым – $d = 98 \pm 5$ нм и иризирующий красным – $d = 231 \pm 5$ нм.

Модельный расчет первого приближения для этих образцов был проведен при $N = 1000$ слоев. Дальнейшее увеличение количества слоев не влияло на суммарную интенсивность и положение интерференционных максимумов. Для первого приближения в разработанной модели были приняты следующие допущения:

- поглощение образца учитывалось путем искусственного ограничения расчетного количества слоев;

– угол δ принимался равным 10° . Результаты расчета интерференционных спектров для эталонных образцов представлены на рис. 4.

Для иризирующих образцов (рис. 4, б–г) в видимой области модельных спектров наблюдаются непрерывные линии максимумов. Положение этих линий соответствует тем областям спектра, в которых наблюдается иризация в каждом конкретном образце. Иначе говоря, расчетные интерференционные картины, полученные для иризирующих образцов эталонной группы, соответствуют (с учетом относительной спектральной чувствительности глаза) наблюдаемым визуально.

Стоит также отметить, что во всех модельных спектрах иризирующих плагиоклазов линии максимумов захватывают краевые части соседних цветовых зон. Это объясняет отсутствие в иризационных картинах плагиоклазов чистых (ньютоновских) цветов. В данном случае происходит сложение длин волн соседних диапазонов, в результате чего синий цвет приобретает оттенки фиолетового и зеленого, зеленый – синего и желтого, желтый – зеленого и красного, красный – желтого, оранжевого и карминного.

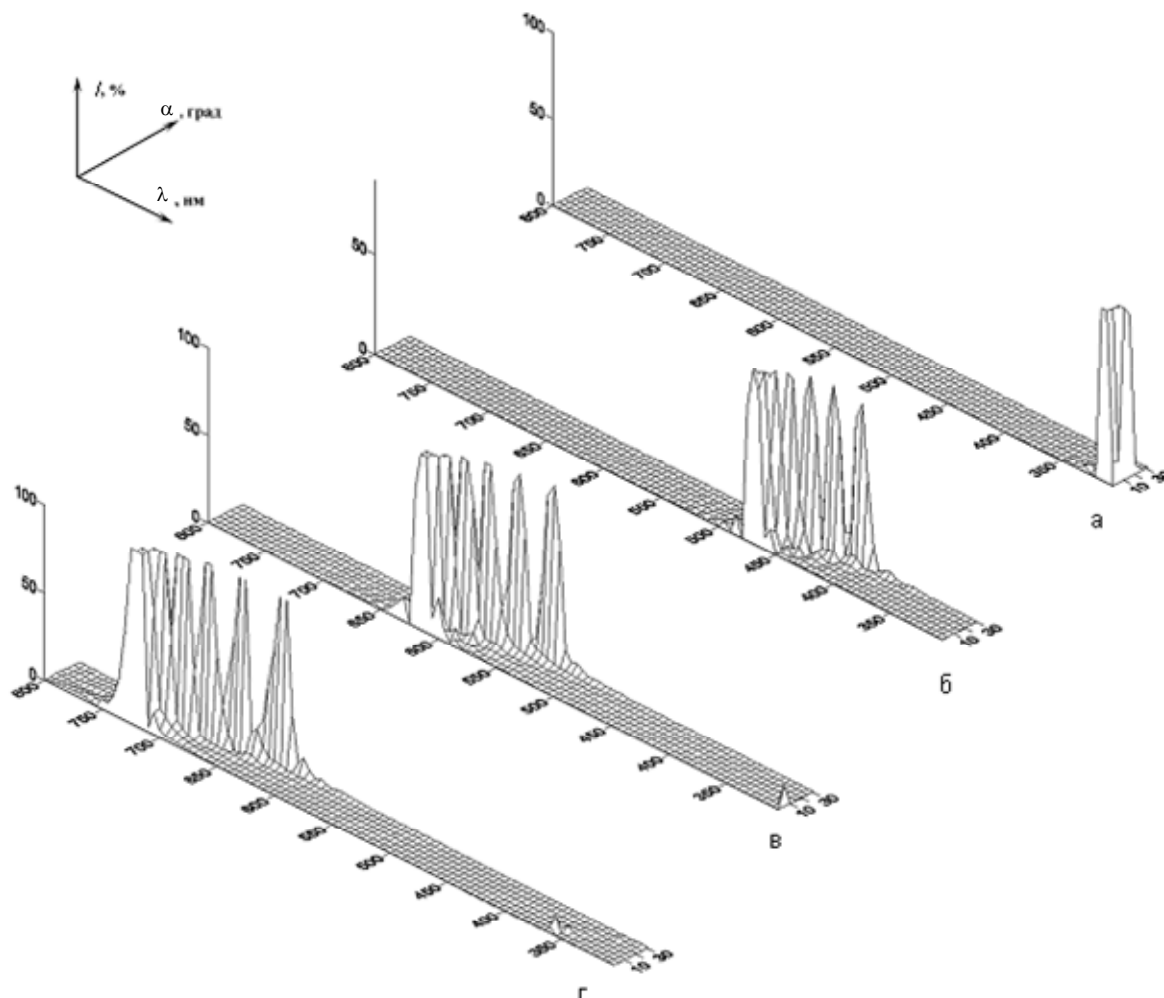


Рис. 4. Модельные спектры эталонной группы проб лабрадора: неиризирующий (а); с фиолетово-синей иризацией (б); с желтовато-зеленой иризацией (в); с карминно-красной иризацией (г)

Особого внимания заслуживают спектры неиризирующего лабрадора и лабрадора с красной (карминно-красной) иризацией (рис. 4, а, г, соответственно). Спектр неиризирующего лабрадора, как и следовало ожидать, не имеет максимумов интерференционного усиления в видимой области. В то же время можно наблюдать четкий пик в интервале 300–325 нм, что соответствует ультрафиолетовой области. Линия максимумов на спектре лабрадора с красной иризацией также частично находится вне видимой области спектра – в ближнем инфракрасном диапазоне.

Для проверки достоверности расчетных диапазонов интерференционного усиления были получены и усреднены по углам экспериментальные спектры иризационных картин (рис. 5). Соответствие диапазонов в данном случае определяется близостью значений λ середины диапазона интерференционного усиления на модельном спектре с положением максимальной по интенсивности точки на экспериментальном спектре [11]. Наложение спектров показало, что области интерференционного усиления на экспериментальных и модельных спектрах совпадают в пределах погрешности 10–15%. Поскольку модель-

ный расчет проводился для идеализированной перистеритовой решетки, можно считать, что разработанная модель вполне адекватно описывает реально наблюдаемые интерференционные картины.

Спектрофотометрические исследования также подтвердили наличие интерференционного эффекта в ближней инфракрасной области для образца с карминно-красной иризацией.

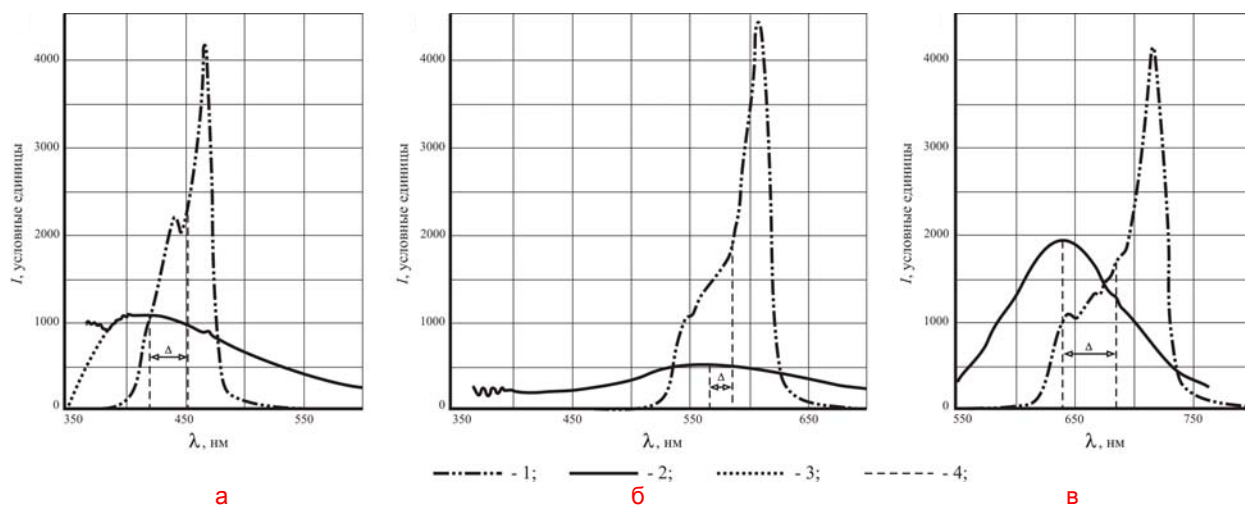


Рис. 5. Сравнение модельных интерференционных спектров с основными пиками экспериментальных спектров для лабрадоров эталонной группы проб: лабрадор с фиолетово-синей иризацией (а); лабрадор с желтовато-зеленой иризацией (б); лабрадор с карминно-красной иризацией (в). 1 – модельный спектр; 2 – усредненный по углам экспериментальный спектр; 3 – экстраполяция экспериментального спектра; 4 – положение середины диапазона модельных спектров и пика экспериментальных спектров; Δ – разность положения сопоставляемых линий. Экспериментальные спектры получены на спектрофотометре «Fluorolog» с микроскопом «Olimpus BX 51» (аналитик Е.А. Васильев)

Заключение

Алгоритм расчета спектров иризации, представленный в данной работе, позволяет перейти от общих представлений о природе иризации к численному моделированию этого явления. В то же время разработанная модель не претендует на полное математическое описание явления иризации в рамках интерференционной концепции.

Приведенные в работе результаты демонстрируют достаточно корректное описание интерференционных спектров и подтверждают зависимость цвета иризации от толщины перистеритов.

Обоснована возможность интерференционного усиления волн в ультрафиолетовой области для плагиоклазов, не проявляющих иризацию в видимом диапазоне. Экспериментально подтверждено наличие интерференционного эффекта в ближней инфракрасной области для плагиоклазов, иризирующих в верхнем красном диапазоне видимого спектра.

Таким образом, представленная модель позволяет расширить имеющиеся представления о природе явления иризации. Она объясняет изменчивость проявления иризационного эффекта в видимом диапазоне. При дальнейшем совершенствовании настоящей модели, возможно, удастся решить и обратную задачу – оценить размеры перистеритов по экспериментально полученному интерференционному спектру.

Литература

1. Baier E., Pense J. Elektronenmikroskopische Untersuchungen an Labradoren // Naturwissenschaften. – 1957. – V. 44. – № 5. – P. 110–111.
2. Laves F., Nissen H.-U., Bollmann W. On schiller and submicroscopical lamellae of labradorite, (Na, Ca) (Si, Al)₃ O₈ // Naturwissenschaften. – 1965. – V. 52. – № 14. – P. 427–428.
3. Крамаренко Н.К. Фазовый состав, пластинчатое строение и иризация плагиоклазов. – Киев, Наукова думка, 1975. – 108 с.
4. Симаков А.П.: Автореф. дис. ... кандидата геолого-минералогических наук [Электронный ресурс]. – Режим доступа: http://www.spmi.ru/system/files/lib/sci/aspirant-doctorant/avtoreferaty/2011/simakov_ar.pdf, свободный. Яз. рус. (дата обращения 06.06.2012).
5. Фурман Ш.А. Тонкослойные оптические покрытия. – Л.: Машиностроение, 1976. – 264 с.
6. Путилин Э.С. Оптические покрытия. – СПб: СПбГУ ИТМО, 2005. – 201 с.
7. Furman Sh.A., Tikhonravov A.V. Basics of optics of multilayer systems. – Edition Frontieres, Gif-sur-Yvette, 1992. – 104 p.
8. Wang Z.L., Liu Y., Zhang Z. Handbook of Nanophase and Nanostructured Materials: Materials Systems and Applications. – N.Y.: Kluwer Academic Press, 2002. – 265 p.

9. Вережников В.Н. Избранные главы коллоидной химии. – Воронеж: Изд-во ВГУ, 2011. – 237 с.
10. Бутиков Е.И. Оптика: Учебное пособие. – СПб: Лань, 2012. – 608 с.
11. Симаков А.П., Иванов М.А., Васильев Е.А. Исследование иризации плагиоклазов в ближней инфракрасной области // Материалы XVII Международного совещания «Кристаллохимия, рентгенография и спектроскопия минералов – 2011». – СПб, 2011. – С. 187–188.

<i>Симаков Антон Петрович</i>	– Национальный минерально-сырьевой университет «Горный», кандидат геолого-минералогических наук, научный сотрудник, simakov@spmi.ru
<i>Смирнов Александр Витальевич</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, доцент, smirnav@phd.ifmo.ru
<i>Федоров Борис Александрович</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, borfedorov@rambler.ru
<i>Иванов Михаил Александрович</i>	– Национальный минерально-сырьевой университет «Горный», доктор геолого-минералогических наук, проректор, ivan@spmi.ru

УДК 621.3.049.77+534.08+681.787

МОДИФИКАЦИЯ СХЕМЫ ОБРАБОТКИ ДАННЫХ ФАЗОВОГО ИНТЕРФЕРОМЕТРИЧЕСКОГО АКУСТИЧЕСКОГО ДАТЧИКА

М.Ю. Плотников, И.Г. Дейнека, И.А. Шарков

Представлены результаты разработки цифровой схемы высокоскоростной записи данных на micro SD-карту памяти с блока электронной обработки сигналов волоконно-оптического акустического датчика интерферометрического типа на базе программируемой логической интегральной схемы с использованием программной среды разработки Xilinx ISE. Приводятся результаты эксперимента, подтверждающие работоспособность предложенной схемы при скорости записи данных 141000 бит/с.

Ключевые слова: фазовый интерферометрический датчик, цифровая обработка, запись данных, карта памяти.

Введение

Одним из активно развиваемых направлений современного оптико-электронного приборостроения является создание волоконно-оптических датчиков [1, 2]. Они обладают существенными преимуществами по сравнению с традиционными пьезоэлектрическими датчиками – высокой чувствительностью, большим динамическим диапазоном, устойчивостью к электромагнитным помехам, а также пригодны к эксплуатации в условиях повышенной температуры, влажности и радиации.

В большинстве современных волоконно-оптических датчиков используются цифровые методы приема и обработки сигналов, позволяющие значительно расширить их функциональные возможности и увеличить производительность операций приема и обработки данных по сравнению с традиционными аналоговыми схемами обработки. В качестве основного элемента схемы обработки все чаще используются программируемые логические интегральные схемы (ПЛИС) [3, 4]. Эта тенденция обусловлена тем, что изменения в функциональности и производительности схемы обработки могут быть обеспечены за счет изменений в конфигурации ПЛИС, что гораздо проще, быстрее и дешевле, чем замена компонент в электрической схеме или разработка новой платы. Таким образом, на основе ПЛИС могут быть реализованы сложные методы цифровой обработки сигналов, модификация которых может осуществляться путем перепрограммирования.

Поскольку современные ПЛИС обладают высоким быстродействием и выполнение операций на их основе может вестись на частотах в сотни МГц, могут возникнуть проблемы с передачей информации с платы обработки конечному пользователю ввиду ограниченной скорости работы традиционных интерфейсов передачи данных. Для решения этих проблем требуется подключение к плате обработки дополнительных высокоскоростных интерфейсов передачи данных, взаимодействие которых со схемой обработки может быть реализовано за счет изменений в конфигурации ПЛИС.

В настоящей работе представлены результаты разработки цифровой схемы высокоскоростной записи данных со схемы обработки сигналов волоконно-оптического акустического датчика интерферометрического типа на micro SD-карту памяти с использованием ПЛИС и программной среды разработки Xilinx ISE, а также приводятся результаты апробации предложенной схемы записи.

Принцип работы волоконно-оптического акустического датчика

Структурная схема фазового волоконно-оптического акустического датчика интерферометрического типа представлена на рис. 1 [5]. Электронный блок обработки является функциональной частью макета волоконно-оптического фазового акустического датчика интерферометрического типа и осущест-

вляет функции приема, обработки и передачи информации на персональном компьютере (ПК) для ее последующего анализа и отображения. Чувствительным элементом датчика служит двулучепреломляющее оптическое волокно с записанными в него брэгговскими решетками.

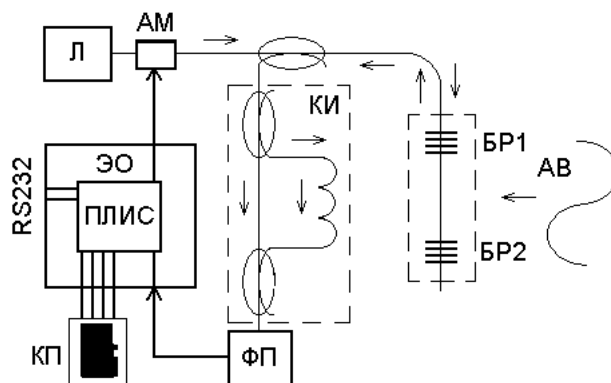


Рис. 1. Структурная схема волоконно-оптического акустического датчика интерферометрического типа:
ФП – фотоприемник; КИ – компенсирующий интерферометр; БР – брэгговская решетка;
ЭО – блок электронной обработки; КР – карта памяти; Л – лазер; АВ – акустическая волна;
АМ – амплитудный модулятор

Принцип работы волоконно-оптического акустического датчика заключается в следующем. Световое излучение лазера проходит через амплитудный модулятор (АМ), формирующий световые импульсы, которые, последовательно отражаясь от брэгговских решеток БР1 и БР2, попадают в компенсирующий интерферометр (КИ). В КИ разность фаз между отраженными импульсами компенсируется за счет разной длины плеч интерферометра, тем самым формируя интерференционную картину на фотоприемнике (ФП). Под действием акустического давления чувствительный элемент датчика деформируется, что вызывает изменение разности фаз интерферирующих импульсов. Эта разность фаз преобразуется ФП в изменение величины тока. Таким образом, обрабатывая сигнал с ФП, можно судить о характере акустического воздействия.

Функцию обработки сигнала с фотоприемника выполняет ПЛИС. Отличием ПЛИС от специализированных микросхем различных типов является ее внутренняя архитектура – она задается не в процессе изготовления, а посредством программирования. Поскольку программирование ПЛИС может производиться многократно, электронные схемы обработки на их основе имеют очень гибкую архитектуру, что позволяет существенным образом изменять и расширять функциональные возможности схемы без ее повторного проектирования.

Взаимодействие электронного блока обработки датчика и ПК осуществляется по асинхронному последовательному интерфейсу RS-232 [6]. Однако использование этого интерфейса имеет существенное ограничение по скорости передачи информации между блоком обработки и ПК, так как максимальная скорость передачи данных при использовании стандартного СОМ-порта ПК составляет 115200 бит/с. Но при увеличении частоты дискретизации звуковых сигналов, необходимости передачи дополнительной информации с платы обработки или при увеличении количества чувствительных волоконно-оптических элементов может потребоваться существенно большая пропускная способность информационного канала.

Ввиду того, что созданный макет волоконно-оптического акустического датчика был изначально рассчитан на акустические частоты до 500 Гц, согласно теореме Котельникова, достаточной для восстановления формы акустического сигнала является частота дискретизации сигнала 1 кГц. В связи с этим, с учетом возможного увеличения количества чувствительных волоконно-оптических элементов в ходе дальнейшей разработки макета, а также для возможности вывода дополнительной отладочной информации с платы обработки сигналов была выбрана частота дискретизации сигналов 3 кГц.

С учетом того, что размер одного пакета данных с блока обработки сигналов вместе с заголовком, окончанием пакета, стартовыми и стоповыми битами составляет 47 бит, для передачи данных с частотой дискретизации 3 кГц требуется пропускная способность канала передачи данных не менее 141000 бит/с. Это более чем на 22% превышает максимальную пропускную способность интерфейса RS-232 – 115200 бит/с. Следовательно, в случае передачи информации с частотой дискретизации 3 кГц по интерфейсу RS-232 часть данных будет теряться.

Таким образом, возникает необходимость высокоскоростной записи информации с фазовых интерферометрических датчиков непосредственно с платы обработки сигналов. В работе представлен один из способов такой записи информации с фазового волоконно-оптического акустического датчика интерферометрического типа.

Реализация записи на карту памяти

В ходе разработки макета волоконно-оптического акустического датчика было проанализировано несколько возможных вариантов реализации высокоскоростной записи со схемы цифровой обработки сигналов. Выбор оптимального интерфейса для высокоскоростной записи производился из существующих интерфейсов последовательной передачи данных.

Основными критериями для выбора интерфейса высокоскоростной записи данных являлись максимально возможная скорость передачи информации, гибкость и простота реализуемой цифровой схемы на базе ПЛИС, а также доступность соответствующей технической документации.

Анализ современных технических решений в этой области показал, что для последовательной передачи данных между микросхемами различного назначения чаще всего используются последовательные интерфейсы SPI (Serial Peripheral Interface) [7] и I²C (Inter Integrated Circuit) [8]. В результате анализа в качестве высокоскоростного интерфейса передачи данных был выбран последовательный периферийный интерфейс SPI. Данный выбор был обусловлен следующими преимуществами SPI-интерфейса над I²C:

- в случае осуществления передачи данных по SPI-интерфейсу максимальная скорость передачи может достигать десятков Мбит/с, в то время как максимальная скорость передачи данных для I²C составляет 3,4 Мбит/с [8];
- все линии шины SPI являются симплексными (однаправленными), что существенно упрощает реализацию цифровой схемы интерфейса на базе ПЛИС;

Для записи данных с платы обработки по SPI-интерфейсу была выбрана micro SD-карта памяти, поскольку современные карты памяти имеют большую емкость при небольших размерах, а скорость записи информации на них может достигать 90 Мбит/с. Это делает карту памяти очень удобным и компактным средством для хранения информации.

В настоящей работе для реализации цифровой схемы SPI-интерфейса на базе ПЛИС использовалась среда разработки Xilinx ISE 13.4. Данная среда позволяет осуществлять весь цикл разработки проекта – от описания логической архитектуры и его моделирования до конфигурирования любой из существующих ПЛИС фирмы Xilinx.

В работе использовалась ПЛИС Spartan 3 XC3S 1000 5FPG256C. Основные характеристики ПЛИС, как правило, отражены в ее наименовании и определяют рабочие характеристики систем на их основе. В данном случае XC3S 1000 обозначает тип ПЛИС, 5 (Speed Grade) – производительность, 256 – общее число вводов-выводов, C (Commercial) – рабочий температурный диапазон 0–85°C [9].

В процессе разработки на базе ПЛИС была реализована схема SPI-интерфейса, представленная на рис. 2. Схема включает в себя набор функциональных блоков и логических примитивов, причем логика работы блоков описывается на языке VHDL (Very high speed integrated circuits Hardware Description Language), в то время как логические примитивы являются стандартными элементами среды разработки ISE Xilinx. В данной схеме ведущим SPI-интерфейса является блок SPI_Master, а ведомым – карта памяти, общение с которой осуществляется посредством четырех портов (SCLK, MOSI, MISO, SS):

1. MOSI – выход ведущего, вход ведомого (Master Out Slave In). Служит для передачи данных от ведущего устройства ведомому;
2. MISO – вход ведущего, выход ведомого (Master In Slave Out). Служит для передачи данных от ведомого устройства ведущему;
3. SCLK – последовательный тактовый сигнал (Serial Clock). Служит для передачи тактового сигнала для ведомых устройств;
4. SS – выбор микросхемы (Slave Select) или выбор ведомого. Необходим в случае работы по SPI-интерфейсу с несколькими ведомыми устройствами.

Блок SPI_Master выполняет следующие функции: анализ и выполнение команд пользователя, прием информации со схемы обработки для ее последующей записи, передача данных для записи на карту памяти, передача тактирующего сигнала на карту памяти и прием ответов с карты памяти. Блок uart_control реализует интерфейс RS-232 между ПК и платой обработки, принимает команды пользователя и передает их блоку SPI_Master, принимает ответы карты памяти от блока SPI_Master и передает их пользователю. Блок ctrl осуществляет контроль процесса записи данных на карту памяти, формирует управляющие сигналы для блоков SPI_Master и uart_control, а также проверяет окончание передачи данных пользователю.

Приведенная схема работает следующим образом. На ПЛИС подается внешний сигнал от высокоточного кварцевого генератора с тактовой частотой 250 МГц, далее этот сигнал делится делителями частоты на сигналы в 100 кГц и 8 МГц, представляющие собой тактовые частоты инициализации карты памяти и записи информации на карту памяти соответственно. Эти сигналы подаются на блок SPI_Master, который осуществляет обмен командами и данными непосредственно с картой памяти. Блок uart_control поддерживает связь между ПЛИС и ПК – это необходимо для передачи стартовых команд записи от пользователя к плате обработки. На входной порт DATA блока SPI_Master подается информация для записи на карту памяти, причем, благодаря гибкости реализованной схемы, формат передаваемой ин-

записи были заложены возможности увеличения скорости записи до 8 Мбит/с. Апробация функционирования макета датчика на этой скорости записи является задачей будущих экспериментов.

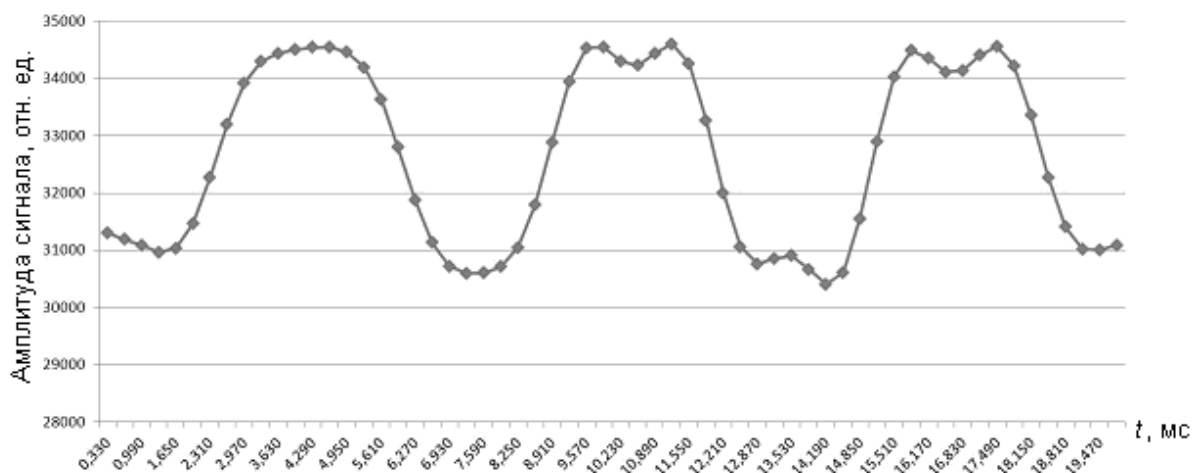


Рис. 3. Данные после обработки, записанные на карту памяти (частота акустического сигнала 150 Гц)

Заключение

В результате выполнения была разработана и программно реализована высокоскоростная запись данных со схемы обработки фазового акустического интерферометрического датчика на micro SD-карту памяти с использованием SPI-интерфейса.

В ходе экспериментальной проверки предложенной схемы была продемонстрирована ее работоспособность на примере записи данных на карту памяти воздействием на датчик акустического сигнала с известной частотой в 150 Гц. При этом расчетное количество цифровых отсчетов, приходящихся на один период звукового колебания, полностью соответствовало количеству записываемых с частотой 3 кГц отсчетов сигнала. Построенный по записанным данным график (рис. 3) повторял форму акустического сигнала. Однако в ходе работы был выявлен существенный недостаток карты памяти – максимальная скорость записи информации ограничивается ее классом быстродействия. По этой причине в предложенную схему были заложены возможности по увеличению скорости записи до 8 Мбит/с.

В настоящее время разработанная схема используется в плате обработки сигналов фазового интерферометрического акустического датчика. В дальнейшем планируется ее тестирование на картах памяти с более высоким классом быстродействия и модификация с целью повышения скорости записи.

Литература

1. Shizhuo Yin, Paul B.Ruffin, Francis T.S.Yu. Fiber Optic Sensors. – 2d ed. – Pennsylvania State University: CRC Press Taylor & Francis Group, 2008. – 477 p.
2. Волоконно-оптические датчики. Вводный курс для инженеров и научных работников / Под ред. Э. Удда – М.: Техносфера, 2008. – 520 с.
3. Максфилд К. Проектирование на ПЛИС. Курс молодого бойца. – М.: Додэка-XXI, 2007. – 408 с.
4. Китаев Ю.В. Использование ПО MAX+PLUS II и E-LAB для сквозного проектирования в лабораторных работах [Электронный ресурс]. – Режим доступа: <http://books.ifmo.ru/book/vip/162.pdf>, свободный. Яз. рус. (дата обращения 20.08.2012).
5. Куликов А.В. Волоконно-оптические акустические сенсоры на брэгговских решетках. Автореф. дис. ... к.т.н. – СПбГУ ИТМО, 2012. – 18 с.
6. Тарасов И.Е. Разработка цифровых устройств на основе ПЛИС Xilinx с применением языка VHDL. – М.: Горячая линия – Телеком, 2005. – 252 с.
7. Communication with MMC_SD card via SPI protocol [Электронный ресурс]. – Режим доступа: http://www.docstoc.com/docs/26093436/Communication-with-MMC_SD-card-via-SPI-protocol, свободный. Яз. англ. (дата обращения 11.05.2012).
8. I²C bus by Philips Semiconductors [Электронный ресурс]. – Режим доступа: <http://tmd.hav.it.cz/Papers/I2C.pdf>, свободный. Яз. англ. (дата обращения 21.07.2012).
9. Spartan 3 FPGA Family Data Sheet [Электронный ресурс]. – Режим доступа: http://www.xilinx.com/support/documentation/data_sheets/ds099.pdf, свободный. Яз. англ. (дата обращения 11.05.2012).

10. SD-Memory Card Specifications [Электронный ресурс]. – Режим доступа: http://www.electfreaks.com/store/download/datasheet/breakout/SD/SPI_Mode.pdf, свободный. Яз. англ. (дата обращения 11.05.2012).

- Плотников Михаил Юрьевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, plotnikov-michael@yandex.ru
- Дейнека Иван Геннадьевич** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, vanoDNK@yandex.ru
- Шарков Илья Александрович** – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, ilya.sharkov@gmail.com

УДК 681.7.063

ОДНОИМПУЛЬСНАЯ ЗАПИСЬ И ВИЗУАЛИЗАЦИЯ ВОЛОКОННЫХ РЕШЕТОК БРЭГГА ТИПА II

С.В. Варжель, А.В. Куликов, В.В. Захаров, В.А. Асеев

Представлены результаты записи брэгговских решеток типа II в двулучепреломляющем оптическом волокне с эллиптической напрягающей оболочкой одиночным импульсом K_rF эксимерного лазера. Дифракционная эффективность решеток, индуцированных одиночным импульсом, превышает 99,9% при ширине спектра отражения на полувысоте около 1 нм. Изображения волоконных брэгговских решеток типа II получены методами дифференциального интерференционного контраста и светлого поля.

Ключевые слова: волоконная брэгговская решетка, двулучепреломление, фазовая маска, эксимерный лазер.

Введение

Тип индуцируемой волоконной брэгговской решетки (ВБР) зависит от того, какой механизм фоточувствительности задействован при ее образовании. Образование ВБР типа I в германо-силикатных оптических волокнах (ОВ) объясняют электрострикционным механизмом [1], для которого характерно монотонное возрастание показателя преломления (ПП) при увеличении дозы облучения [2]. Решетки типа II, образованные термоупругими напряжениями [3], записываются при плотности энергии лазерного импульса, близкой к порогу разрушения кварцевого стекла 1 Дж/см² [2].

Для записи ВБР в данной работе применялся эксимерный лазер Lambda Physik Compex 102 с энергией в импульсе 150 мДж при использовании газовой смеси K_rF. Схема записи ВБР методом фазовой маски (ФМ) представлена в работе [4]. Плотность энергии записывающего излучения на ОВ регулировалась аттенуатором без изменения параметров фокусировки.

В настоящей работе применялось двулучепреломляющее ОВ с эллиптической напрягающей оболочкой, выполненное по технологии, описанной в работах [5, 6]. Для повышения фоторефрактивности ОВ концентрация диоксида германия (GeO₂) в его сердцевине увеличена до 16%.

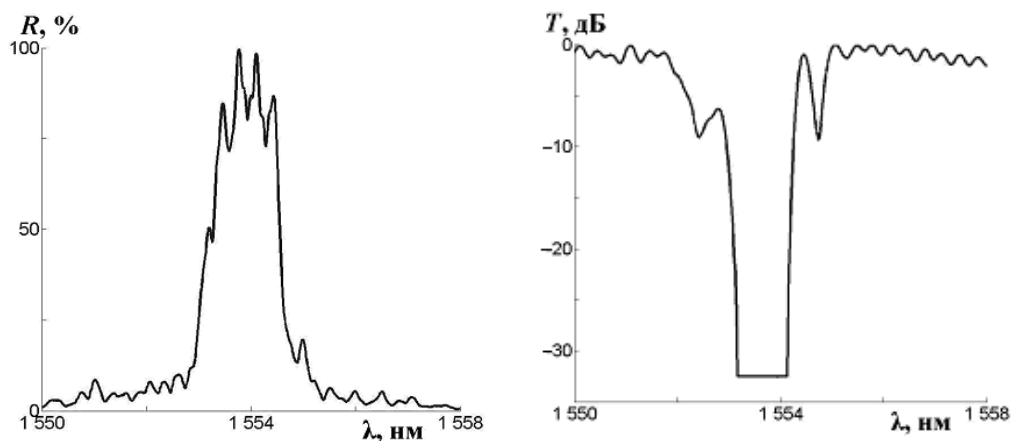
Результаты записи решеток

В работе [4] были представлены результаты записи ВБР типа I с эффективностью около 10% и шириной спектра отражения на полувысоте около 0,1 нм. Данные ВБР соответствуют электрострикционной модели образования решетки ПП [1]. Аналогичные образцы решеток типа I в ходе настоящей работы были записаны одиночным лазерным импульсом при энергии на выходе с аттенуатора 75,9 мДж.

При несколько большей энергии – 79,3 мДж – была записана решетка типа II, обусловленная термоупругим механизмом [3]. Ее дифракционная эффективность составляет около 100% при ширине спектра на полувысоте около 1 нм. Спектры отражения и пропускания данной ВБР представлены на рис. 1.

Как было сказано выше, в работе использовалось двулучепреломляющее ОВ, а так как эффективные ПП для каждой из выделенных осей анизотропного волокна различны, то и длина волны брэгговского резонанса будет отличаться для света, распространяющегося по быстрой и медленной осям двулучепреломляющего световода. Однако вследствие уширения спектров (рис. 1) произошло слияние пиков отражения двух ортогональных поляризаций. Уширение спектров связано с тем, что ВБР типа I имеет наведенную модуляцию ПП порядка 10⁻⁴ [1], в то время как наведенная модуляция ПП ВБР типа II может достигать 6×10⁻³ [7]. Увеличение модуляции ПП ведет к уменьшению рабочей длины решетки, что, в свою очередь, приводит к уширению спектра.

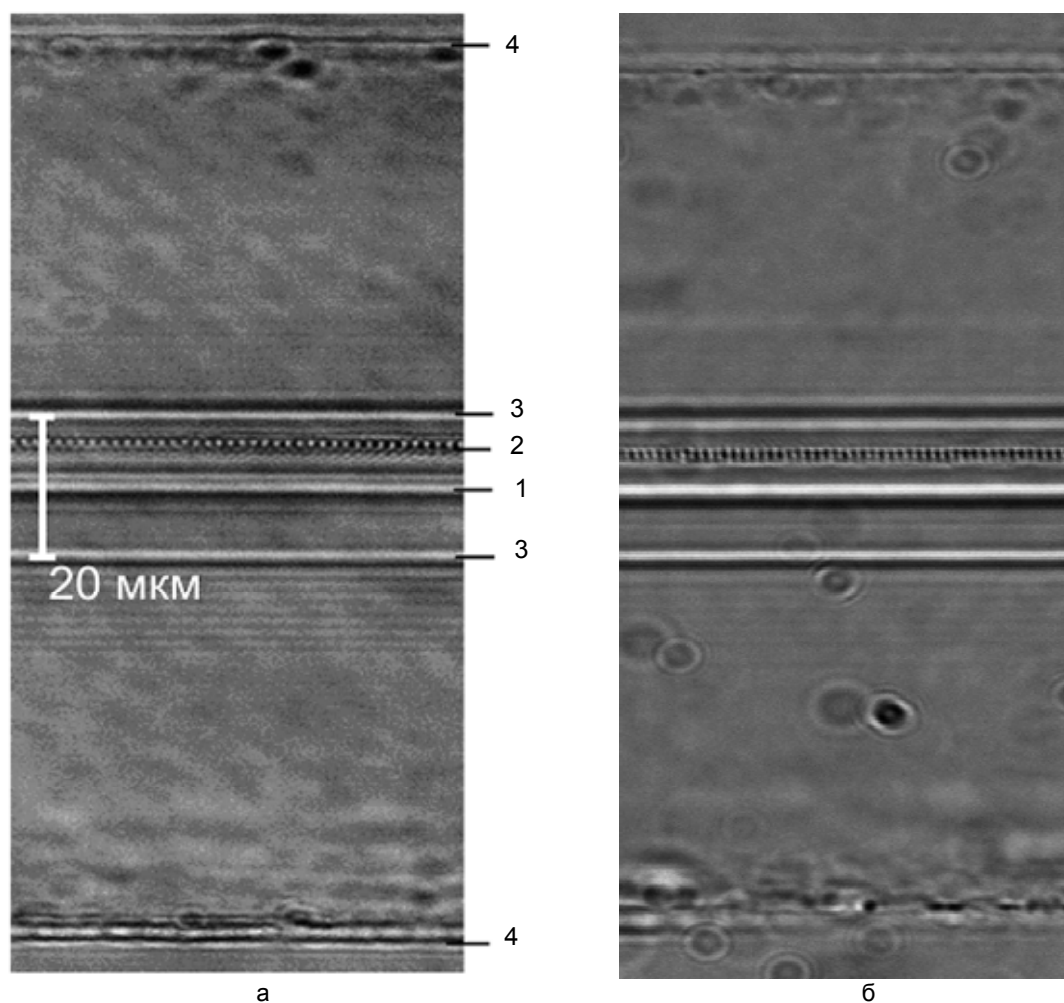
Наблюдаемая зависимость типа индуцируемой решетки от энергии в импульсе согласуется с экспериментальными данными, представленными в работах [7, 8].



а б
Рис. 1. Спектры ВБР: отражения (а); пропускания (б)

Визуализация волоконных решеток Брэгга

Большинство опубликованных микроскопических изображений ВБР относится к изотропным ОВ с обычной структурой, включающей сердцевину и окружающую ее оболочку [8–11]. В настоящей работе продемонстрировано изображение решетки Брэгга, записанной в двулучепреломляющем ОВ с эллиптической напрягающей оболочкой.



а б
Рис. 2. Изображения ВБР, полученные методами ДИК с лазером 405 нм (а), светлого поля с лампой накаливания (б): 1 – сердцевина; 2 – ВБР; 3 – граница между изолирующей и напрягающей оболочками; 4 – внешняя граница волокна

Для получения изображений брэгговских решеток типа II в двулучепреломляющих ОВ с эллиптической напрягающей оболочкой применен оптический микроскоп Zeiss Axio Imager.Z1 с объективом LD «Plan-Neofluar» 63×/0,75 Corr M27. Объектив обладает увеличенным рабочим расстоянием и корректировочным кольцом, позволяющим уменьшить искажения при наблюдении объектов, находящихся под прозрачным слоем.

Образующая ВБР типа II система микротрещин в ОВ сильно рассеивает свет и поэтому видна на изображениях, полученных при использовании обычной микроскопии светлого поля. Однако метод дифференциального интерференционного контраста (ДИК), предназначенный для визуализации градиентов ПП, позволяет получить более контрастное изображение решетки. На рис. 2 представлены изображения ВБР типа II в двулучепреломляющем ОВ с эллиптической напрягающей оболочкой, полученные методами светлого поля и ДИК в свете лампы накаливания и лазера с длиной волны 405 нм. На рисунке хорошо видны внешние границы ОВ диаметром 125 мкм, а также границы между его сердцевинной диаметром около 4 мкм и круговой оболочкой диаметром около 20 мкм, изолирующей сердцевину от борсодержащего стекла напрягающей зоны – эллиптической оболочки, границы которой видны хуже. Это объясняется тем, что напрягающая оболочка, наряду с B_2O_3 , легирована GeO_2 для выравнивания значений ПП оболочек, окружающих сердцевину. Периодическая структура, локализованная вблизи границы между сердцевинной и 20-микронной оболочкой ОВ, отчетливо видна как на изображении, полученном методом ДИК, так и на изображении, полученном традиционным методом светлого поля. Это объясняется тем, что система микротрещин, образующая ВБР типа II, столь сильно рассеивает свет, что она хорошо видна на изображениях, полученных различными методами контрастирования, включая традиционную микроскопию светлого поля.

Измеренный по полученным в настоящей работе изображениям пространственный период решеток составляет 1 мкм, что соответствует периоду ФМ (1065,3 нм), использованной для записи ВБР. Это обстоятельство подтверждает, что записанные ВБР являются решетками типа II, так как решетки типа I имеют вдвое меньший период, чем у используемых для их записи ФМ, оптимизированных под +1/-1 порядок дифракции [8].

Заключение

В работе представлены результаты записи брэгговских решеток типа II в двулучепреломляющем оптическом волокне с эллиптической напрягающей оболочкой одиночным импульсом KrF эксимерного лазера. Дифракционная эффективность решеток, индуцированных одиночным импульсом, превышает 99,9% при ширине спектра отражения на полувывоте около 1 нм. Продемонстрированы микроскопические изображения волоконной брэгговской решетки типа II, индуцированные в двулучепреломляющем оптическом волокне с эллиптической напрягающей оболочкой. Визуализация волоконной брэгговской решетки типа II проведена методами дифференциального интерференционного контраста и светлого поля на оптическом микроскопе Zeiss Axio Imager Z1.

Литература

1. Неуструев В.Б. Электрострикционный механизм образования брэгговской решетки в германосиликатных световодах // Квантовая электроника. – 2001. – Т. 31. – № 11. – С. 1003–1006.
2. Васильев С.А., Медведков О.И., Королев И.Г., Божков А.С., Курков А.С., Дианов Е.М. Волоконные решетки показателя преломления и их применение // Квантовая электроника. – 2005. – Т. 35. – № 12. – С. 1085–1103.
3. Колдунов М.Ф., Маненков А.А., Покотило И.Л. Термоупругий и абляционный механизмы лазерного повреждения поверхности прозрачных твердых тел // Квантовая электроника. – 1998. – Т. 25. – № 3. – С. 277–281.
4. Варжель С.В., Куликов А.В., Асеев В.А., Брунов В.С., Калько В.Г., Артеев В.А. Запись узкополосных волоконных брэгговских отражателей одиночным импульсом эксимерного лазера методом фазовой маски // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 5 (75). – С. 27–30.
5. Дукельский К.В., Ероньян М.А., Комаров А.В., Кондратьев Ю.Н., Ромашова Е.И., Серков М.М., Хохлов А.В. Тонкие анизотропные одномодовые волоконные световоды с эллиптической напрягающей оболочкой // Оптический журнал. – 2000. – Т. 57. – № 10. – С. 104–106.
6. Буреєв С.В., Дукельский К.В., Ероньян М.А., Злобин П.А., Комаров А.В., Левит Л.Г., Страхов В.И., Хохлов А.В. Технология крупногабаритных заготовок анизотропных одномодовых световодов с эллиптической оболочкой // Оптический журнал. – 2007. – Т. 74. – № 4. – С. 85–87.
7. Reekie L., Archambault J.-L., Russell P.St.J. 100% reflectivity fibre gratings produced by a single excimer laser pulse // OSA/OFC. – 1993. – PD14. – P. 327–330.

8. Malo B., Johnson D.C., Bilodeau F., Albert J., Hill K.O. Single-excimer-pulse writing of fiber gratings by use of a zero-order nulled phase mask: grating spectral response and visualization of index perturbations // Opt. Lett. – 1993. – V. 18. – № 15. – P. 1277–1279.
9. Liu H.Y., Liu H.B., Peng G.D., Chu P.L. Observation of type I and type II gratings behavior in polymer optical fiber // Optics Communications. – 2003. – V. 220. – P. 337–343.
10. Dragomir N.M., Baxter G., Collins S.F., Farrell P.M., Stevenson A.J., Garchev D.D., Roberts A. DIC imaging of an optical fiber Bragg grating // OSA Trends in Optics and Photonics. – 2002. – V. 70. – ThGG62.
11. Dragomir N.M., Rollinson C., Wade S.A., Stevenson A.J., Collins S.F., Baxter G.W., Farrell P.M., Roberts A. Nondestructive imaging of a type I optical fiber Bragg grating // Optics Letters. – 2003. – V. 28. – № 10. – P. 789–791.

- | | |
|------------------------------------|--|
| Варжель Сергей Владимирович | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, vsv187@gmail.com |
| Куликов Андрей Владимирович | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, a.kulikov86@gmail.com |
| Захаров Виктор Валерьевич | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, Viktor-zah@yandex.ru |
| Асеев Владимир Анатольевич | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, НИИ нанофотоники и оптоинформатики, научный сотрудник, aseev@oi.ifmo.ru |

УДК 535.338.1

ОСОБЕННОСТИ ГЕНЕРАЦИИ ФЕМТОСЕКУНДНОГО СПЕКТРАЛЬНОГО СУПЕРКОНТИНУУМА В КРИСТАЛЛИЧЕСКИХ СРЕДАХ

С $\chi^{(2)}$ -НЕЛИНЕЙНОСТЬЮ

С.С. Налегает, В.Г. Беспалов, С.Э. Путилин

Приведены результаты экспериментов по генерации суперконтинуума при возбуждении оптическими импульсами с центральной длиной волны 830 нм, длительностью 40–50 фс и энергией до 1,0–1,5 мДж. В кристалле DKDP и других кристаллах с нелинейностью $\chi^{(2)}$ зарегистрировано излучение суперконтинуума в области от 400 нм до 2200 нм, причем излучение в синей области (400–480 нм) и в ИК области (900–2200 нм) обусловлено в основном параметрическими процессами. Полученное излучение не уступает по ширине выходного спектра излучению генераторов суперконтинуума на основе микроструктурированных волокон.

Ключевые слова: фемтосекундные импульсы, нелинейность $\chi^{(2)}$, нелинейность $\chi^{(3)}$, генерация суперконтинуума, параметрическая генерация.

Введение

Генерация белого света, или генерация спектрального суперконтинуума (СК) [1], заключается в получении оптического излучения с непрерывным или иногда дискретным спектром, ширина которого составляет обычно не менее одной октавы (т.е. частоты фурье-компонент в спектре излучения отличаются в два раза), с использованием импульсных лазеров пико- и фемтосекундной длительности. Такое спектрально уширенное излучение, как правило, является результатом нелинейного оптического пространственно-временного преобразования поля сверхкоротких лазерных импульсов высокой интенсивности в диэлектрических средах [2].

Как правило, в генераторах СК используются преимущественно среды с большей нелинейностью $\chi^{(3)}$, однако при этом относительно малый вклад вносят параметрические процессы. Применение в качестве генераторов СК кристаллических сред, обладающих одновременно существенными нелинейностями $\chi^{(2)}$ и $\chi^{(3)}$, позволит увеличить вклад параметрических процессов и получить более широкий выходной спектр.

Генераторы СК в настоящее время применяются в экспериментах по «pump-probe» спектроскопии [3], в оптической метрологии [4], оптической когерентной томографии [5] и др. [1]. Расширение спектра СК очень важно для технологических целей, так как одна из возможных сфер использования генерации СК, это передача информации, в том числе по оптоволоконным линиям связи, где используется ИК диапазон длин волн (1300–1600 нм). Расширение спектра СК увеличивает диапазон частот и соответственно – количество каналов передачи информации, повышая, таким образом, общую скорость передачи [6].

Следует отметить, что планомерных исследований генерации белого света при использовании нелинейно-квадратичных кристаллических сред и импульсного излучения накачки высокой плотности мощности и фемтосекундной длительности, в том числе и возможности расширения в ИК область спектра, не проводилось.

Целью настоящей работы является определение оптимальных условий генерации сверхширокополосного спектрального СК в средах с нелинейностью $\chi^{(2)}$, обладающих также существенной нелинейностью $\chi^{(3)}$, а также выявление особенностей его излучения. В качестве сред с нелинейностью $\chi^{(2)}$ использовались образцы кристаллов DKDP, KDP и LiIO₃. Для сравнения были проведены измерения спектров излучения СК на образце кристалла LiF с кубическим типом решетки.

Методика проведения измерений и описание экспериментального макета

Кристаллы KDP и DKDP широко используются для генерации второй гармоники лазерного излучения, обладают высокой лазерной прочностью и достаточно высокими нелинейными коэффициентами, как $\chi^{(2)}$, так и $\chi^{(3)}$, одновременно [7]. Более высокая лучевая прочность позволила безопасно повысить плотность мощности излучения накачки (относительно других кристаллов – LiF и LiIO₃), а вместе с этим и поднять эффективность преобразования в СК.

Образцы LiF были добавлены в исследуемую группу в качестве объектов сравнения, по которым можно оценить вклад квадратично-нелинейных процессов при генерации СК, так как вследствие кубической centrosymmetричной решетки все элементы тензора нелинейной восприимчивости второго порядка $\chi^{(2)}$ данного кристалла равны нулю.

Кристалл LiIO₃, также как и KDP или DKDP, используется для генерации второй гармоники и в параметрических генераторах, но его лучевая прочность заметно ниже, поэтому необходимо было

уменьшать плотность мощности лазерного излучения накачки в процессе экспериментов. Поэтому одной из задач исследования было выяснение влияния интенсивности накачки на ширину спектра СК.

Под оптимальными условиями генерации в работе понимается изменение условий фазировки различных компонент СК при варьировании угла распространения преломленного луча накачки в объеме исследуемого кристалла, что осуществлялось путем поворота кристалла в двух плоскостях относительно оси накачки.

Следует отметить, что выбранные для эксперимента образцы кристаллов KDP, DKDP и LiIO_3 были предназначены для генерации второй гармоники на длине волны 1,064 мкм, и для обеспечения синхронизма на центральной длине волны накачки $\lambda = 830$ нм их было необходимо повернуть на легко определяемый угол [7].

Характеристики фемтосекундной лазерной системы на кристаллах сапфира, активированных титаном, использовавшиеся в экспериментах, подробно описаны в [8]. Энергия и длительность одиночного импульса составляли соответственно 1,0–1,5 мДж и 40–50 фс, а средняя мощность лазерного излучения – 50–75 мВт (при частоте следования импульсов 50 Гц).

В экспериментальном макете (рис. 1) излучение накачки, приходящее слева, фокусируется линзой 1, за которой на некотором расстоянии от фокуса устанавливается исследуемый образец кристалла 3 на столике 2. Прошедшее сквозь образец излучение далее собирается линзой 4 и направляется на матовую пластинку 5, установленную в ее фокусе, после чего проходит несколько оптических фильтров 6 и попадает на вход спектрографа 7 «ASP-100» (диапазон измерений 0,19–1,1 мкм) или спектрографа «EPP2000-NIRX-SR-512» (0,9–2,3 мкм).

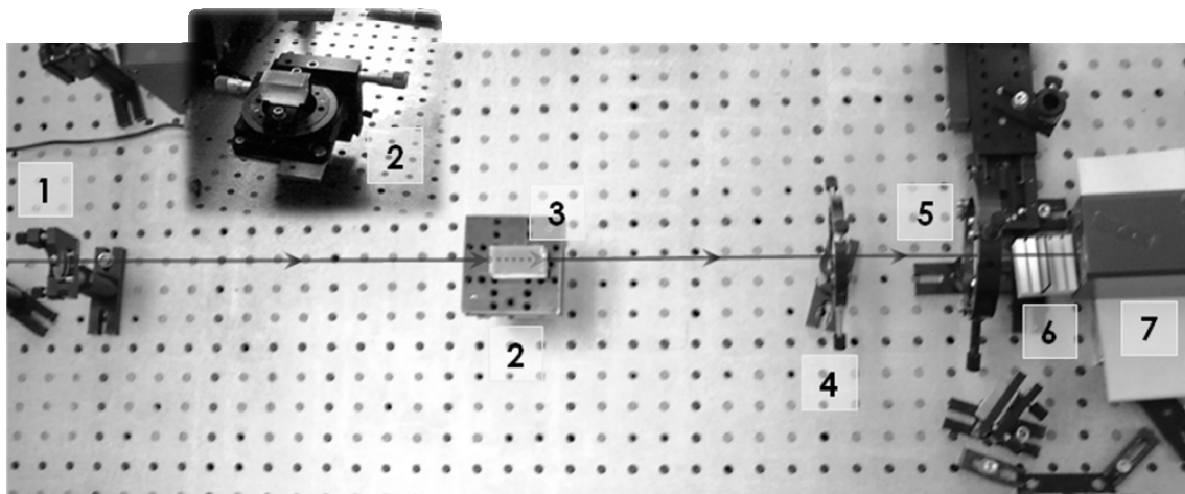


Рис. 1. Экспериментальный макет для генерации спектрального СК: 1, 4 – линзы; 2 – двухкоординатный столик-вращатель; 3 – исследуемый образец кристалла; 5 – матовая пластинка; 6 – светофильтры; 7 – спектрограф «ASP-100» (0,19–1,1 мкм) или «EPP2000-NIRX-SR-512» (0,9–2,3 мкм)

Данные со спектрографов в ходе экспериментов сохранялись в памяти персонального компьютера; для минимизации погрешностей при каждом эксперименте записывалось около сотни спектров генерации СК и спектров шумового сигнала. Для обработки результатов было разработано специальное программное обеспечение, позволяющее сократить время анализа записанных данных.

Результаты эксперимента

На рис. 2 представлены графики излучения спектрального СК при использовании кристалла DKDP для нескольких измеренных экспериментально углов отклонения преломленного пучка накачки относительно оптической оси кристалла. Цифрами и вертикальными линиями (со значениями длин волн) отмечены области спектра, в которых происходит процесс преобразования излучения (с отмеченными значениями длин волн) в излучение второй гармоники. При повороте кристалла изменяются условия генерации второй гармоники – фазовый синхронизм начинает выполняться для излучения на другой, смещенной (в зависимости от угла) частоте фурье-компоненты СК. На рис. 3 представлен график излучения спектрального СК для одного из углов отклонения в ИК диапазоне спектра. Аналогичные экспериментальные результаты были получены и с использованием кристалла KDP.

Суперконтинуум, полученный при генерации с использованием кристалла LiIO_3 , достаточно заметно уступает выходному спектрально-уширенному излучению при использовании кристаллов KDP и DKDP как по общей ширине выходного спектра СК, так и по интенсивности в ближней ИК области

спектра, поскольку интенсивность накачки в данном случае была в 5–6 раз меньше, чем при использовании KDP и DKDP, из-за вероятности оптического пробоя поверхности кристалла.

Спектр генерации СК при использовании кристалла LiF также существенно уже, чем при использовании сред с выраженной квадратичной нелинейностью, как по общей ширине выходного спектра СК, так и по интенсивности в ближней ИК области спектра. Мощность излучения спектрального СК в ИК области более чем вдвое ниже, чем соответствующая мощность на выходе кристаллов KDP и DKDP, что объясняется происходящими в них процессами, основанными на $\chi^{(2)}$ -нелинейности, среди которых наибольший вклад имеют процессы параметрической генерации света, а именно – генерация разностной и суммарной частот. Более подробный анализ происходящих процессов будет представлен в следующей работе.

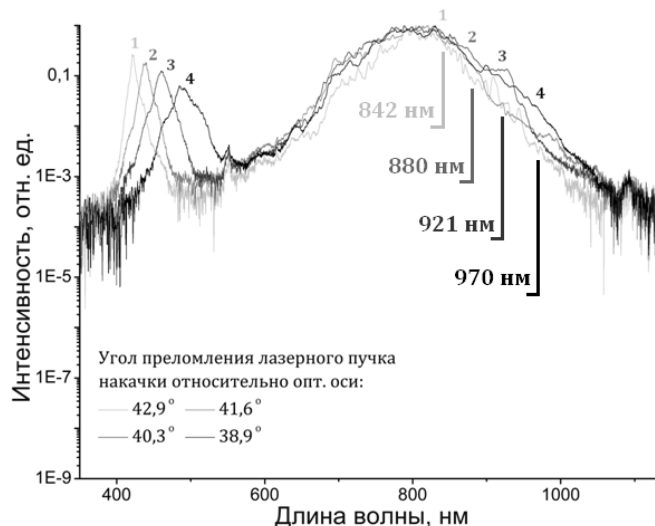


Рис. 2. Спектральные зависимости генерации СК от угла распространения преломленного пучка излучения накачки в объеме кристалла DKDP (энергия импульса накачки 1,0–1,5 мДж, длительность импульса 40–50 фс). Цифрами 1–4 отмечены длины волн, соответствующие генерации второй гармоники от соответствующих фурье-компонент СК

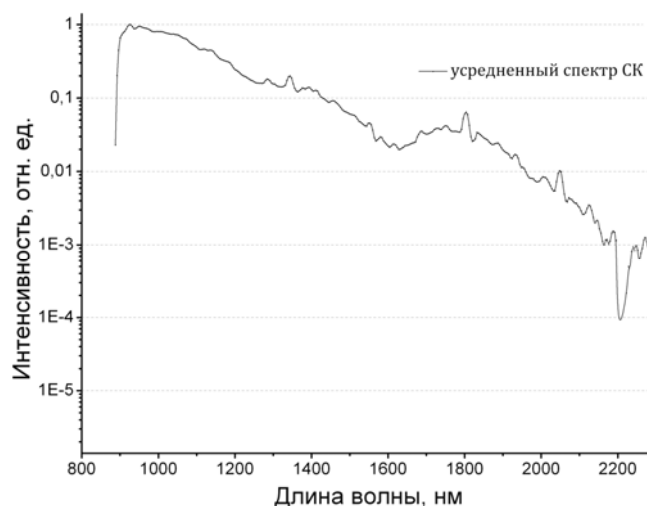


Рис. 3. Спектр генерации СК DKDP в ИК области спектра (энергия импульса накачки 1,0–1,5 мДж, длительность импульса 40–50 фс)

Заключение

Полученное экспериментально излучение суперконтинуума в средах с нелинейностью $\chi^{(2)}$ практически не уступает по ширине выходного спектра широко распространенным в настоящее время генераторам на основе оптических волокон различных типов. Следует отметить, что объемные нелинейно-квадратичные среды позволяют достаточно эффективно генерировать излучение суперконтинуума в синей (400–480 нм) и ИК (900–2200 нм) областях спектра, причем, изменяя угол падения накачки на кристалл, можно регулировать интенсивность суперконтинуума по спектру, выделяя отдельные спектральные полосы.

Использование в качестве сред для генерации кристаллов с квадратичной нелинейностью позволяет заметно расширить выходной спектр излучения (в том числе в коротковолновой области – за счет генерации второй гармоники), а также существенно повысить эффективность генерации спектрального суперконтинуума в ИК области спектра в сравнении со средами, у которых преобладает нелинейность, связанная с $\chi^{(3)}$. В связи с этим метод получения излучения сверхширокополосного спектрального суперконтинуума на основе нелинейных кристаллов $\chi^{(2)}$ имеет перспективы использования в областях современной науки и технологий, например, как источник излучения для оптической томографии, метрологии или «pump-probe» спектроскопии.

Литература

1. Alfano R.R. The Supercontinuum Laser Source: Fundamentals with Updated References. – N.Y.: Springer-Verlag, 2005. – 552 p.
2. Беспалов В.Г., Козлов С.А., Сутягин А.Н., Шполянский Ю.А. Сверхуширение спектра интенсивных фемтосекундных лазерных импульсов и их временное сжатие до одного колебания светового поля // Оптический журнал. – 1998. – Т. 65. – № 10. – С. 85–88.
3. Крылов В.Н., Беспалов В.Г., Стаселько Д.И., Лобанов С.А., Милогладов Э.В., Seyfang G. Спектральные особенности взаимодействия разночастотных фемтосекундных импульсов света в приграничных областях керровской среды // Оптика и спектроскопия. – 2005. – Т. 99. – № 5. – С. 853–858.
4. Багаев С.Н., Денисов В.И., Захарьяш В.Ф., Каширский А.В., Клементьев В.М., Кузнецов С.А., Корель И.И., Пивцов В.С. Фемтосекундные оптические часы // Квант. электроника. – 2004. – Т. 34. – № 12. – С. 1096–1100.
5. Гуров И.П. Оптическая когерентная томография: принципы, проблемы и перспективы // Проблемы когерентной и нелинейной оптики / Под ред. И.П. Гурова, С.А. Козлова. – СПб: СПбГУ ИТМО. – 2004. – С. 6–30.
6. Беспалов В.Г., Васильев В.Н., Козлов С.А., Шполянский Ю.А. Использование фемтосекундного суперконтинуума в системах сверхплотной передачи информации // Сборник статей «Оптические и лазерные технологии». – СПб: СПб ГИТМО (ТУ), 2001. – С. 214–219.
7. Дмитриев В.Г., Тарасов Л.В., Прикладная нелинейная оптика. – М.: ФИЗМАТЛИТ, 2004. – 512 с.
8. Беспалов В.Г., Киселев В.М., Кисляков И.М. и др. Антистоксов самосдвиг и уширение спектра излучения фемтосекундного лазера в сильно поглощающей среде // Оптика и спектроскопия. – 2009. – Т. 106. – № 4. – С. 670–679.

<i>Налегаев Сергей Сергеевич</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, fioi@inbox.ru
<i>Беспалов Виктор Георгиевич</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, victorbespaloff@gmail.com
<i>Путилин Сергей Эдуардович</i>	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, ст. научный сотрудник, SEPutilin@yandex.ru

УДК 535.371, 537.9

ИССЛЕДОВАНИЕ КИНЕТИКИ ЛЮМИНЕСЦЕНЦИИ КВАНТОВЫХ ТОЧЕК СУЛЬФИДА СВИНЦА

А.П. Литвин, П.С. Парфенов, Е.В. Ушакова, А.В. Баранов

Исследованы размерные зависимости спектральных и кинетических параметров люминесценции квантовых точек сульфида свинца в широком спектральном диапазоне (0,8–2 мкм). Зарегистрированные времена жизни лежат в диапазоне 0,25–2,7 нс и уменьшаются с увеличением размера нанокристалла. Для объяснения больших времен жизни люминесценции рассмотрена трехуровневая модель с энергетическим уровнем в запрещенной зоне. Показано, что аномальная размерная зависимость времен жизни люминесценции может быть связана с существованием переходов как с понижением, так и повышением энергии при комнатной температуре.

Ключевые слова: квантовые точки, люминесценция, ближний ИК диапазон, времена жизни.

Введение

В последние несколько лет интенсивно исследуются квантовые точки (КТ) узкозонных полупроводников, в частности, КТ сульфида свинца (PbS). Оптические переходы в КТ PbS, в зависимости от размера КТ, охватывают как видимую, так и всю ближнюю инфракрасную (ИК) область спектра. Это объясняет тот факт, что КТ PbS являются многообещающим материалом для использования в самых раз-

ных приложениях. В первую очередь, это волоконно-оптические телекоммуникационные системы, солнечная энергетика, визуализация и мониторинг, биомедицинские приложения [1]. Также КТ халькогенидов свинца могут выступать в качестве хорошей модели КТ, поскольку относительно малые и практически одинаковые величины эффективных масс электронов и дырок позволяют исследовать режим сильного пространственного ограничения (конфайнмента) уже при достаточно больших размерах нанокристаллов.

Однако, несмотря на интенсивные исследования, имеющиеся экспериментальные данные во многом противоречивы и плохо согласуются с общепринятой теорией релаксации в КТ. Вследствие этого отсутствует детальное описание динамики процессов релаксации, крайне важное для разработки элементов фотоники и оптоэлектроники ИК диапазона на базе КТ. Самый наглядный пример расхождения – в оценке времен затухания люминесценции. Известно, что, вследствие эффекта диэлектрического экранирования, ожидаемые времена затухания люминесценции КТ PbS должны быть больше, чем, например, для КТ полупроводников группы A_2B_6 . В органических растворителях эта величина должна составлять порядка 250 нс. Однако имеются сообщения о регистрации необычно больших времен затухания, порядка 1–2 мкс [2, 3].

Для объяснения столь больших времен жизни была предложена модель, согласно которой люминесценция наблюдается с долгоживущего энергетического состояния вблизи дна зоны проводимости с энергией меньшей, чем энергия фундаментального перехода КТ PbS [4]. На существование такого состояния указывают эксперименты по наведенному поглощению [5] и значительный стоксов сдвиг [6]. Исследования, однако, ограничивались, в основном, КТ достаточно малого размера, не более 4–5 нм, что связано с трудностями проведения люминесцентных измерений в спектральном диапазоне далее 1,4 мкм. Данный факт значительно осложняет объяснение общих закономерностей динамики электронных возбуждений в КТ PbS. В связи с этим основной целью работы стало исследование спектральных и кинетических параметров люминесценции КТ PbS в диапазоне 0,8–2 мкм.

Эксперимент

Для изготовления нанокристаллов был применен высокотемпературный органометаллический синтез в органическом растворе – метод горячей инъекции. Реакция проводилась при температуре 150–300°C в атмосфере аргона, что позволило синтезировать КТ PbS диаметром 2,5–9 нм. Все образцы были приготовлены в виде растворов в тетрахлорметане с концентрацией порядка 10^{-9} моль/м³.

Для исследования энергетической структуры, поглощательной способности и концентрации растворов КТ PbS использовался спектрофотометр Shimadzu UV-3600. Для проведения спектрофлуориметрических измерений был использован оригинальный экспериментальный комплекс для получения спектров люминесценции в ближнем ИК диапазоне [7, 8], построенный на основе монохроматора Acton SP-2558 с апертурой $f/6,5$, фокусным расстоянием 500 мм и дифракционной решеткой 150 мм^{-1} . Сфокусированное излучение твердотельного лазера с длиной волны 532 нм направляется на кювету с раствором. Люминесцентное излучение собирается по 90°-ной схеме. В качестве приемника используется InGaAs фотодиод (G5852-21) фирмы Hamamatsu расширенного (0,9–2,1 мкм) спектрального диапазона, охлаждаемый до –20°C и включенный в фотогальванический режим. Для получения истинных спектров люминесценции была проведена прецизионная калибровка экспериментального комплекса по спектру абсолютно черного тела [8].

Для измерений времен затухания люминесценции была создана оригинальная установка на основе пикосекундного импульсного лазера (PicoQuant LDH-P-C-640B, 100 пс, 0,4 нДж), скоростного InGaAs фотодиода (Femto® HCA-S-200M-IN) и осциллографа (Tektronix TDS-2002), позволяющего регистрировать времена затухания в диапазоне 20 нс–10 мкс на длинах волн 800–1700 нм [9]. Для достижения отношения сигнал/шум, необходимого для уверенного измерения времен затухания вплоть до нескольких микросекунд, в среде программирования LabView разработана программа, считывающая с осциллографа результаты аппаратного усреднения и позволяющая проводить процедуру дополнительного накопления сигналов вплоть до 10^6 измерений.

Результаты и обсуждение

Для КТ PbS диаметром 2,5–9 нм проведены измерения спектров поглощения и люминесценции. Для всех образцов по спектрам поглощения и люминесценции определялся стоксов сдвиг. При этом наблюдалась характерная зависимость: значительный стоксов сдвиг вплоть до 350 мэВ для наименьших КТ и незначительный стоксов сдвиг для самых больших нанокристаллов.

Для КТ PbS диаметром 2,8–7,5 нм проведены измерения времен затухания люминесценции. В связи с тем, что средняя плотность мощности возбуждающего излучения не превышала 200 мкВ/см^2 , что необходимо для обеспечения одноэлектронного режима возбуждения, запись кривых затухания проводилась при усреднении по 1 млн измерений. Кривые затухания хорошо описывались одноэкспоненциальной зависимостью с постоянной времени 0,25–2,7 мкс, сильно зависящей от размера нанокристалла. Зарегистрированные времена жизни на порядок превышают предсказываемые теорией. Кроме того, на-

блюдается аномальная зависимость уменьшения времени жизни люминесценции с увеличением размера КТ (увеличением длины волны соответствующего энергетического перехода).

Наблюдаемые размерные зависимости времен жизни люминесценции и стоксового сдвига представлены на рис. 1. Большие величины стоксового сдвига могут свидетельствовать о том, что люминесценция протекает не с первого возбужденного состояния, а с некоторого энергетического уровня, располагающегося внутри запрещенной зоны (below-gap state). Положение этого уровня находится в зависимости от величины энергии конфинмента и соответственно размера КТ. Большие времена жизни люминесценции могут быть связаны именно с этим энергетическим состоянием, собственное время жизни которого может намного превышать время жизни фундаментального перехода.

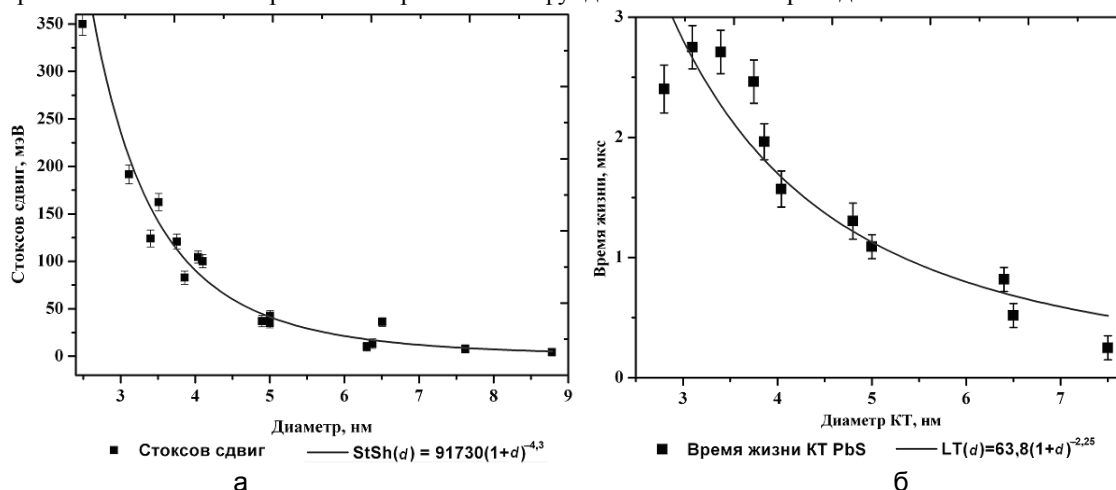


Рис. 1. Размерные зависимости стоксового сдвига (а) и времен жизни люминесценции (б)

Моделирование

Для объяснения наблюдаемых экспериментально результатов авторами рассмотрена модель трехуровневой системы, в которой излучательная рекомбинация может происходить с уровней S_1 и S_2 . Собственное время жизни уровня S_1 предполагается соответствующим расчетному для фундаментального перехода, уровня S_2 – на порядок больше, что связано с регистрацией микросекундных времен жизни люминесценции для КТ малого размера. Схематически модель трехуровневой системы изображена на рис. 2, а. Уровни S_1 и S_2 связаны скоростями безызлучательных переходов k_{12} и k_{21} . Поскольку измерения проводились при комнатной температуре, следует учитывать переходы как с понижением, так и с повышением энергии. Другими словами, при малой величине энергетического зазора может иметь место обратное термическое возбуждение электрона на уровень S_1 . Так, скорость k_{21} может быть выражена в виде

$$k_{21} = A \cdot \exp\left(-\frac{dE}{kT}\right),$$

где A – константа, dE – величина энергетического зазора, k – постоянная Больцмана, T – температура в кельвинах. Можно получить выражение, описывающее отношение эффективностей люминесценции с каждого из уровней S_1 и S_2 :

$$\frac{\varphi_{S_1}}{\varphi_{S_2}} = \frac{k_{10}}{k_{12}} + A \cdot \exp\left(-\frac{dE}{kT}\right) \cdot \frac{k_{10} \cdot (2k_{12} + k_{10})}{k_{20} \cdot k_{12} \cdot (k_{10} + k_{12})},$$

где φ_{S_1} и φ_{S_2} – эффективности люминесценции с уровней S_1 и S_2 соответственно; k_{10} и k_{20} – скорости излучательной рекомбинации с уровней S_1 и S_2 соответственно. Здесь φ_{S_2} включает в себя как «собственную» люминесценцию, так и люминесценцию, связанную с термической активацией электронов с уровня S_2 . При этом под эффективностью люминесценции понимается величина $\varphi = \frac{\tau}{\tau_r}$, где τ – время жизни;

τ_r – радиационное время жизни данного энергетического уровня, т.е. фактически квантовый выход с данного энергетического состояния. Эффективность этих процессов во многом зависит от скоростей безызлучательных переходов k_{12} и k_{21} , при расчете положено $A = k_{12} = 10^{10} \text{ c}^{-1}$. На основании экспериментальных данных о стоксовом сдвиге можно сделать предположение об относительном положении энергетических уровней S_1 и S_2 и построить зависимость эффективностей люминесценции с двух энергетических уровней от размера нанокристалла (рис. 2, б). Как видно, для КТ малого размера будет превалировать люминесценция с нижнего энергетического состояния, для КТ большого размера – с верхнего. С учетом того, что время жизни энергетического уровня S_2 на порядок больше, нежели время жизни основ-

ного уровня S_1 , полученный результат согласуется с наблюдаемой экспериментально размерной зависимостью времен жизни люминесценции КТ PbS. Наличие двух каналов люминесценции должно проявляться в спектрах люминесценции в виде уширения и асимметрии полос люминесценции. Однако, как видно на рис. 2, б, отношения эффективностей имеют очень резкие зависимости, поэтому заметный вклад обе компоненты могут вносить лишь для КТ среднего размера, что и наблюдается в эксперименте (рис. 2, в).

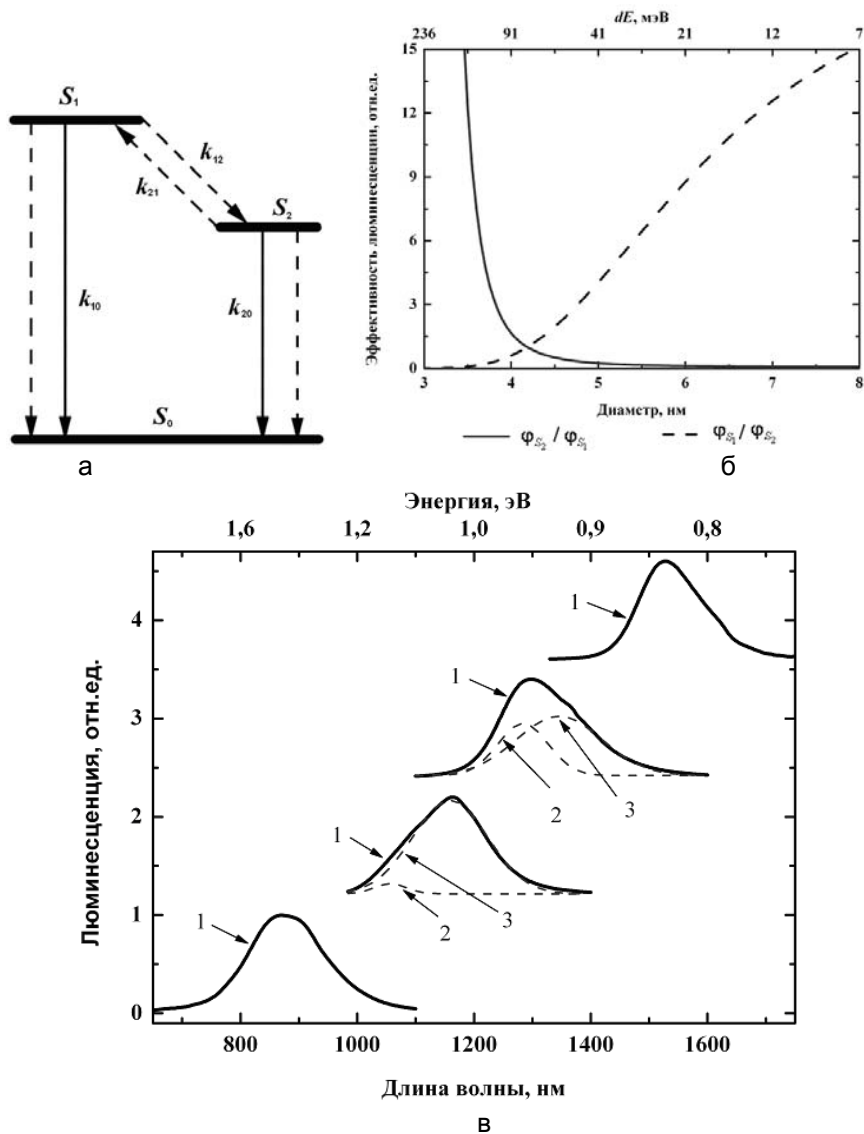


Рис. 2. Модель трехуровневой системы (а); отношения эффективностей люминесценции с уровнями S_1 и S_2 (б); контуры полос люминесценции (1) и разложение их двумя гауссовыми функциями (2, 3) (в)

Измерение времен затухания люминесценции проводилось при комнатной температуре и с использованием нерезонансного возбуждения, поэтому возможно описание процессов без применения строгого квантового формализма. Кроме того, многократно наблюдается ансамбль идентичных (за исключением дисперсии по размерам) КТ. По этой причине возможно введение населенностей состояния и составление скоростных уравнений [10]. Уравнения для населенностей в данной системе могут быть представлены в следующем виде:

$$\begin{cases} \frac{dN_1}{dt} = -k_1 \cdot N_1 + k_{21} \cdot N_2, \\ \frac{dN_2}{dt} = k_{12} \cdot N_1 - k_2 \cdot N_2, \end{cases},$$

причем $k_1 = k_{10} + k_{12}$, $k_2 = k_{20} + k_{21}$. Система имеет точное решение:

$$N_{1,2}(t) = N_{1,2}(t_0) \cdot \exp(s_{1,2}t) + N_{2,1}(t_0) \frac{k_2 - k_1 \pm \sqrt{(k_1 - k_2)^2 + 4k_{12}k_{21}}}{2k_{12}} \cdot \exp(s_{2,1}t),$$

где $N_{1,2}(t_0)$ – населенность энергетических уровней S_1 и S_2 в начальный момент времени. Видно, что в общем случае для каждого уровня наблюдается биэкспоненциальная зависимость распада населенности, однако скорости совпадают, и закон затухания фотолюминесценции может быть представлен как

$$PL(t) = A_1 \cdot \exp\left(-\frac{t}{\tau_1}\right) + A_2 \cdot \exp\left(-\frac{t}{\tau_2}\right),$$

$$\tau_{1,2} = 2 \cdot \left[k_1 + k_2 \pm \sqrt{(k_1 - k_2)^2 + 4k_{12}k_{21}} \right]^{-1}.$$

Для того чтобы выразить собственные времена жизни для уровней S_1 и S_2 , используем «золотое правило» Ферми и учтем также эффект диэлектрического экранирования:

$$k_{10} = \tau_0(S_1)^{-1} = \frac{2e^2 \sqrt{\varepsilon_1} \omega}{3mc^3 \hbar} \cdot \frac{2P^2}{m} \cdot \left(\frac{3\varepsilon_1}{\varepsilon_2 + 2\varepsilon_1} \right)^{-2},$$

$$k_{20} = \tau_0(S_2)^{-1} = C \cdot \frac{2e^2 \sqrt{\varepsilon_1} \frac{E - dE}{\hbar}}{3mc^3 \hbar} \cdot \frac{2P^2}{m} \cdot \left(\frac{3\varepsilon_1}{\varepsilon_2 + 2\varepsilon_1} \right)^{-2},$$

где e и m – заряд и масса электрона; ε_1 и ε_2 – диэлектрические проницаемости PbS (17,2) и CCl_4 (2,2); ω – частота фундаментального перехода; c – скорость света; $\hbar$ – постоянная Планка; $\frac{E - dE}{\hbar}$ соответ-

ствует частоте перехода с уровня S_2 ; параметр $\frac{2P^2}{m}$ равен 3 эВ для PbS; C – константа.

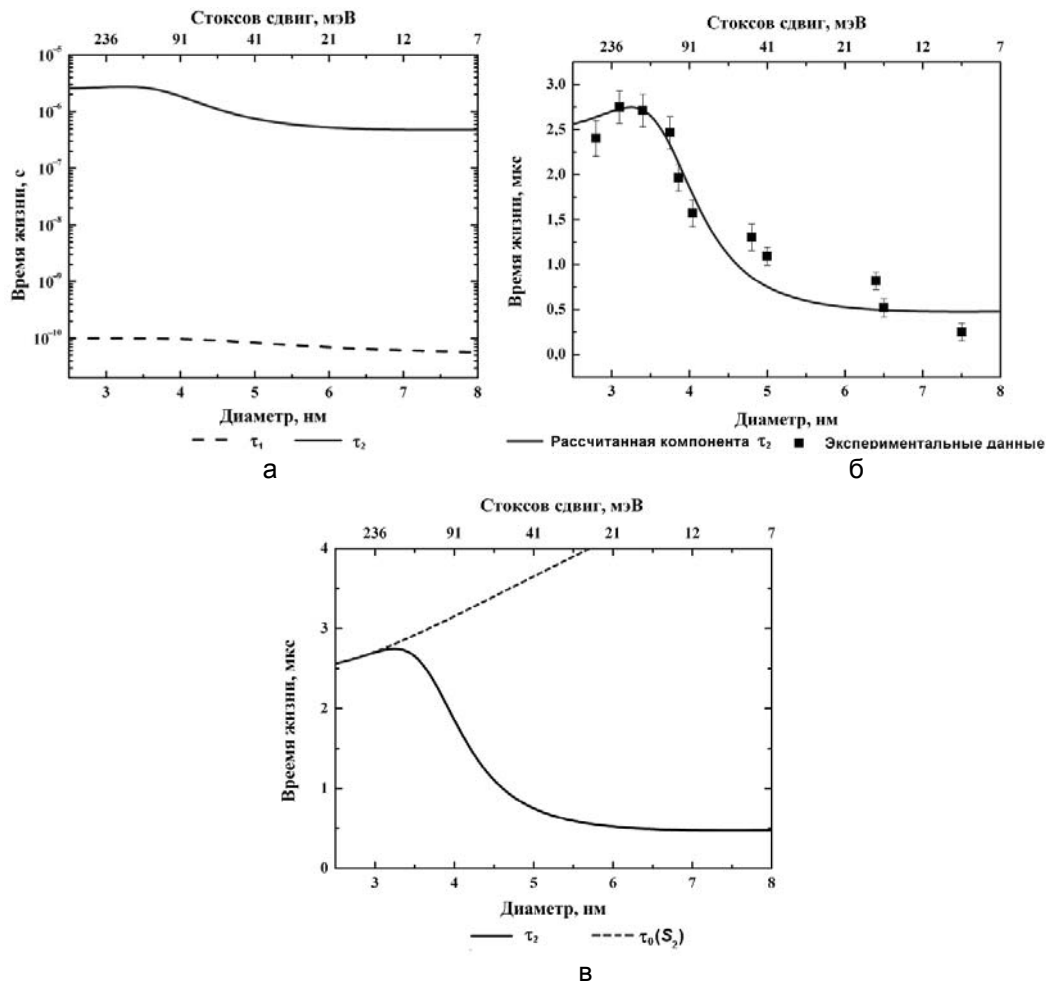


Рис. 3. Зависимость быстрой и медленной компонент времени затухания от размера КТ (а); сравнение экспериментальных данных и рассчитанной медленной компоненты (б); сравнение размерных зависимостей τ_2 , полученных при учете только «золотого правила» и эффекта диэлектрического экранирования (штриховая линия) и при учете также кинетических процессов в трехуровневой системе (сплошная линия) (в)

Расчет обеих компонент времени в зависимости от диаметра КТ и соответствующей величины стокового сдвига представлены на рис. 3, а. Как видно, время компоненты τ_1 мало и может быть не зарегистрировано в эксперименте, в то время как компонента τ_2 меняется в широких пределах и хорошо описывает экспериментальные результаты (рис. 3, б).

На рис. 3, в, приведено сравнение размерных зависимостей компоненты времени жизни τ_2 , полученных при учете только «Золотого правила» и эффекта диэлектрического экранирования (штриховая линия), и учета также кинетических процессов в трехуровневой системе (сплошная линия). Видно, что при слишком большом энергетическом зазоре между уровнями (КТ наименьшего диаметра) трансфер электронов обратно на уровень S_1 с уровня S_2 практически не имеет места, поэтому время жизни τ_2 определяется в основном естественной зависимостью от частоты перехода (кривые совпадают). Затем, при сокращении энергетического зазора термическая активация носителей становится более вероятной, и появляется дополнительный канал для снижения населенности уровня S_2 , что приводит к перепутыванию двух этих энергетических состояний, биэкспоненциальному распаду и плавному уменьшению времени жизни τ_2 . Таким образом, аномальная размерная зависимость времен жизни люминесценции связана с возможностью переходов как с понижением, так и повышением энергии при комнатной температуре.

Возможным объяснением природы такой системы в КТ является существование поверхностных состояний и излучательная рекомбинация с их участием. Как известно из теории полупроводников, неспаренные химические связи атомов, находящихся на поверхности, могут приводить к возникновению поверхностных состояний Шокли, могут образовывать дискретные или квазинепрерывные энергетические уровни в запрещенной зоне. Их концентрация может быть сопоставима по величине с концентрацией поверхностных атомов [11]. В КТ доля атомов, связанных с поверхностью, становится существенной и играет все большую роль с уменьшением размера КТ. На возможное влияние поверхностных состояний на процессы фотолюминесценции в КТ PbS указывается и в работах [12, 13].

Заключение

Впервые исследованы размерные зависимости спектральных и кинетических параметров люминесценции квантовых точек сульфида свинца в широком спектральном диапазоне 0,8–2 мкм. Результаты указывают на существование эффективно люминесцирующего энергетического уровня внутри запрещенной зоны, положение которого зависит от размера квантовых точек. На основе трехуровневой системы показано, что аномальная размерная зависимость времен жизни люминесценции может быть связана с существованием переходов, как с понижением, так и повышением энергии при комнатной температуре.

Работа выполнена при финансовой поддержке РФФИ (грант 12-02-01263 и 12-02-00938) и Минобрнауки РФ (проекты 11.519.11.3020, 11.519.11.3026 и 14.740.11.1366).

Литература

1. Sargent E.H. Infrared quantum dots // *Adv. Mater.* – 2005. – V. 17. – № 5. – P. 515–522.
2. Clark S.W., Harbold J.M., Wise F.W. Resonant energy transfer in PbS quantum dots // *J. Phys. Chem. C.* – 2007. – V. 111. – P. 7302–7305.
3. Гапоненко М.С., Юмашев К.В., Онущенко А.А. Люминесценция нанокристаллов сульфида свинца в матрице из силикатного стекла // *Журнал прикладной спектроскопии.* – 2010. – Т. 77. – № 5. – С. 720–724.
4. Fernee M.J., Thomsen E., Jensen P., Rubinsztein-Dunlop H. Highly efficient luminescence from a hybrid state found in strongly quantum confined PbS nanocrystals // *Nanotechnology.* – 2006. – V. 17. – P. 956–962.
5. Zhang J., Jiang X. Confinement-dependent below-gap state in PbS quantum dot films probed by continuous-wave photoinduced absorption // *J. Phys. Chem. B.* – 2008. – V. 112. – № 32. – P. 9557–9560.
6. Warner J.H., Thomsen E., Watt A.R. et al. Time-resolved photoluminescence spectroscopy of ligand-capped PbS nanocrystals // *Nanotechnology.* – 2005. – V. 16. – P. 175–179.
7. Парфенов П.С., Баранов А.В., Вениаминов А.В., Орлова А.О. Комплекс для люминесцентного анализа макро- и микрообразцов в ближнем инфракрасном диапазоне // *Оптический журнал.* – 2011. – Т. 78. – С. 48–52.
8. Парфенов П.С., Литвин А.П., Баранов А.В. и др. Калибровка спектральной чувствительности приборов в ближней инфракрасной области // *Журнал прикладной спектроскопии.* – 2011. – Т. 78. – № 3. – С. 460–466.
9. Парфенов П.С., Литвин А.П., Баранов А.В. и др. Измерение времен затухания люминесценции квантовых точек PbS в ближнем инфракрасном диапазоне // *Оптика и спектроскопия.* – 2012. – Т. 6. – С. 939–944.
10. Новотный Л., Хехт Б. Основы нанооптики. – М.: ФИЗМАТЛИТ, 2009. – 484 с.
11. Зеегер К. Физика полупроводников. – М.: Мир, 1977. – 615 с.
12. Kim D., Kuwabara T., Nakayama M. Photoluminescence properties related to localized states in colloidal PbS quantum dots // *Journal of luminescence.* – 2006. – V. 119–120. – P. 214–218.

13. Lifshitz E., Sirota M., Porteanu H. Continuous and time-resolved photoluminescence study of lead sulfide nanocrystals, embedded in polymer film // Journal of crystal growth. – 1999. – V. 196. – P. 126–134.

- | | |
|-------------------------------------|---|
| <i>Литвин Александр Петрович</i> | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, litvin88@gmail.com |
| <i>Парфенов Петр Сергеевич</i> | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, qrs peter@pochta.ru |
| <i>Ушакова Елена Владимировна</i> | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, el.ushakova@gmail.com |
| <i>Баранов Александр Васильевич</i> | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, ст. научный сотрудник, профессор, a_v_baranov@yahoo.com |

УДК 535.2

МЕТОД КОМПЕНСАЦИИ СПЕКТРАЛЬНОЙ НЕОДНОРОДНОСТИ ИСТОЧНИКА ИЗЛУЧЕНИЯ ПОДСТРОЙКОЙ ЕГО ЦВЕТОВОЙ ТЕМПЕРАТУРЫ ДЛЯ ГИПЕРСПЕКТРАЛЬНЫХ ПРИЛОЖЕНИЙ В МИКРОСКОПИИ

И.П. Гуров, А.И. Лопатин, А.В. Мельников

Рассмотрен метод компенсации спектральной неоднородности источника излучения монохромного осветителя микроскопа-гиперспектрофотометра для повышения эффективности использования динамического диапазона регистрирующей видеоматрицы.

Ключевые слова: микроскопия, гиперспектральное изображение, спектрофотометрия.

Введение

Гиперспектральные приложения в микроскопии основаны на последовательной записи набора изображений, полученных при освещении микрообъекта монохромным излучением на различных длинах волн [1]. Для получения набора изображений, записанных в широком спектральном диапазоне, необходимо учитывать спектральную неоднородность излучения источника света, спектральную чувствительность приемника, спектральное пропускание оптических элементов микроскопа-гиперспектрофотометра и распределение коэффициента отражения (пропускания) по площади исследуемого образца, так как аналого-цифровой преобразователь видеоматрицы имеет ограниченный динамический диапазон [2]. Известны методы расширения эффективного динамического диапазона видеоматриц [2, 3], наиболее популярные из которых сводятся к подстройке времени экспозиции [3–5]. Экспозиция видеоматрицы $H(\lambda)$ определяется произведением освещенности светочувствительного элемента на время экспозиции [6]:

$$H(\lambda) = E(\lambda)t_{exp},$$

где $E(\lambda)$ – освещенность в плоскости видеоматрицы; t_{exp} – время экспозиции.

Спектральную неоднородность излучения, освещающего плоскость видеоматрицы, можно компенсировать изменением времени экспозиции на различных спектральных интервалах [1]. Применение этого метода компенсации спектральной неоднородности приводит к усложнению учета темнового тока видеоматрицы при вычислениях коэффициента пропускания или отражения, так как уровень темнового тока видеоматрицы зависит от значения времени экспозиции. В этом случае требуются дополнительные операции при выполнении процедуры записи гиперспектрального изображения, которые приводят к увеличению времени ее проведения. В настоящей работе рассматривается другой подход к повышению эффективности использования динамического диапазона регистрирующей видеоматрицы, который заключается в подстройке уровня освещенности изменением цветовой температуры источника излучения монохромного осветителя микроскопа-гиперспектрофотометра при регистрации гиперспектральных изображений на различных длинах волн.

Теоретические основы метода

Источник излучения для гиперспектральных приложений микроскопии должен обеспечивать возможность получения изображений микрообъектов в узких спектральных интервалах во всем рабочем спектральном диапазоне, который имеет границы примерно 0,4–1,0 мкм [1], с учетом спектральной чувствительности видеоматрицы [7]. Этим условиям соответствует галогеновая лампа накаливания, которая имеет номинальную цветовую температуру около 2850 К [8].

При использовании гиперспектральных методов в микроскопии возникает задача повышения уровня точности измерений коэффициента пропускания, так как для применения спектральных методов анализа состава микрообъектов необходимо обеспечить возможность измерения коэффициента пропускания с точностью до четвертого знака [9]. Динамический диапазон обычной монохромной видеоматрицы равен трем порядкам [2]. Спектральная неоднородность излучения монохромного осветителя микроскопа-гиперспектрофотометра и спектральная неоднородность чувствительности элементов видеоматрицы приводят к снижению точности измерений коэффициента пропускания. В настоящей работе предлагается метод компенсации спектральной неоднородности, обеспечивающий повышение точности измерений.

Значение интенсивности излучения, освещающего элементы видеоматрицы, можно записать в виде уравнения

$$I(\lambda) = I_{src}(\lambda)T_{dev}(\lambda)S_v(\lambda)T_s(\lambda),$$

где $I_{src}(\lambda)$ – спектр излучения источника света; $T_{dev}(\lambda)$ – коэффициент пропускания оптических элементов микроскопа-гиперспектрофотометра; $S_v(\lambda)$ – спектральная чувствительность элементов видеоматрицы, $T_{sample}(\lambda)$ – коэффициент пропускания исследуемого микрообъекта.

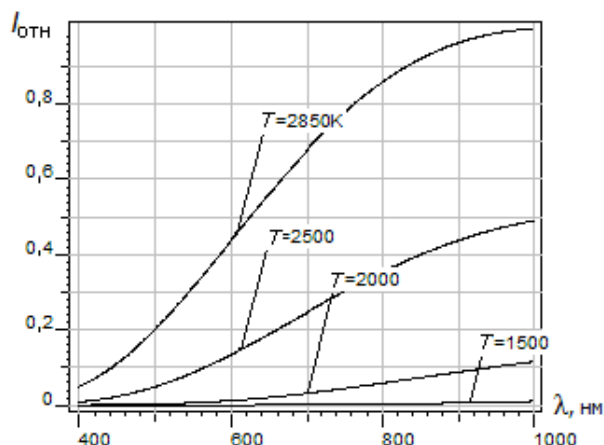


Рис. 1. Спектр излучения галогеновой лампы с различными цветовыми температурами нити накаливания

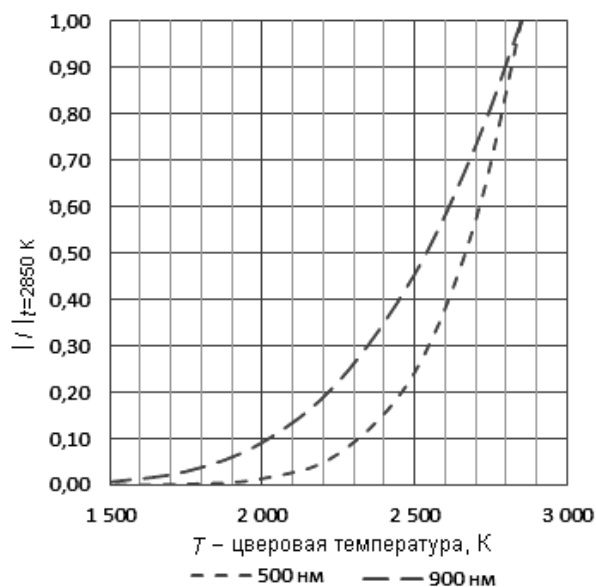


Рис. 2. Графики зависимости отношения интенсивности излучения галогеновой лампы от значения цветовой температуры нити накаливания к интенсивности, соответствующей цветовой температуре, равной 2850 K, на длинах волн 500 нм и 900 нм

Гиперспектральные методы микроскопии сводятся к получению значений величины коэффициента пропускания $T_{sample}(\lambda)$ смежных участков микрообъекта, площадь которых ограничивается размером одного элемента видеоматрицы, а их количество – числом элементов видеоматрицы [10]. Повысить точность определения коэффициента пропускания участка микрообъекта $T_{sample}(\lambda)$ можно подстройкой

множителя $I_{src}(\lambda)$ таким образом, чтобы величина, равная произведению $I_{src}(\lambda)T_{dev}(\lambda)S_v(\lambda)$, оставалась постоянной. В этом случае изменение освещенности элементов видеоматрицы будет происходить только в результате изменения величины коэффициента пропускания $T_{sample}(\lambda)$. Величина коэффициента пропускания оптических элементов микроскопа-гиперспектрофотометра $T_{dev}(\lambda)$ и спектральная чувствительность видеоматрицы $S_v(\lambda)$ считаются постоянными, а интенсивность источника излучения можно изменять различными методами в зависимости от его типа [8]. Изменять спектр излучения $I_{src}(\lambda)$ галогеновой лампы можно изменением цветовой температуры ее нити накаливания [8]. Форма спектра излучения галогеновой лампы соответствует спектру абсолютно черного тела, который описывается формулой Планка [11]. На рис. 1 показаны спектры галогеновой лампы с различными цветовыми температурами нити накаливания. Из рис. 1 видно, что отношение интенсивностей излучения при различных цветовых температурах в отдельных узких спектральных интервалах различно. На рис. 2 показаны графики отношения интенсивностей излучения, соответствующих значениям цветовых температур, равных 1500 К, 2000 К и 2500 К, к интенсивности излучения, соответствующей цветовой температуре, равной 2850 К, на длинах волн, равных 500 нм и 900 нм.

Из рис. 2 видно, что интенсивность излучения галогеновой лампы на различных длинах волн при разных цветовых температурах нити накаливания изменяется с разными значениями их отношений к интенсивности излучения на той же длине волны при цветовой температуре, равной 2850 К. В связи с этим подстройка цветовой температуры должна выполняться на каждой длине волны по своей калибровочной кривой.

Экспериментальные результаты

Для записи набора калибровочных кривых была собрана установка, состоящая из монохроматора МДР-206 с дифракционной решеткой 1200 штр/мм, осветителя с галогеновой лампой КГМ 12-20, лабораторного источника питания и фотоприемного устройства на основе фотодиода ФДУК-2УТ, спектральная чувствительность которого наиболее близко соответствует спектральной чувствительности монохромной видеоматрицы OV9121, которая установлена в микроскопе-спектрофотометре [1]. Оптическая схема установки показана на рис. 3. Задание различных значений цветовой температуры нити накаливания галогеновой лампы осветителя осуществлялось с помощью изменения напряжения источника питания.

При получении калибровочных кривых не учитывался коэффициент пропускания оптических элементов микроскопа. Был выполнен анализ спектров пропускания микрообъективов из каталога фирмы Carl Zeiss [12]. Коэффициент пропускания микрообъективов варьируется от 80% до 95% в спектральном интервале 0,4–1,0 мкм. Уровень неоднородности коэффициента пропускания микрообъектива существенно не изменяет значения освещенности элементов видеоматрицы, поэтому значение коэффициента пропускания оптических элементов микроскопа не учитывалось при расчете калибровочных кривых, его полагали равным единице во всем спектральном диапазоне.

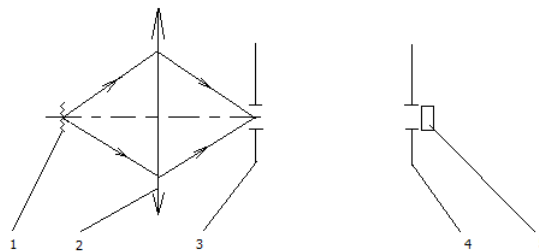


Рис. 3. Оптическая схема установки для получения набора калибровочных кривых: 1 – нить накаливания галогеновой лампы; 2 – конденсор осветителя; 3 – входная щель монохроматора; 4 – выходная щель монохроматора; 5 – фотоприемное устройство, установленное вблизи выходной щели

На рис. 4 показан набор графиков полученных зависимостей произведения $I_{src}(\lambda)T_{dev}(\lambda)S_v(\lambda)$ от длины волны, которые отнормированы на максимальное значение при различных напряжениях питания галогеновой лампы. Диапазон значений напряжения питания галогеновой лампы изменялся от номинального уровня, который равен 12 В, до уровня, равного 2 В. Исходя из данных, представленных на рис. 4, установлено, что минимальное значение величины $I_{src}(\lambda)T_{dev}(\lambda)S_v(\lambda)$ на длине волны, равной 400 нм, при напряжении питания галогеновой лампы 12 В выше уровня, соответствующего длине волны 1000 нм при напряжении питания галогеновой лампы 2 В. Таким образом, изменением цветовой температуры можно подстроить уровень освещенности элементов видеоматрицы, чтобы он оставался постоянным во всем рабочем спектральном диапазоне, используя полученные калибровочные кривые.

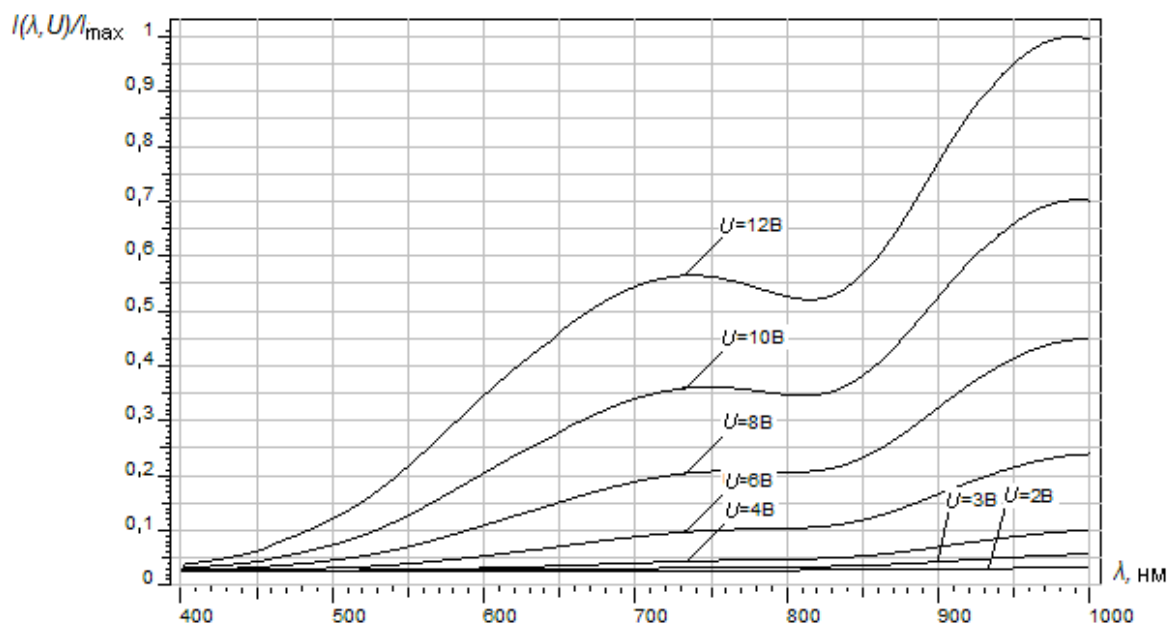


Рис. 4. Графики зависимостей произведения $I_{src}(\lambda)T_{dev}(\lambda)S_v(\lambda)$ от напряжения питания галогеновой лампы

На рис. 5 показаны калибровочные кривые для двух значений длин волн, равных 500 нм и 900 нм, которые построены с использованием линейной интерполяции.

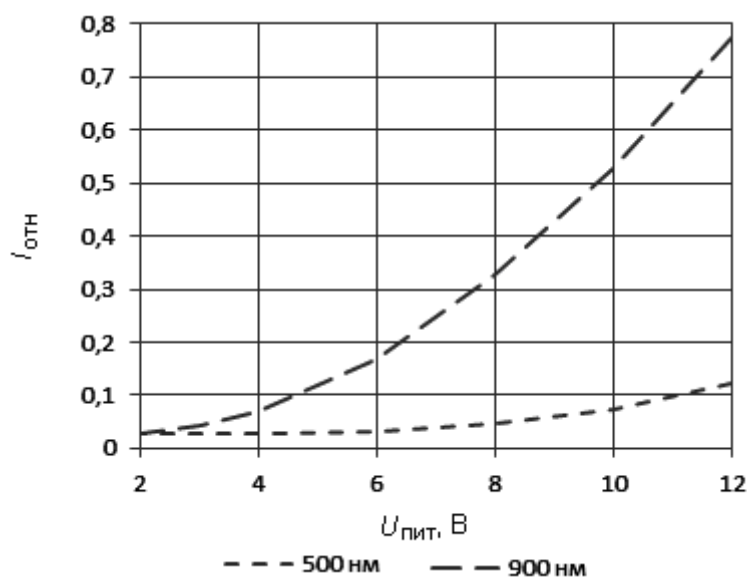


Рис. 5. Графики полученных калибровочных кривых для двух значений длин волн, равных 500 нм и 900 нм

Заключение

Полученные экспериментальные результаты показали возможность применения предложенного метода компенсации спектральной неоднородности источника излучения во всем рабочем спектральном диапазоне микроскопа-гиперспектрофотометра 0,4–1,0 мкм. Метод позволил избежать дополнительных операций при записи гиперспектрального изображения микрообъекта, связанных с учетом зависимости уровня темнового тока видеоматрицы от времени экспозиции, так как время экспозиции в рассматриваемом методе остается постоянным. Алгоритмы построения калибровочной кривой по полученным данным на установке, описанной в работе, для заданной длины волны и расчета напряжения питания галогеновой лампы по заданному значению относительной интенсивности реализованы в программном обеспечении микроскопа-гиперспектрофотометра и позволяют полностью автоматизировать описанный метод.

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

Литература

1. Белашенков Н.Р., Гуров И.П., Лопатин А.И., Мельников А.В. Микроскоп-спектрофотометр с матричным фотоприемником // Научно-технический вестник СПбГУ ИТМО. – 2007. – Т. 43. – С. 260–265.
2. Nayar S.K., Branzoi V. Adaptive dynamic range imaging: optical control of pixel exposure over space and time // Proc. of the Ninth IEEE International Conference on Computer Vision (ICCV). – 2003. – V. 2. – P. 1168–1175.
3. Nuske S., Roberts J., Wyeth G. Extending the dynamic range of robotic vision // Proc. of the 2006 IEEE International Conference on Robotics and Automation. – 2006. – P. 162–167.
4. Burt P., Kolczynski R.J. Enhanced image capture through fusion // Proc. of International Conference on Computer Vision (ICCV). – 1993. – P. 173–182.
5. Nayar S.K., Mitsunaga T. High dynamic range imaging: spatially varying pixel exposures // Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. – 2000. – V. 1. – P. 472–479.
6. Понс Ж., Форсайт Д. Компьютерное зрение. Современный подход. – М.: Вильямс, 2004. – 928 с.
7. Гуров И.П., Лопатин А.И., Мельников А.В. Метод цветовой визуализации изображений микрообъектов в инфракрасной области // Научно-технический вестник СПбГУ ИТМО. – 2007. – Т. 43. – С. 266–271.
8. Ишанин Г.Г., Панков Э.Д., Андреев А.Л., Польшиков Г.В. Источники и приемники излучения. – СПб: Политехника, 1991. – 240 с.
9. Зайдель А.Н., Островская Г.В., Островский Ю.И. Техника и практика спектроскопии. – М.: Наука, 1972. – 376 с.
10. Smith R. Introduction to Hyperspectral Imaging. – NY: MicroImages, 2006. – 24 p.
11. Ландсберг Г.С. Оптика. – М.: ФИЗМАТЛИТ, 2003. – 848 с.
12. Каталог микрообъективов фирмы Carl Zeiss [Электронный ресурс]. – Режим доступа: <http://www.micro-shop.zeiss.com>, свободный. Яз. англ. (дата обращения 28.07.2012)

Гуров Игорь Петрович

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

Лопатин Александр Иосифович

– ОАО «ЛОМО», кандидат физ.-мат. наук, главный оптик, ailopatin@gmail.com

Мельников Алексей Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, melnikov.alexey@gmail.com

УДК 53.086

ИССЛЕДОВАНИЕ РАСТИТЕЛЬНЫХ ТКАНЕЙ МЕТОДОМ ОПТИЧЕСКОЙ КОГЕРЕНТНОЙ МИКРОСКОПИИ

И.П. Гуров, М.А. Волынский, Е.В. Жукова, Н.Б. Маргарянц

Обсуждаются возможности изучения внутренней микроструктуры тканей растительного происхождения методом оптической когерентной микроскопии с перестраиваемой длиной волны (1305 ± 75 нм). Приводятся изображения томограмм (В-сканов) и 3D-изображения областей исследования для тканей древовидного алое (*Aloe arborescens*), рипсалидопсиса (*Hatiora gaertneri*), пеларгонии зональной (*Pelargonium zonale*). Анализируются результаты исследования опробковения, деградации поверхностных слоев. Приведены спектры диффузного отражения здоровых, больных и сухих тканей листьев в области 900–1700 нм.

Ключевые слова: оптическая когерентная микроскопия, томограмма, 3D-изображение, структура ткани растения.

Введение

Применение метода оптической когерентной микроскопии для изучения микроструктуры оптически неоднородных и слоистых сред убедительно продемонстрировало диагностические возможности данного неразрушающего метода для контроля материалов и веществ в решении разнообразных фундаментальных и прикладных задач. Возможности применения метода к изучению растительных тканей были проанализированы в работе [1], в которой подробно рассмотрена специфика изучения клеточного строения тканей арабидопсиса и обсуждаются возможности трехмерной визуализации изучаемой области. Авторами было выполнено детальное сопоставление реконструированных изображений с данными, получаемыми при помощи сканирующего электронного микроскопа, отмечены преимущества метода оптической когерентной микроскопии перед традиционными методами исследования растительных тканей. В работах [2, 3] приведены сравнительные результаты исследования тканей листьев традесканции методами оптической когерентной микроскопии и томографии, отмечается несомненное преимущество в изучении глубинных слоев тканей, неразрушающий характер измерений. Особый интерес в изучении тканей вызывают работы, посвященные анализу влияния внешних факторов на развитие клеточных структур и ткани, таких как, например, влияние водоснабжения на протекание физиологических процес-

сов в клетках, на изменение структуры и деформацию слоев. В работах [4, 5] при помощи метода оптической когерентной микроскопии исследованы влияния вирусной инфекции на строение ткани цветка орхидеи, изучена морфология и развитие в клетках хлоропластов листьев табака. Практическая реализация метода позволила авторам [6] создать оптический сканер, работающий на длине волны 830 нм, обеспечивающий ширину сканирования до 40 мм и предназначенный для диагностики поверхности тканей овощных и фруктовых культур. Цель настоящей работы – проанализировать возможности применения оптической когерентной микроскопии с перестраиваемой длиной волны в районе 1300 нм для изучения поверхностных тканевых структур растительного происхождения.

Экспериментальная установка

Экспериментальные исследования были выполнены с использованием возможностей оптического когерентного микроскопа (ОКМ) EX1301 с перестраиваемой длиной волны, изготовленного фирмой Michelson Diagnostics (Великобритания). В конструкции используется микроинтерферометр малой когерентности. Источником излучения служит лазер HLS-2000-MDL производства фирмы SANTEC (Япония). В процессе измерения томограмм происходит изменение длины волны излучения в пределах 1305 ± 75 нм. Мощность излучения источника составляет 15 мВт. Оптическая схема фокусировки – многоканальная, что обеспечивает на одной длине волны одновременную фокусировку излучения сразу на четыре глубины сканирования в пределах поверхностного слоя объекта. Результаты совместной обработки отраженных сигналов учитываются при регистрации томограмм. Прибор характеризуется разрешением по глубине не хуже 10 мкм и в боковом направлении не хуже 7 мкм, что создает условия для использования его в исследованиях тканей растительного происхождения. Подробно методика измерения томограмм на данном приборе рассмотрена в [7].

Для сравнительного анализа отражающих свойств поверхностей тканей растений были выполнены измерения спектров диффузного отражения. Для измерений в области 900–1700 нм использовался спектрометр NIRQUEST и приставка для измерения спектров диффузного отражения производства фирмы Ocen Optics. В качестве источника излучения использовалась галогенная лампа (360–2000 нм) мощностью 20 Вт.

Экспериментальные результаты

Анализ экспериментальных результатов целесообразно начать с демонстрации общих возможностей используемого метода в изучении растительных тканей. На рис. 1 представлены результаты исследования листа алоэ древовидного (*Aloe arborescens*).

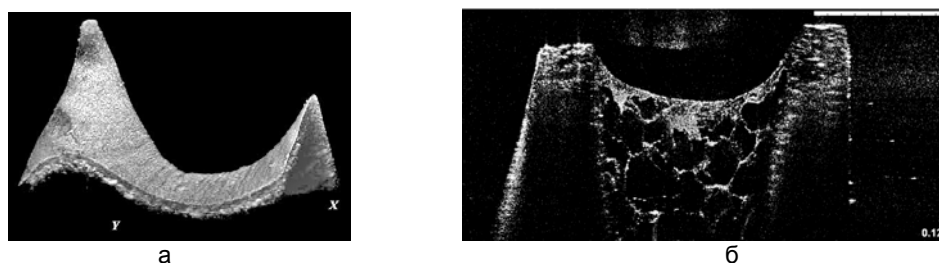


Рис. 1. 3D-изображение и томограмма листа алоэ (*Aloe arborescens*): 3D-изображение в области шипа ($x=2,47$ мм; $y=3,0$ мм; $z=1,69$ мм) (а); томограмма (В-скан) поверхности поперечного среза листа (б).
Размер В-скана – 3,63 мм×2 мм

На приведенном реконструированном 3D-изображении шипа листа алоэ (рис. 1, а), на поверхности видны темные области, появление которых связано с локальной деформацией поверхностных тканевых слоев, что и повлияло на величину отраженного сигнала при записи томограмм. Это – пример изучения на ОКМ растительного образца со сложной формой. Результат визуализации показывает, что анатомическое поверхностное изучение тканей растений важно для выявления патологических областей на начальной стадии развития. На рис. 1, б, приведена томограмма поперечного среза листа алоэ. На томограмме хорошо видна ткань мезофилла, которая выполняет водоносную и запасную функции. Ткань состоит из крупных тонкостенных клеток с внутриклеточным гелем. Анализ изображения поперечного сечения листа показал, что общую толщину слоев эпидермиса и хлоренхимы для среза внешней части листа (левая часть изображения на томограмме) можно оценить в 370 мкм, клетки мезофилла занимают пространство для данного сечения около 1580 мкм. Стрелка прогиба поверхности в области геля после разрезания листа составила около 410 мкм, форма поверхности геля определяется эффектом поверхностного натяжения. Размер клеток мезофилла в диаметре составляют более 310 мкм. Глубина проникновения излучения в области хлоренхимы составила около 315 мкм, а в области мезофилла – более 1250 мкм. Методика регистрации томографических изображений позволяет оценивать внутриклеточные размеры отдельных структур тканей, как это видно на примере изучения листа алоэ.

Учитывая, что разрешение прибора не позволяет наблюдать по глубине органические структуры размером менее 10 мкм, были выполнены исследования более крупных структурных видоизменений поверхностных тканей. В качестве образца использовались листья рипсалидопсиса (*Hatiora gaertneri*). Приведено изображение поверхности здоровой ткани растения (рис. 2, а). Излучение, проникая вглубь, позволяет визуализировать структуру тканей в пределах толщины слоя, равного примерно 310 мкм. Сильно отражающий поверхностный слой – это кутикула, содержащая воскообразную пленку. На томограмме не видны границы между слоем кутикулы и эпидермиса, а также границы слоя полисадных клеток, находящегося под слоем эпидермиса. Но попиксельный анализ увеличенного изображения позволил оценить общую протяженность этих трех слоев (около 30 мкм).

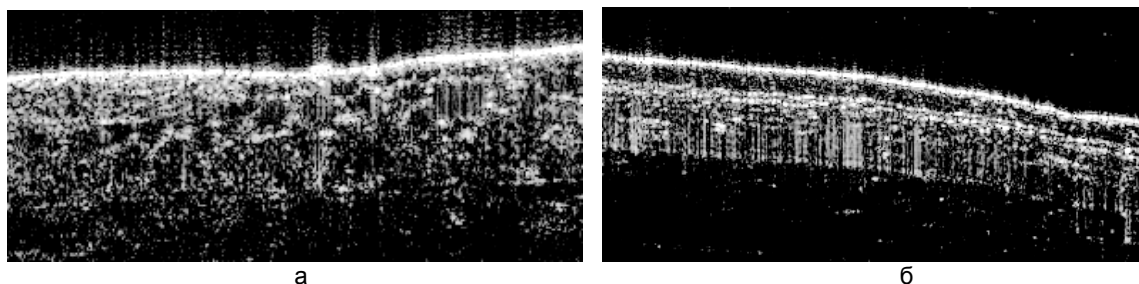


Рис. 2. Томограммы (В-сканы) поверхности листа рипсалидопсиса (*Hatiora gaertneri*): здоровая ткань, размер В-скана – 1,28 мм×0,56 мм (а); поверхность с областью опробковения, размер В-скана – 1,58 мм×0,73 мм (б)

На томограмме здоровой ткани хорошо видны клетки губчатой паренхимы, диаметр которых составляет более 70 мкм. На рис. 2, б, представлена томограмма поверхности листа в области опробковения, которая образовалась за счет возрастного накопления воскообразного вещества суберина. Толщину этого слоя можно оценить в пределах 65 мкм. В области воскового слоя глубина проникновения излучения составила около 260 мкм, о сильно отражающих свойствах границы воздух/восковой слой свидетельствует появление участка с артефактами, что не наблюдается для естественной границы воздух/кутикула. Это – примеры изучения поверхностного процесса, обусловленного физиологическими явлениями, протекающими в растении.

Метод оптической когерентной томографии может быть использован и для изучения процессов поверхностной деградации тканей растений. Результаты показаны на рис. 3, а–г. Рис. 3, а, б, демонстрируют процессы видоизменения тканей под воздействием бактерий. Ткани рипсалидопсиса могут быть подвержены заражению группой бактерий *Erwinia*, которая приводит к изменению структуры поверхностного слоя. Глубина проникновения излучения в поверхность составила около 280 мкм, в пределах которой четко визуализированы участки, отличающиеся по своей структуре от здоровой ткани. На рис. 3, а, размер образовавшихся пор в диаметре составил около 90–140 мкм, на томограмме (рис. 3, б) наблюдаются поры размером в 350–450 мкм вдоль поверхности листа. Вокруг пор видны сильно рассеивающие участки, появление которых указывает, что это – области с быстротекающими процессами метаболизма, обусловленного действием бактерий, так как образуются структуры с мелкими органеллами, они и вызывают сильное рассеяние фотонов. Деградация тканей бактериальными колониями сопровождается деформацией поверхности листа, появлением углублений. На рис. 3, в, г, приведен другой пример изменения строения тканей. На поверхности листа рипсалидопсиса произошла локальная деформация поверхности, обусловленная образованием нароста. Природа таких явлений для данных растений может быть вызвана действием вирусов. Из томограммы видно, что ступенька по высоте нароста составляет около 600 мкм, а протяженность нароста более 3200 мкм. Произошло видоизменение структуры в тонких поверхностных слоях (кутикула/эпидермис). В центре нароста наблюдается уплотнение поверхностных слоев, суммарная толщина которых уже составляет около 70 мкм, это связано с необратимыми изменениями в тканях эпидермы. На рис. 3, г, показано реконструированное 3D-изображение нароста. Центральная часть нароста на изображении, имеющая более темный цвет, свидетельствует об отличии в отражающих свойствах на данном участке поверхности, это результат локального уплотнения и более длительного процесса видоизменения тканей. Представленные примеры наглядно демонстрируют возможности изучения поверхностных процессов, происходящих в тканях кутикулы и эпидермиса, в частности, воздействия бактерий, вирусов, развития онкологических процессов, внешних механических поверхностных повреждений.

На рис. 4, а, б, показана томограмма листа пеларгонии зональной (*Pelargonium zonale*) и реконструированное трехмерное изображение участка стебля листа. Толщину слоя кутикулы можно оценить менее 10 мкм, общую толщину слоев верхнего эпидермиса и полисадной паренхимы – около 85 мкм для данного образца, а размеры клеток губчатой паренхимы – менее 10 мкм. На трехмерном изображении образца видны как проводящие слои внутренних тканей стебля, так и поверхностные железистые волоски (трихомы),

на конце которых находятся головки-емкости с эфирным маслом. Этот пример демонстрирует возможности применения метода для изучения морфологического строения отдельных участков тканей.

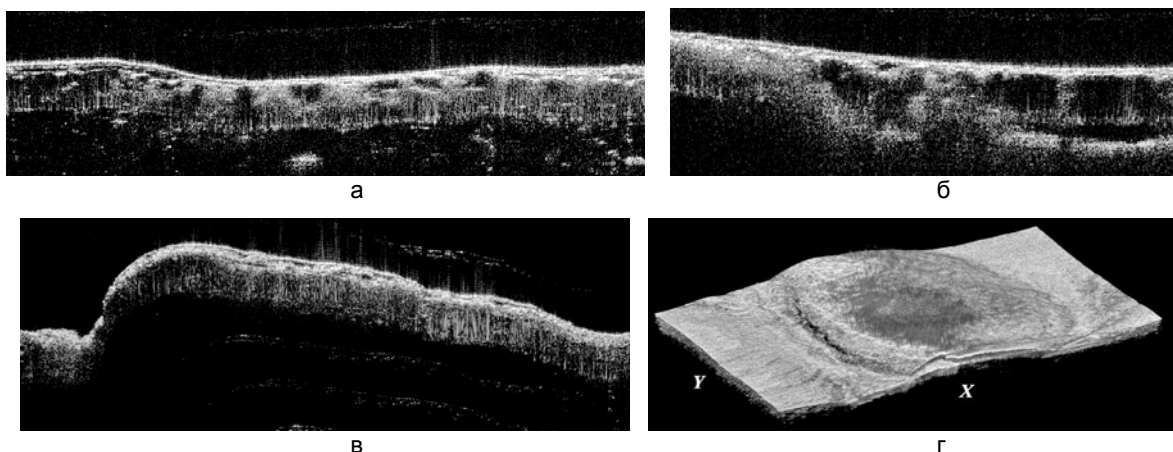


Рис. 3. Поверхность листа рипсалидопсиса (*Hattoria gaertneri*): в области разрушения ткани группой бактерий *Erwinia*, размеры В-сканов – 3,58 мм×0,92 мм (а); 2,64 мм×0,85 мм (б); в области нароста, размер В-скана – 3,73 мм×1,29 мм (в); 3D-изображение нароста на поверхности листа ($x=5,0$ мм; $y=3,0$ мм; $z=1,69$ мм) (г)

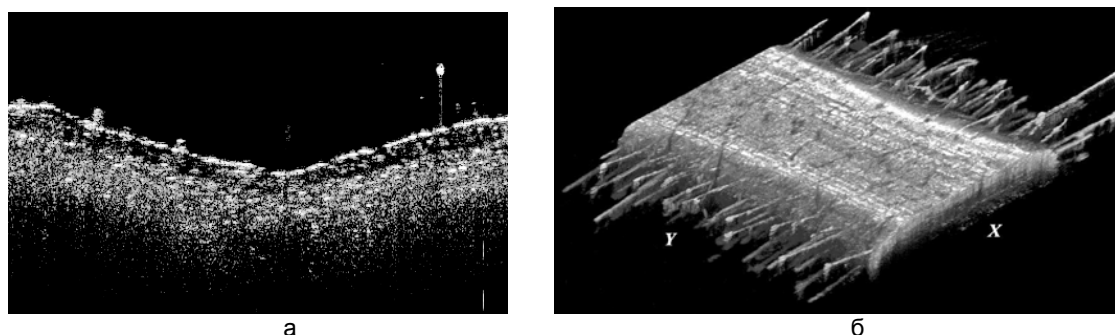


Рис. 4. Томограмма (В-скан) поверхности листа пеларгонии (*Pelargonium zonale*) и 3D-изображение участка стебля листа: поверхность листа в области верхней эпидермы, размер В-скана – 3,73 мм×1,29 мм (а); стебель листа ($x=5,0$ мм; $y=3,0$ мм; $z=0,86$ мм) (б)

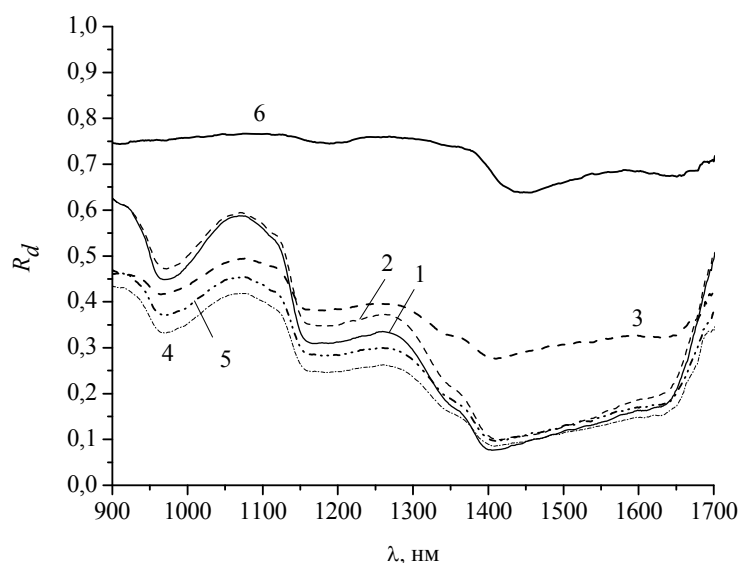


Рис. 5. Спектры диффузного отражения, измеренные от поверхности листьев рипсалидопсиса (*Hattoria gaertneri*): 1 – здоровый лист; 2 – увядший (бледно желтый); 3 – сухой участок на краю листа; 4 – область нароста на поверхности; 5 – участок опробковения; 6 – высохший полностью лист

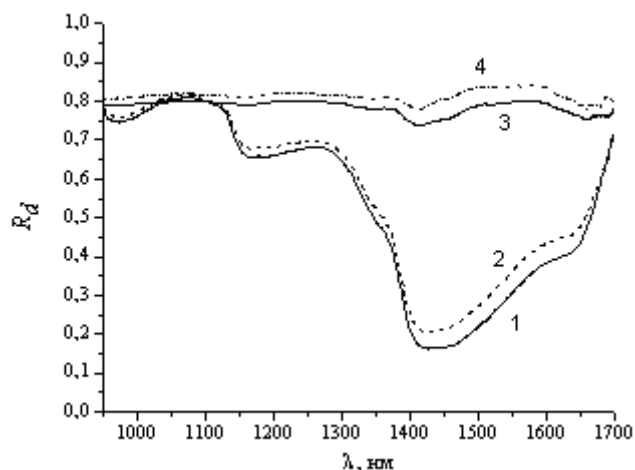


Рис. 6. Спектры диффузного отражения, измеренные от поверхности здорового листа пеларгонии (*Pelargonium zonale*): 1 – верхняя поверхность; 2 – нижняя поверхность; 3 – сухой участок листа со стороны внешней поверхности; 4 – сухой участок со стороны нижней поверхности листа

На рис. 5, 6 приведены спектры диффузного отражения, измеренные для разных растительных тканей. Объектами служили листья ропсидопсиса (*Hatiora gaertneri*) и пеларгонии зональной (*Pelargonium zonale*). Измерения были выполнены на поверхности здоровой ткани, а также на участках локальной дегенерации биотканей, вызванных, в том числе, и процессами высыхания. На рисунках видно, что на спектральный ход изменения коэффициентов диффузного отражения R_d влияют полосы поглощения воды, особенно в области длины волны $\lambda = 1400$ нм.

Образование локальных областей высыхания меняет естественную структуру листа. Процессы высыхания листьев связаны с уменьшением содержания влаги в поверхности, сморщиванием поверхностных тканей, видоизменением рельефа листьев, что сопровождается ростом коэффициента отражения и уменьшением интенсивности полос поглощения. Это хорошо видно на примере опыта с листом пеларгонии.

Приведенные спектры полезны для понимания возможностей применения метода оптической когерентной томографии в изучении поверхностных тканевых структур. Видно, что область 1230–1380 нм, которая охватывает диапазон изменения длин волн источника оптического когерентного микроскопа, обеспечивает необходимую чувствительность для диагностики структуры тканей и может быть использована для дальнейших исследований, но с большей разрешающей способностью.

Заключение

Представленные в работе результаты изучения микроструктуры растительных тканей методом оптической когерентной микроскопии показали, что в области длины волны 1300 нм можно успешно выполнять экспериментальные исследования по изучению поверхностных структур биоматериалов. Несмотря на расположение по обеим сторонам диапазона 1230–1380 нм интенсивных полос поглощения воды, регистрируемые томограммы содержат информацию о микроструктуре тканей и отдельных слоев.

В соответствии с результатами исследования разных образцов тканей следует отметить, что метод оптической когерентной микроскопии может быть использован для сканирования поверхности биотканей в данном спектральном диапазоне с целью последующей трехмерной визуализации отдельных участков. При соответствующей организации эксперимента метод может быть рекомендован для наблюдения в режиме *in vivo* за зарождением новых листьев, формированием зародышей, контролем пространственной направленности отдельных участков поверхности тканей, для проведения вспомогательных исследований при выполнении таксонометрических исследований и классификации растительных объектов.

Метод изучения поверхностных структур тканей растений может быть использован для диагностики состояния слоистых структур, для анализа влияния внешних факторов на развитие, дегенерацию сред на клеточном и тканевом уровне, как это было показано на примерах изучения влияния бактерий, процессов старения и локального перерождения клеточных структур. Это представляет важное направление в экспериментальной практике, так как толщина поверхностных слоев тканей и их клеточное строение свидетельствуют о возрастных свойствах растения, а также о характере адаптации тканей к среде обитания.

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

Литература

1. Hettinger J.W., Mattozzi M., Myer W.R., Williams M.E., Reeves A., Parsons R.L., Haskell R.C., Petersen D.C., Wang R., Medford J.I. Optical Coherence Microscopy. A technology for rapid, *in vivo*, non-destructive visualization of plants and plant cells // Plant Physiology. – 2000. – V. 123. – P. 3–15.
2. Сапожникова В.В., Каменский В.А., Куранов Р.В. Визуализация растительных тканей методом оптической томографии // Физиология растений. – 2003. – Т. 50. – № 2. – С. 316–320.
3. Кутис И.С., Сапожникова В.В., Куранов Р.В., Каменский В.А. Исследование методами оптической когерентной микроскопии и оптической когерентной томографии формо-функционального состояния тканей высших растений // Физиология растений. – 2005. – Т. 52. – № 4. – С. 635–640.
4. Chow T.H., Tan K.M., Ng B.K., Razul S.G., Tay C.M. Diagnosis of virus infection in orchid plants with highresolution optical coherence tomography // Journal of Biomedical Optics. – 2009. – V. 14. – № 1. – P. 014006–1–014006–6.
5. Boccara A.C., De Paepe R., Dubois A., Boccara M. Full field optical coherence tomography (OCT) and early alterations in chloroplast morphology // Proc. of SPIE. – 2007. – V. 6755. – 67550E1–67550E5.
6. Shiina T., Kishiwaki D., Ito M., Honda T., Okamura Y. OCT system for plant measurement // Proc. of SPIE. – 2005. – V. 5948. – 59482D1–59482D9.
7. Гуров И.П., Жукова Е.В., Маргарянц Н.Б. Исследование внутренней микроструктуры материалов методом оптической когерентной микроскопии с перестраиваемой длиной волны // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 3 (79). – С. 40–44.

Гуров Игорь Петрович

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

Волынский Максим Александрович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, научный сотрудник, maxim.volynsky@gmail.com

Жукова Екатерина Владимировна

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, zhukova-ifmo@mail.ru

Маргарянц Никита Борисович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ст. преподаватель, fosp@grv.ifmo.ru

УДК 681.513.6

АДАПТИВНЫЙ РЕГУЛЯТОР СТАБИЛИЗАЦИИ ПРОСТОЙ СТРУКТУРЫ¹

В.О. Никифоров, Д.Н. Герасимов

Для линейного стационарного многомерного объекта, представленного нижней треугольной канонической формой, предложен адаптивный регулятор стабилизации. Особенностью предложенного регулятора является простота его структуры – регулятор содержит всего один настраиваемый параметр и один простой алгоритм адаптации интегрального типа. Представлены робастные модификации регулятора, приведены результаты моделирования.

Ключевые слова: линейный объект, нижняя треугольная форма, адаптивное и робастное управление.

Введение

В современной теории адаптивного управления предложено решение многих сложных задач: адаптивного управления линейными объектами по выходу с использованием схемы расширенной ошибки [1–3] или алгоритмов адаптации высокого порядка [4–6], адаптивного управления нелинейными объектами с использованием итеративной процедуры синтеза «обратного обхода интеграторов» [3, 7, 8], робастного управления в условиях внешних возмущений [3, 9], адаптивной компенсации заранее неизвестных детерминированных возмущений [6, 10, 11] и т.д.

Одним из основных недостатков большинства предложенных систем адаптивного управления является сложная структура адаптивных регуляторов. Так, динамический порядок адаптивного регулятора с расширенной ошибкой равен $2n(n-m+2)-1$ (где n – степень знаменателя передаточной функции объекта управления, а m – степень числителя передаточной функции), а для расчета сигнала управления требуется выполнение $10n+3$ операций умножения и одной операции деления [3]. Очевидно, что реализация на практике регуляторов высокого динамического порядка, содержащих большое число нелинейных элементов (перемножения и деления), весьма затруднительна, несмотря на практические успехи развития микропроцессорных систем управления. В связи с этим одной из актуальных задач современной теории адаптивного управления является синтез достаточно простых и эффективных алгоритмов адаптивного управления.

Для линейного стационарного многомерного объекта с неизвестными параметрами, представленного нижней треугольной канонической формой, предложен регулятор адаптивной стабилизации простой структуры: адаптивный регулятор содержит всего один настраиваемый параметр и один простой алгоритм адаптации интегрального типа. Принцип построения регулятора основан на использовании идеи «большого коэффициента усиления». Представлены также робастные модификации регулятора, обеспечивающие ограниченность траекторий замкнутой системы, действующей в условиях внешних возмущений.

Постановка задачи

Рассматривается линейный объект управления вида

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{e}_n \mathbf{u}, \quad (1)$$

где $\mathbf{x} \in R^n$ – вектор состояния, доступный для прямых измерений, $\mathbf{e}_i$ – n -мерный единичный вектор с единицей на i -м месте, u – сигнал управления, а матрица системы имеет вид

$$\mathbf{A} = \begin{bmatrix} a_{11} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ a_{n-11} & a_{n-12} & a_{n-13} & \dots & 1 \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}. \quad (2)$$

Параметры матрицы a_{ij} считаются неизвестными. Пара $(\mathbf{A}, \mathbf{e}_n)$ является полностью управляемой, а пара $(\mathbf{e}_1^T, \mathbf{A})$ – полностью наблюдаемой. Требуется синтезировать стабилизирующее управление в форме обратной связи по состоянию, обеспечивающее выполнение целевого условия

$$\lim_{t \rightarrow \infty} \mathbf{x}(t) = \mathbf{0}. \quad (3)$$

¹ Работа выполнена при финансовой поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (соглашение №14.В37.21.0406)

Вспомогательный результат

Утверждение 1. Для пары $n \times n$ матриц $\mathbf{B}$ и $\mathbf{G}(\theta)$, где $\mathbf{B}$ – произвольная нижняя треугольная матрица, т.е.

$$\mathbf{B} = \begin{bmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix},$$

с постоянными коэффициентами $b_{i,j}$, а $\mathbf{G}(\theta)$ – сопровождающая матрица характеристического полинома от комплексной переменной p в форме бинома Ньютона

$$D(p, \theta) = (p + \theta)^n = p^n + C_{n-1}\theta p^{n-1} + C_{n-2}\theta^2 p^{n-2} + \dots + C_1 \theta^{n-1} p + \theta^n, \quad (4)$$

т.е.

$$\mathbf{G}(\theta) = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ -\theta^n & -C_1 \theta^{n-1} & \dots & -C_{n-1} \theta \end{bmatrix}, \quad (5)$$

существует такое число $\theta_0 > 0$, что для всех $\theta > \theta_0$ матрица

$$\mathbf{S} = \mathbf{B} + \mathbf{G}(\theta)$$

является гурвицевой.

В формулах (4) и (5) C_{n-j} – коэффициент бинома Ньютона степени n , стоящий при j -й степени переменной θ (при этом $C_n = 1$ и $C_0 = 1$).

Доказательство. Пусть

$$R(p) = \det(p\mathbf{I} - \mathbf{B}) = p^n + r_{n-1}p^{n-1} + \dots + r_1 p + r_0 \quad (6)$$

– характеристический полином матрицы $\mathbf{B}$, а r_j – его постоянные коэффициенты. Тогда можно показать, что характеристический полином $N(p, \theta)$ матрицы $\mathbf{S}$ можно представить в следующем виде:

$$\begin{aligned} N(p, \theta) &= \det(p\mathbf{I} - \mathbf{S}) = \det(p\mathbf{I} - \mathbf{B} - \mathbf{G}(\theta)) = p^n + (C_{n-1}\theta + r_{n-1})p^{n-1} + \\ &+ (C_{n-2}\theta^2 + r_{n-2} + \alpha_{n-2}(\bar{r}, \bar{\theta}_1, \bar{C}_1))p^{n-2} + (C_{n-3}\theta^3 + r_{n-3} + \alpha_{n-3}(\bar{r}, \bar{\theta}_2, \bar{C}_2))p^{n-3} + \dots \\ &+ (C_1\theta^{n-1} + r_1 + \alpha_1(\bar{r}, \bar{\theta}_{n-2}, \bar{C}_{n-2}))p + \theta^n + r_0 + \alpha_0(\bar{r}, \bar{\theta}_{n-1}, \bar{C}_{n-1}), \end{aligned}$$

где $\bar{r}$ – множество всех коэффициентов характеристического полинома (6); $\bar{\theta}_j$ – множество всех степеней θ^j ($j = 0, n$); $\bar{C}_j$ – множество всех коэффициентов C_{n-j} ($j = 1, n-1$); $\alpha_{n-j}(\bar{r}, \bar{\theta}_{j-1}, \bar{C}_{j-1})$ – сумма соответствующих коэффициентов r_i , степеней θ^i и коэффициентов C_i .

Так как члены α_{n-j} содержат настраиваемые коэффициенты θ степени $j-1$, а стоящие с ними в скобках при тех же степенях комплексной переменной p коэффициенты гурвицева характеристического полинома $D(p, \theta)$ содержат настраиваемые коэффициенты θ степени j , то справедливо следующее равенство:

$$\lim_{\theta \rightarrow \infty} N(p, \theta) = D(p, \theta).$$

Последнее означает, что полином $N(p, \theta)$ является гурвицевым при бесконечно больших значениях коэффициента θ . Однако свойство экспоненциальной устойчивости (задаваемое гурвицевым полиномом $D(p, \theta)$) является грубым, т.е. выполняется при некоторых отклонениях коэффициентов характеристического полинома от их номинальных значений. По этой причине полином $N(p, \theta)$ будет гурвицевым при конечных значениях достаточно большого коэффициента θ . Утверждение доказано.

Структура регулятора

Сформируем адаптивный регулятор в виде

$$u = -\theta^n x_1 - C_1 \theta^{n-1} x_2 - \dots - C_{n-2} \theta^2 x_{n-2} - C_{n-1} \theta x_n, \quad (7)$$

где постоянные коэффициенты C_j – коэффициенты характеристического полинома (4) (т.е. коэффициенты бинома Ньютона), а настраиваемый параметр θ генерируется алгоритмом адаптации

$$\dot{\theta} = \gamma x_1^2 \quad (8)$$

с постоянным положительным коэффициентом усиления γ .

Замечание. Управление (7) может быть записано в «свернутой» форме, минимизирующей при практической реализации число операций взаимного перемножения настраиваемого параметра θ :

$$u = -\theta \left(C_{n-1}x_n + \theta \left(C_{n-2}x_{n-1} + \theta \left(C_{n-3}x_{n-2} + \dots \theta x_1 \right) \right) \right). \quad (9)$$

Тогда справедливо следующее утверждение.

Утверждение 2. Адаптивный регулятор (7), (8) при любых $\gamma > 0$ и $\theta(0)$ обеспечивает для объекта управления (1) с произвольными начальными условиями $x(0)$ выполнение целевого условия (3).

Доказательство. Нетрудно убедиться, что матрица замкнутой системы, состоящей из объекта (1) и регулятора (7), имеет вид

$$\mathbf{A} + \mathbf{e}_n \begin{bmatrix} -\theta^n, -C_1\theta^{n-1}, \dots, -C_{n-2}\theta^2, -C_{n-1}\theta \end{bmatrix} = \begin{bmatrix} a_{11} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ a_{n-11} & a_{n-12} & a_{n-13} & \dots & 1 \\ a_{n1} - \theta^n & a_{n2} - C_1\theta^{n-1} & a_{n3} - C_2\theta^{n-2} & \dots & a_{nn} - C_{n-1}\theta \end{bmatrix}.$$

Иначе говоря, по своей структуре она соответствует матрице $\mathbf{S}$ из утверждения 1. В силу полной наблюдаемости пары $(\mathbf{e}_1^T, \mathbf{A})$ измерения переменной x_1 , используемого в алгоритме адаптации, достаточно для установления факта неустойчивости объекта управления. Так как переменная x_1 входит в алгоритм адаптации в четной (второй) степени, то любое ненулевое значение x_1 будет приводить к росту настраиваемого параметра θ , который, в соответствии с утверждением 1, будет расти до тех пор, пока характеристический полином замкнутой системы не станет гурвицевым. Объект управления является линейным, поэтому никаких требований на скорость настройки параметров не накладывается, и стабилизация замкнутого объекта управления может быть обеспечена при любых $\gamma > 0$. Утверждение доказано.

Таким образом, предложен достаточно простой регулятор (7) (или (9)) всего с одним настраиваемым параметром θ , генерируемым алгоритмом адаптации (8). Другими словами, показано, что для адаптивной стабилизации n -мерного динамического объекта, представленного в канонической нижней треугольной форме, можно использовать динамический регулятор первого порядка (в соответствии с порядком алгоритма адаптации (8)), содержащий в своей структуре n операций умножения (для «свернутой» формы (9)).

Робастная модификация

Рассмотрим объект управления, подверженный влиянию внешнего возмущения:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{e}_n \mathbf{u} + \delta, \quad (10)$$

где $\delta = \delta(t)$ – заранее неизвестное ограниченное возмущение, а матрица $\mathbf{A}$ имеет вид (2). Поставим задачу обеспечения ограниченности всех траекторий объекта (1) и выполнения целевого условия

$$|x(t)| \leq \Delta \quad \text{для всех } t > T_0 > 0, \quad (11)$$

где Δ – произвольное наперед заданное положительное число.

Используя идеи робастного адаптивного управления [3], в случае возмущенного объекта (10) для достижения целевого условия (11) можно использовать регулятор (7) с модифицированным алгоритмом адаптации с параметрической обратной связью:

$$\dot{\theta} = -\sigma\theta + \gamma x_1^2, \quad (12)$$

где σ – произвольное (малое) положительное число (коэффициент параметрической обратной связи).

Можно также показать, что для достижения целевого условия (11) можно использовать статический регулятор (7), параметры которого рассчитываются по формуле

$$\theta = \gamma x_1^2.$$

Результаты моделирования

Рассмотрим неустойчивый объект управления

$$\dot{x}_1 = 0, 1x_1 + x_2,$$

$$\dot{x}_2 = x_1 - 2x_2 + x_3,$$

$$\dot{x}_3 = 0, 2x_1 + 0, 1x_2 - x_3 + u,$$

где числовые значения параметров при координатах вектора состояния считаются неизвестными. Для адаптивной стабилизации объекта используем адаптивный регулятор вида

$$\begin{aligned} u &= -\theta^3 x_1 - 3\theta^2 x_2 - 3\theta x_3, \\ \dot{\theta} &= \gamma x_1^2. \end{aligned} \quad (13)$$

Результаты моделирования процессов стабилизации при $\gamma = 1$ и $x(0) = [1, 0, 0]^T$ приведены на рис. 1 и демонстрируют выполнение целевого условия (3).

Пусть теперь на вход объекта в соответствии с формулой (10) вместе с управлением поступает неизмеряемое возмущение $\delta(t) = 3 \sin 2t$. Тогда для выполнения целевого условия (11) предлагается регулятор (13) с робастным алгоритмом адаптации (12). Результаты моделирования переходных процессов в замкнутой системе, подверженной влиянию внешнего возмущения при $\gamma = 1$, $\sigma = 0,1$ и $x(0) = [1, 0, 0]^T$, приведены на рис. 2 и демонстрируют небольшую установившуюся ошибку стабилизации, которая может быть уменьшена за счет увеличения коэффициента γ .

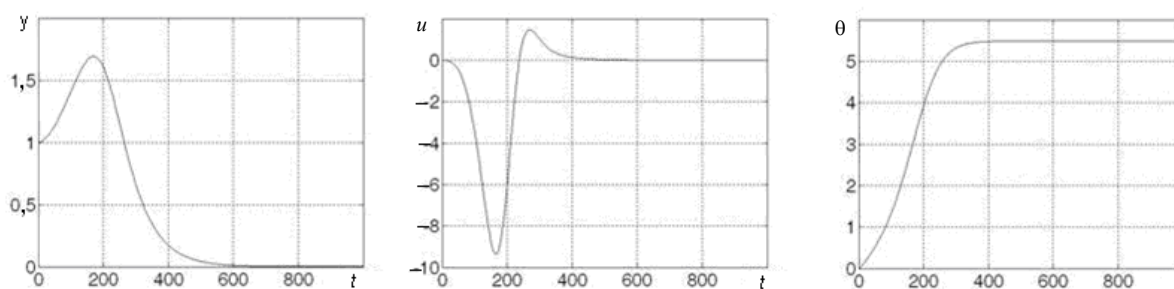


Рис. 1. Процессы в адаптивной системе стабилизации выходной переменной объекта (11) с регулятором (12) и алгоритмом адаптации (13) при $\gamma = 1$ и $x(0) = [1, 0, 0]^T$

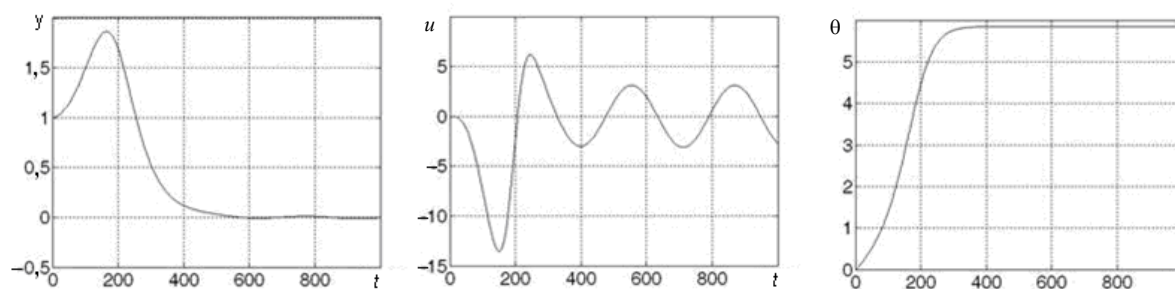


Рис. 2. Процессы в адаптивной системе стабилизации выходной переменной объекта (11) с регулятором (12) и алгоритмом адаптации (13) при действии возмущения $\delta(t) = 3 \sin 2t$ и $\gamma = 1$ и $x(0) = [1, 0, 0]^T$

Заключение

В работе показано, что для адаптивной стабилизации n -мерного динамического объекта, представленного в канонической нижней треугольной форме, можно использовать адаптивный динамический регулятор первого порядка, содержащий в своей структуре n операций умножения. Насколько известно авторам, это наиболее простой (по критерию динамического порядка и числа используемых нелинейных операций) адаптивный регулятор стабилизации. Дальнейшее развитие предложенного подхода должно состоять в его расширении на задачи адаптивного слежения, а также управления динамическими объектами с неизмеряемым вектором состояния.

Литература

1. Narendra K.S., Annaswamy A.M. Stable Adaptive Systems. – Englewood Cliffs. N.J.: Prentice-Hall, 1989. – 495p.
2. Никифоров В.О., Фрадков А.Л. Системы адаптивного управления с расширенной ошибкой // Автоматика и телемеханика. – 1994. – № 9. – С. 3–22.
3. Мирошник И.В., Никифоров В.О., Фрадков А.Л. Нелинейное и адаптивное управление сложными динамическими системами. Серия «Анализ и синтез нелинейных систем» / Под ред. Г.А. Леонова и А.Л. Фрадкова. – СПб: Наука, 2000. – 549 с.
4. Morse A.S., Isidor A., Tarn T.J. (Eds.) High-order parameter tuners for adaptive control of nonlinear systems // Systems, Models and Feedback: Theory and Applications. – Birkhauser, 1992. – P. 339–264.

5. Nikiforov V.O. Robust high-order tuner of simplified structure // Automatica. – 1999. – V. 35. – № 8. – P. 1409–1415.
6. Никифоров В.О. Адаптивное и робастное управление с компенсацией возмущений. – СПб: Наука, 2003. – 282 с.
7. Krstic M., Kanellakopoulos I., Kokotovic P.V. Nonlinear and Adaptive control design. – N.Y.: John Wiley & Sons, 1995. – 563 p.
8. Никифоров В.О. Робастное управление линейным объектом по выходу // Автоматика и телемеханика. – 1998. – № 9. – С. 87–99.
9. Ioannou P.A., Kokotovic P.V. Instability analysis and improvement of robustness of adaptive control // Automatica. – 1984. – № 5. – P. 583–594.
10. Никифоров В.О. Адаптивная стабилизация линейного объекта, подверженного внешним детерминированным возмущениям // Известия РАН. Теория и системы управления. – 1997. – № 2. – С. 103–106.
11. Никифоров В.О. Адаптивная компенсация внешних детерминированных возмущений // Мехатроника, автоматизация и управление. – 2003. – № 5. – С. 8–12.

Никифоров Владимир Олегович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, проректор, nikiforov@mail.ifmo.ru

Герасимов Дмитрий Николаевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, gerasimovdn@mail.ru

УДК 681.5.01

СТРУКТУРА СОБСТВЕННЫХ ВЕКТОРОВ МАТРИЦ СОСТОЯНИЯ МНОГОКАНАЛЬНЫХ СИСТЕМ КАК ВЫРОЖДАЮЩИЙ ФАКТОР¹

Н.А. Дударенко, А.В. Ушаков

Рассматривается задача системного вырождения динамических многоканальных систем, порождаемого структурой собственных векторов матриц состояния. Задача решается применительно к ситуациям контроля управляемости, его идентифицируемости и оценке качества по норме вектора состояния.

Ключевые слова: многоканальные системы, управляемость, вырождение, идентифицируемость.

Введение. Постановка задачи

Парадигма физического вырождения в ее системной интерпретации предполагает существование матрицы, которая характеризуется близостью к математическому вырождению. Если физическое вырождение некоторой системы или процесса хорошо коррелирует с математическим вырождением некоторой матрицы, то правомерно назвать эту матрицу критериальной. Термин «критериальная матрица» авторы позаимствовали из работ [1–3]. Известно [4–7], что развитие инструментария контроля вырождения критериальных матриц идет по пути «контроль изменения ранга этой матрицы – контроль изменения числа обусловленности матрицы – контроль изменения спектра функционалов вырождения этой матрицы». Каждый из перечисленных инструментов решает свою задачу, характеризуется различным уровнем функциональной связи используемого критерия вырождения с параметрами критериальной матрицы, вычислительным удобством и степенью полноты описания такого явления как вырождение. Авторы в своих исследованиях [6–11] затронули проблемы, связанные с аппаратом функционалов вырождения, которые представляют собой отношение i -го сингулярного числа критериальной матрицы системы к ее максимальному сингулярному числу. Возможности данного аппарата описаны в работах авторов [6, 7].

Основные усилия современной теории управления в задачах конструирования объектов управления и их идентификации в основном тратятся пока на формирование структуры собственных значений матриц состояния, в то время как потеря контроля над структурой собственных векторов может привести к неожиданным системным эффектам, одним из которых является вырождение. Важно повысить внимание разработчиков и исследователей к структуре собственных векторов с тем, чтобы их конкретная реализация не приводила к вырожденным системным ситуациям. В работе приводятся решения как в классе дискретных, так и непрерывных систем.

Структура собственных векторов матрицы состояния объекта и проблема управляемости

Проблему управляемости применительно к структуре собственных векторов матрицы состояния объекта будем решать для случая дискретного объекта, имеющего векторно-матричное описание процессов по вектору состояния в форме

¹ Работа выполнена при финансовой поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (соглашение №14.В37.21.0406)

$$\mathbf{x}(k+1) = \bar{\mathbf{A}}\mathbf{x}(k) + \bar{\mathbf{B}}\mathbf{u}(k); \mathbf{x}(0), \quad (1.1)$$

где $\mathbf{x}$, $\mathbf{u}$ – вектора состояния и управления соответственно, $\mathbf{x} \in R^n$, $\mathbf{u} \in R^r$; $\bar{\mathbf{A}}$, $\bar{\mathbf{B}}$ – матрицы состояния и управления соответственно, такие, что $\bar{\mathbf{A}} \in R^{n \times n}$, $\bar{\mathbf{B}} \in R^{n \times r}$; k – дискретное время, выраженное в числе интервалов дискретности длительности Δt , так что описание (1.1) оказывается справедливым для точек $t = k\Delta t$ непрерывного времени t .

Сделаем следующие системные предположения:

1. $\text{rang} \bar{\mathbf{A}} = n$,
2. $r = \arg \{1 \leq r \leq n\}$, $\text{rang} \bar{\mathbf{B}} = \arg \{1 \leq \text{rang} \bar{\mathbf{B}} \leq r\}$.

Введем следующее определение.

Определение 1.1. Матрицей управляемости индекса ν объекта управления (1.1) будем называть матрицу, удовлетворяющую следующим условиям:

$$\bar{\mathbf{W}}_{\nu} = [\bar{\mathbf{B}} \quad \bar{\mathbf{A}}\bar{\mathbf{B}} \quad \dots \quad \bar{\mathbf{A}}^{\nu-1}\bar{\mathbf{B}}], \dim(\bar{\mathbf{W}}_{\nu}) = n \times \nu n.$$

Примечание 1.1. Очевидно, что:

1. $\nu = n$ в случае, когда $r = 1$;
2. $\nu = 1$ в случае, когда $r = n$.

Таким образом, ν удовлетворяет неравенству $1 \leq \nu \leq n$.

Определение 1.2. Объект (1.1) с парой матриц $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ обладает управляемостью индекса ν , если выполняется условие

$$\text{rang} \bar{\mathbf{W}}_{\nu} = n = \dim(\mathbf{x}). \quad (1.3)$$

Примечание 1.2. Если условие (1.3) выполняется при $\nu = n$, то матрица управляемости с индексом n не сопровождается этим индексом, и выполняется соотношение

$$\bar{\mathbf{W}}_{\nu|\nu=n} = \bar{\mathbf{W}}_y = [\bar{\mathbf{B}} \quad \bar{\mathbf{A}}\bar{\mathbf{B}} \quad \dots \quad \bar{\mathbf{A}}^{n-1}\bar{\mathbf{B}}]. \quad (1.4)$$

Рассмотрим структуру собственных векторов матрицы $\bar{\mathbf{A}}$ на предмет принадлежности столбцов матрицы $\bar{\mathbf{B}}$ линейным оболочкам, натянутым на собственные векторы матрицы $\bar{\mathbf{A}}$. В связи с этим сформулируем следующие утверждения.

Утверждение 1.1. Пусть $r = 1$, а матрица $\bar{\mathbf{B}}$ есть собственный вектор матрицы $\bar{\mathbf{A}}$, тогда становятся справедливыми следующие положения:

1. $\text{rang} \bar{\mathbf{W}}_y = 1$.
2. Пара матриц $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ оказывается не полностью управляемой, при этом подпространство управляемости совпадает с линейной оболочкой, натянутой на матрицу $\bar{\mathbf{B}}$.

Доказательство. В соответствии с определением собственного вектора матрицы, если $\bar{\mathbf{B}}$ есть собственный вектор матрицы $\bar{\mathbf{A}}$, то существует такое $\bar{\lambda}$: $\det(\bar{\lambda}\mathbf{I} - \bar{\mathbf{A}}) = 0$, что

$$\bar{\mathbf{A}}\bar{\mathbf{B}} = \bar{\lambda}\bar{\mathbf{B}}. \quad (1.5)$$

Соотношение (1.5) порождает следующую систему равенств:

$$\begin{aligned} \bar{\mathbf{A}}^2\bar{\mathbf{B}} &= \bar{\mathbf{A}}\bar{\mathbf{A}}\bar{\mathbf{B}} = \bar{\mathbf{A}}\bar{\lambda}\bar{\mathbf{B}} = \bar{\lambda}^2\bar{\mathbf{A}}\bar{\mathbf{B}} = \bar{\lambda}^2\bar{\lambda}\bar{\mathbf{B}} \\ \bar{\mathbf{A}}^3\bar{\mathbf{B}} &= \bar{\mathbf{A}}\bar{\mathbf{A}}^2\bar{\mathbf{B}} = \bar{\mathbf{A}}\bar{\lambda}^2\bar{\mathbf{B}} = \bar{\lambda}^2\bar{\mathbf{A}}\bar{\mathbf{B}} = \bar{\lambda}^3\bar{\mathbf{B}} \\ &\vdots \\ \bar{\mathbf{A}}^{n-1}\bar{\mathbf{B}} &= \bar{\mathbf{A}}\bar{\lambda}^{n-2}\bar{\mathbf{B}} = \bar{\lambda}^{n-1}\bar{\mathbf{B}}. \end{aligned} \quad (1.6)$$

Подстановка (1.6) в выражение (1.4) дает

$$\bar{\mathbf{W}}_y = [\bar{\mathbf{B}} \quad \bar{\lambda}\bar{\mathbf{B}} \quad \dots \quad \bar{\lambda}^{n-1}\bar{\mathbf{B}}]. \quad (1.7)$$

Таким образом, $\text{rang} \bar{\mathbf{W}}_y = \text{rang} \bar{\mathbf{B}} = 1$, и подпространство управляемости совпало с линейной оболочкой, натянутой на собственный вектор.

Примечание 1.3. $n \times n$ – матрица управляемости (1.7), обладая рангом $\text{rang} \bar{\mathbf{W}}_y = 1$, оказалась вырожденной, она характеризуется бесконечным значением числа обусловленности и только одним ненулевым функционалом вырождения.

Утверждение 1.2. Пусть $\text{rang} \bar{\mathbf{B}} = r < n$, при этом столбцы матрицы $\bar{\mathbf{B}}_j, j = \overline{1, r}$ есть собственные вектора матрицы $\bar{\mathbf{A}}$, тогда становятся справедливыми следующие положения:

1. $\text{rang} \bar{\mathbf{W}}_y = r < n$.
2. Пара матриц $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ оказывается не полностью управляемой, при этом подпространство управляемости совпадает с линейной оболочкой, натянутой на столбцы матрицы $\bar{\mathbf{B}}$.

Доказательство. Доказательство сформулированного утверждения опирается на доказательство утверждения 1.1, в соответствии с которым и положениями рассматриваемого случая матрица управляемости (1.4) по аналогии с (1.7) принимает вид

$$\bar{W}_y = [\bar{B} \quad \bar{A}\bar{B} \quad \dots \quad \bar{A}^{n-1}\bar{B}] = \text{row}\{\bar{\lambda}_1^i \bar{B}_1 \quad \bar{\lambda}_2^i \bar{B}_2 \quad \dots \quad \bar{\lambda}_j^i \bar{B}_j \quad \dots \quad \bar{\lambda}_r^i \bar{B}_r\}, i = \overline{0, n-1} \quad (1.8)$$

Примечание 1.4. $n \times n$ – матрица управляемости (1.8), обладая рангом $\text{rang} \bar{W}_y = r < n$, оказалась вырожденной, она характеризуется бесконечным значением числа обусловленности и r ненулевыми функционалами вырождения.

Ситуация коренным образом меняется, когда ранг матрицы управления $\bar{B}$ удовлетворяет условию $\text{rang}(\bar{B}) = r = n$, при этом все столбцы его являются собственными векторами матрицы $\bar{A}$. В этом случае становится справедливым утверждение 1.3, которое приводится без доказательства.

Утверждение 1.3. Пусть $\text{rang} \bar{B} = n$, при этом столбцы матрицы $\bar{B}_j, j = \overline{1, n}$ суть собственные вектора матрицы $\bar{A}$, тогда становятся справедливыми следующие положения:

$$1. \bar{W}_y = [\bar{B}]. \quad (1.9)$$

$$2. \text{rang} \bar{W}_y = n.$$

3. Пара матриц $(\bar{A}, \bar{B})$ оказывается полностью управляемой, при этом подпространство управляемости, совпадая с линейной оболочкой, натянутой на столбцы матрицы $\bar{B}$, совпадает с пространством состояния объекта.

Примечание 1.5. $n \times n$ – матрица управляемости (1.9), обладая рангом $\text{rang} \bar{W}_y = n$, оказалась невырожденной, она характеризуется n ненулевыми функционалами вырождения.

Нетрудно видеть, что ключевым моментом в исследовании влияния фактора структуры собственных векторов матрицы состояния объекта $\bar{A}$ на его управляемость являются положения утверждения 1.1. Проиллюстрируем это на примере.

Пример 1.1. Покажем, что дискретный объект (1.1) с матрицами $\bar{A} = \begin{bmatrix} 0 & 1 \\ -0,16 & 1 \end{bmatrix}; \bar{B} = \begin{bmatrix} 3 \\ 0,6 \end{bmatrix};$
 $x(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ может быть переведен в состояние $x(k) = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, а $x(k) = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ – нет.

Нетрудно видеть, что матрица $\bar{B}$ есть собственный вектор матрицы $\bar{A}$, соответствующий собственному значению $\lambda_1 = 0,2$. Действительно,

$$\bar{A}\bar{B} = \begin{bmatrix} 0 & 1 \\ -0,16 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 0,6 \end{bmatrix} = \begin{bmatrix} 0,6 \\ 0,12 \end{bmatrix} = 0,2 \cdot \begin{bmatrix} 3 \\ 0,6 \end{bmatrix}.$$

Как следствие, матрица управляемости $\bar{W}_y = [\bar{B} \quad \bar{A}\bar{B}] = \begin{bmatrix} 3 & 0,6 = 3 \cdot 0,2 \\ 0,6 & 0,12 = 0,6 \cdot 0,2 \end{bmatrix}$. Таким образом, подпространство управляемости есть линейная оболочка $L\left\{\begin{bmatrix} 3 \\ 0,6 \end{bmatrix}\right\}$, натянутая на собственный вектор

$\bar{\xi}_1 = \bar{B}$. Нетрудно видеть, что вектор $x(k) = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ принадлежит линейной оболочке $L\left\{\begin{bmatrix} 3 \\ 0,6 \end{bmatrix}\right\}$. Действительно, вектор $x(k) = \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0,6 \end{bmatrix} \cdot \frac{5}{3}$.

Воспользуемся теперь рекуррентной процедурой (1.1) для вычисления управления $u(k)$, переводящего объект из состояния $x(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ в состояние $x(k) = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$. В силу рекуррентной процедуры оказывается справедливым соотношение

$$x(1) = \bar{A}x(0) + \bar{B}u(0) = \bar{B}u(0) = \begin{bmatrix} 3 \\ 0,6 \end{bmatrix} u(0).$$

Нетрудно видеть, что при любом $u(0)$ вектор $x(1)$ будет принадлежать векторной оболочке $L\left\{\begin{bmatrix} 3 \\ 0,6 \end{bmatrix}\right\}$, которой принадлежит и вектор $x(k) = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, что позволяет приравнять его к вектору $x(1)$.

Соотношение $\begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0,6 \end{bmatrix} u(0)$ выполняется при $u(0) = \frac{5}{3}$. Таким образом, объект может быть переведен из нулевого состояния в состояние $\mathbf{x}(k) = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ за один такт управлением $u(0) = \frac{5}{3}$.

Рассмотрим теперь возможность перевода объекта управления из состояния $\mathbf{x}(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ в состояние $\mathbf{x}(k) = \begin{bmatrix} 1 \\ 5 \end{bmatrix} \notin L \left\{ \mathbf{B} = \begin{bmatrix} 3 \\ 0,6 \end{bmatrix} \right\}$. Для вычисления управления, которое должно осуществлять этот перевод, воспользуемся рекуррентной процедурой (1.1), тогда получим

$$\mathbf{x}(1) = \bar{\mathbf{A}}\mathbf{x}(0) + \bar{\mathbf{B}}u(0) = \bar{\mathbf{B}}u(0) = \begin{bmatrix} 3 \\ 0,6 \end{bmatrix} u(0).$$

Нетрудно видеть, что не существует такое $u(0)$, которое бы удовлетворяло условию $u(0) = \arg \left\{ \begin{bmatrix} 3 \\ 0,6 \end{bmatrix} u(0) = \begin{bmatrix} 1 \\ 5 \end{bmatrix} \right\}$. За один такт перевод из состояния $\mathbf{x}(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ в состояние $\mathbf{x}(k) = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ невозможен.

Перейдем в рекуррентной процедуре (1.1) ко второму такту, тогда получим

$$\mathbf{x}(2) = \bar{\mathbf{A}}\mathbf{x}(1) + \bar{\mathbf{B}}u(1) = \bar{\mathbf{A}}\bar{\mathbf{B}}u(0) + \bar{\mathbf{B}}u(1) = \begin{bmatrix} \bar{\mathbf{A}}\bar{\mathbf{B}} & \bar{\mathbf{B}} \end{bmatrix} \cdot \begin{bmatrix} u(0) \\ u(1) \end{bmatrix} = \begin{bmatrix} 0,6 & 3 \\ 0,12 & 0,6 \end{bmatrix} \cdot \begin{bmatrix} u(0) \\ u(1) \end{bmatrix}.$$

Из полученного выражения следует, что не существует такое $\begin{bmatrix} u(0) \\ u(1) \end{bmatrix}$, которое бы удовлетворяло условию $\begin{bmatrix} u(0) \\ u(1) \end{bmatrix} = \arg \left\{ \begin{bmatrix} 0,6 & 3 \\ 0,12 & 0,6 \end{bmatrix} \cdot \begin{bmatrix} u(0) \\ u(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix} \right\}$ в силу необратимости матрицы управляемости $\begin{bmatrix} 0,6 & 3 \\ 0,12 & 0,6 \end{bmatrix}$, которая является матрицей управляемости исходного объекта.

Объект оказался вырожденным с подпространством управляемости в виде линейной оболочки $L \left\{ \mathbf{B} = \begin{bmatrix} 3 \\ 0,6 \end{bmatrix} \right\}$.

Примечание 1.6. В силу того, что матричные компоненты непрерывного и дискретного объектов связаны преобразованием вида «матричная функция от матрицы», относительно которого структура собственных векторов матрицы состояния является инвариантной, полученные результаты в полной мере переносимы на случай непрерывных объектов, так что для непрерывного объекта с парой матриц $(\mathbf{A}, \mathbf{B})$ матрица управляемости принимает вид (1.4) с точностью до замены матриц $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ на $(\mathbf{A}, \mathbf{B})$.

Структура собственных векторов матрицы состояния системы и проблема качества сходящихся процессов

Проблему, вынесенную в заголовок раздела, будем решать для случая многоканальной непрерывной системы. Для исследования влияния структуры собственных векторов матрицы состояния системы на качество сходящихся процессов свободного движения рассмотрим замкнутую многоканальную непрерывную систему

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t); \mathbf{x}(0), \quad (2.1)$$

где $\mathbf{F} \in R^{n \times n}$ – гурвицева матрица системы. Свободная составляющая системы (2.1) имеет вид

$$\mathbf{x}(t) = e^{\mathbf{F}t} \mathbf{x}(0). \quad (2.2)$$

Переход в (2.2) к согласованным матричным и векторным нормам позволяет записать:

$$\|\mathbf{x}(t)\| = \|e^{\mathbf{F}t} \mathbf{x}(0)\| \leq \|e^{\mathbf{F}t}\| \|\mathbf{x}(0)\|. \quad (2.3)$$

Для дальнейших исследований воспользуемся условием преобразования подобия матриц [11] и представим матрицу $\mathbf{F}$ посредством диагональной $n \times n$ – матрицы Λ на основе соотношения

$$\mathbf{F} = \mathbf{M}\Lambda\mathbf{M}^{-1}. \quad (2.4)$$

С учетом соотношения (2.4) и свойства матричной функции от матрицы [11] компонент $e^{\mathbf{F}t}$ неравенства (2.3) может быть представлен в виде

$$\|e^{Ft}\| = \|Me^{At}M^{-1}\| \leq \|M\| \|e^{At}\| \|M^{-1}\| \leq \|M\| \|M^{-1}\| \|e^{At}\|. \quad (2.5)$$

Согласно определению [5], число обусловленности $C\{*\}$ матрицы $(*)$ задается соотношением

$$C\{*\} = \frac{\Lambda}{\|(*)\| \|(*)^{-1}\|}.$$

Тогда соотношение (2.5) может быть представлено в форме

$$\|e^{Ft}\| \leq C\{M\} \|e^{At}\| = C\{M\} \|e^{At} = \text{diag}\{e^{\lambda_i t}, i = \overline{1, n}\}\| = C\{M\} e^{\lambda_M t}, \quad (2.6)$$

где λ_M – собственное значение матрицы F с отрицательной вещественной частью, максимально приближенное к началу координат комплексной плоскости. Подстановка (2.6) в (2.3) позволяет записать:

$$\|x(t)\| = \|e^{Ft}x(0)\| \leq C\{M\} e^{\lambda_M t} \|x(0)\|. \quad (2.7)$$

Нетрудно видеть, что соотношение (2.7) представляет собой экспоненциальное покрытие свободного движения системы по вектору состояния, представленного его нормой. Таким образом, оценка качества сходящихся процессов в системе при фиксированной структуре собственных значений может быть получена с помощью числа $C\{M\}$ обусловленности матрицы M приведения подобия или ее глобального функционала $J_G = J_G\{M\}$ вырождения [6, 7], связанных соотношением $J_G\{M\} = C^{-1}\{M\}$, при этом следует иметь в виду, что число обусловленности произвольной матрицы $(*)$ в силу его свойств удовлетворяет неравенству

$$1 \leq C\{M\} \leq \infty. \quad (2.8)$$

Известно [11], что матрица M преобразования подобия (2.4) имеет своими столбцами собственные вектора диагонализируемой матрицы F . Следовательно, качество процессов в системе напрямую связано со структурой собственных векторов матрицы состояния этой системы: в силу (2.8), если собственные вектора близки к коллинеарности, то в аperiodической системе, т.е. системе с вещественными некрратными собственными значениями ее матрицы состояния, возможны выбросы движения по норме ее вектора состояния, достигающей сколь угодно больших значений.

Пример 2.1. Рассмотрим две системы вида (2.1) с подобными матрицами состояния F и $\tilde{F}$, имеющие одинаковый спектр собственных значений: $\sigma\{F\} = \{\lambda = \arg\{\det(\lambda I - F) = 0\} : \lambda_1 = -2, \lambda_2 = -9\}$ и $\sigma\{\tilde{F}\} = \{\tilde{\lambda} = \arg\{\det(\tilde{\lambda} I - \tilde{F}) = 0\} : \tilde{\lambda}_1 = -2, \tilde{\lambda}_2 = -9\}$. Матрицы F и $\tilde{F}$, характеризуются соответственно спектрами собственных векторов

$$\left\{ \xi_i = \arg\{F\xi_i = \lambda_i \xi_i : \xi_1 = [0, 707 \ 0, 707]^T, \xi_2 = [-0, 707 \ 0, 707]^T \} \right\},$$

$$\left\{ \tilde{\xi}_i = \arg\{\tilde{F}\tilde{\xi}_i = \tilde{\lambda}_i \tilde{\xi}_i : \tilde{\xi}_1 = [1 \ 0]^T, \tilde{\xi}_2 = [-1 \ 0, 05]^T \} \right\}.$$

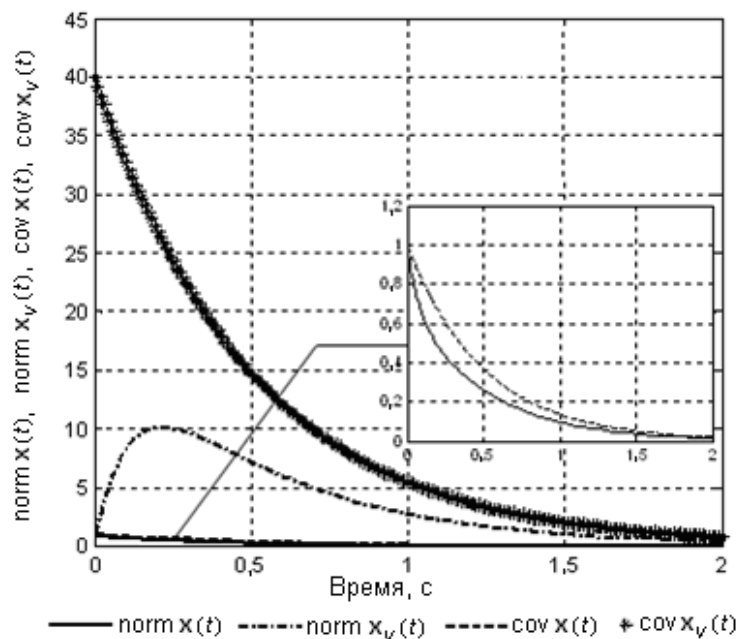


Рисунок. Кривые процессов по нормам $\|x(t)\|$ и $\|\tilde{x}(t)\|$ при $x(0) = \tilde{x}(0) = [0 \ 1]^T$ и их экспоненциальные покрытия

Нетрудно видеть, что матрицы диагонализации $\mathbf{M}$ и $\tilde{\mathbf{M}}$, приводящие $\mathbf{F}$ и $\tilde{\mathbf{F}}$ к диагональному виду $\Lambda = \text{diag}\{-2, -9\}$, сконструированные на спектре собственных векторов, принимают вид $\mathbf{M} = \begin{bmatrix} 0,707 & -0,707 \\ 0,707 & 0,707 \end{bmatrix}$; $\tilde{\mathbf{M}} = \begin{bmatrix} 1 & -1 \\ 0 & 0,05 \end{bmatrix}$. Также вычислим матрицы $\mathbf{F}$ и $\tilde{\mathbf{F}}$ согласно (2.4): $\mathbf{F} = \begin{bmatrix} -5,5 & 3,5 \\ 3,5 & -5,5 \end{bmatrix}$; $\tilde{\mathbf{F}} = \begin{bmatrix} -2 & 140 \\ 0 & -9 \end{bmatrix}$. Матрицы $\mathbf{M}$ и $\tilde{\mathbf{M}}$ характеризуются соответственно числами обусловленности $C\{\mathbf{M}\} = 1$; $C\{\tilde{\mathbf{M}}\} = 40,025$. В силу (2.7) экспоненциальные покрытия процессов по норме вектора состояния в системе с матрицей $\mathbf{F}$ и матрицей $\tilde{\mathbf{F}}$ будут существенно различаться. Проиллюстрируем это прямым вычислением норм $\|\mathbf{x}(t)\|$ и $\|\tilde{\mathbf{x}}(t)\|$ в силу соотношений $\|\mathbf{x}(t)\| = \|e^{\mathbf{F}t} \mathbf{x}(0)\|$ и $\|\tilde{\mathbf{x}}(t)\| = \|e^{\tilde{\mathbf{F}}t} \tilde{\mathbf{x}}(0)\|$ соответственно, положив в них $\mathbf{x}(0) = \tilde{\mathbf{x}}(0) = [0 \ 1]^T$, и их экспоненциальных покрытий вида (2.7). Графики полученных процессов (рисунок) наглядно иллюстрируют тот факт, что отсутствие контроля за назначением структуры собственных векторов может приблизить систему к вырождению.

Структура собственных векторов матрицы состояния объекта и проблема идентифицируемости

Проблему идентифицируемости объекта управления будем решать, пользуясь определением идентифицируемого объекта [12].

Определение 3.1. Объект

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t); \mathbf{x}(0) \quad (3.1)$$

называется идентифицируемым, если по измерениям координат состояния объекта можно определить матрицу системы $\mathbf{A}$.

Свяжем задачу идентификации объекта (3.1) с удачностью назначения ненулевого начального состояния $\mathbf{x}(0)$ с учетом структуры собственных векторов матрицы $\mathbf{A}$. Прежде чем решать данную задачу, воспользуемся определением П. Эйкоффа [12] критерия идентифицируемости объекта (3.1), опирающегося на следующее определение матрицы идентифицируемости объекта.

Определение 3.2. Матрицей идентифицируемости $\mathbf{W}_1$ объекта, построенной на паре $\{\mathbf{A}, \mathbf{x}(0)\}$, называется матрица, сформированная в соответствии с правилом

$$\mathbf{W}_1 = [\mathbf{x}(0) \ \mathbf{A}\mathbf{x}(0) \ \dots \ \mathbf{A}^{n-1}\mathbf{x}(0)].$$

Тогда можно сформулировать критерий идентифицируемости [12] с использованием матрицы идентифицируемости.

Определение 3.3. Критерием идентифицируемости объекта (3.1), состоящей в определении матрицы $\mathbf{A}$, по измерениям его координат, является выполнение условия

$$\text{rang} \mathbf{W}_1 = \text{rang} \left\{ [\mathbf{x}(0) \ \mathbf{A}\mathbf{x}(0) \ \dots \ \mathbf{A}^{n-1}\mathbf{x}(0)] \right\} = n = \dim \mathbf{x}. \quad (3.2)$$

Нетрудно видеть, что выражение (3.2) совпадает с выражением (1.4) для матрицы управляемости линейного объекта с точностью до замены матрицы $\bar{\mathbf{B}}$ на $\mathbf{x}(0)$, матрицы $\bar{\mathbf{A}}$ – на $\mathbf{A}$. Очевидно, что все требования к матрице $\bar{\mathbf{B}}$, сформулированные в предыдущем разделе, как функции геометрического спектра собственных векторов матрицы $\bar{\mathbf{A}}$ переносятся на вектор $\mathbf{x}(0)$, чтобы гарантировать невырожденность матрицы идентифицируемости. Таким образом, задача идентификации получает корректное решение в случае, если вектор $\mathbf{x}(0)$ имеет представление $\mathbf{x}(0) = \sum_{i=1}^n \alpha_i \xi_i$, где все $\alpha_i \neq 0$, ξ_i – собственные векторы матрицы $\mathbf{A}$. При таком начальном состоянии возбуждаются все моды ($e^{\lambda_i t}$, $\lambda_i = \arg(\det(\lambda \mathbf{I} - \mathbf{A})) = 0, i = \overline{1, n}$) объекта, и задача идентификации получает корректное решение. При наличии хотя бы одного $\alpha_i = 0$ условие (3.2) не выполняется, матрица идентифицируемости оказывается вырожденной, при таком начальном состоянии возбуждаются не все моды объекта, а задача идентификации не получает полного решения.

Заключение

Проведенные авторами исследования показывают, что структура собственных векторов должна быть подробно исследована разработчиками системы при ее проектировании. Анализ структуры собственных векторов позволяет уже на этапе проектирования системы выявить такие сочетания ее параметров, которые могут привести к вырожденным системным ситуациям.

Литература

1. Sheng-Guo Wang. Robust Schur Stability and Eigenvectors of Uncertain Matrices // Proceedings of the American Control Conference. – 1997. – V. 5. – P. 3449–3454.
2. Baarda W. S-transformations and criterion matrices // Netherlands geodetic commission, 1981. – V. 5. – № 1. – 168 p.
3. Годованный П.А. Моделирование процессов нарушения проводимой политики безопасности в РВС // Сборник научных трудов НГТУ. – 2003. – № 4 (34). – С. 13–18.
4. Голуб Дж., Ван Лоун Ч. Матричные вычисления: Пер. с англ. – М.: Мир, 1999. – 548 с.
5. Wilkinson J.H. The algebraic eigenvalue problem. – Oxford: Clarendon Press, 1965. – 570 p.
6. Дударенко Н., Ушаков А. Анализ многомерных динамических систем: технология контроля вырождения. – Saarbrücken: LAP LAMBERT Academic Publishing, 2011. – 232 с.
7. Дударенко Н.А., Ушаков А.В. Вырождение сложных дискретных динамических систем: проблема контроля с помощью частотных чисел обусловленности // Научно-технический вестник СПбГУ ИТМО. – 2004. – № 14. – С. 62–66.
8. Дударенко Н.А., Ушаков А.В., Полякова М.В. Алгебраическая организация условий обобщенной синхронизируемости многоагрегатных динамических объектов // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 2 (66). – С. 30–36.
9. Дударенко Н.А., Ушаков А.В., Полякова М.В. Формирование интервальных векторно-матричных модельных представлений антропокомпонентов-операторов в составе сложных динамических систем // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 6 (70). – С. 32–36.
10. Дударенко Н.А., Бирюков Д.С., Ушаков А.В., Полякова М.В., Акунов Т.А. Формирование спектра сингулярных чисел квадратной матрицы простой структуры // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 6 (76). – С. 53–58.
11. Дударенко Н.А., Слита О.В., Ушаков А.В. Математические основы современной теории управления: аппарат метода пространства состояний: Учебное пособие / Под ред. А.В. Ушакова. – СПб: СПбГУ ИТМО, 2009. – 325 с.
12. Эйкофф П. Основы идентификации систем управления. – М.: Мир, 1975. – 683 с.

Дударенко Наталия Александровна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, dudarenko@yandex.ru

Ушаков Анатолий Владимирович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, Ushakov-AVG@yandex.ru

УДК 681.5

ОПТИМАЛЬНЫЙ СИНТЕЗ ДЛЯ ДИНАМИЧЕСКИХ СИСТЕМ С ЗАПАЗДЫВАНИЕМ ПО УПРАВЛЕНИЮ

Д.А. Музыка, Р.О. Пешеров, В.Ю. Тертычный-Даури

Рассмотрена задача формирования закона оптимального управления для нелинейных динамических систем с запаздыванием по времени в канале управления. В соответствии с принципом оптимальности обосновывается необходимое условие оптимальности (уравнение Беллмана) для систем с запаздыванием в канале по управлению. Выводы анализа подкрепляются результатами численного моделирования в задаче оптимальной стабилизации вращения твердого тела.

Ключевые слова: запаздывание в канале управления, оптимальное управление, принцип оптимальности, беллмановская оптимизация.

Введение

Основной поток публикаций по регулируемым динамическим системам с запаздыванием касается вопросов устойчивости и стабилизации изучаемых процессов (например, работы [1–4] и содержащаяся там библиография). Полученные результаты можно рассматривать как обобщение результатов теории дифференциальных уравнений с запаздывающим аргументом в фазовой переменной.

В некоторых работах решены задачи с запаздыванием по управлению применительно к общей (но не оптимальной) адаптивной задаче управления с возмущениями [5–7]. Значительно более скромным выглядит список работ по оптимизации управляемых динамических систем с запаздыванием по управлению [1–3]. Данные публикации в основном посвящены принципу максимума с учетом эффекта запаздывания.

В настоящей работе, по-видимому, впервые ставится и решается задача синтеза оптимального управления в непрерывных динамических системах с запаздыванием в канале управления с использова-

нием беллмановского оптимизационного подхода (метода динамического программирования). На рис. 1 условно изображена схема формируемой системы управления.

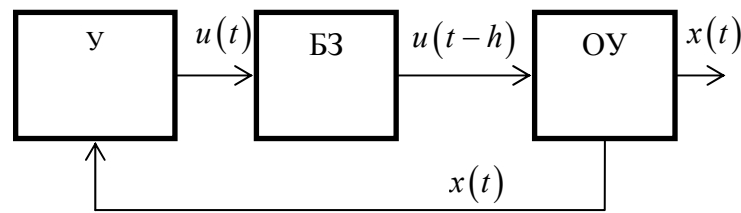


Рис. 1. Общая блок-схема системы управления с инерционным запаздыванием: ОУ – объект управления; БЗ – блок запаздывания; У – управление

Ставится основная цель – построить оптимальное управление объектом, которое бы решало задачу минимизации функционала качества в условиях запаздывания по управлению.

Постановка задачи

Пусть объект управления задается векторным уравнением

$$\dot{x} = f[x(t), u(t-h), t], \quad (1)$$

где $x(t) \in R^n$ – состояние системы в момент времени t , где $t \in [t_0, t_1]$ – заданный интервал, $h = \text{const} > 0$ – запаздывание в управлении (так называемое инерционное запаздывание); при этом предполагается, что в самом объекте (1) запаздывания нет, но оно есть в регуляторе $u(t-h) \in R^n$. Интегрируя уравнение (1), получим равносильное ему векторное интегральное уравнение Вольтерра:

$$x(t) = x_0 + \int_{t_0}^t f[x(s), u(s-h), s] ds, \quad (2)$$

где $x_0 = x(t_0)$ – заданный вектор начального состояния системы. Уравнение (2) показывает, что $x(t)$ – состояние системы в момент времени t – зависит от значений управления $u(s-h)$ в предыдущие моменты времени $s-h$, где $t_0 < s < t(t_0 \geq 0, h > 0)$.

Далее, управление $u(\theta) \in R^n$ входит в уравнения (1)–(2) в виде значения $u(\theta)$ в запаздывающий момент времени $\theta = s-h$, где $h > 0$. При малых $s > t_0$ запаздывающий момент $\theta = s-h$ может оказаться отрицательным. В связи с этим, чтобы подынтегральное выражение в уравнении (2) имело смысл, управление $u(t)$ следует задавать и при отрицательных t , а именно при $t \in [t_0-h, 0]$, когда $t_0 < h$. Таким образом, управление $u(t)$ надо задавать на более широком интервале времени $t \in [t_0-h, t_1]$, причем состояние $x(t)$ должно быть определено на более узком интервале времени $t \in [t_0, t_1]$.

Будем считать, что на управляющие силы $u \in R^n$ наложены некоторые ограничения: $u \in U \subset R^n$, где U – некоторое заданное множество допустимых управлений. Требуется выбором управления $u(t)$, $t \in [t_0-h, t_1]$ обеспечить минимум функционала качества

$$J = V(x(t_1), t_1) + \int_{t_0}^{t_1} F[x(s), u(s-h), s] ds \rightarrow \min_{u \in U} \quad (3)$$

и перевести систему (1) из начального состояния $x(t_0)$ в конечное $x(t_1)$. Полагаем, что в системе (1) с функционалом (3) вектор-функция f и скалярные функции V, F непрерывно дифференцируемы по всем своим аргументам.

Напомним, что принцип оптимальности Беллмана, лежащий в основе метода динамического программирования, применим для систем, последующее движение которых полностью определяется состоянием этих систем в любой текущий момент времени [1]. Согласно Беллману, оптимальная стратегия определяется только начальным условием и конечной целью, т.е. принцип оптимальности утверждает, что для любого первоначального состояния и стратегии (управления) в начальный момент последующие стратегии должны составлять оптимальное движение относительно состояния, полученного в результате применения начальной стратегии. Указанная формулировка принципа оптимальности останется справедливой и для систем с запаздыванием, если в понятие состояния системы в текущий момент времени t' включить и предысторию изменения фазовых координат системы на промежутке времени последствия: $t'-h < t < t'$.

Отметим также, что отличительной особенностью метода динамического программирования, использующего принцип оптимальности, является то, что отрезки оптимальной траектории определяются в обратной последовательности, начиная с заданного конечного (целевого) состояния $x(t_1)$.

Необходимое условие оптимальности

Принцип оптимальности Беллмана позволяет сформулировать необходимое условие оптимальности для динамических систем с последствием по управлению вида (1) с функционалом качества (3).

Допустим, что $x^0(t)$ – оптимальная траектория системы (1) с заданным начальным $x(t_0)$ и конечным состоянием $x(t_1)$. Требуется перевести систему (1) из векторной точки $x(t_0)$ в векторную точку $x(t_1)$ по траектории $x^0(t)$, выбрав оптимальное управление $u^0(t-h)$, минимизирующее функционал (3). Можно показать, что функционал качества (3) с запаздыванием по времени в управлении можно подходящим функциональным преобразованием свести к функционалу с управлением без запаздывания по времени, но с запаздыванием по индексу [3]. Тем самым возникает возможность использовать стандартные оптимизационные процедуры метода динамического программирования и к системам с запаздыванием по управлению.

Теорема. Пусть поставлена задача синтеза оптимального управления для системы (1) с функционалом (3) с оговоренными выше требованиями непрерывности и гладкости для всех входящих скалярных функций и вектор-функций.

Тогда, если $x^0(t)$ – оптимальная траектория системы (1) с заданными значениями $x(t_0)$ и $x(t_1)$, оптимальное управление $u^0(t-h)$ удовлетворяет уравнение Беллмана (уравнение Гамильтона–Якоби–Беллмана) вида

$$\min_{u \in U} \left\{ \frac{dS(x^0(t), t)}{dt} + F[x^0(t), u(t-h), t] \right\} = 0, \quad (4)$$

либо

$$\frac{dS(x^0(t), t)}{dt} + F[x^0(t), u^0(t-h), t] = 0, \quad (5)$$

где обозначено

$$S(x^0(t), t) = V|_{t_1} + \min_{u \in U} \int_t^{t_1} F[x^0(s), u(s-h), s] ds, \quad (6)$$

причем

$$S(x^0(t_1), t_1) = V|_{t_1} = V(x(t_1), t_1), \quad (7)$$

а для подынтегральной функции $F(\cdot)$ имеет место равенство (5), (6).

Доказательство. Обозначим через $S(x^0(t_0), t_0)$ минимум функционала J (3). Из принципа оптимальности следует, что часть траектории с концами $x^0(t)$ (в начале при $t=t$) и $x^0(t_1)$ (в конце при $t=t_1$), удовлетворяющая уравнению (1), также оптимальна. Значит, минимальное значение порождаемого этой частью траектории функционала равно $S(x^0(t), t)$ (6) с граничным значением $S(x^0(t_1), t_1) = V|_{t_1}$ (7). Приходим тем самым к так называемому функциональному уравнению Беллмана (6).

Пусть $t' = t + \Delta t$, где Δt – достаточно малый интервал времени. Тогда минимальное значение функционала по части оптимальной траектории с начальным состоянием $x^0(t') = x^0(t + \Delta t)$ и конечным состоянием $x^0(t_1)$ определяется равенством

$$S(x^0(t'), t') = V|_{t_1} + \min_{u \in U} \int_{t'}^{t_1} F[x^0(s), u(s-h), s] ds. \quad (8)$$

Разобьем интервал интегрирования на два: от t до $t' = t + \Delta t$ и от t' до t_1 . Тогда, сравнивая интегралы (6) и (8), получим, что

$$S(x^0(t), t) = V|_{t_1} + \min_{u \in U} \left(\int_t^{t+\Delta t} F[x^0(s), u(s-h), s] ds + \int_{t+\Delta t}^{t_1} F[x^0(s), u(s-h), s] ds \right), \quad (9)$$

или с точностью до малых $\alpha_1(\Delta t)$ более высокого порядка, чем Δt , можно написать (с учетом оптимальности на втором интервале):

$$S(x^0(t), t) = V|_{t_1} + \min_{u \in U} \left(F[x^0(t), u(t-h), t] \Delta t + \min_{u \in U} \int_t^{t_1} F[x^0(s), u(s-h), s] ds \right) + \alpha_1(\Delta t),$$

где с точностью до $\alpha_1(\Delta t)$ имеем в соотношении (9) для первого интеграла справа

$$\int_t^{t+\Delta t} F[x^0(s), u(s-h), s] ds = F[x^0(t), u(t-h), t] \Delta t + \alpha_1(\Delta t),$$

$$\lim_{\Delta t \rightarrow 0} \frac{\alpha_1(\Delta t)}{\Delta t} = 0.$$

Таким образом, имеем запись

$$S(x^0(t), t) = \min_{u \in U} \{ F[x^0(t), u(t-h), t] \Delta t + S(x^0(t'), t') \} + \alpha_1(\Delta t). \quad (10)$$

Пусть, ради простоты записи, $x(t) = x^0(t)$. Тогда, разлагая $x(t')$ в ряд Тейлора, получим

$$x(t') = x(t + \Delta t) = x(t) + \dot{x}(t) \Delta t + \alpha_2(\Delta t) = x(t) + f[x(t), u(t-h), t] \Delta t + \alpha_2(\Delta t), \quad (11)$$

где $\alpha_2(\Delta t)$ – остаточный член выше первого порядка малости от Δt . Подставим это разложение $x(t')$ (11) в выражение для $S(x(t'), t')$. При соответствующем разложении в ряд Тейлора, полагая при этом, что существуют частные производные $\partial S / \partial x_i$, $i = \overline{1, n}$, и $\partial S / \partial t$, получим

$$\begin{aligned} S(x(t'), t') &= S[x(t + \Delta t), t + \Delta t] = S\{x(t) + f[x(t), u(t-h), t] \Delta t + \alpha_2(\Delta t), t + \Delta t\} = \\ &= S(x(t), t) + \sum_{i=1}^n \frac{\partial S(x(t), t)}{\partial x_i} f_i[x(t), u(t-h), t] \Delta t + \frac{\partial S(x(t), t)}{\partial t} \Delta t + \alpha_3(\Delta t), \end{aligned} \quad (12)$$

где $\alpha_3(\Delta t)$ – это остаточный член выше первого порядка малости по Δt , причем здесь

$$\begin{aligned} \sum_{i=1}^n \frac{\partial S(x(t), t)}{\partial x_i} f_i[x(t), u(t-h), t] + \frac{\partial S(x(t), t)}{\partial t} &= \frac{dS(x(t), t)}{dt}, \\ \frac{\partial S}{\partial x} &= \left(\frac{\partial S}{\partial x_1}, \frac{\partial S}{\partial x_2}, \dots, \frac{\partial S}{\partial x_n} \right)^* = \text{grad} S, \end{aligned}$$

(*) сверху по-прежнему означает операцию транспонирования. Следовательно, для $S(x(t'), t')$ имеем

$$S(x(t'), t') = S(x(t), t) + \left(\frac{\partial S(x(t), t)}{\partial x} \right)^* f[x(t), u(t-h), t] \Delta t + \frac{\partial S(x(t), t)}{\partial t} \Delta t + \alpha_3(\Delta t). \quad (13)$$

Подставим затем выражение (13) в правую часть соотношения (10), полагая $x(t) = x^0(t)$. Поскольку выражения $S(x(t), t)$ и $\partial S / \partial t$ не зависят от $u(\cdot) = u(t-h)$, то их можно вынести за знак $\min_{u \in U}$. После сокращения и деления обеих частей на Δt получим

$$-\frac{\partial S(x^0(t), t)}{\partial t} = \min_{u \in U} \left\{ \left(\frac{\partial S(x^0(t), t)}{\partial x^0} \right)^* f[x^0(t), u(t-h), t] + F[x^0(t), u(t-h), t] \right\} + \frac{\alpha_4(\Delta t)}{\Delta t}, \quad (14)$$

где $\alpha_4(\Delta t)$ – остаточный член выше первого порядка малости по Δt , т.е. $\alpha_4(\Delta t) / \Delta t \rightarrow 0$ при $\Delta t \rightarrow 0$.

При $\Delta t \rightarrow 0$ из уравнения (14) получим уравнение Беллмана для управляемых систем с запаздыванием в управлении:

$$-\frac{\partial S(x^0(t), t)}{\partial t} = \min_{u \in U} \left\{ \left(\frac{\partial S(x^0(t), t)}{\partial x^0} \right)^* f[x^0(t), u(t-h), t] + F[x^0(t), u(t-h), t] \right\}, \quad (15)$$

либо

$$-\frac{\partial S(x^0(t), t)}{\partial t} = \left(\frac{\partial S(x^0(t), t)}{\partial x^0} \right)^* f[x^0(t), u^0(t-h), t] + F[x^0(t), u^0(t-h), t]. \quad (16)$$

С помощью полной производной dS / dt последние два уравнения (15) и (16) можно записать в виде соотношений (4) и (5) соответственно из формулировки теоремы. Тем самым утверждение полностью доказано.

Модельный пример

В качестве простейшего модельного примера можно взять управляемую линейную систему с уравнением движения

$$\dot{x}(t) = x(t) + u[x(t), x_p(t), t-h], \quad x, x_p \in R,$$

с целевым функционалом качества вида (3):

$$J = V(y(t_1)) + \int_{t_0}^{t_1} (V(y(s)) + u^2(s-h)) ds \rightarrow \min_{u \in U},$$

где $V(y(t)) = y^2(t)$ – функция Беллмана, $u(t-h) = u[x(t), x_p(t), t-h]$, $y(t) = x(t) - x_p(t)$, $x_p(t)$ – программное движение системы, и стабилизационным условием $\lim_{t \rightarrow t_1} |x(t) - x_p(t)| < \delta$, где $\delta > 0$ – заданная достаточно малая постоянная. Применяя описанный выше метод оптимальной стабилизации с помощью теоремы, получим необходимое условие оптимальности в виде уравнения Беллмана

$$\min_{u \in U} (\dot{V}(y) + y^2 + u^2) = 0.$$

С учетом исходного уравнения движения $\dot{y} = \dot{x} - \dot{x}_p = x + u - \dot{x}_p$ это уравнение можно записать в развернутом виде:

$$\frac{\partial V}{\partial y} (x - \dot{x}_p) + y^2 + \min_{u \in U} \left(\frac{\partial V}{\partial y} u + u^2 \right) = 0, \quad \frac{\partial V}{\partial y} = 2y,$$

откуда следует формула для выбора оптимального управления

$$u^0(t-h) = u^0[x(t), x_p(t), t-h] = -y(t) = -(x(t) - x_p(t)).$$

После подстановки u^0 в уравнение движения получим

$$\dot{x} = x_p,$$

а при подстановке u^0 в уравнение Беллмана будем иметь

$$2y(x - \dot{x}_p) + y^2 - 2y^2 + y^2 = 0,$$

или $2y(x - \dot{x}_p) = 0$. Чтобы уравнение Беллмана имело место, выберем $x_p(t)$, полагая $x - \dot{x}_p = 0$.

Таким образом, приходим к системе двух уравнений первого порядка относительно $x(t)$ и $x_p(t)$:

$$\dot{x} = x_p, \quad \dot{x}_p = x.$$

Очевидно, что эта система равносильна системе двух уравнений второго порядка

$$\ddot{x}_p = x_p, \quad \ddot{x} = x$$

с общими решениями

$$x_p(t) = C_1 e^{t-t_0} + C_2 e^{-(t-t_0)}, \quad x(t) = C_1 e^{t-t_0} - C_2 e^{-(t-t_0)},$$

где C_1, C_2 – произвольные постоянные. Выбирая начальные условия

$$x_p(t_0) = C_1 + C_2 = \dot{x}(t_0), \quad \dot{x}_p(t_0) = C_1 - C_2 = x(t_0)$$

так, чтобы $C_1 = 0$, т.е.

$$x_p(t_0) = C_2 = \dot{x}(t_0), \quad x(t_0) = -C_2 = \dot{x}_p(t_0)$$

(это обеспечивается выбором программной траектории $x_p(t) = C_2 e^{-(t-t_0)}$, $t \in [t_0, t_1]$), придем к задаче оптимального торможения или, в противном случае, т.е. когда $C_1 \neq 0$, к задаче оптимального разгона движения исходного объекта управления.

Оптимальная стабилизация вращения твердого тела

В качестве примера синтеза оптимального управления рассмотрим движение твердого тела вокруг неподвижного центра инерции под действием управляющего момента M :

$$I \dot{\omega} + \omega \times I \omega = M, \quad (17)$$

либо в скалярной форме

$$A \dot{p} + (C - B) q r = M_x, \quad B \dot{q} + (A - C) p r = M_y, \quad C \dot{r} + (B - A) p q = M_z,$$

Здесь A, B, C – главные центральные моменты инерции тела; p, q, r – проекции вектора угловой скорости ω твердого тела на главные центральные оси инерции связанной с телом системы координат $Oxyz$, $I = \text{diag}(A, B, C)$ – тензор инерции. Уравнения Эйлера (17) можно записать в нормальном виде:

$$\dot{\omega} = -I^{-1}(\omega \times I\omega) + u, \quad u = I^{-1}M,$$

или

$$\dot{p} = k_1 qr + u_1, \quad \dot{q} = k_2 pr + u_2, \quad \dot{r} = k_3 pq + u_3,$$

где обозначено

$$u_p = \frac{M_x}{A}, \quad u_q = \frac{M_y}{B}, \quad u_r = \frac{M_z}{C}, \quad k_1 = \frac{B-C}{A}, \quad k_2 = \frac{C-A}{B}, \quad k_3 = \frac{A-B}{C}.$$

Тем самым имеем

$$\dot{\omega} = f(\omega) + u, \quad (18)$$

$$\omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{pmatrix} p \\ q \\ r \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad f(\omega) = \begin{pmatrix} k_1 qr \\ k_2 pr \\ k_3 pq \end{pmatrix}.$$

Зададим также программную траекторию

$$\dot{\omega}_p = f(\omega_p).$$

Введем в рассмотрение вектор-функцию $y = \omega - \omega_p$, где $\omega_p = \omega_p(t)$ – программное движение. Цель управления – минимизация разницы между движением системы и программной траекторией. Необходимо выбрать закон оптимального стабилизирующего управления u^0 в функции измеряемых значений $\omega(t)$, $t \in [t_0, t_1]$ так, чтобы обеспечивались следующие целевые условия:

$$J(u, y, t_0, t_1) = y^* y|_{t=t_1} + \int_{t_0}^{t_1} (y^* y + u^* u) dt \rightarrow \min_{u \in U}, \quad \lim_{t \rightarrow t_1} \left(\frac{y(t)}{y(t_0)} \right) < \delta, \quad (19)$$

где $\delta > 0$ – заданная малая постоянная, а $\|y(t)\|$ – евклидова норма вектора $y(t)$. Для решения задачи (17)–(19) воспользуемся полученными ранее результатами. Зададим стационарную функцию Беллмана

$$V(y) = y^* y, \quad V|_{t=t_1} = y^* y|_{t=t_1}$$

как решение уравнения Беллмана (4):

$$\min_{u \in U} (\dot{V}(y) + V(y) + u^* u) = 0, \quad (20)$$

с функционалом качества (19). С учетом выражения (18) выражение (20) запишется в виде

$$2y^* f(y) + y^* y + \min_{u \in U} (2y^* u + u^* u) = 0, \quad (21)$$

откуда будет следовать формула задания оптимального управления: $u_0 = -y$. При таком значении управления уравнение движения примет следующий вид:

$$\dot{y} + y = F(\omega, t), \quad F(\omega, t) = f(\omega, t) - \dot{\omega}_p,$$

или

$$\dot{\omega} - \dot{\omega}_p + \omega - \omega_p = f(\omega, t) - \dot{\omega}_p.$$

Если положить $F(\omega, t) = 0$, то $V(y) = y^* y \rightarrow 0$ ($t \rightarrow \infty$), откуда следует, что $y \rightarrow 0$, $\omega \rightarrow \omega_p$ ($t \rightarrow \infty$). Таким образом, имеем ограничение на выбор ω_p : $F(\omega, t) = 0 \Leftrightarrow \dot{\omega}_p = f(\omega, t)$, где $\omega(t_0) \neq \omega_p(t_0)$, т.е. $y_0 \neq 0$, $y(t) = y(t_0)e^{-(t-t_0)}$. Подставляя это выражение, получим дифференциальное уравнение для определения $\omega_p(t)$:

$$\dot{\omega}_p(t) = f(\omega_p(t) + y(t_0)e^{-(t-t_0)}, t).$$

Приведем данные численных расчетов для модели (17)–(19). В примере задавались следующие значения: главных моментов инерции: $A = 3 \text{ кг} \cdot \text{м}^2$; $B = 1 \text{ кг} \cdot \text{м}^2$; $C = 2 \text{ кг} \cdot \text{м}^2$; $[t_0, t_1] = [0, 10] \text{ с}$; $p(0) = 5 \text{ рад/с}$; $q(0) = 6 \text{ рад/с}$; $r(0) = 7 \text{ рад/с}$; $\delta = 0,05$.

После подстановки $u^0 = -y$ обратно в уравнение Беллмана (20), (21) получим

$$\dot{V} = -2V, \quad V = y^* y \rightarrow 0, \quad y \rightarrow 0,$$

при $t \rightarrow \infty$ по экспоненциальному закону. Из графиков видно, что цель управления достигнута и произведена стабилизация вращения твердого тела. При данном оптимальном управлении u^0 функционал качества J принимает минимальное постоянное значение $J = V(t_0) = y^*(t_0)y(t_0) = 110 \text{ (рад/с)}^2$, где $y^*(t_0) = (5, 6, 7) \text{ рад/с}$.

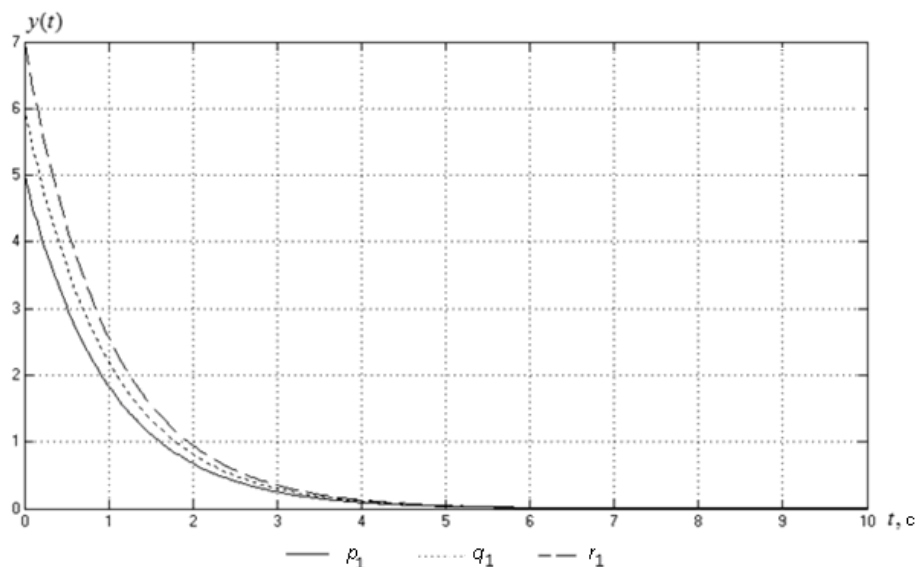


Рис. 2. Графики зависимостей угловых скоростей вращения твердого тела;

$$y(t) = \omega(t) - \omega_p(t) = (p_1(t), q_1(t), r_1(t))$$

Заключение

Основным результатом проделанной работы следует считать формирование алгоритма оптимального стабилизирующего управления для нелинейных динамических систем с запаздыванием в канале обратной связи. Отметим важные особенности данного алгоритма:

1. Уравнение Беллмана (4) обосновано в той степени, в которой имеют место требования гладкости функции Беллмана, т.е. в той мере, в которой справедливо допущение о существовании частных производных $\partial S / \partial x$, $\partial S / \partial t$ функции $S(x(t), t)$;
2. Уравнение Беллмана (4) позволяет выразить оптимальное управление $u^0 = u^0(t-h)$ в момент времени $t-h$ в функции вектора состояния $x(t)$ в момент времени t и самого времени t . Отметим, что формирование блока запаздывания (рис. 1), указывающего на зависимость между управлениями $u^0(t)$, $u^0(t-h)$, является самостоятельной задачей и в данной статье не рассматривается.

Литература

1. Габасов Р., Кириллова Ф.М. Качественная теория оптимальных процессов. – М.: Наука, 1971. – 508 с.
2. Колмановский В.Б., Носов В.Р. Устойчивость и периодические режимы регулируемых систем с последействием. – М.: Наука, 1981. – 448 с.
3. Матвеев А.С., Якубович В.А. Оптимальные системы управления: обыкновенные дифференциальные уравнения. Специальные задачи. – СПб: Изд-во СПбГУ, 2003. – 540 с.
4. Тертычный-Даури В.Ю. Галамех. Оптимальная механика. В 4-х томах. – М.: ФИЗМАТЛИТ, 2008. – Т. 4. – 607 с.
5. Бобцов А.А., Пыркин А.А. К задаче управления параметрически неопределенным линейным объектом с запаздыванием в канале управления // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 3 (73). – С. 138.
6. Бобцов А.А., Пыркин А.А. Компенсация гармонического возмущения в условиях запаздывания по управлению // Изв. РАН. Теория и системы управления. – 2008. – № 4. – С. 19–23.
7. Бобцов А.А., Колюбин С.А., Пыркин А.А. Компенсация неизвестного мультигармонического возмущения для нелинейного объекта с запаздыванием по управлению // Автоматика и телемеханика. – 2010. – № 11. – С. 136–148.

Музыка Дмитрий Александрович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, магистрант, 146038@niuitmo.ru

Пещеров Руслан Олегович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, магистрант, greshcherov@mail.ru

Тертычный-Даури Владимир Юрьевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, tertychny-dauri@mail.ru

УДК 681.586.772:531.14

ЕМКОСТНЫЙ ДАТЧИК ПОЛОЖЕНИЯ ПЛАНАРНОГО ЭЛЕКТРОПРИВОДА

А.П. Балковой, М.Г. Тяпкин

Планарный электропривод на базе планарного линейного шагового двигателя получил применение в машинах, требующих высокой точности и скорости перемещения в плоскости XU . Описан принцип действия емкостного датчика положения планарного электропривода и методы обработки его сигналов, представлен разработанный прототип емкостного датчика положения, интегрированный в якорь планарного линейного шагового двигателя. В результатах представлена оценка разрешения и точности датчика.

Ключевые слова: планарный электропривод, емкостный датчик положения.

Введение

Современное производство микросхем, печатных плат, лазерная обработка, а также устройства автоматического тестирования требуют высокой производительности и микронной точности позиционирования в плоскости. Электропривод на базе планарного линейного шагового двигателя (ПЛШД) может быть использован для таких применений [1].

Распространенная конструкция ПЛШД использует совмещенную взаимно перпендикулярную нарезку зубцов на рабочей поверхности статора и ортогональное размещение модулей линейных шаговых двигателей, образующих пары осевых двигателей (X и Y) на подвижном якоре. Каждая пара осевых двигателей обеспечивает перемещение по одной из взаимно перпендикулярных осей X и Y . Число двигателей каждой оси вдвое больше принципиально достаточного минимума, обеспечивающего линейное перемещение, что объясняется необходимостью баланса нормальных и тяговых сил модулей и компенсации разворота якоря относительно его центра масс (ось φ). Подвижный якорь ПЛШД удерживается над поверхностью статора с помощью интегрированных аэроэластических опор.

ПЛШД может применяться в микрошаговом режиме, но погрешность (30 мкм/100 мм) и повторяемость (10 мкм) позиционирования в этом режиме недостаточны в ряде применений. Недостатками микрошагового режима также являются большая колебательность и динамическая ошибка, невысокое значение максимальной скорости и невозможность компенсации возмущений.

Увеличение точности позиционирования, скорости и поворотной устойчивости планарного электропривода обеспечивается вентильным режимом пар осевых двигателей и замыканием электропривода по положению якоря в трех координатах (X , Y , φ) [2] с помощью датчика положения (ДП). ДП должен реализовать обратную связь по координатам $X1$, $X2$ центров масс пары осевых двигателей X и по ортогональной координате Y . Требуемые параметры датчика – субмикронное разрешение, микронная точность, высокая чувствительность (доли угловой минуты) к развороту.

В качестве ДП для ПЛШД могут использоваться датчики различных типов. В [3], например, представлен ДП электромагнитного типа с разрешением 0,2 мкм для осей X , Y и 0,0014° для оси φ , точность датчика без калибровки ± 25 мкм, с калибровкой ± 9 мкм, частота опроса 14 кГц. Недостатком датчика является чувствительность к намагничиванию статора токами двигателя и паразитная модуляция при движении по ортогональной оси. Для устранения этих недостатков предлагается сложная программная коррекция. В [4] представлен оптический ДП с повторяемостью 1 мкм, основным достоинством которого является нечувствительность к электромагнитному полю, а недостатком – высокая чувствительность к чистоте поверхности статора. В [5, 6] представлен ДП емкостного типа. Емкостной датчик имеет высокий уровень усреднения ошибки за счет большой площади чувствительного элемента, точность ± 10 мкм. К недостаткам емкостного ДП можно отнести высокую чувствительность к влажности и температуре окружающей среды.

Темой настоящей работы является разработка прецизионного и надежного емкостного датчика положения для планарного электропривода.

Принцип работы емкостного датчика положения

ДП состоит из измерительного первичного преобразователя (ПП) и вторичного преобразователя, или интерполятора (И), преобразующего аналоговые сигналы датчика в цифровой код положения.

Принцип действия датчика основан на изменении емкостей, образованных электродами датчика и заземленными зубцами статора планарного двигателя. Емкость такого конденсатора C пропорциональна площади S перекрытия электродов датчика и зубцов статора:

$$C = \frac{\varepsilon \cdot \varepsilon_0 \cdot S}{h},$$

где h – воздушный зазор между статором и поверхностью электрода; ε – относительная диэлектрическая проницаемость воздуха; ε_0 – электрическая постоянная.

Датчик положения планарного двигателя выполняется по принципу синусно-косинусного датчика (СКД). На рис. 1 представлен внешний вид элементарного ПП в составе СКД, содержащего группу из

четырёх металлических электродов ($\cos$, $/\sin$, $/\cos$, $\sin$) толщиной h_s , нанесенных на непроводящее термостабильное основание. Период τ_m зубцов статора составляет 640 мкм, ширина зубцов w_m и пазов – по 320 мкм, глубина пазов h_m – 200 мкм, пазы заполнены эпоксидной смолой. Зубцы статора заземлены и являются ответной частью чувствительного элемента датчика. Период электродов датчика τ_s составляет 480 мкм. Таким образом, на 3 зубца статора приходится 4 электрода, а каждый электрод сдвинут относительно предыдущего на $3/4$ периода зубцов. Для уменьшения погрешности датчика, вызванной погрешностями изготовления зубцов статора, длина электродов l_s выбирается по возможности большей, а элементарные ПП соединяются в параллельные группы.

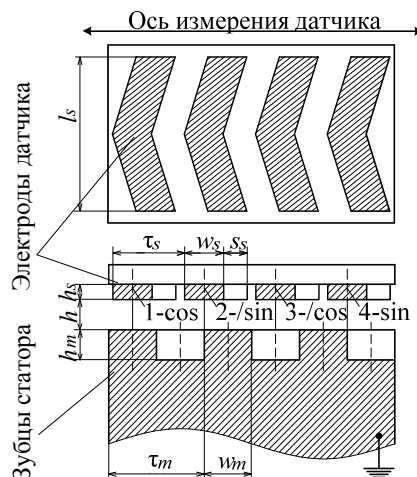


Рис. 1. Схематичный вид элементарного ПП и зубцов статора планарного двигателя

Для того чтобы емкость, образованная электродами датчика и заземленными зубцами статора, не изменялась при движении ПП по ортогональной координате, длина электродов l_s должна быть кратной периоду зубцов статора:

$$l_s = n \cdot \tau_m, \quad n = 1, 2, 3 \dots$$

Приблизительно емкости электродов датчика зависят от положения следующим образом:

$$C_1 = C_0 + C_{\max 1} \cos(\theta); \quad C_2 = C_0 - C_{\max 1} \sin(\theta);$$

$$C_3 = C_0 - C_{\max 1} \cos(\theta); \quad C_4 = C_0 + C_{\max 1} \sin(\theta),$$

где C_0 – постоянная составляющая емкости; $C_{\max 1}$ – амплитуда первой гармоники; электрический угол θ соответствует линейному положению датчика внутри периода x :

$$\theta = x \cdot \frac{2\pi}{\tau_m}.$$

Электроды датчика выполнены в виде «шеврона». Ширина w_s , скос s_s и толщина h_s электрода, а также высота расположения электрода над поверхностью h статора были оптимизированы для минимизации состава высших гармоник емкости, образованной электродом и зубцами статора.

Аппаратная часть емкостного датчика положения

Структурная схема емкостного ДП представлена на рис. 2. Электроды датчика и заземленные зубцы статора образуют переменные емкости, включенные вместе с сопротивлениями R в измерительные мосты. Верхняя точка моста подключена к высокочастотному генератору возбуждения.

При движении якоря емкости и баланс напряжений на плечах моста изменяются так, что напряжения на выходах дифференциальных усилителей (ДУ) представляют собой пару квадратурных амплитудно-модулированных сигналов $U_{\sin}$, $U_{\cos}$:

$$U_{\text{exc}}(t) = U_{\max \text{exc}} \sin(2\pi f_{\text{exc}} t);$$

$$U_{\sin}(\theta, t) = U_{\max} \sin(\theta) \sin(2\pi f_{\text{exc}} t + \psi);$$

$$U_{\cos}(\theta, t) = U_{\max} \cos(\theta) \sin(2\pi f_{\text{exc}} t + \psi),$$

где U_{exc} – напряжение возбуждения датчика; $U_{\max \text{exc}}$ – амплитуда возбуждения; f_{exc} – частота возбуждения (частота несущей); $U_{\max}$ – амплитуда сигналов $U_{\sin}$, $U_{\cos}$; ψ – фаза сигналов относительно напряжения возбуждения. Сигналы ДП $U_{\sin}$, $U_{\cos}$ поступают в аналого-цифровой преобразователь (АЦП), входящий в состав интерполятора, реализованного на базе цифрового сигнального процессора (ЦСП). После цифро-

вой обработки сигналы передаются в блок управления планарным электроприводом в аналоговом (синус/косинус, размах 1В) и цифровом виде.

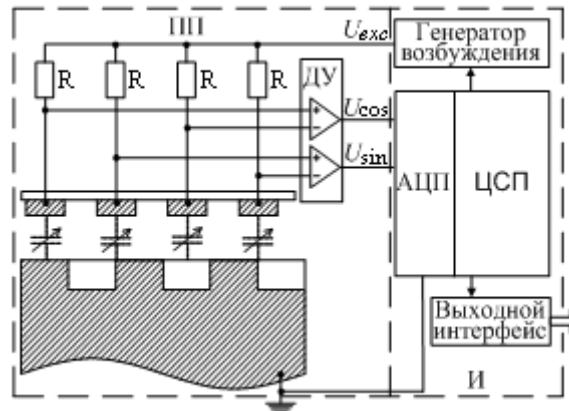


Рис. 2. Структурная схема емкостного ДП

Обработка сигналов емкостного датчика положения

Первый этап обработки сигналов ДП – импульсное детектирование, т.е. выборка сигналов датчика в моменты времени t_k , когда несущая составляющая равна единице:

$$\begin{aligned} \sin(2\pi f_{exc} t_k + \psi) &= 1; \\ U_{\sin}(\theta_k, t_k) &= U_{\max} \sin(\theta_k) \sin(2\pi f_{exc} t_k + \psi) = U_{\max} \sin(\theta_k) = U_{\sin}(\theta_k); \\ U_{\cos}(\theta_k, t_k) &= U_{\max} \cos(\theta_k) \sin(2\pi f_{exc} t_k + \psi) = U_{\max} \cos(\theta_k) = U_{\cos}(\theta_k). \end{aligned} \quad (1)$$

Частота выборки разработанного емкостного ДП составляет

$$f_s = f_{exc} = 18300 \text{ Гц}.$$

Анализ фигур Лиссажу ДП показал, что детектированные сигналы отличаются от идеальных (1). Различные амплитуды сигналов $A_{\sin}$, $A_{\cos}$ обусловлены отличными от идеальных коэффициентами усиления ДУ, постоянные составляющие сигналов $O_{\sin}$, $O_{\cos}$ вызваны паразитными емкостями и неточностью установки ДП, нелинейные возмущения $U_{d\sin}(\theta_k)$, $U_{d\cos}(\theta_k)$ – нелинейностью (высшими гармониками) емкости ДП, шумы $N_{\sin k}$, $N_{\cos k}$ – дрейфом электронных компонентов и помехами от токов двигателя:

$$\begin{aligned} U_{\sin}(\theta_k) &= O_{\sin} + A_{\sin} \sin(\theta_k) + U_{d\sin}(\theta_k) + N_{\sin k}; \\ U_{\cos}(\theta_k) &= O_{\cos} + A_{\cos} \cos(\theta_k) + U_{d\cos}(\theta_k) + N_{\cos k}. \end{aligned}$$

Для вычисления положения внутри периода ДП используется арктангенсное преобразование, которое, в случае идеальных сигналов, дает следующий результат:

$$x_k = \frac{\tau_m}{2\pi} \arctan \left(\frac{U_{\sin}(\theta_k, t_k)}{U_{\cos}(\theta_k, t_k)} \right) = \frac{\tau_m}{2\pi} \arctan \left(\frac{\sin(\theta_k)}{\cos(\theta_k)} \right) = \frac{\tau_m}{2\pi} \theta_k. \quad (2)$$

Возмущения в сигналах ДП приводят к погрешности датчика. По этой причине детектированные сигналы датчика перед вычислением положения подвергаются коррекции. Основными параметрами, ухудшающими точность, являются постоянные составляющие и неравенство амплитуд сигналов датчиков. Коррекция сигналов состоит в определении постоянной составляющей и амплитуды и изменении на основе полученных данных сигналов датчика. Для определения параметров коррекции в процесс цифровой обработки сигналов включен алгоритм поиска максимума и минимума.

Значения постоянных составляющих и амплитуд получаются из экстремумов сигналов датчика внутри одного периода по формулам

$$\begin{aligned} A_{\sin}^{calc} &= \frac{\max(U_{\sin}(\theta_k)) - \min(U_{\sin}(\theta_k))}{2}; & A_{\cos}^{calc} &= \frac{\max(U_{\cos}(\theta_k)) - \min(U_{\cos}(\theta_k))}{2}; \\ O_{\sin}^{calc} &= \frac{\max(U_{\sin}(\theta_k)) + \min(U_{\sin}(\theta_k))}{2}; & O_{\cos}^{calc} &= \frac{\max(U_{\cos}(\theta_k)) + \min(U_{\cos}(\theta_k))}{2}. \end{aligned}$$

В результате коррекции получаются следующие сигналы:

$$U_{\sin}^{cor}(\theta_k) = \frac{U_{\sin}(\theta_k) - O_{\sin}^{calc}}{A_{\sin}^{calc}} = \frac{O_{\sin} + A_{\sin} \sin(\theta_k) + U_{d \sin}(\theta_k) + N_{\sin k} - O_{\sin}^{calc}}{A_{\sin}^{calc}},$$

$$U_{\cos}^{cor}(\theta_k) = \frac{U_{\cos}(\theta_k) - O_{\cos}^{calc}}{A_{\cos}^{calc}} = \frac{O_{\cos} + A_{\cos} \cos(\theta_k) + U_{d \cos}(\theta_k) + N_{\cos k} - O_{\cos}^{calc}}{A_{\cos}^{calc}}. \quad (3)$$

Недостатком метода является невозможность устранить ошибку вычисления положения, вызванную нелинейными возмущениями и шумами.

Реализация

На рис. 3 представлено размещение первичных преобразователей ДП в якоре планарного двигателя. ПП Y устанавливается в центре масс якоря, совпадающем с его геометрическим центром. ПП $X1$ и $X2$ устанавливаются симметрично относительно центра масс. Интерполятор выполнен в виде отдельного устройства на базе 12-разрядного АЦП и ЦСП TMS320F2812.

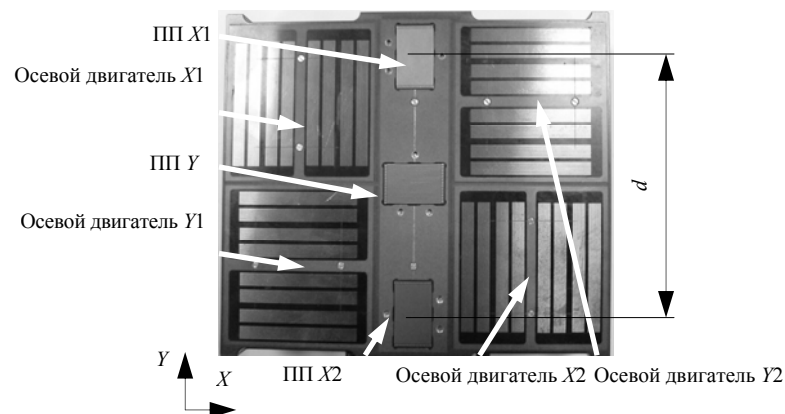


Рис. 3. Размещение первичных преобразователей в якоре планарного двигателя

Линейные координаты и угол разворота вычисляются как

$$X_k = (x_{kX1} + x_{kX2})/2, Y_k = y_k, \varphi_k = \arctan(x_{kX1} - x_{kX2})/d,$$

где x_{kX1} – положение ПП $X1$; x_{kX2} – положение ПП $X2$; y_k – положение ПП Y ; d – расстояние между центрами ПП $X1$ и $X2$.

Результаты

Разрешение датчика определялось как среднеквадратичное отклонение σ шумовой составляющей вычисленного положения. Лазерный интерферометр SP 2000-TR фирмы SIOS был использован для проверки неподвижности якоря (рис. 4).

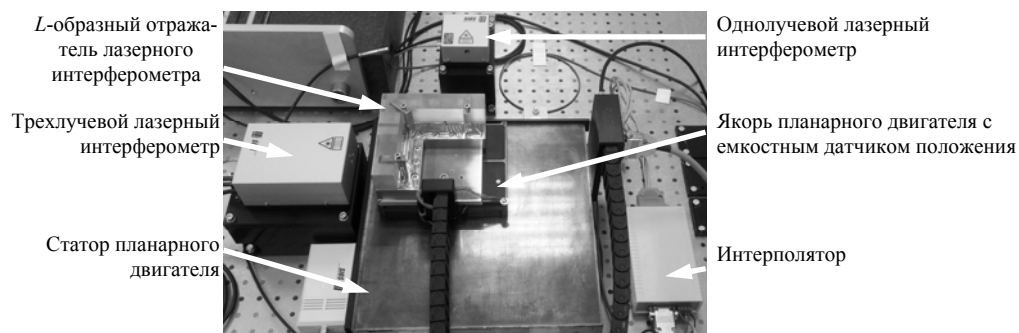


Рис. 4. Аттестационный стенд

Сигналы датчика были записаны с частотой опроса 18300 Гц при неподвижном якоре (рис. 5). Положение было вычислено с использованием арктангенсного преобразования (2). Разрешение каждого датчика было рассчитано по формуле

$$\sigma = \sqrt{\frac{1}{n_s - 1} \cdot \sum_{k=1}^{n_s} (x_k - \bar{x})^2}, \quad \bar{x} = \frac{1}{n_s} \cdot \sum_{k=0}^{n_s} x_k,$$

где k – номер выборки; $n_s = 4000$ – число выборок.

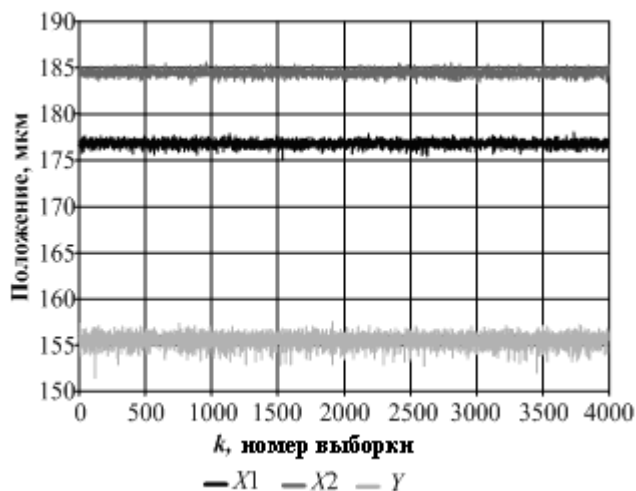


Рис. 5. Сигналы датчика при неподвижном якоре

Для проверки инвариантности шума к положению данный опыт был проделан в 8 точках, отстоящих друг от друга на 80 мкм по осям X и Y . Опыты подтвердили инвариантность шума к положению якоря. В результате разрешение датчиков $X1$ и $X2$ составило 0,35 мкм, датчика Y – 0,68 мкм.

Разрешение по вычислению разворота определяется из показаний $X1$ и $X2$:

$$\sigma_{\varphi} = \arctan\left(\frac{\max(\sigma_{X1}, \sigma_{X2})}{d}\right) = 3,1 \cdot 10^{-6} \text{ радиан } (0,0018^\circ).$$

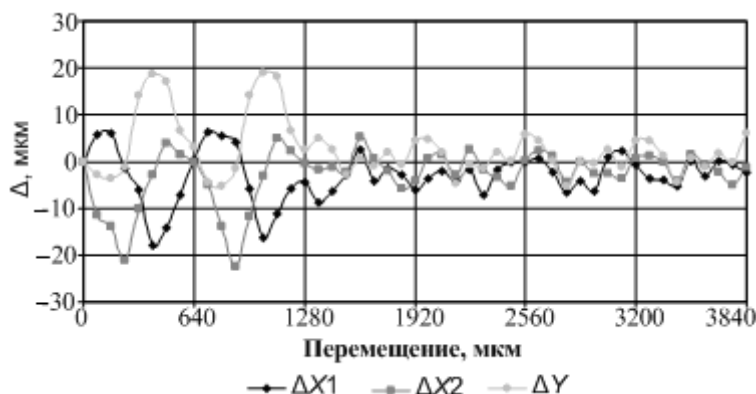


Рис. 6. Позиционная ошибка емкостного ДП без коррекции (0–1280 мкм) и после коррекции (1280–3840 мкм)

Для определения статической точности ДП был также использован лазерный интерферометр. Якорь перемещался с заданным шагом 80 мкм одновременно по осям X и Y , сигналы ДП и показания интерферометра записывались после перемещения в моменты полной остановки якоря. Погрешность Δ вычислялась как разность между положением, вычисленным через арктангенсное преобразование сигналов датчика (2), и показаниями интерферометра. На рис. 6 представлена позиционная ошибка емкостного ДП. Первые два периода (0–1280 мкм) сигналы ДП не были подвергнуты коррекции. Следующие четыре периода (1280–3840 мкм) сигналы датчика были скорректированы по (3), что позволило уменьшить погрешность в 3 раза.

Заключение

В работе представлен емкостный датчик планарного электропривода. Датчик положения обеспечивает измерение положения центра масс якоря планарного линейного шагового двигателя в трех координатах (X , Y , φ). Экспериментально получено разрешение датчика положения: по оси X – 0,35 мкм, по оси Y – 0,68 мкм, по оси φ – 0,0018°. Основное влияние на погрешность датчика положения оказывают постоянные составляющие сигналов, дефекты зубцов статора и шумы. Коррекция сигналов датчика положения увеличивает его точность в 3 раза, статическая погрешность с коррекцией составляет 10 мкм.

Для увеличения точности емкостного датчика положения в дальнейшем предполагается объединение первичных преобразователей и интерполятора, а также интеграция датчика положения внутрь планарного линейного шагового двигателя.

Литература

1. Saweyr B.A. Linear magnetic drive system. – U.S. Patent 3,735,231, May 22, 1973.
2. Quaid E., Hollis Ralph L. 3-DOF Closed-loop control for planar linear motors // Proceeding of the 1998 IEEE International Conference on Robotics & Automation. – Leuven, Belgium. – May, 1998. – P. 2488–2493.
3. Butler Zack J., Rizzi Alfred A., Hollis Ralph L. Integrated Precision 3-DOF Position Sensor for Planar Linear Motors // Proceeding of the 1998 IEEE International Conference on Robotics & Automation. – Leuven, Belgium. – May, 1998. – P. 2652–2658.
4. Fries Gregory A., Rizzi Alfred A., Hollis Ralph L. Fluorescent Dye Based Optical Position Sensing for Planar Linear Motors // Proceedings of the 1999 IEEE International Conference on Robotics & Automation. – Detroit, Michigan. – May, 1999. – P. 1614–1619.
5. Miller G.L. Capacitively incremental position measurement and motion control. – U.S. Patent 4,893,071, January 09, 1990.
6. Мухаметгалеев Т.Х. Разработка замкнутого по положению планарного дискретного электропривода. Кандидатская диссертация. – М.: МЭИ (ТУ), 1994. – 171 с.

Балковой Александр Петрович

– Национальный исследовательский университет «Московский энергетический институт», кандидат технических наук, ст. научный сотрудник, balk1954@yahoo.com

Тяпкин Михаил Геннадьевич

– Национальный исследовательский университет «Московский энергетический институт», аспирант, tiapkinmg@mail.ru

УДК 519.632.4

ИСПОЛЬЗОВАНИЕ БЫСТРОГО МЕТОДА ГРАНИЧНЫХ ЭЛЕМЕНТОВ ДЛЯ РЕШЕНИЯ ЗАДАЧ МАГНИТОСТАТИКИ

И.М. Ступаков, М.Э. Рояк

Рассматривается схема решения задач магнитостатики методом граничных элементов и его ускорение, основанное на быстром методе мультиполей. Такой подход позволяет значительно снизить вычислительные затраты и решать задачи большей размерности. Благодаря использованию симметричной галеркинской постановки метод можно использовать совместно с методом конечных элементов для решения нелинейных задач магнитостатики. Приводятся результаты вычислительных экспериментов, демонстрирующие эффективность рассматриваемого подхода.

Ключевые слова: метод граничных элементов, метод конечных элементов, магнитостатика, неполный скалярный магнитный потенциал, сферические гармоники.

Введение

Задачи магнитостатики часто возникают при моделировании стационарных магнитных процессов. К ним относится расчет магнитных полей, возбуждаемых постоянным током или постоянными магнитами. Основным подходом для численного решения таких задач в настоящее время является метод конечных элементов (МКЭ) [1]. Для его применения область, в которой решается задача, должна быть разбита на элементы, в роли которых обычно выступают многогранники. Результат решения задачи сильно зависит от качества построения этого разбиения. По этой причине в ситуациях, когда необходимо решать задачи с большим количеством трехмерных объектов (или с объектами сложной формы), построение достаточно качественной конечно-элементной сетки может стать нетривиальной проблемой. В таких случаях может быть целесообразным использовать метод граничных элементов (МГЭ), в котором требуется сетка только на границе между объектами, что значительно упрощает ее построение.

Основными недостатками МГЭ по сравнению с МКЭ является невозможность эффективного учета нелинейных свойств объектов и плотная структура системы линейных алгебраических уравнений (СЛАУ), получаемой в результате применения метода. Первый недостаток можно обойти путем совместного использования МКЭ и МГЭ. Для устранения второго недостатка было предложено несколько различных схем быстрого МГЭ [2]: крестовая аппроксимация матриц, использование вейвлетов в качестве базисных функций, использование быстрого метода мультиполей. Далее рассматривается именно последний вариант, основанный на разложении фундаментального решения в ряд по сферическим функциям и иерархическом разбиении пространства. Впервые такой подход был предложен для быстрого моделирования систем частиц в [3], дальнейшее развитие метода изложено в [4]. Благодаря использованию этого метода можно значительно снизить вычислительные затраты в МГЭ.

Математическая модель задачи

Задача магнитостатики может быть описана системой уравнений

$$\begin{cases} \operatorname{rot} \mathbf{H} = \mathbf{J}, \\ \operatorname{div} \mathbf{B} = 0, \end{cases} \quad (1)$$

где $\mathbf{H}$ – вектор напряженности магнитного поля; $\mathbf{J}$ – вектор сторонних токов; $\mathbf{B} = \mu\mathbf{H} + \mathbf{M}$ – вектор индукции магнитного поля, μ – магнитная проницаемость, а $\mathbf{M}$ – намагниченность. Для ее решения используем подход с неполным скалярным магнитным потенциалом [1, 5]. Обозначим $\mathbf{H}_0$ – решение задачи в однородном пространстве, т.е. удовлетворяющее системе уравнений

$$\begin{cases} \operatorname{rot} \mathbf{H}_0 = \mathbf{J}, \\ \operatorname{div} \mathbf{H}_0 = 0. \end{cases} \quad (2)$$

Вычитая первое уравнение системы (2) из первого уравнения системы (1), получаем $\operatorname{rot}(\mathbf{H} - \mathbf{H}_0) = 0$,

значит, $\mathbf{H}$ можно представить в виде

$$\mathbf{H} = \mathbf{H}_0 - \operatorname{grad} u. \quad (3)$$

Подставляя (3) во второе уравнение системы (1), с учетом связи между $\mathbf{H}$ и $\mathbf{B}$ получаем

$$\operatorname{div}(\mu \operatorname{grad} u) = \operatorname{div}(\mu \mathbf{H}_0 + \mathbf{M}). \quad (4)$$

Для решения задачи удобно разбить область на несколько непересекающихся подобластей так, чтобы всем границам между материалами соответствовали границы разбиения. Из уравнения (4) можно получить уравнения связи, которые должны выполняться на границе между подобластями:

$$\begin{cases} [u]_{\Gamma_{ij}} = 0, \\ \left[\mu \frac{\partial u}{\partial \mathbf{n}} \right]_{\Gamma_{ij}} = [(\mu \mathbf{H}_0 + \mathbf{M}) \cdot \mathbf{n}]_{\Gamma_{ij}}, \end{cases} \quad (5)$$

где Γ_{ij} – граница между i -ой и j -ой подобластями.

Если внутри подобласти магнитную проницаемость и намагниченность можно считать константами, уравнение (4) вырождается в уравнение Лапласа, и для построения аппроксимации решения в этой подобласти можно использовать МГЭ. Такие подобласти почти всегда присутствуют в задаче, например, подобласть, содержащая воздушную среду. В остальных подобластях можно использовать МКЭ.

Метод граничных элементов

Рассмотрим применение метода граничных элементов к краевой задаче для уравнения Лапласа $\Delta u = 0$. Пусть Ω – область трехмерного пространства с границей S . Фундаментальное решение уравнения Лапласа определяется формулой

$$U(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi} \frac{1}{|\mathbf{r} - \mathbf{r}'|}.$$

Если подставить фундаментальное решение в формулу Грина

$$\int_{\Omega} v \Delta u - u \Delta v dV = \int_S v \frac{\partial u}{\partial \mathbf{n}} - u \frac{\partial v}{\partial \mathbf{n}} dS,$$

в качестве функции v можно получить интегральное представление гармонической (т.е. удовлетворяющей уравнению Лапласа) функции

$$u(\mathbf{r}) = \int_S U(\mathbf{r}, \mathbf{r}') \frac{\partial u}{\partial \mathbf{n}}(\mathbf{r}') dS_{\mathbf{r}'} - \int_S \frac{\partial U}{\partial \mathbf{n}_{\mathbf{r}'}}(\mathbf{r}, \mathbf{r}') u(\mathbf{r}') dS_{\mathbf{r}'}, \quad (6)$$

где первый интеграл называется потенциалом простого слоя, а второй – потенциалом двойного слоя. При получении уравнения (6) используется тот факт, что $\Delta u = 0$ и $\Delta U(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}')$.

Функция, удовлетворяющая уравнению (6), автоматически удовлетворяет уравнению Лапласа, значит, для решения задачи остается найти значения функции u и ее потоки на границе области так, чтобы выполнялись уравнения (5). Для этого в МГЭ используют граничные интегральные уравнения, которые связывают значение и потоки на границе.

Для получения первого граничного уравнения точку, в которой вычисляется значение функции u в формуле (6), устремим к границе S области Ω и возьмем предел. При этом необходимо учесть, что потенциал двойного слоя терпит разрыв на границе. Запишем результат в операторной форме [2]:

$$u = V \frac{\partial u}{\partial \mathbf{n}} + \left(\frac{1}{2} I - K \right) u, \quad (7)$$

где потенциалы простого и двойного слоя обозначены соответственно символами V и K , тождественный оператор – символом I .

Для получения второго уравнения перед вычислением предела возьмем от выражения (6) производную по направлению нормали к границе S . Учитывая разрыв производной потенциала простого слоя на границе, можно получить второе граничное уравнение

$$\frac{\partial u}{\partial \mathbf{n}} = \left(\frac{1}{2} I + K' \right) \frac{\partial u}{\partial \mathbf{n}} + Du, \quad (8)$$

где символами K' и D обозначены сопряженный потенциал двойного слоя и гиперсингулярный интегральный граничный оператор соответственно [2].

Если выразить из уравнения (7) потоки и подставить в уравнение (8), можно получить оператор Стеклова–Пуанкаре в симметричной форме

$$\frac{\partial u}{\partial \mathbf{n}} = \left(D + \left(\frac{1}{2} I + K' \right) V^{-1} \left(\frac{1}{2} I + K \right) \right) u,$$

на основе которого наиболее удобно производить решение задачи, разбитой на несколько подобластей. На практике вместо обращения оператора V решают систему уравнений (7) и (8), предполагая, что потоки в левой части уравнения (8) даны. Для ее решения будем использовать метод Галеркина.

Запишем систему уравнений (7) и (8) в слабой форме:

$$\begin{cases} (Vw, v) - \left(\left(\frac{1}{2} I + K \right) u, v \right) = 0, \forall v \in H^{-1/2}(S), \\ \left(w, \left(\frac{1}{2} I + K \right) v \right) + (Du, v) = \left(\frac{\partial u}{\partial \mathbf{n}}, v \right), \forall v \in H^{1/2}(S), \end{cases} \quad (9)$$

где скалярное произведение $(u, v) = \int_S u v dS$, а $H^{-1/2}(S)$ и $H^{1/2}(S)$ – пространства Соболева–Слободецкого [2]. Для построения дискретного аналога системы (9) представим функции w и u в виде разложения по финитным полиномиальным базисным функциям

$$w(\mathbf{r}) = \sum_i \mathbf{p}_i \varphi_i(\mathbf{r}), \quad u(\mathbf{r}) = \sum_i \mathbf{q}_i \psi_i(\mathbf{r}).$$

Подставляя это разложение в систему (9) и выбирая в качестве пробных функций v , функции $\{\varphi_i\}$ для первого уравнения и $\{\psi_i\}$ для второго, получаем СЛАУ

$$\begin{pmatrix} \mathbf{V} & -\left(\frac{1}{2} \mathbf{M} + \mathbf{K} \right) \\ \left(\frac{1}{2} \mathbf{M} + \mathbf{K} \right)^T & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{F} \end{pmatrix}, \quad (10)$$

где

$$\mathbf{V}_{ij} = (V\varphi_j, \varphi_i) = \iint_S \iint_S \varphi_i(\mathbf{r}') U(\mathbf{r}, \mathbf{r}') \varphi_j(\mathbf{r}) dS_{\mathbf{r}} dS_{\mathbf{r}'}, \quad (11)$$

$$\mathbf{K}_{ij} = (K\psi_j, \varphi_i) = \iint_S \iint_S \varphi_i(\mathbf{r}') \frac{\partial U}{\partial \mathbf{n}_{\mathbf{r}}}(\mathbf{r}, \mathbf{r}') \psi_j(\mathbf{r}) dS_{\mathbf{r}} dS_{\mathbf{r}'}, \quad (12)$$

$$\mathbf{D}_{ij} = (D\psi_j, \psi_i) = \iint_S \iint_S U(\mathbf{r}, \mathbf{r}') \operatorname{curl} \psi_i(\mathbf{r}') \cdot \operatorname{curl} \psi_j(\mathbf{r}) dS_{\mathbf{r}} dS_{\mathbf{r}'}, \quad (13)$$

$$\mathbf{M}_{ij} = (\psi_j, \varphi_i) = \iint_S \varphi_i(\mathbf{r}) \psi_j(\mathbf{r}) dS_{\mathbf{r}},$$

а $\mathbf{F}$ – вектор правой части, определяемый условиями сопряжения на границе подобласти (или краевыми условиями). Система (10) удобна тем, что для стыковки метода на границе с МКЭ (или другой подобластью граничных элементов) достаточно выбрать одинаковые базисные функции $\{\psi_i\}$ и подставить полученные дискретизации в уравнение неразрывности потоков.

Структура СЛАУ (10) является плотной, а значит, вычислительные затраты на ее сборку растут квадратично с ростом числа элементов. Чтобы этого избежать, используем быстрый мультипольный метод. Он основан на разложении фундаментального решения уравнения Лапласа в ряд

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \sum_{n=0}^{\infty} \frac{|\mathbf{r}|^n}{|\mathbf{r}'|^{n+1}} \sum_{m=-n}^n Y_n^m(\mathbf{r}) Y_n^{-m}(\mathbf{r}'), \quad (14)$$

где Y_n^m – сферические гармоники порядка n [3]. Подставляя это разложение в формулу (11), можно избавиться от вычисления двойного интеграла и считать интегралы по $\mathbf{r}$ и по $\mathbf{r}'$ независимо:

$$\mathbf{V}_{ij} = \sum_{n=0}^{\infty} \sum_{m=-n}^n \left(\int_S |\mathbf{r}|^n Y_n^m(\mathbf{r}) \varphi_j(\mathbf{r}) dS_{\mathbf{r}} \right) \left(\int_S \frac{Y_n^{-m}(\mathbf{r}')}{|\mathbf{r}'|^{n+1}} \varphi_i(\mathbf{r}') dS_{\mathbf{r}'} \right).$$

Матрицы (12) и (13) можно вычислять аналогично. Разумеется, необходимо учитывать область сходимости ряда (14), который сходится только при $|\mathbf{r}| < |\mathbf{r}'|$, что приводит к необходимости древовидного разбиения пространства [3], в котором вклады от пар элементов, расположенных в одном листе дерева, считаются напрямую по формулам (11)–(13), а в остальных случаях используется вычисление вкладов через суммирование ряда (14). Это позволяет вычислять интегралы для каждой базисной функции только один раз и, следовательно, избежать квадратичного роста затрат.

Результаты вычислительных экспериментов

Для оценки точности разрабатываемого подхода сравним результаты вычисления искажений магнитного поля вокруг металлической балки, изображенной на рис. 1, при постоянном внешнем поле с результатами, полученными с помощью МКЭ. На рис. 2 приводятся графики напряженности магнитного поля, нормированные на модуль внешнего поля. Как видно из графиков, решение, полученное МГЭ на грубой сетке, точнее, чем решение, полученное МКЭ на удвоенной сетке.

Отметим, что если брать сферические гармоники до 10-го порядка, то решения, полученные быстрым и обычным МГЭ, практически не отличаются.

Для тестирования эффективности предложенной схемы решим задачу о вычислении искажений магнитного поля земли металлическими объектами в стенах помещения. Расположение металлических объектов представлено на рис. 3.

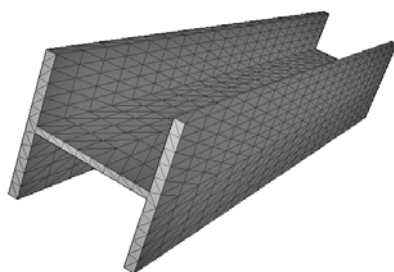


Рис. 1. Металлическая балка (с габаритами 0,2×0,2×1,0 м) и гранично-элементная сетка на ее поверхности

В модели задано 319 металлических прутьев с квадратным сечением 2×2 см, магнитная проницаемость металла принята равной 1000.

При решении этой задачи даже на очень грубой сетке обычным МГЭ затраты времени и памяти становятся значительно больше, чем при использовании ускоренного подхода. Так, прямой метод требует только на сборку матриц более часа времени и более 4 ГБ памяти. Использование МКЭ для решения этой задачи также затруднительно, поскольку необходимо построение сетки в воздухе между объектами. Применение же быстрого МГЭ позволяет решить эту задачу даже на относительно подробной сетке за приемлемое время. Как видно из таблицы, для быстрого МГЭ рост затрат времени и памяти при дроблении сетки близок к линейному, тогда как у обычного МГЭ он является квадратичным.

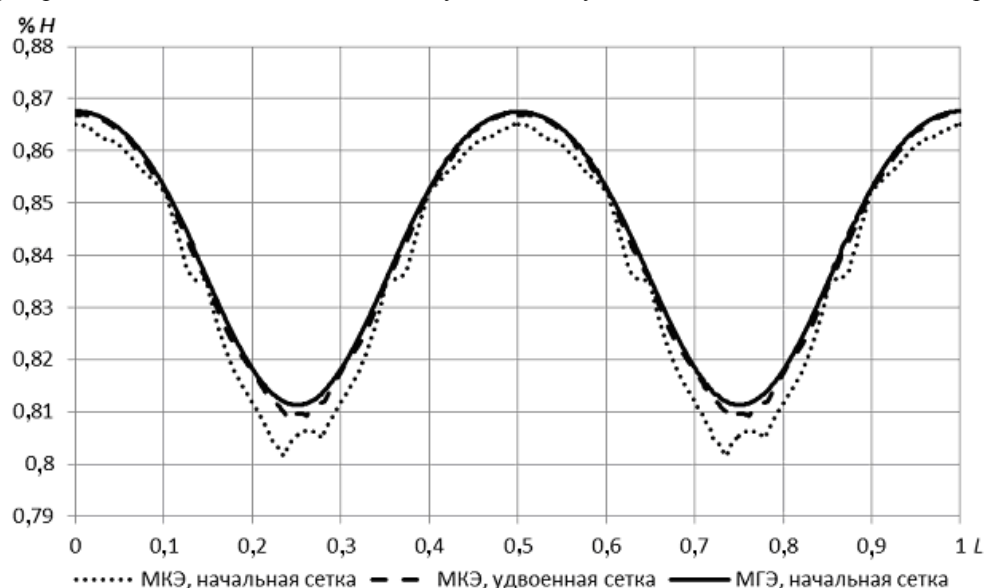


Рис. 2. Нормированные модули напряженности магнитного поля, полученные из расчетов МКЭ и МГЭ

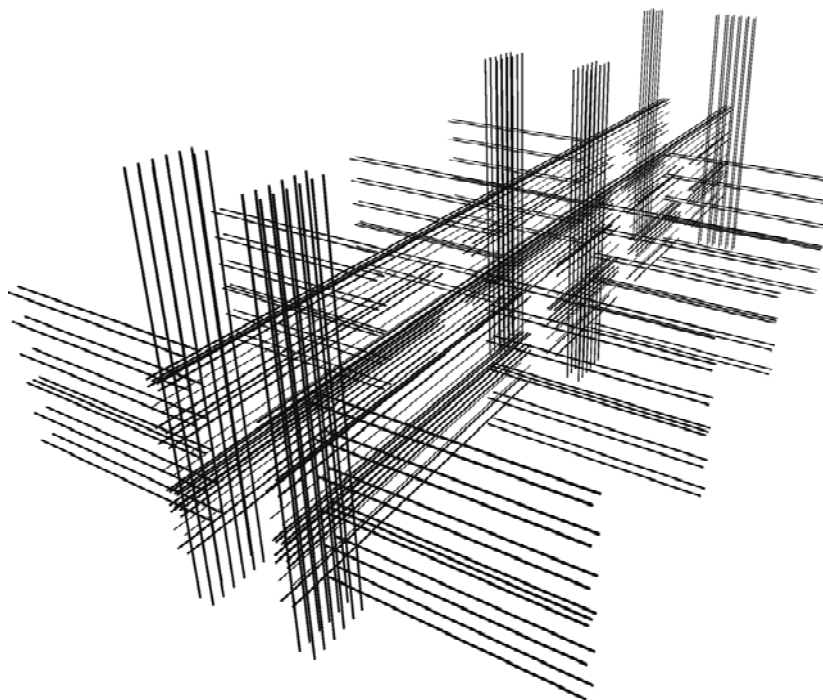


Рис. 3. Металлические объекты в стенах помещения, общая длина – 15 м; ширина – 9 м; высота – 6 м

Количество степеней свободы	Время сборки матриц, с	Полное время решения, с	Память, МБ
22364	561	1338	2150
41156	845	2007	3014
77768	1435	3023	5433
151676	2730	5944	9699

Таблица. Вычислительные затраты быстрого метода граничных элементов

Заключение

Полученные результаты подтверждают, что метод граничных элементов с использованием для ускорения быстрого мультипольного метода эффективен для решения линейных задач магнитостатики. Он позволяет даже на достаточно грубых сетках получать решение с высокой точностью и при этом существенно упрощает процедуры построения сетки по сравнению с методом конечных элементов. При этом он остается достаточно перспективным для решения и нелинейных задач совместно с методом конечных элементов, поскольку в этом случае процедура генерации сетки усложнится несущественно – сетки во всех ферромагнитных объектах могут строиться независимо друг от друга.

Литература

1. Соловейчик Ю.Г., Рояк М.Э., Персова М.Г. Метод конечных элементов для решения скалярных и векторных задач: Учеб. пособие. – Новосибирск: Изд-во НГТУ, 2007. – 896 с.
2. Steinbach O. Numerical approximation methods for elliptic boundary value problems. – New York: Springer Science, 2008. – 386 p.
3. Greengard L., Rokhlin V. A fast algorithm for particle simulations // J. Comput. Phys. – 1987. – V. 73. – P. 325–348.
4. Cheng H., Greengard L., Rokhlin V. A Fast Adaptive Multipole Algorithm in Three Dimensions // J. Comput. Phys. – 1999. – V. 155. – P. 468–498.
5. Ступаков И.М., Корсун М.М., Рояк М.Э. Об учете источников электромагнитного поля в совместном методе конечных и граничных элементов // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 5 (69). – С. 67–71.

Ступаков Илья Михайлович

– Новосибирский государственный технический университет, аспирант, istupakov@gmail.com

Рояк Михаил Эммануилович

– Новосибирский государственный технический университет, доктор технических наук, доцент, профессор, royak@fpm.ami.nstu.ru

УДК 517.521: 004.046

СВОЙСТВА ПРЕОБРАЗОВАНИЯ n -МЕРНЫХ ЦИФРОВЫХ СИГНАЛОВ ПО БАЗИСУ ПРЯМОУГОЛЬНЫХ ВСПЛЕСКОВ

А.Ю. Гришенцев

Рассмотрены свойства способа декомпозиции n -мерных цифровых сигналов по базису прямоугольных всплесков. Показано выполнение свойства линейного преобразования и условия сохранения энергии сигнала при переходе от пространственного к частотно-пространственному представлению. Сформулирован переход к ортогональной форме преобразования.

Ключевые слова: декомпозиция n -мерных сигналов, спектральный анализ, цифровая обработка сигналов.

Введение

На сегодняшний день наиболее востребованными способами взаимного преобразования цифрового сигнала из частотной в пространственную область являются Фурье- и вейвлет-преобразование [1, 2]. Вейвлет-преобразование, также как и оконное преобразование Фурье, позволяет не только получить спектр сигнала, но и локализовать его в пространстве. В настоящей работе рассматриваются свойства способа преобразования по базису прямоугольных всплесков (БПВ) [3, 4], который также позволяет получить пространственную локализацию спектра цифрового сигнала, при этом достаточно просто реализуется с помощью программных или только аппаратных средств.

Свойства преобразования по базису прямоугольных всплесков

Основной задачей рассматриваемого способа декомпозиции по БПВ является получение спектра n -мерного цифрового сигнала и его локализация в пространстве R^n , фактически отображение цифрового сигнала в фазовое пространство. Эта задача решается в ходе прямого преобразования (декомпозиции) путем последовательных итеративных вычислений в соответствии с выражением

$$f_{k-1} = f_k - S_k, \quad (1)$$

где k – номер спектрального элемента декомпозиции S_k , выделяемого по масштабному признаку из сигнала f_k . Значение индекса k соответствует протяженности взаимно перпендикулярных и параллельных элементов спектральных компонент, f_k – остаточный сигнал. Максимальное (исходное) значение k равно K , S_k – спектральные элементы декомпозиции, отобранные по масштабному признаку. Сумма всех полученных в ходе прямого преобразования спектральных элементов декомпозиции S_k является результатом обратного преобразования (синтеза) и равна

$$f_K = \sum_{k=1}^K S_k. \quad (2)$$

Таким образом, разложение n -мерного сигнала происходит не по выбранному заранее n -мерному базису (базисной функции), а по взаимно параллельным и перпендикулярным элементам исходного сигнала, которые образуют множество уникальных для данного сигнала n -мерных базисов, являющихся частью исходного сигнала. Основой для формирования таких базисов разложения n -мерного сигнала служит меандр-подобный сигнал, называемый в рамках рассматриваемого способа элементарным всплеском. Под элементарным всплеском будем понимать дискретную структуру, имеющую размерность, равную размерности исходного сигнала с отличной от нуля амплитудой. В направлении выделения элементарного всплеска его протяженность может иметь любое отличное от нуля значение, но не более размера исходного сигнала в данном направлении. По другим направлениям размеры элементарного всплеска равны единице дискретизации соответствующих направлений [3, 4].

На рис. 1 показаны некоторые свойства конфигурации элементов декомпозиции на примере одномерных сигналов. Приведены варианты четырех одномерных сигналов $f[x]$ и некоторые возможные способы декомпозиции. Стрелками обозначены переходы к корректным вариантам декомпозиции, подчеркнутые стрелки обозначают некорректные варианты декомпозиции (присутствуют на рис. 1, в, г) с последующим переходом к корректным. В ходе наблюдения за возможными вариантами декомпозиции и разделением их на корректные и некорректные можно сделать некоторые обобщения:

- (1) менее протяженные элементарные всплески конфигурационно могут быть расположены только полностью над непрерывным более протяженным, либо над нулевым (имеется в виду его отсутствие) всплеском;
- (2) соседние всплески не могут быть расположены неразрывно, между всплесками по оси положения в пространстве должен присутствовать разрыв, минимальная протяженность которого равна единице дискретизации сигнала в данном направлении.

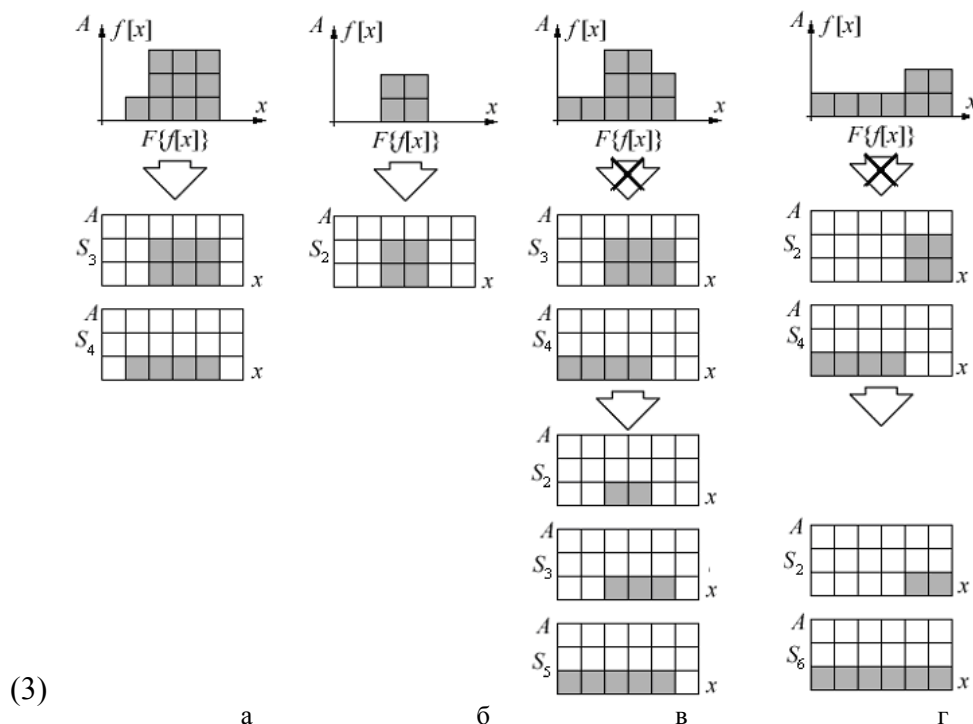


Рис. 1. Допустимые и недопустимые варианты конфигурации элементов декомпозиции по БПВ:
 $f[x]$ – исходные сигналы с амплитудой A ; S_k – спектральный элемент декомпозиции, полученный
 выделением всплесков протяженностью k

Указанные свойства (1)–(2) являются следствием декомпозиции сигнала в соответствии с алгоритмом, рассмотренным в [3, 4], и могут быть обобщены на случай многомерного сигнала. Из свойства (2) можно вывести понятие периода $T = k + 1$ (или минимального периода) элементарного всплеска как минимально допустимого периода повторения элементарных всплесков заданной протяженности k . Покажем, что декомпозиция по БПВ является линейным преобразованием, т.е. обладает свойствами линейной системы – аддитивностью и однородностью [5].

Аддитивность декомпозиции по БПВ обусловлена тем, что суммирование сигналов $f[x] + g[x]$ в пространственной области эквивалентно суммированию сигналов в пространственно-частотной области $F\{f[x]\} + F\{g[x]\}$, причем при суммировании сигналов в пространственно-частотной области необходимо приводить результат суммирования к конечному виду в соответствии со свойствами (1)–(2). Аддитивность преобразования является следствием равенства суммы $f[x] + g[x] = F^{-1}\{F\{f[x]\} + F\{g[x]\}\}$ для каждого конкретного значения x и одновременной инвариантности декомпозиции по БПВ. Однородность декомпозиции по БПВ $F\{m \cdot f[x]\} = m \cdot F\{f[x]\}$ является следствием равенства

$$m \sum_{k=1}^K S_k[x] = m \cdot f[x], \text{ где } m - \text{рациональное число.}$$

Для n -мерного пространства R^n можно записать свойства линейности преобразования по БПВ:
 аддитивность – $f[R^n] + g[R^n] = F^{-1}\{F\{f[R^n]\} + F\{g[R^n]\}\}$,

$$\text{однородность – } m \sum_{k=1}^K S_k[R^n] = m \cdot f[R^n].$$

Отметим, что декомпозиция по БПВ не является инвариантом относительно сдвига сигнала $f[R^n]$, так как сдвиг исходного сигнала вызывает соответствующее смещение положения элементарных всплесков. При сдвиге сигнала $f[R^n]$ инвариантом является спектральная плотность p , рассчитываемая как отношение суммы всех элементов каждого спектрального элемента декомпозиции $S_k[R^n]$ к числу всех дискретных элементов, в котором задана функция $f[R^n]$. Покажем, что при преобразовании по БПВ неизменной остается энергия. В случае одномерного сигнала $f[x]$, заданного на интервале X , его энергия может быть определена как

$$E = \sum_X f^2[x], \quad (3)$$

из выражения (2) [1] следует, что

$$f^2[x] = \left(\sum_{k=1}^K S_k[x] \right)^2. \quad (4)$$

Раскрывая выражение (4) как полиномиальный многочлен второй степени, получаем

$$\left(\sum_{k=1}^K S_k[x] \right)^2 = \sum_{k=1}^K S_k^2[x] + 2 \sum_{i=1}^{K-1} \sum_{j=i+1}^K S_i[x] S_j[x],$$

или, в конечном виде,

$$\sum_X f^2[x] = \sum_X \left[\sum_{k=1}^K S_k^2[x] + 2 \sum_{i=1}^{K-1} \sum_{j=i+1}^K S_i[x] S_j[x] \right]. \quad (5)$$

Правую часть, стоящую под общей суммой выражения (5), можно записать в виде

$$\sum_{k=1}^K S_k^2[x] + 2 \sum_{i=1}^{K-1} \sum_{j=i+1}^K S_i[x] S_j[x] = \sum_{i=1}^K \sum_{j=1}^K S_i[x] S_j[x]. \quad (6)$$

На рис. 2 приведен пример произведений ненулевых спектральных элементов декомпозиции S_k сигнала $f[x]$ в соответствии с выражением (6). Заметим, что произведения спектральных элементов декомпозиции $S_i S_j$ имеют протяженность наименьшего значения $\min(i, j)$ и размерность квадрата амплитуды A^2 . Извлекая корень из суммы произведений $S_i S_j$, выделенных по признаку равных протяженностей, получаем взаимно ортогональные формы спектральных элементов декомпозиции:

$$\bar{S}_i[x] = \sqrt{S_i^2[x] + 2 \sum_{j=i+1}^K S_i[x] S_j[x]}.$$

В силу ортогональности (выполняется равенство Парсеваля как обобщение теоремы Пифагора для n -мерного случая) обратное преобразование будет иметь вид

$$f[x] = \sqrt{\sum_{i=1}^K \bar{S}_i^2[x]}, \quad (7)$$

а эквивалент выражения (3) записывается в форме

$$\sum_X f^2[x] = \sum_X \left[\sum_{k=1}^K \bar{S}_k^2[x] \right] \quad (8)$$

для $\bar{S}_i[x]$ и в форме

$$\sum_{R^n} f^2[R^n] = \sum_{R^n} \left[\sum_{k=1}^K \bar{S}_k^2[R^n] \right]. \quad (9)$$

для пространства R^n . Отметим, что выражение (9) для преобразования по БПВ можно рассматривать как аналог уравнения Парсеваля для преобразования Фурье [5–7].

Рассмотрим в качестве примера переход к ортогональной форме $\bar{S}_i[x]$ для сигнала $f[x]$ на рис. 2. Вначале производится расчет множества спектральных элементов декомпозиции $S = \langle S_5, S_3, S_2, S_1 \rangle$ в соответствии с (1), подробное описание декомпозиции можно найти в [3, 4]. Далее формируются элементы произведений $S \times S$ (упорядоченные пары), для которых, впрочем, выполняется условие коммутативности $S_i \times S_j = S_j \times S_i$. Результат произведений отображен в центральной части рис. 2. Произведем запись значений амплитуд спектральных элементов декомпозиции в ортогональной форме для 4-го отсчета (отсчет производится слева, начиная с нуля по оси x , шкала отображена на графике $f[x]$): $\bar{S}_5[4] = \sqrt{2^2} = 2$, $\bar{S}_3[4] = \sqrt{3^2 + 2 \cdot 3 \cdot 2} = \sqrt{21}$, $\bar{S}_2[4] = \sqrt{1^2 + 2 \cdot 1 \cdot 2 + 2 \cdot 1 \cdot 3} = \sqrt{11}$, $\bar{S}_1[4] = \sqrt{2^2 + 2 \cdot 2 \cdot 2 + 2 \cdot 2 \cdot 3 + 2 \cdot 2 \cdot 1} = 2\sqrt{7}$. Результаты расчетов $\bar{S}_i[x]$ представлены на рис. 2 (столбец справа). Очевидно, что амплитуды исходного сигнала по значениям $\bar{S}_i[x]$ могут быть восстановлены в соответствии с выражением (7). Далее рассчитаем полную энергию E в соответствии с (8): $\sum_X \left[\sum_{k=1}^K \bar{S}_k^2[x] \right] = 5 \cdot 4 + 3 \cdot 21 + 2 \cdot 11 + 1 \cdot 28 = 133$, вычисление по выражению (3) дает результат $\sum_X f^2[x] = 0^2 + 2^2 + 5^2 + 6^2 + 8^2 + 2^2 + 0^2 + 0^2 = 133$.

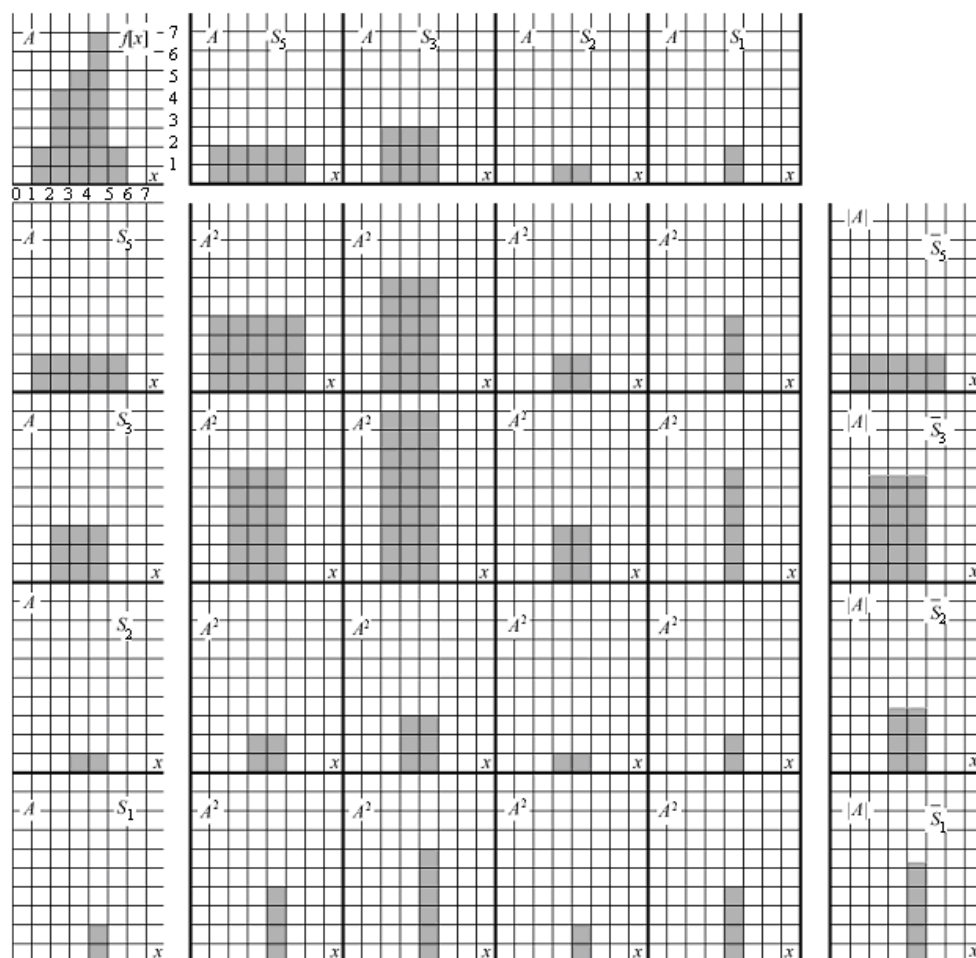


Рис. 2. Получение ортогональной формы декомпозиции по БПВ:
 $f[x]$ – исходный сигнал с амплитудой A ; S_k – спектральный элемент декомпозиции, полученный
 выделением всплесков протяженностью k

Порядок вычислительной сложности алгоритма прямого преобразования по БПВ для одномерного случая можно оценить как $O(N)$, где N – размер массива данных.

Для прямого преобразования по БПВ в n -мерном пространстве R^n сигнала $f[R^n]$, ограниченного размерами пространства, в котором задан сигнал $X_1 \cdot X_2, \dots, \cdot X_n$, порядок вычислительной сложности можно оценить как $O(n \cdot X_1 \cdot X_2, \dots, \cdot X_n)$.

Отметим, что все преобразование в соответствии с (1) и (2) может быть выполнено на кольце целых чисел, что обеспечивает высокое быстродействие и достаточно простую реализацию способа преобразования по БПВ полностью аппаратными средствами.

Заключение

В работе показаны свойства преобразования по базису прямоугольных всплесков – линейность, сохранение энергии сигнала, ортогональная форма преобразования. Рассмотрен ряд примеров, выполнена оценка вычислительной сложности. Приведены выражения, готовые к непосредственному применению в практических вычислениях.

Литература

1. Чобану М. Многомерные многоскоростные системы обработки сигналов. – М.: Техносфера, 2009. – 480 с.
2. Шарк Г.Г. Применение вейвлетов для ЦОС. – М.: Техносфера, 2007. – 192 с.
3. Гришенцев А.Ю., Коробейников А.Г. Декомпозиция n -мерных цифровых сигналов по базису прямоугольных всплесков // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 4 (80). – С. 75–79.

4. Заявка на изобретение. Способ построения спектра n -мерных неразделимых цифровых сигналов. Гришенцев А.Ю., Коробейников А.Г. № 2011126856, от 29.06.2011.
5. Смит С. Цифровая обработка сигналов. Практическое руководство для инженеров и научных работников. – М.: Додека-XXI, 2011. – 720 с.
6. Оппенгейм А., Шафер Р. Цифровая обработка сигналов. – 2-е изд. – М.: Техносфера, 2009. – 856 с.
7. Лайонс Р. Цифровая обработка сигналов. – 2-е изд. Пер. с англ. – М.: Бином пресс, 2009. – 656 с.

Гришенцев Алексей Юрьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, tigerpost@ya.ru

УДК 004.75

ОПТИМИЗАЦИЯ РАСПРЕДЕЛЕНИЯ ЗАПРОСОВ В СИСТЕМЕ КЛАСТЕРОВ ПРИ СОЧЕТАНИИ АНАЛИТИЧЕСКОГО И ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

И.Ю. Голубев, В.А. Богатырев

Предложена многоэтапная процедура оптимизации распределения потока запросов между кластерами вычислительной системы, использующая аналитическое и имитационное моделирование. Процедура позволяет найти оптимальную долю перераспределяемого потока запросов при различных законах распределения интервалов между поступающими в систему запросами и времени их обслуживания.

Ключевые слова: распределение нагрузки, имитационное моделирование, кластер, оптимизация.

Введение

Основными требованиями, предъявляемыми к распределенным вычислительным системам, являются их надежность, отказоустойчивость и производительность [1]. Высокая отказоустойчивость и производительность распределенных систем достигается в результате эффективного распределения запросов (нагрузки) между их узлами [2–8]. В распределенных вычислительных системах, объединяющих множество кластеров, перераспределение запросов может осуществляться между узлами как одного, так и различных кластеров, соединенных через сеть. Во втором случае увеличиваются издержки на межмашинный обмен, но возрастают возможности балансировки загрузки и сохранения работоспособности при накоплении отказов, что обуславливает актуальность оптимизации процесса распределения запросов и разработки соответствующих процедур оптимизации.

Λ

Постановка задачи

Цель представленной работы – разработка процедуры оптимизации распределения запросов между кластерами вычислительной системы при различных законах распределения интервалов между поступающими в систему запросами и времени их обслуживания.

Структура исследуемой распределенной вычислительной системы кластеров представлена на рис. 1.

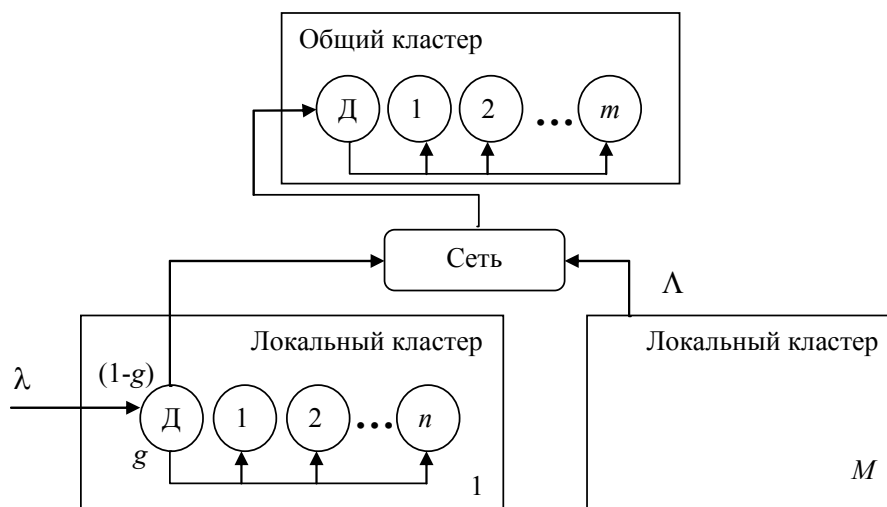


Рис. 1. Структура распределенной вычислительной системы

В системе имеется M локальных кластеров (по n серверов в каждом) и группа из m общедоступных серверов (общий кластер), обеспечивающая возможность адаптации системы к перегрузкам отдельных локальных кластеров в случае отказов входящих в их состав серверов или к возрастанию потока запросов к локальным кластерам. Компьютеры кластеров связывает сеть, включающая N резервированных коммутационных узлов (маршрутизаторов или коммутаторов). Распределение потока запросов осуществляется диспетчерами (Д), направляющими запросы на выполнение внутри кластера или через сеть в общий кластер.

Ставится задача оптимизации процесса распределения потока запросов между кластерами при различных законах распределения интервалов между поступающими в систему запросами и времени их обслуживания. В результате оптимизации распределения запросов для заданной структуры системы требуется найти их долю g_i , перераспределяемую через сеть в общий кластер, при которой достигается минимум времени пребывания запросов T в системе кластеров. Поиск проводится для заданных вариантов значений интенсивностей запросов (λ_i) и их вероятностей (b_i).

Оптимизация процесса распределения запросов

При проектировании вычислительных систем применяется как аналитическое, так и имитационное моделирование. Результаты аналитического моделирования при законах распределения интервалов между поступающими в систему запросами и времени их обслуживания общего вида могут иметь существенную погрешность, а имитационное моделирование не ориентировано на решение оптимизационных задач и требует значительного времени и ресурсов компьютера для проведения имитационных экспериментов [9].

Для решения оптимизационной задачи, если законы распределения интервалов между поступающими в систему запросами и времени их обслуживания – не экспоненциальные, предлагается комбинированный подход, предполагающий совместное применение аналитических и имитационных моделей.

В рамках комбинированного подхода разработана процедура оптимизации распределения запросов, включающая следующие этапы:

- предварительное определение оптимальной доли перераспределяемых через сеть запросов в предположении простейшего потока запросов и экспоненциального распределения времени обслуживания с использованием аналитического моделирования;
- уточнение результатов оптимизации на основе проведения имитационных экспериментов в области значений, полученных в ходе аналитического моделирования.

Если законы распределения интервалов между поступающими в систему запросами и времени их обслуживания известны, то эксперименты проводятся в условиях соответствующих законов распределения; в противном случае эксперименты проводятся при варьировании законов распределения нагрузочных параметров; решение определяется по среднему результату или по известным критериям принятия решений.

Аналитическая модель системы с перераспределением запросов

При оптимизации на основе аналитического моделирования воспользуемся результатами работы [4], в которой показана эффективность динамической оптимизации процесса перераспределения запросов между кластерами вычислительной системы.

Критерий оптимальности определен как

$$T = \min_{(g_i)} \sum_{i=0}^h b_i T(g_i, \lambda_i),$$

$$T(g_i, \lambda_i) = g_i \left(\frac{v_0}{1 - g_i \lambda_i v_0 / n} \right) + (1 - g_i) \left(\frac{2v_1}{1 - ((1 - g_i) + \beta) 2\lambda_i v_1 / N} + \frac{v_2}{1 - ((1 - g_i) + \beta) \lambda_i v_2 / m} \right).$$

Здесь h – число возможных значений интенсивности входного потока; v_0, v_1, v_2 – средние времена выполнения запросов в серверах локального кластера, в коммутационных узлах и в серверах общего кластера. Нагрузка общедоступного кластера от множества локальных кластеров моделировалась потоком запросов с интенсивностью $\Lambda_i = \beta \lambda_i$, N – число (кратность резервирования) коммутационных узлов в сети, через которые возможно перераспределение запросов.

Оптимизация проводилась для $i = 0, 1, \dots, h$ и при условии стационарности:

$$(g_i \lambda_i v_0 / n < 1) \wedge (((1 - g_i) + \beta) 2\lambda_i v_1 / N < 1) \wedge (((1 - g_i) + \beta) \lambda_i v_2 / m < 1).$$

В предположении простейшего потока запросов и экспоненциального распределения времени обслуживания определяется зависимость значений доли перераспределяемых через сеть запросов g_i от интенсивности запросов λ_i , при которой среднее время пребывания запросов в системе минимально [4].

В результате оптимизации на аналитической модели найден [4] вектор значений доли перераспределяемых запросов (0,721; 0,494; 0,456; 0,446; 0,446; 0,451).

Построение имитационной модели

На втором этапе оптимизации в среде AnyLogic 6 построена имитационная модель рассматриваемой системы, изображенная на рис. 2.

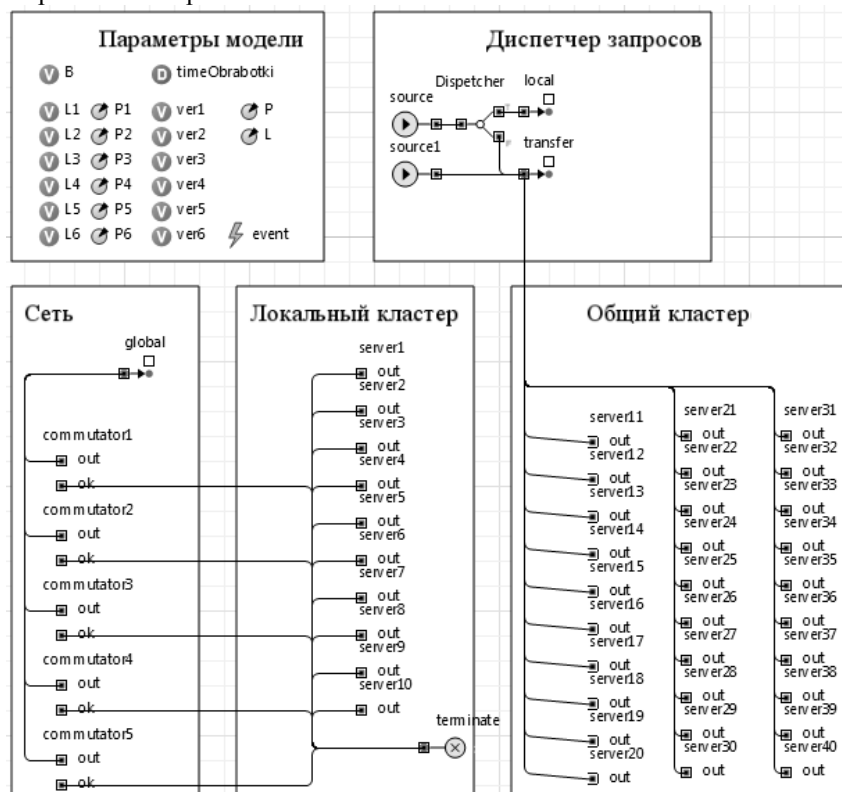


Рис. 2. Создание имитационной модели

Серия оптимизационных экспериментов проведена в области значений, полученных в результате аналитического моделирования, при следующих законах распределения интервалов между поступающими в систему запросами: экспоненциальный закон; равномерный закон; закон Эрланга 2-го порядка; гиперэкспоненциальный закон (с коэффициентом вариации 1,202).

Оптимизация проведена для $n = 10$ шт., $N = 5$ шт., $m = 30$ шт.; $v_0 = 10$ с, $v_1 = 1$ с, $v_2 = 10$ с, $\beta = 1$. Варианты возможных значений интенсивностей запросов и их вероятности представлены векторами (0,1; 0,3; 0,5; 0,7; 0,9; 1,1) и (0,1; 0,1; 0,15; 0,15; 0,2; 0,3), а вероятности соответствия закона распределения интервалов между поступающими в систему запросами моделируемому закону – (0,2; 0,25; 0,25; 0,3).

Сравнение аналитической и имитационной моделей

Результаты анализа моделей при экспоненциальном распределении интервалов между поступающими в систему запросами и времени их обслуживания представлены на рис. 3, а, б. Кривые 1 соответствуют интенсивности входного потока $0,1 \text{ с}^{-1}$, кривые 2 – $0,3 \text{ с}^{-1}$, кривые 3 – $0,5 \text{ с}^{-1}$, кривые 4 – $0,7 \text{ с}^{-1}$, кривые 5 – $0,8 \text{ с}^{-1}$. Из представленных графиков видно, что существует оптимальное значение доли запросов (g), перераспределяемых через сеть на выполнение в общий кластер, и для построенных моделей оно совпадает.

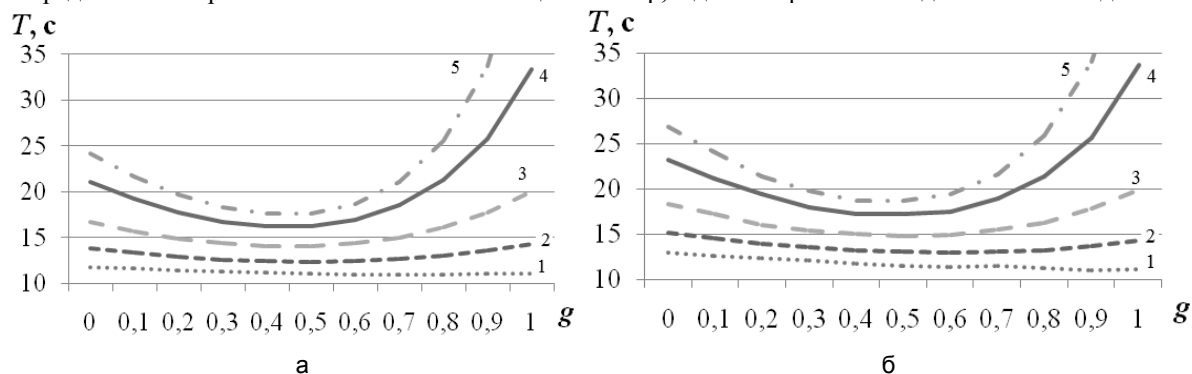


Рис. 3. Среднее время пребывания запроса в системе: аналитическая модель (а) и имитационная модель (б). Кривые 1–5 соответствуют интенсивности входного потока $0,1$; $0,3$; $0,5$; $0,7$; $0,8 \text{ с}^{-1}$.

Результаты второго этапа оптимизации представлены в таблице. Согласно критерию Гермейера, оптимальным является решение в строке 4.

№	Вектор значений доли перераспределяемых запросов (g_i)	Среднее время пребывания запросов в системе T , с				
		Закон распределения интервалов между поступающими в систему запросами				Результат (критерий Гермейера)
		Экспоненциальный	Равномерный	Эрланга	Гиперэкспоненциальный	
1	(0,716; 0,508; 0,454; 0,453; 0,447; 0,464)	20,985	20,748	20,786	21,433	6,4299
2	(0,72; 0,484; 0,464; 0,47; 0,445; 0,458)	21,101	20,722	20,917	21,378	6,4134
3	(0,738; 0,576; 0,452; 0,455; 0,457; 0,459)	21,013	20,807	20,76	21,369	6,4107
4	(0,724; 0,505; 0,485; 0,467; 0,468; 0,475)	21,052	20,739	20,883	21,367	6,4101

Таблица. Результаты второго этапа оптимизации

На рис. 4 показаны отклонения ε результатов, полученных в ходе имитационных экспериментов, от результатов аналитического моделирования: кривая 1 соответствует серии экспериментов для экспоненциального закона распределения интервалов между поступающими в систему запросами; кривая 2 – для равномерного закона; кривая 3 – для закона Эрланга, кривая 4 – для гиперэкспоненциального закона.

Максимальное отклонение результатов аналитического и имитационного моделирования при рассмотренных законах распределения интервалов между поступающими в систему запросами на всей рассматриваемой области значений интенсивности входящего потока запросов составило 16,6%, а среднее отклонение – 2,9% (0,082 и 0,015 в абсолютных значениях соответственно). Сужение области поиска оптимального значения до двух максимальных отклонений (0,164) на этапе уточнения результатов позволило сократить время, затраченное на имитационное моделирование, в 6 раз.

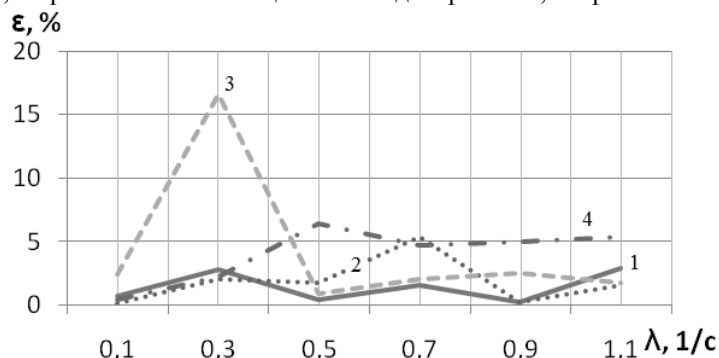


Рис. 4. Отклонения результатов имитационного моделирования. Кривые 1–4 соответствуют экспоненциальному, равномерному, эрланговскому, гиперэкспоненциальному законам распределения интервалов между поступающими в систему запросами

Заключение

Предложена процедура оптимизации процесса распределения потока запросов между кластерами вычислительной системы при различных законах распределения интервалов между поступающими в систему запросами и времени их обслуживания.

Предложенная процедура оптимизации предусматривает выполнение этапа оптимизации на основе аналитического моделирования для простейшего входного потока и экспоненциального распределения времени обслуживания и уточнение результатов моделирования на основе имитационных экспериментов при реальных законах распределения интервалов между поступающими в систему запросами и длительности их обслуживания.

Литература

1. Таненбаум Э., Ван Стеен М. Распределенные системы. Принципы и парадигмы. – СПб: Питер. – 2003. – 877 с.
2. Gaeta M., Kononov M., Shorgin S. Development of mathematical models and methods of task distribution in distributed computing system // Reliability: Theory & Applications. – 2006. – V. 1. – № 4. – P. 16–21.

3. Богатырев В.А., Богатырев С.В. К анализу и оптимизации серверных систем кластерной архитектуры с балансировкой нагрузки // Приборы и системы. Управление, контроль, диагностика. – 2010. – № 2. – С. 4–9.
4. Bogatyrev V.A., Bogatyrev S.V., Golubev I.Yu. Optimization and the Process of Task Distribution between Computer System Clusters // Automatic Control and Computer Sciences. – 2012. – V. 46. – № 3. – P. 103–111.
5. Богатырев В.А. К повышению надежности вычислительных систем на основе динамического распределения функций // Изв. вузов. Приборостроение. – 1981. – С. 62–64.
6. Богатырев В.А. Децентрализованное динамическое распределение запросов в многомашинных вычислительных системах // Электронное моделирование. – 1994. – Т. 16. – № 3. – С. 38.
7. Богатырев В.А., Богатырев С.В. Критерии оптимальности многоустойчивых отказоустойчивых компьютерных систем // Научно-технический вестник СПбГУ ИТМО. – 2009. – № 5 (63). – С. 92–97.
8. Богатырев В.А. Оптимальное резервирование системы разнородных серверов // Приборы и системы. Управление, контроль, диагностика. – 2007. – № 12. – С. 30–36.
9. Алиев Т.И. Основы моделирования дискретных систем: Учебное пособие. – СПб: СПбГУ ИТМО. – 2009. – 363 с.

Голубев Иван Юрьевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, www.golubev@mail.ru

Богатырев Владимир Анатольевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, Vladimir.bogatyrev@gmail.com

УДК 629.7.017

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОЦЕНКИ НАДЕЖНОСТИ ИЗДЕЛИЯ АВИОНИКИ С «ХОЛОДНЫМ» РЕЗЕРВОМ

О.А. Кузнецова

Построена математическая модель оценки безотказности изделия авионики, представленного в виде нагруженного дублирования с дополнительным «холодным» резервом. Модель учитывает неполный контроль работоспособности входящих устройств и время задержки во включении «холодного» устройства. Выполнен сравнительный анализ вариантов структурного резервирования.

Ключевые слова: надежность, безотказность, «холодный» резерв, контроль работоспособности.

Введение

Учитывая требования по отказобезопасности гражданской авиации в части недопущения возникновения на борту самолета сложной ситуации, показатель отказобезопасности аппаратуры самолета (вероятность отказа за 1 ч полета) должен быть не выше 10^{-5} . Для обеспечения указанного требования, учитывая современные показатели безотказности авионики, в проектных решениях структур изделий необходимо предусматривать не менее чем двукратное резервирование. Структурное резервирование в простейшем варианте может быть реализовано в виде постоянно включенной группы изделий, резервирующих друг друга по выполняемым функциям, либо с применением «холодного» резерва. В обоих вариантах резервирования равной кратности резервированные группы состоят из одного основного и двух резервирующих изделий, отличающихся тем, что в одной из групп второе резервное изделие до момента возникновения первого отказа находится в выключенном состоянии, т.е. используется в режиме «холодного» резерва (рис. 1). Применение избыточности как способа повышения надежности имеет широкое применение при проектировании сложных технических систем [1–4]. Учитывая, что за время полета ремонт возникших отказов авионики не может быть выполнен сразу после обнаружения, на первый взгляд, для рассматриваемой резервированной группы с «холодным» резервом необходимо было бы воспользоваться моделью оценки надежности невосстанавливаемых систем из справочников [1, 2]. Но эта модель позволяет оценить надежность системы без учета того, что у основного устройства, помимо «холодного» резерва, имеется нагруженный резерв. При попытке заменить группу с нагруженным резервом на нерезервированное устройство с аналогичным показателем безотказности возникает трудность в использовании предлагаемой модели, так как интенсивности отказов входящих устройств становятся неравными (что противоречит исходным условиям применения модели [1, 2]). Наличие же задержки во включении «холодного» резерва ведет к еще большему затруднению, так как в этом случае предлагается перейти к рассмотрению восстанавливаемой системы без резерва, у которой время восстановления равно времени переключения на резерв. Учитывая изложенные недостатки имеющейся модели, для анализа применения «холодного» резерва возникла необходимость построения более точной модели оценки вероятности отказа группы с «холодным» резервом, учитывающей время задержки включения в работу «холодного» резерва и полноту контроля работоспособности устройств в резервных группах, так как включение «хо-

лодного» резерва в работу осуществляется только в том случае, если возникший в нагруженных устройствах отказ обнаружен средствами контроля.

Рассматриваемые варианты структурного резервирования

Первый вариант структурного построения резерва (рис. 1, а) состоит из дублированной группы идентичных изделий с неполным контролем работоспособности и дополнительного изделия, используемого в режиме «холодного» резерва. Для включения «холодного» резерва в работу требуется время подготовки, равное Δt . Второй вариант структурного построения (рис. 1, б) состоит из группы двукратного резервирования идентичных изделий с неполным контролем работоспособности.

В дублированной группе первое устройство функционирует приоритетно, второе выполняет функции управления после контролируемого отказа первого. До отказа приоритетного изделия второе находится во включенном состоянии. После возникновения отказа в одном из включенных изделий и обнаружения его средствами контроля подается сигнал о включении изделия, находящегося в ненагруженном «холодном» резерве. Контроль работоспособности изделий, как правило, основывается на результатах [3]:

- контроля отдельного устройства собственными встроенными программно-аппаратными средствами, полнота контроля при этом ограничивается возможностями внутренних программных и аппаратных средств;
- функционального контроля – контроля на уровне изделия в целом при включении специальных режимов контроля или при непосредственном рабочем функционировании;
- органолептического контроля со стороны экипажа или обслуживающего персонала. Результаты данного контроля могут повысить надежность только в том случае, если переключение между резервом может осуществляться оператором.

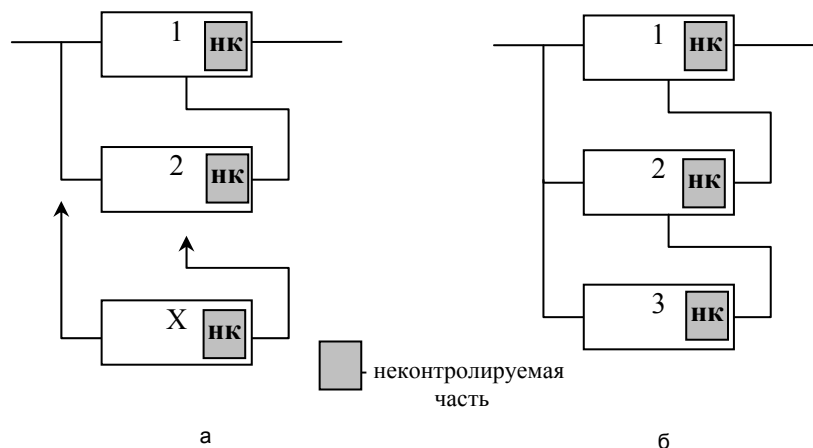


Рис. 1. Рассматриваемые варианты структурного резервирования: нагруженное дублирование с дополнительным «холодным» резервом (а); нагруженное двукратное резервирование (б). 1 – первое нагруженное изделие; 2 – второе нагруженное изделие; 3 – третье нагруженное изделие, X – «холодный» резерв

Построение модели расчета безотказности

Первая модель. Для формализации вероятности отказа рассматриваемой структуры с «холодным» резервом все возможные состояния отказа представлены схематично в виде временных диаграмм на рис. 2–5.

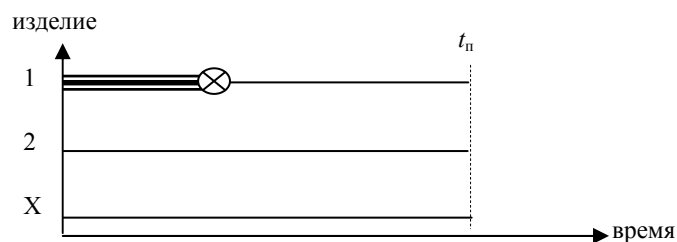


Рис. 2. Временная диаграмма, изображающая отказ основного изделия в неконтролируемой части

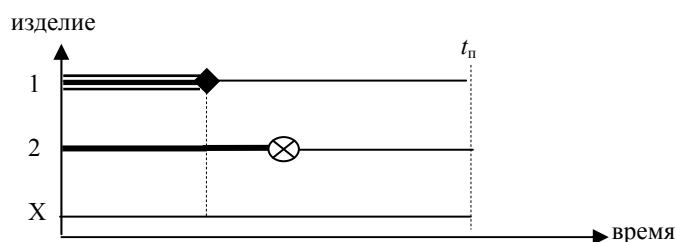


Рис. 3. Временная диаграмма, изображающая отказ в неконтролируемой части резервного изделия и контролируемый отказ основного изделия

Рис. 2 отображает случай с отказом основного, приоритетного изделия в неконтролируемой части: из-за неполного контроля отказов переход на резерв не осуществится, и, хотя резерв может в это время быть в работоспособном состоянии, группа будет отказавшей. Похожая ситуация представлена на рис. 3, но теперь отказ произошел в неконтролируемой части второго изделия. При этом отказ второго изделия скажется на отказе всей группы только после того, как второе изделие приняло или должно было принять функции управления (т.е. после контролируемого отказа первого изделия).

Ситуации, представленные на рис. 4, ведут к отказу резервированной группы из-за недостаточности времени между первым и вторым отказом для включения «холодного» резерва.

На рис. 5 представлены ситуации, когда за время до окончания полета произойдут отказы всех трех изделий резервированной группы в следующей последовательности: контролируемый отказ первого изделия, затем контролируемый отказ второго и включенного «холодного» резерва (рис. 5, а), или контролируемый отказ второго изделия, затем контролируемый отказ первого и отказ включенного «холодного» резерва (рис. 5, б).

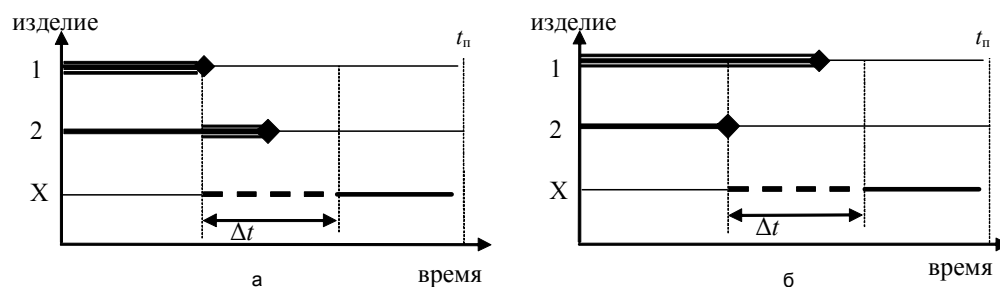


Рис. 4. Временные диаграммы, изображающие отказ резервного изделия после контролируемого отказа основного изделия за время, меньшее Δt (а), и отказ основного изделия после контролируемого отказа резервного изделия за время, меньшее Δt (б)

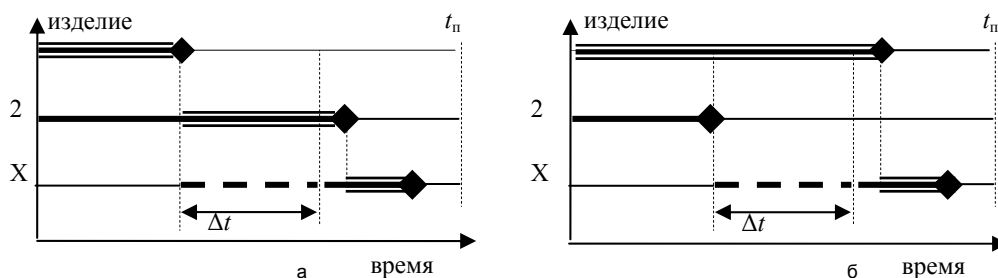


Рис. 5. Временные диаграммы, изображающие отказ включенного «холодного» резерва после контролируемого отказа первого, а затем и контролируемого отказа второго изделия (а), а также после контролируемого отказа второго, а затем и контролируемого отказа первого изделия (б)

Формализация вероятности отказа за время полета резервированной группы, возможные состояния отказов которой представлены на рис. 2–5, при условии, что в полет группа отправляется с полным резервом, а вероятность безотказной работы переключающих элементов и элементов, включающих «холодное» изделие, равна 1 (для периода нормальной эксплуатации) имеет вид суммы вероятностей:

$$Q_{cx,x}(t_n) = Q_{cx2}(t_n) + Q_{cx3}(t_n) + Q_{cx4a}(t_n) + Q_{cx4b}(t_n) + Q_{cx5a}(t_n) + Q_{cx5b}(t_n). \quad (1)$$

Здесь $Q_{cx2}(t_n)$, $Q_{cx3}(t_n)$, $Q_{cx4a}(t_n)$, $Q_{cx4b}(t_n)$, $Q_{cx5a}(t_n)$, $Q_{cx5b}(t_n)$ – вероятности возникновения ситуаций, схематично представленных на рис. 2–5 соответственно:

$$Q_{cx2}(t_n) = Q_{1нк}(t_n)P_{1к}(t_n), \quad (2)$$

$$Q_{cx3}(t_n) = P_{1нк}(t_n)Q_{1к}(t_n)Q_{2нк}(t_n)P_{2к}(t_n), \quad (3)$$

$$Q_{cx4a}(t_n) = \int_0^{t_n} P_{1нк}(\tau) P_{2нк}(\tau + \Delta t) \frac{d[Q_{1к}(\tau)]}{d\tau} \int_{\tau}^{\tau + \Delta t} \frac{dQ_{2к}(f)}{df} df d\tau, \quad (4)$$

$$Q_{cx4б}(t_n) = \int_0^{t_n} P_{2нк}(\tau) P_{1нк}(\tau + \Delta t) \frac{d[Q_{2к}(\tau)]}{d\tau} \int_{\tau}^{\tau + \Delta t} \frac{dQ_{1к}(f)}{df} df d\tau, \quad (5)$$

$$Q_{cx5a}(t_n) = \int_0^{t_n - \Delta t} Q_x(t_n - \tau) P_{1нк}(\tau) P_{2нк}(\tau + \Delta t) \frac{d[Q_{1к}(\tau)]}{d\tau} \int_{\tau + \Delta t}^{t_n} \frac{dQ_{2к}(f)}{df} df d\tau, \quad (6)$$

$$Q_{cx5б}(t_n) = \int_0^{t_n - \Delta t} Q_x(t_n - \tau) P_{2нк}(\tau) P_{1нк}(\tau + \Delta t) \frac{d[Q_{2к}(\tau)]}{d\tau} \int_{\tau + \Delta t}^{t_n} \frac{dQ_{1к}(f)}{df} df d\tau. \quad (7)$$

$P_i(t)$, $P_{ик}(t)$, $P_{инк}(t)$ – вероятность безотказной работы i -го изделия, контролируемой части и неконтролируемой части i -го изделия соответственно за время t . Индекс $i = 1, 2, 3, x$ используется для обозначения в резервированных структурах изделий в соответствии с рис. 1, а.

$Q_i(t)$, $Q_{ик}(t)$, $Q_{инк}(t)$ – вероятность отказа i -го изделия, отказа в контролируемой и неконтролируемой частях соответственно за время t .

$$P_i(t) = \exp(-\lambda_i t), \quad Q_i(t) = 1 - \exp(-\lambda_i t),$$

$$P_{ик}(t) = \exp(-\omega_i \lambda_i t), \quad Q_{ик}(t) = 1 - \exp(-\omega_i \lambda_i t),$$

$$P_{инк}(t) = \exp(-[1 - \omega_i] \lambda_i t), \quad Q_{инк}(t) = 1 - \exp(-[1 - \omega_i] \lambda_i t),$$

где λ_i – интенсивность отказов i -го изделия; ω_i – полнота контроля i -го изделия.

Вторая модель. Для аналогичной структуры, но с включенным резервом (рис. 1, б), формула расчета вероятности отказа, учитывающая полноту контроля работоспособного состояния изделий, имеет вид [4]:

$$Q_{cx,н}(t_n) = Q_{1нк}(t_n) P_{1к}(t_n) + Q_{1к}(t_n) P_{1нк}(t_n) Q_{2нк}(t_n) P_{2к}(t_n) + Q_{1к}(t_n) P_{1нк}(t_n) Q_{2к}(t_n) P_{1нк}(t_n) Q_3(t_n). \quad (8)$$

Анализ полученных математических моделей

Для примера по формулам (1)–(8) рассчитаны и на рис. 6 представлены графики зависимостей вероятностей отказов сравниваемых структур от времени полета t_n .

Для расчета и построения графика на рис. 6 использовались следующие данные: $\lambda_i = 250 \times 10^{-6} 1/ч$; $\omega_i = 0,99$, $\Delta t = 0$.

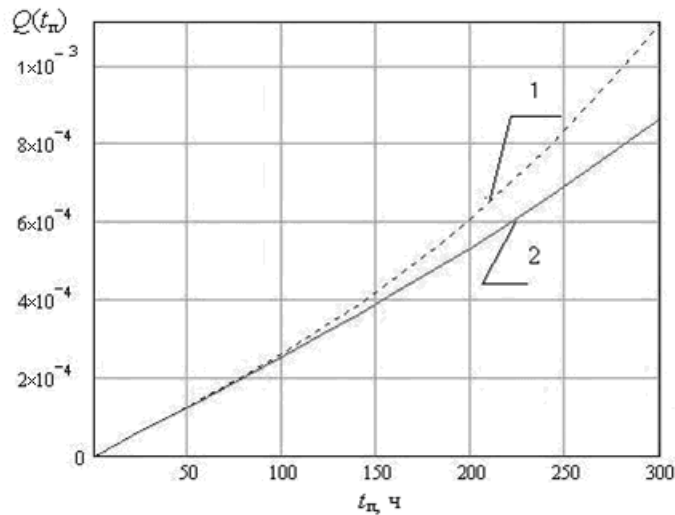


Рис. 6. Графики зависимости вероятностей отказов сравниваемых структур от времени полета t_n :
1 – вероятность отказа нагруженного двукратного резервирования (рис. 1, б, модель 2);
2 – вероятность отказа нагруженного дублирования с дополнительным «холодным» резервом (рис. 1, а, модель 1)

Согласно полученным результатам для изделия в двух вариантах построения на времени полета более 50 ч наблюдается постепенное расхождение кривых 1 и 2, при этом растет скорость ухудшения надежности изделия во втором варианте исполнения – с нагруженными резервами.

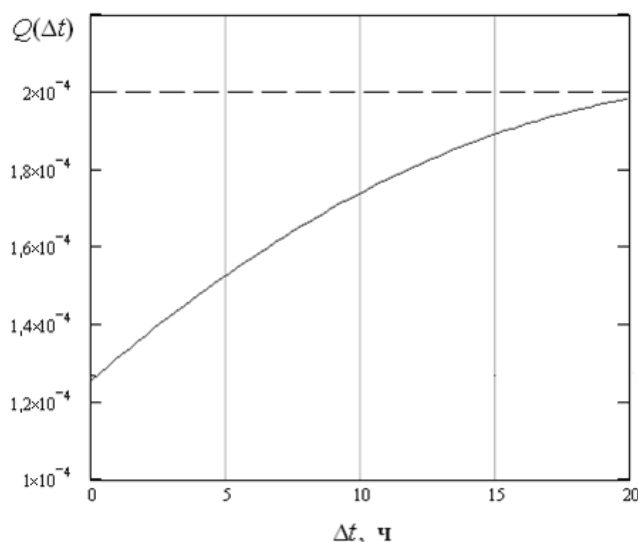


Рис. 7. Влияние значения времени задержки включения «холодного» резерва – Δt на вероятность отказа изделия, построенного с «холодным» резервом $Q(\Delta t)$

На рис. 7 представлена зависимость вероятности отказа изделия с «холодным» резервом от времени задержки во включении Δt (при $t_n = 50$ ч). Пунктиром на рисунке отмечено значение вероятности отказа обычной дублированной группы, к которому стремится значение вероятности отказа изделия с «холодным» резервом при увеличении времени Δt , т.е. с ростом Δt эффективность введенного «холодного» резерва полностью исчезает.

Заключение

Построена математическая модель оценки вероятности отказа изделия авионики, представленного в виде дублированной группы с дополнительным «холодным» резервом, учитывающая время задержки включения в работу «холодного» резерва и полноту контроля работоспособности входящих устройств.

Построенная модель позволит на начальных этапах разработки изделия авионики выполнить сравнительную оценку надежности возможных вариантов построения и принять решение о необходимом виде резервирования изделия, позволяющем обеспечить требования по надежности и отказобезопасности.

Литература

1. Козлов Б.А., Ушаков И.А. Справочник по расчету надежности аппаратуры радиоэлектроники. – М.: Советское радио, 1975. – 472 с.
2. Беляев Ю.К., Богатырев В.А., Болотин В.В. и др. Надежность технических систем. Справочник / Под ред. И.А. Ушакова. – М.: Радио и связь, 1985. – 608 с.
3. Кузнецова О.А. Оценка надежности структурно избыточных комплексов авионики с учетом среднего времени между восстановлениями при отказах // Изв. вузов. Приборостроение. – 2012. – № 3. – С. 65–70.
4. Кузнецова О.А., Гатчин Ю.А., Лобов В.В. Оценка надежности структурно избыточных изделий при проектировании сложных технических систем // Научно-технический вестник СПбГУ ИТМО. – 2009. – № 1 (59). – С. 44–52.

Кузнецова Ольга Александровна – ФГУП СПб ОКБ «Электроавтоматика», начальник подразделения, аспирант, kuzola@ya.ru

УДК 004.891

**АНАЛИЗ ВЛИЯНИЯ СОСТОЯНИЯ НАРКОТИЧЕСКОГО ОПЬЯНЕНИЯ
НА ХАРАКТЕРИСТИКИ ГОЛОСОВ ДИКТОРОВ****А.Н. Раев, Ю.Н. Матвеев, Т.И. Голощапова**

Проведен анализ фонограмм речи дикторов до и после лечения от наркотической зависимости, который показал отсутствие статистически значимых зависимостей искажения речевых характеристик диктора от разных групп наркотических средств и степени наркотического опьянения. Изменения основного тона не носят регулярного, общего для всех, характера. Главной причиной изменения основного тона является изменение эмоционального состояния диктора, а не выход из состояния наркотического опьянения. Исследование влияния длительности употребления наркотических средств на речь дикторов показали, что для дикторов, имеющих длительный стаж употребления наркотических средств героиновой группы, обнаружена тенденция к снижению основного тона примерно на 3% в год.

Ключевые слова: диктор, частота основного тона, наркотическое опьянение.

Введение

В большинстве известных работ состояние наркотического опьянения связывают с изменением эмоционального и функционального состояния человека.

В работе [1] проведен анализ общетеоретических и прикладных исследований по оценке влияния эмоциональных и функциональных состояний человека на акустические характеристики его речи. Показано, что в большинстве случаев в качестве наиболее информативных акустических коррелятов эмоциональных и функциональных состояний рассматривается ряд частотных, временных и мощностных характеристик голосового сигнала. Отмечено, что, как правило, стенические состояния ведут к возрастанию, а астенические – к понижению частот основного тона (ОТ) и формант, а также интенсивности. Под частотой ОТ понимается частота вибрации голосовых связок. У каждого говорящего базовая частота ОТ индивидуальна и обусловлена особенностями строения гортани. Для оценки частоты ОТ разработано множество различных алгоритмов [2]. Результаты настоящей работы получены при использовании оригинального алгоритма, разработанного в ООО «ЦРТ» [3].

В работе [4] отмечено, что в качестве характеристик голосового сигнала, являющихся индикаторами действия психотропных препаратов и антидепрессантов, могут использоваться, например, частота ОТ, распределение энергии в спектре, длительность пауз, которые изменяются при изменении состояния (настроения) пациентов в различные периоды их лечения.

В работе [5] проводился анализ следующих акустических параметров: среднее значение и стандартные отклонения звуковой мощности, частоты ОТ, частот первой и второй формант, длительность произнесения. В результате сравнения значений параметров каждой из рассматриваемых эмоциональных интонаций с соответствующими параметрами нейтральной интонации было обнаружено, что наиболее выраженными в процентном отношении были отклонения частот ОТ и первой форманты. Было показано, что направленность этих отклонений не зависит от семантического содержания речевого высказывания и его длительности, от возраста, пола говорящего, хотя на абсолютные значения частот индивидуальные особенности говорящих оказывали влияние. В работе [6] исследовано изменение границ частоты ОТ и смещение формантных частот голоса при воздействии на человека различных видов стрессов и эмоций.

Проблеме влияния состояния наркотического опьянения на характеристики голосов дикторов посвящено относительно малое число исследований, отсутствуют исследования зависимостей искажения речевых характеристик диктора от разных групп наркотических средств и степени наркотического опьянения, а также стажа употребления наркотических средств. Это в основном связано с отсутствием необходимых для проведения таких исследований речевых баз данных.

Для обеспечения проведения перечисленных исследований авторами была собрана специализированная база речевых данных образцов речи дикторов, находящихся в состоянии наркотического опьянения (далее – РБДН). Сбор РБДН осуществлялся ООО «ЦРТ» по заказу Федеральной службы Российской Федерации по контролю над оборотом наркотиков и проводился в ряде специализированных медицинских учреждений Санкт-Петербурга и Ленинградской области [7].

В настоящей работе приводятся результаты исследований влияния на частоту ОТ состояния наркотического опьянения на основе собранной РБДН. Результаты этих исследований важны для экспертной практики [7, 8], а также должны учитываться при разработке систем автоматической идентификации личностей по голосу [9, 10] и идентификации эмоционального состояния дикторов по записям их устной речи [11].

Описание базы образцов речи дикторов, находящихся в состоянии наркотического опьянения

При сборе РБДН выполнялась регистрация дикторов, при которой каждому диктору присваивался его личный порядковый номер, под которым в отчете делалась запись, содержащая следующую информацию:

- пол диктора;
- предположительная степень наркотического опьянения;
- предположительный тип наркотического средства, под воздействием которого он находится.

В качестве дикторов привлекались лица в состоянии наркотического опьянения. Определение состояния и степени наркотического опьянения осуществлялось специалистами-наркологами (врачами наркологической больницы, экспертами наркологических диспансеров).

Длительность произнесенного диктором речевого материала составляла не менее 32 с. Каждого диктора записывали либо один раз (в состоянии наркотического опьянения), либо два раза (один раз в состоянии наркотического опьянения, второй раз – после выхода из состояния наркотического опьянения). Запись образцов речи дикторов производилась на цифровой диктофон, находящийся на расстоянии 1–2 м от губ диктора, под наблюдением специально обученных операторов, в соответствии с инструкцией оператору по записи фонограмм.

После завершения записи записанный речевой материал полностью прослушивался и сегментировался в звуковом редакторе. Из записей удалялась речь оператора, паузы и внешние шумы.

Общее количество фонограмм в РБДН – 156 (89 мужчин и 67 женщин), в которой представлены образцы речи дикторов, имеющих следующие степени наркотического опьянения (табл. 1): легкая, средняя, тяжелая. Степень опьянения определялась экспертно. В РБДН также представлены образцы речи дикторов, находящихся в состоянии наркотического опьянения под действием следующих наркотических средств (табл. 1): растительных (анаша, гашиш), полусинтетических (героин), синтетических (амфетамин). Кроме того, в РБДН представлены дикторы разного пола и с разным стажем употребления наркотических средств (табл. 2).

	Степень наркотического опьянения			Тип наркотического средства		
	легкая	средняя	тяжелая	растительные (анаша, гашиш)	полусинтетические (героин)	синтетические (амфетамин)
Количество дикторов	40	64	52	10	141	5

Табл. 1. Распределение количества дикторов по степени наркотического опьянения и типу употребляемых наркотических средств

	Стаж употребления наркотических средств, лет				
	1	2	5	10	25
Количество дикторов	66	21	24	34	11

Табл. 2. Распределение количества дикторов по стажу употребления наркотических средств

Анализ фонограмм дикторов, находящихся в состоянии наркотического опьянения

На основе собранной РБДН был проведен анализ фонограмм на наличие зависимости искажений речевых характеристик диктора от степени наркотического опьянения, группы наркотических средств и длительности употребления наркотических средств. Результаты анализа приведены на рис. 1–5, где по вертикальной оси отложена относительная частота встречаемости частот ОТ в фонограмме диктора, по горизонтальной оси – частота ОТ в Гц. Гистограммы нормированы таким образом, чтобы площадь под каждой гистограммой была равна 1.

Исследование фонограмм речи дикторов до и после лечения от наркотической зависимости показывает, что речь дикторов изменяется незначительно и нерегулярным образом. На рис. 1 изображены усредненные гистограммы ОТ голосов четырех мужчин до и после лечения от наркотической зависимости. В среднем параметры гистограммы изменились мало: среднее значение и медиана распределения ОТ по частотам остались теми же, а дисперсия уменьшилась. Но это произошло, скорее всего, из-за изменения эмоционального состояния.

Аналогичная ситуация наблюдается и для женщин: мода распределения ОТ по частотам выросла с 180 Гц до почти 200 Гц, но среднее значение и медиана распределения почти не изменились. Статистически достоверных регулярных изменений не наблюдается. Это заставляет предположить, что главной

причиной изменения ОТ является изменение эмоционального состояния дикторов, а не процесс избавления от наркотической зависимости.

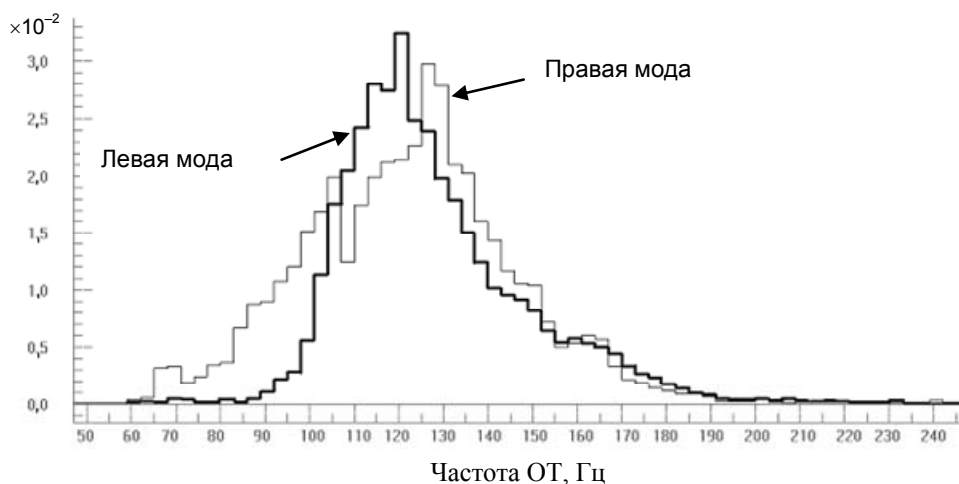


Рис. 1. Усредненные гистограммы ОТ голосов мужчин до (правая мода) и после (левая мода) лечения от наркотической зависимости. Шаг по частоте – 3 Гц

Было также проведено статистическое исследование влияния на ОТ следующих параметров:

- степени наркотического опьянения;
- типа наркотического вещества;
- длительности употребления наркотиков.

Поскольку, по заявлениям врачей, объективная методика определения степени наркотического опьянения в настоящее время отсутствует, была выполнена субъективная (экспертная) оценка степени наркотического опьянения и продолжительности употребления наркотиков.

Для определения степени наркотического опьянения, в первую очередь, рассматривались следующие факторы: нарушения речи («заплетающийся язык»); неспособность диктора понять, чего от него хотят врачи; пропуск или непонимание отдельных пунктов задания; собственное заявление диктора о том, что он недавно употреблял наркотики. Все эти признаки, включая заявление о недавнем употреблении наркотиков, являются лишь косвенными доказательствами того, что диктор находится в состоянии наркотического опьянения. В этом случае некоторые фонограммы, помеченные как принадлежащие дикторам, находящимся в состоянии наркотического опьянения, могут принадлежать трезвым дикторам, и наоборот. Тем не менее, выводы, сделанные по этим выборкам, будут достоверны в статистическом смысле.

Длительность употребления наркотиков оценивалась на основании следующих факторов:

- заявление диктора о том, что он состоит на учете в наркологическом диспансере с такого-то года,
- заявление диктора о том, что он употребляет наркотики с такого-то года,
- заявление диктора о том, что у него имеется гепатит В или С с такого-то года,
- заявление диктора о том, что у него имеется ВИЧ с такого-то года.

Все эти признаки позволяют лишь приблизительно оценить длительность употребления наркотиков. Многие дикторы многократно лечились от наркозависимости и у них, по их словам, были ремиссии (длительный полный отказ от употребления наркотиков), в том числе многолетние.

Почти все записи голосов дикторов, находящихся в состоянии тяжелого наркотического опьянения, – очень тихие, со средней амплитудой около 1–2 тысяч отсчетов. Однако такой параметр, как энергия, зависит гораздо больше от расстояния от микрофона до губ диктора, чем от состояния диктора. В связи с этим в дальнейшем данный параметр не рассматривался.

Гистограммы ОТ фонограмм голосов женщин приведены на рис. 2. На этих гистограммах никаких статистически достоверных аномалий не наблюдается. Медиана распределения ОТ находится в интервале 190–220 Гц, что вполне укладывается в допустимые пределы для нормальных женских голосов. С некоторой натяжкой аномальной можно считать одну гистограмму, которая имеет двухмодовое распределение с модами в областях 120 Гц и 225 Гц. Однако, по оценкам экспертов, несколько процентов женских голосов имеет двухмодовое распределение.

На рис. 3 приведены примеры гистограмм ОТ голосов 6 мужчин, находящихся в состоянии тяжелого наркотического опьянения. На этих гистограммах не видно никаких статистически достоверных аномалий. Медиана ОТ находится в интервале 110–130 Гц, что вполне укладывается в допустимые пределы для нормальных мужских голосов. Аномальной можно считать только одну гистограмму из шести, что не позволяет сделать статистически достоверных выводов.

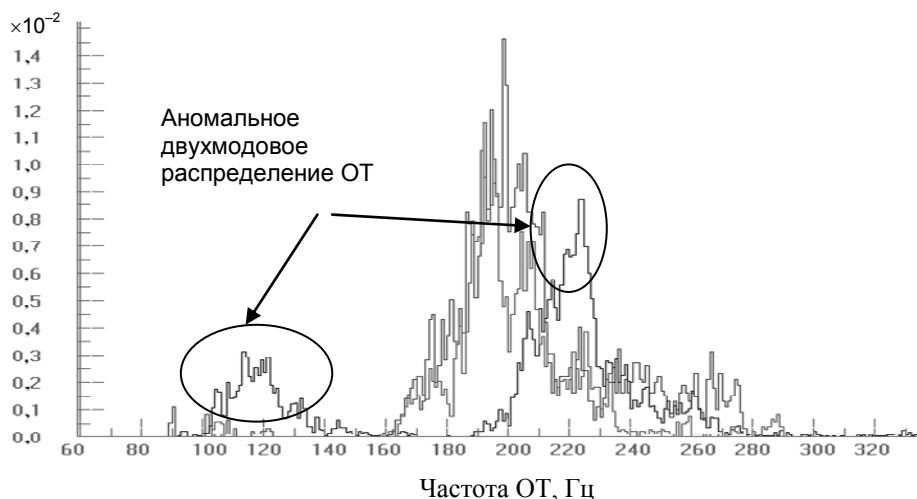


Рис. 2. Гистограммы частот ОТ фонограмм голосов женщин в состоянии тяжелого наркотического опьянения. Шаг по частоте – 1 Гц

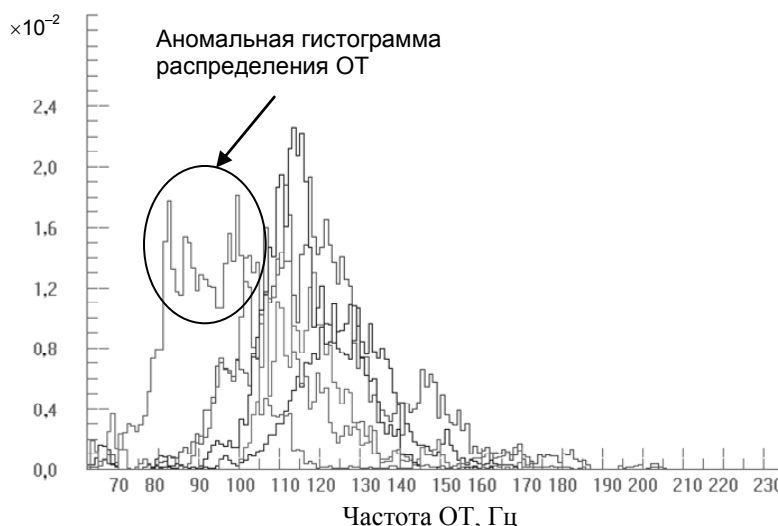


Рис. 3. Гистограмма частот ОТ фонограмм голосов 6 мужчин в состоянии тяжелого наркотического опьянения. Шаг по частоте – 1 Гц

Всего же из трех гистограмм ОТ голосов женщин и 22 гистограмм ОТ голосов мужчин, находящихся в состоянии тяжелого наркотического опьянения, обнаружены только одна аномальная гистограмма у женщин и 6 аномальных гистограмм у мужчин. Аномалия для мужчин состоит в низком среднем ОТ у мужчин (ниже 100 Гц), что редко встречается у мужчин, не принимающих наркотики (по оценкам экспертов – не более чем у 5% мужчин). Следовательно, состояние наркотического опьянения непосредственно не влияет на основной тон голоса.

Анализ аномалий, в то же время, обращает внимание на то, что из шести аномалий у мужчин, в четырех случаях она обнаруживается у дикторов, принимавших наркотики пять и более лет. Единственная аномалия у женщин принадлежит женщине, принимавшей наркотики в течение четырех лет. Это позволяет предположить другую причину аномалий, а именно – не состояние наркотического опьянения, а длительность употребления наркотиков.

На рис. 4 изображены гистограммы ОТ 15 фонограмм, в которых записаны голоса мужчин, употреблявших наркотики пять и более лет. Здесь наблюдается совершенно другая картина. Более чем у 2/3 дикторов основная мода статистически достоверно переместилась из области 120–130 Гц в область 95–100 Гц. У оставшихся 1/3 дикторов перемещение основной моды не наблюдается, что можно объяснить тем, что продолжительность употребления наркотиков известна не точно, а получена на основании косвенных данных. Кроме того, часть дикторов лечилась от наркотической зависимости, и у них наблюдались длительные ремиссии.

Если просуммировать все гистограммы, изображенные на рис. 4, то после нормировки получится картина, изображенная на рис. 5, из которой видно, что мода распределения ОТ находится на частоте примерно 97 Гц (см. положение курсора).

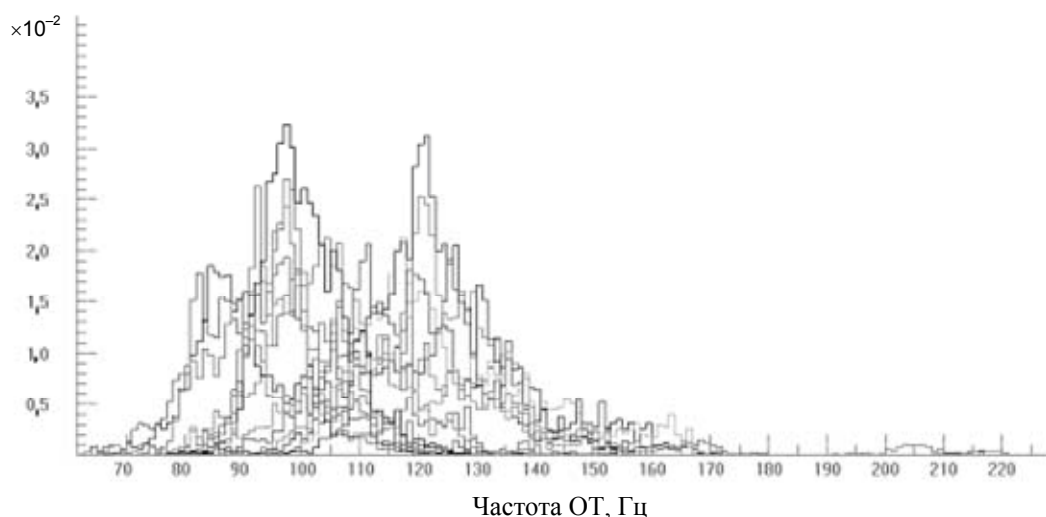


Рис. 4. Гистограммы частоты ОТ 15 фонограмм, в которых записаны голоса мужчин, употреблявших наркотики пять и более лет. Шаг по частоте – 1 Гц

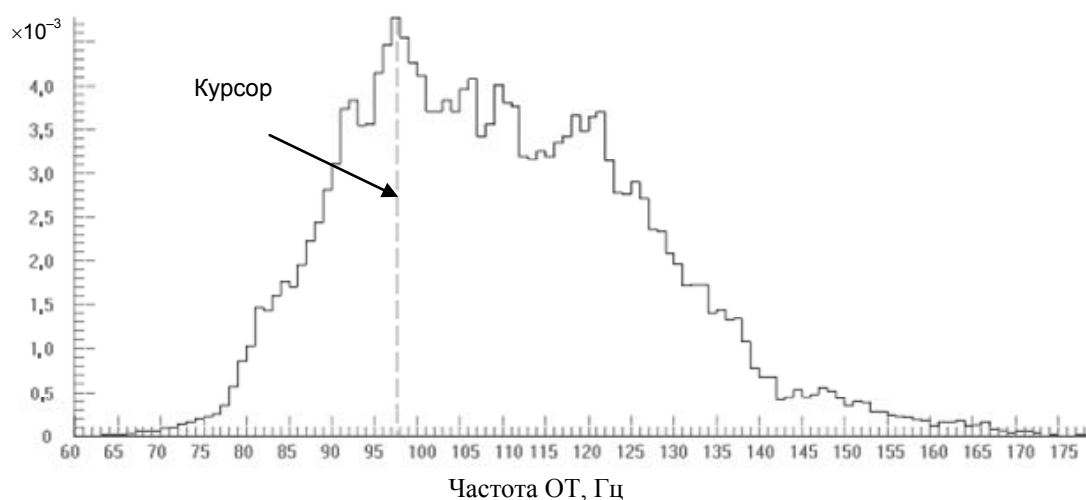


Рис. 5. Усредненная гистограмма частот ОТ 15 фонограмм, в которых записаны голоса мужчин, употреблявших наркотики пять и более лет. Шаг по частоте – 1 Гц

Все это позволяет сделать вывод, что длительное употребление наркотиков героиновой группы приводит к снижению основного тона.

Средняя скорость снижения может быть очень приблизительно оценена как 4–5 Гц (3%) в год. Линейная зависимость для оценки снижения высоты ОТ выбрана как простейшая. Ее применимость ограничивается 2–8-летним стажем интенсивного употребления наркотиков героиновой группы. Для исследования более сложных зависимостей и для оценки влияния еще большего стажа употребления наркотиков полученных данных недостаточно.

Аналогичный вывод по другим типам наркотиков сделан быть не может из-за недостатка данных. Скорость снижения высоты ОТ может быть несколько занижена и может быть неравномерной.

Анализ типа наркотика и продолжительности его употребления позволяет предположить, что наркоманы постепенно переходят на героин с более «легких» наркотиков.

Заключение

На основе собранной базы речевых данных, в которой представлены образцы речи дикторов, находящихся в состоянии различной степени наркотического опьянения, проведен анализ фонограмм речи дикторов до и после лечения от наркотической зависимости и исследование влияния длительности употребления наркотических средств на речь дикторов. Анализ показал отсутствие статистически значимых зависимостей «искажения» речевых характеристик диктора от разных групп наркотических средств и степени наркотического опьянения. Изменения основного тона не носят регулярного, общего для всех характера. Главной причиной изменения основного тона является изменение эмоционального состояния диктора, а не выход из состояния наркотического опьянения.

Исследования влияния длительности употребления наркотических средств на речь дикторов показали, что для дикторов, имеющих длительный стаж употребления наркотических средств героиновой группы, обнаружена тенденция к снижению основного тона примерно на 3% в год.

Литература

1. Адашинская Г.А., Чернов Д.Н. Акустические корреляты индивидуальных особенностей функциональных и эмоциональных состояний // *Авиакосмическая и экологическая медицина*. – 2007. – Т. 41. – № 2. – С. 3–13.
2. de Cheveigné A., Kawahara H. Comparative evaluation of F0 estimation algorithms // *Proc. Interspeech-2001*. – Aalborg, Denmark, 2001. – P. 2451–2454.
3. Labutin P., Koval S., Raev A. Speaker identification based on the statistical analysis of f0 // *Proc. IAFPA 16th Annual Conference*. – Plymouth, UK, 2007 [Электронный ресурс]. – Режим доступа: http://www.iafpa.net/abstracts07/Labutin_et_al_-_IAFPA_2007.pdf, свободный. Яз. англ. (дата обращения 20.08.2012).
4. Helffrich H., Standke R., Scherer K.R. Vocal indicators of psychoactive drug effects // *Speech Communication*. – 1984. – № 3. – P. 245–252.
5. Гельман В.Я., Дмитриева Е.С., Зайцева К.А., Орлов А.М. Влияние индивидуальных особенностей человека на акустические корреляты эмоциональной интонации речи // *Журнал высшей нервной деятельности им. И.П. Павлова*. – 2009. – Т. 59. – № 5. – С. 538–546.
6. Tolkmitt F.J., Scherer K.R. Effect of Experimentally Induced Stress on Vocal Parameters // *Journal of Experimental Psychology and Human Perceptual Performance*. – 1986. – V. 12. – № 3. – P. 302–313.
7. Raev A.N., Matveev Y.N., Goloshapova T.I. The Effect of Use of Drugs on Speaker's Fundamental Frequency // *Proc. SPECOM-2011. 14th International conference on SPEECH and COMPUTER*. – Kazan, 2011. – P. 308–314.
8. Smirnova N., Starshinov A., Oparin I., Goloshchapova T. Speaker Identification Using Selective Comparison of Pitch Contour Parameters // *Proc. ICPhS XVI*. – Saarbrücken, 2007. – P. 1801–1804.
9. Матвеев Ю.Н., Симончик К.К. Система идентификации дикторов по голосу для конкурса NIST SRE 2010 // *20-я Международная конференция по компьютерной графике и зрению (GraphiCon)*. – СПб: СПбГУ ИТМО, 2010. – P. 315–319.
10. Белых И.Н., Капустин А.И., Козлов А.В., Лоханова А.И., Матвеев Ю.Н., Пеховский Т.С., Симончик К.К., Шулипа А.К. Система идентификации дикторов по голосу для конкурса NIST SRE 2010 // *Информатика и ее применения*. – 2012. – Т. 6. – Вып.1. – С. 24–31.
11. Шолохов А.В. Классификация эмоционального состояния человека по записям устной речи // *Научно-технический вестник информационных технологий, механики и оптики*. – 2012. – № 3. – С. 150.

Раев Андрей Николаевич

– ООО «ЦРТ», директор научно-исследовательского департамента, raev@speechpro.com

Матвеев Юрий Николаевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ООО «ЦРТ-инновации», доктор технических наук, главный научный сотрудник, профессор, matveev@speechpro.com

Голощапова Татьяна Ивановна

– Экспертно-криминалистическое управление ФСКН России, кандидат филол. наук, начальник отдела

УДК 004.052, 004.75, 004.89, 004.942

**ТЕХНОЛОГИЯ ОЦЕНКИ СОСТОЯНИЙ НАДЕЖНОСТИ МНОГОАГЕНТНЫХ
РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ
НА ОСНОВЕ НЕЧЕТКОЙ МАРКОВСКОЙ МОДЕЛИ****А.В. Маслобоев, В.Н. Богатиков**

Разработана технология оценки состояний надежности многоагентных распределенных информационных систем, основанная на использовании нечеткой марковской модели, узлами которой являются интегральные показатели надежности. В качестве показателя используется оценка, получаемая как степень нечеткого равенства текущей ситуации, которая описывается системой показателей надежности, ситуации, соответствующей области безопасного функционирования системы. Технология позволяет определить распределение в системе информационных и аппаратных ресурсов, обеспечивающее требуемый уровень надежности.

Ключевые слова: технология, моделирование, оценка состояний надежности, распределенные информационные системы, нечеткая марковская модель, многоагентные системы.

Введение

Надежность функционирования информационно-коммуникационных систем (ИКС) является одним из основных условий их эффективного применения при решении задач управления функционированием и развитием социально-экономических систем различного уровня. Особое значение надежность ИКС приобретает при управлении территориями, которые характеризуются экстремальными природно-климатическими условиями и очаговой хозяйственной деятельностью.

Территориальная компьютерная сеть Арктических регионов Российской Федерации (РФ) может быть представлена как ассоциация территориально рассредоточенных вычислительных комплексов различного типа и назначения, взаимодействующих друг с другом через коммуникационную подсеть – сеть передачи данных – с целью предоставления пользователям, независимо от их месторасположения, услуг по хранению, передаче и обработке информации, а также обеспечения доступа к информационным и вычислительным ресурсам компьютерной сети. Разнородность региональных факторов компонентов ИКС, на которые они оказывают влияние, и характеристик влияния существенно осложняют построение формальной модели оценки надежности ИКС в Арктической зоне РФ. Для разработки этой модели был проведен концептуальный анализ особенностей Арктической зоны, влияющих на характеристики надежности ИКС. Выделены четыре группы факторов различной природы:

- геофизические (солнечный ветер, вертикальная компонента межпланетного магнитного поля, солнечная радиация);
- природно-климатические (температура, давление, влажность воздуха, скорость перемещения воздушных масс, фотопериодический режим);
- социально-экономические (ресурсное обеспечение компонентов сложной информационной системы – кадры, энергетика, оборудование и системы управления);
- территориально-географические (географическое расположение, условия рельефа и др.).

Необходимо отметить, что степень надежности технических средств обеспечения информационных систем в Арктическом регионе практически не изучена. Вместе с тем показано, что солнечная активность и обусловленные ею геомагнитные возмущения влияют на распространение радиоволн и в период геомагнитных возмущений создают серьезные помехи для радиосвязи. В Арктическом регионе, где радиосвязь в отдельных ситуациях является единственным средством коммуникации, такие помехи могут приводить к роковым последствиям. И именно в Арктическом регионе, в силу широтных особенностей строения магнитосферы и ионосферы Земли, помехи в радиосвязи в период геомагнитных возмущений наиболее выражены. Кроме того, в высоких широтах отмечается и наибольшее число сбоев в работе технических систем и информационного обеспечения. Эти сбои могут быть вызваны широтным возрастанием интенсивности космических лучей, порождающих ошибки в работе компьютерных систем и программного обеспечения.

Учитывая возрастание в последние годы интереса мирового сообщества к Арктическим территориям, возросла актуальность решения задач повышения уровня системной надежности в рамках формирования единых виртуальных информационных пространств (ЕВИП) Арктических регионов. Под ЕВИП понимается целостная информационно-аналитическая среда, представляющая собой комплекс проблемно-ориентированных, взаимоувязанных и взаимодействующих информационных и аналитических ресурсов, а также технологическую и организационную инфраструктуру их создания и использования [1]. Ввиду разнородности компонентов, составляющих ЕВИП, решение задачи оценки и управления его надежностью представляется сложной проблемой, требующей комплексного решения.

В качестве технологической платформы для практической реализации и развертывания открытых расширяемых многофункциональных ЕВИП предлагается использовать современные технологии одно-ранговых мультиагентных распределенных информационных систем [2] и принципы сетецентрического управления [3]. Использование технологии мультиагентных систем (МАС) [4] позволяет создать адекватную информационно-аналитической среде поддержки управления развитием Арктических регионов, учитывая распределенность, динамичность и структурную сложность образующих их подсистем. При таком подходе представляется возможность реализовать виртуализацию функций управления отдельными аспектами регионального развития за счет делегирования их интеллектуальным про-активным агентам, а на основе проблемно-ориентированных коалиционных взаимодействий агентов можно будет обеспечить эффективное функционирование самоорганизующегося ЕВИП и его компонентов. Под агентом в данном случае понимается аппаратная или программная сущность, действующая либо от лица пользователя, либо от лица системы, делегировавшей агенту полномочия на выполнение тех или иных действий в интересах достижения целей при решении пользовательских задач [5].

Актуальность исследований в области МАС, согласно работам [6, 7], определяется сложностью современных организационных и технических систем, разнообразием, сложностью и распределенностью решаемых задач, огромными объемами потоков информации и высокими требованиями к времени ее обработки. Теоретические исследования в области МАС ведутся в основном по следующим направлениям: теория агентов; коллективное поведение агентов; архитектура агентов и МАС; методы, языки и средства коммуникации агентов; языки реализации агентов; средства поддержки миграции агентов по сети. Анализ современного состояния исследований отечественных и зарубежных научных школ в рассматриваемой проблемной области показал, что вопросы обеспечения надежности технических и организационных компонентов мультиагентных информационных систем и создаваемых на их основе агентно-ориентированных ЕВИП мало изучены. Проблема моделирования и оценки состояний для обеспечения надежности МАС и мультиагентных ЕВИП на основе исследования структур и механизмов проявления разнородных потенциальных опасностей является важной фундаментальной задачей, которую необходимо решать при разработке МАС и ЕВИП, ориентированных на использование в различных областях. Следует также отметить, что получение комплексной оценки надежности МАС зависит от оценок состояний образующих МАС компонентов, в частности, агентов, которые сами по себе являются самостоятельными интеллектуальными системами. В связи с этим моделирование и оценка надежности агентов является отдельной задачей.

Наряду с технической компонентой надежности работы мультиагентных информационных систем, человеческий фактор также влияет на их надежность. От психофизиологического состояния пользователя-оператора во многом зависит своевременность и правильность принятия решений и, следовательно, надежность работы информационных систем. Исследования последних лет показали, что вариации геомагнитного поля, также как и интенсивность космических лучей, существенно модулируют психофизиологическое состояние человека. Другими словами, для комплексной оценки степени надежности работы используемых в Арктических регионах проблемно-ориентированных мультиагентных информационных систем необходимо учитывать также и состояние пользователя-оператора.

Работа посвящена разработке нечеткой модели оценки состояний для анализа надежности компонентов распределенных МАС, ориентированной на поддержку принятия решений в области управления формированием, функционированием и развитием мультиагентного ЕВИП Арктических регионов. В качестве перспективного подхода к количественной оценке надежности МАС предложено использование аппарата нечеткой логики. Проведен анализ возможности адаптации существующих нечетких методов оценки надежности технологических процессов к задачам оценки надежности МАС и ЕВИП. Разработан вариант построения моделей оценки надежности функционирования МАС и ЕВИП на основе интегрального показателя надежности и развития метода марковских цепей.

Модель оценки состояния надежности функционирования агента

В настоящее время активно развиваются технологии управления на основе методологий искусственного интеллекта с целью повышения потенциала самоорганизации и качества принимаемых решений [8]. Основополагающими составляющими направлений развития таких технологий являются агентные технологии. Эффективность управления современными мультиагентными информационными системами во все большей степени зависит от уровня интеллектуальности агентов, их способности к самоорганизованному поиску резервов и ресурсов развития виртуального пространства, способности видеть перспективы и последствия принимаемых решений.

Механизмы функционирования мультиагентных виртуальных сред должны обеспечивать процесс их эволюции таким образом, чтобы минимизировать риски возникновения проблемных ситуаций без разрушительных последствий. Это требует разработки специальной теории и методологии управления виртуальными организационными системами, основанными на обучении, адаптации и самоорганизации.

Далее рассмотрим один из возможных вариантов поведения агентно-ориентированной системы, основное целеустремление которой подчиняется минимизации риска последствий принимаемых решений и различных влияний внешних возмущений.

Процесс функционирования любой системы можно рассматривать как последовательную схему смены ее состояний на некотором интервале времени $(t_0; t_k)$. Состояние системы в каждый момент времени t из этого интервала характеризуется набором параметров этой системы, на которые накладываются ограничения $\varphi(T, K, U) \leq 0$, зависящие от множеств параметров $\{T_i, K_j, U_{pr}\}$ (технологических – $\{T_i, i = \overline{1..I}\}$; конструктивных – $\{K_j, j = \overline{1..J}\}$; управления – $\{U_{pr}, l = \overline{1..L}\}$). Выход за эти ограничения означает переход процесса во внештатную ситуацию. Эти ограничения «вырезают» на множестве всех состояний процесса n -мерную область (рис. 1), в которой процесс не выходит во внештатные ситуации – это область всех работоспособных состояний процесса $S_p \subseteq S$. На основе оценки свойств системы агент формирует с помощью когнитивных механизмов субъективное представление о системе. Это представление включает область возможных состояний, законы поведения системы, оценки риска.

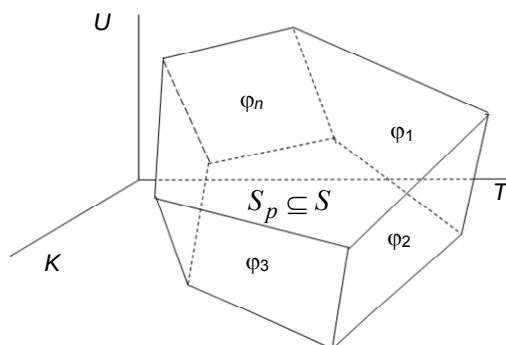


Рис. 1. Область существования состояний системы

В большинстве случаев управление складывается из целеустремлений к определенным состояниям, которые в конкретных ситуациях являются наиболее предпочтительными. Основой такого управления является представление о некоем центре – точке в пространстве состояний, которая доминирует по своим свойствам над остальной областью (рис. 2). Задача управления в этом случае понимается как задача перевода свойств системы в точку, как можно более близкую к области центра. Используемые на рис. 2 обозначения расшифрованы в табл. 1.

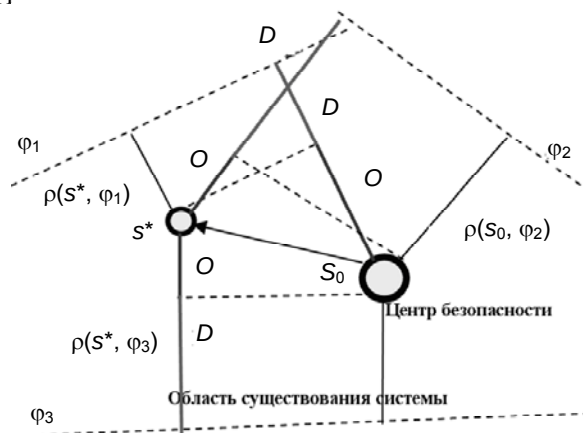


Рис. 2. Оценка индекса безопасности функционирования агента в пространстве состояний

Агент оценивает ситуацию, возникающую в системе, и ставит в соответствие каждой ситуации $\tilde{S}_i$ из определенного набора ситуаций S_s , характеризующего все возможные состояния объекта, некоторое управляющее решение R_i . Перечень ситуаций, входящих в набор S_s , формируется агентом на основе своих знаний. Будем называть эти ситуации эталонами представлений агента. Входная ситуация ${}^T S_0$ сравнивается с эталонными ситуациями $\tilde{S}_i \in S_s$, и определяется эталонная нечеткая ситуация, в некотором смысле наиболее близкая к входной нечеткой ситуации. Модель операции сравнения можно построить, используя операцию нечеткой эквивалентности [9].

Обозначения и формулы для определения индекса безопасности	Описание
S	множество всевозможных состояний системы
S_p	множество состояний агента
$\varphi_i, i = \overline{1..n}$	множество границ области состояния агента («вырезает» на S множество S_p)
S_0	точка процесса, соответствующая оценке центра безопасности
s^*	текущая точка
$l_i = \min(s^*, \varphi_i)$	расстояние от точки s^* до границ области φ_i состояния агента
$\Delta_i^* = \min \rho(s^*, s_0)_{\varphi_i}$	расстояние от точки s^* до центра безопасности S_0 относительно границ
$\delta_i = \min \rho(s_0, \varphi_i)$	расстояние от центра безопасности S_0 до границ области возможных состояний $\varphi_i, i = \overline{1..n}$
$O_i = d_i - l_i$	степень удаленности данного состояния s^* относительно границ $\varphi_i, i = \overline{1..n}$
$D_i = \delta_i - O_i$	степень удаленности для данного состояния s^* относительно границ $\varphi_i, i = \overline{1..n}$

Таблица 1. Обозначения и формулы к рис. 2

Формальное определение «нечеткой» ситуации. Пусть $X = \{x_1, x_2, \dots, x_n\}$ – множество признаков, которые используются агентом. Каждый признак x_i описывается соответствующей лингвистической переменной $\langle \beta_i, E_i, F_i \rangle$, где β_i – название лингвистической переменной; $E_i = \{E_i^1, E_i^2, \dots, E_i^{M_i}\}$ – термножество лингвистической переменной β_i ; F_i – базовое множество лингвистической переменной β_i .

Нечеткое равенство или эквивалентность. В качестве меры близости между ситуациями обычно рассматриваются два критерия: степень нечеткого включения и степень нечеткого равенства.

Степень нечеткого включения ситуации в ситуацию определяется выражением

$$v(\tilde{S}_{X_i}, \tilde{S}_{X_j}) = \&_X v(\mu_{S_{X_i}}(X), \mu_{S_{X_j}}(X)),$$

где

$$v(\mu_{S_{X_i}}(X), \mu_{S_{X_j}}(X)) = \&_{E_k} \left(\mu_{S_{X_i}}(X)(E_k) \rightarrow \mu_{S_{X_j}}(X)(E_k) \right);$$

$$\mu_{S_{X_i}}(X)(E_k) \rightarrow \mu_{S_{X_j}}(X)(E_k) = \max \left\{ 1 - \mu_{S_{X_i}}(X)(E_k), \mu_{S_{X_j}}(X)(E_k) \right\}.$$

Обычно считают, что ситуация $\tilde{S}_i$ нечетко включается в ситуацию $\tilde{S}_j$, $\tilde{S}_i \subseteq \tilde{S}_j$, если степень включения $\tilde{S}_i$ в $\tilde{S}_j$ не меньше некоторого порога включения $t_{inc} \in [0, 6:1]$, определяемого условиями управления, т.е. $v(\tilde{S}_i, \tilde{S}_j) \geq t_{inc}$.

Степень нечеткого равенства. Если множество текущих ситуаций $\tilde{S}_X$ содержит такие ситуации $\tilde{S}_{X_i}$ и $\tilde{S}_{X_j}$, что $\tilde{S}_{X_i}$ нечетко включается в $\tilde{S}_{X_j}$, а $\tilde{S}_{X_j}$ нечетко включается в $\tilde{S}_{X_i}$, то ситуации $\tilde{S}_{X_i}$ и $\tilde{S}_{X_j}$ нужно воспринимать как одну ситуацию. Это означает, что при данном пороге включения t_{inc} ситуации $\tilde{S}_{X_i}$ и $\tilde{S}_{X_j}$ примерно одинаковы. Такое сходство ситуаций называется нечетким равенством, при этом степень нечеткого равенства $\mu(\tilde{S}_{X_i}, \tilde{S}_{X_j}) = v(\tilde{S}_{X_i}, \tilde{S}_{X_j}) \& v(\tilde{S}_{X_j}, \tilde{S}_{X_i})$.

В отличие от набора ${}^T S_s = \{ {}^T S_1, {}^T S_2, \dots, {}^T S_n \}$ текущих ситуаций, набор $S_s = \{ \tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_n \}$, ($n \leq N$) эталонных ситуаций агента не содержит нечетко равных при заданном пороге равенства ситуаций. Предполагается, что множество S_s полно. Таким образом, ситуация $\tilde{S}_i$ существует для любой входной ситуации S_0 . По решающей таблице для этой эталонной ситуации определяется управляющее решение. Данный подход построен на основании метода ситуационного управления [10].

Для определения оценки уровня безопасности введем понятие индекса. Для оценки агентом текущего состояния системы необходимо сравнить на нечеткое равенство входную нечеткую ситуацию с нечеткой ситуацией, которая характеризует центр безопасности. При этом степень нечеткого равенства $In(\tilde{S}_x^*) = v(\tilde{S}_x^*, \tilde{S}_{x_0}) \& v(\tilde{S}_{x_0}, \tilde{S}_x^*)$ покажет величину, которую можно определить как субъективный индекс идеала агента.

Тестовые результаты расчета центра безопасности агента представлены на рис. 3 и рис. 4, а–г.

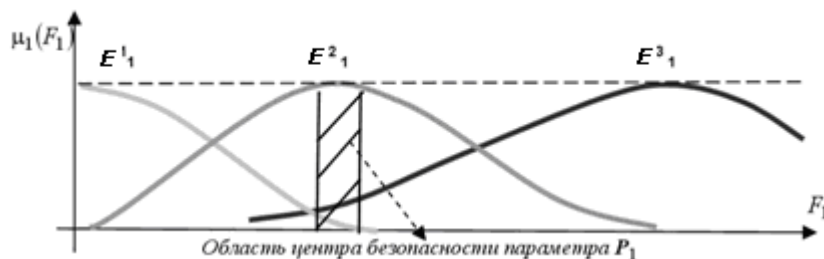


Рис. 3. Функции принадлежности лингвистической переменной $\langle \beta_i, E_i, F_i \rangle$

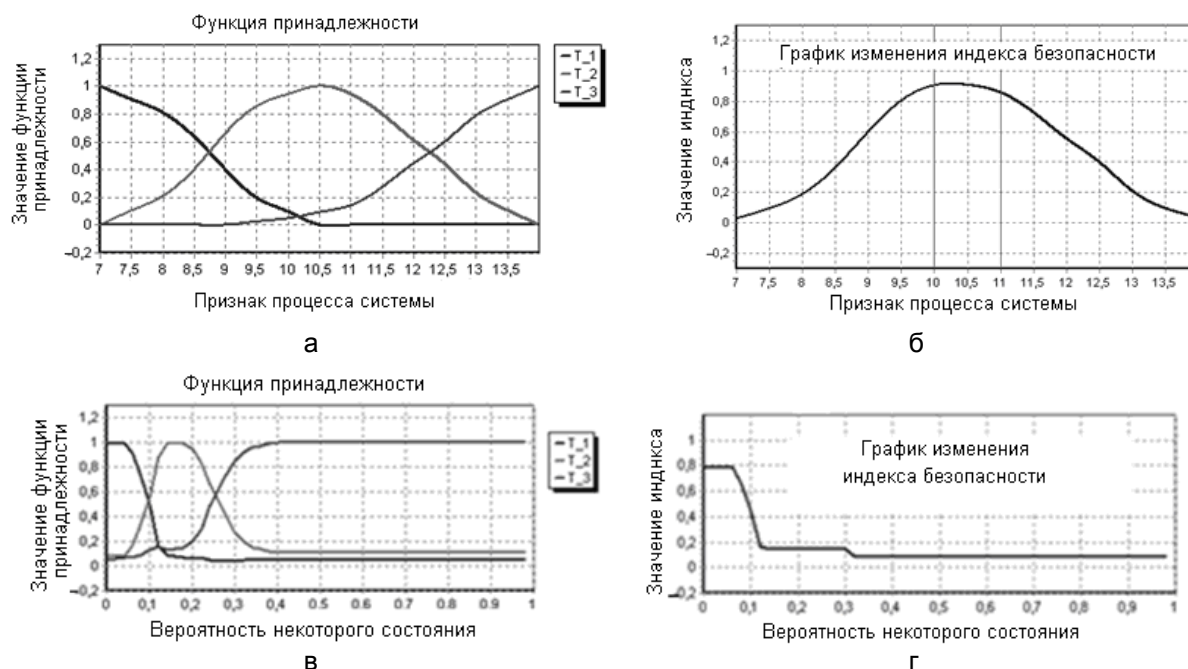


Рис. 4. Результаты расчета индекса безопасности: изменение функции принадлежности лингвистической переменной $\langle \beta_i, E_i, F_i \rangle$ в зависимости от значений признака процесса системы (а); изменение индекса безопасности в зависимости от значений признака процесса системы (б); изменение функции принадлежности лингвистической переменной $\langle \beta_i, E_i, F_i \rangle$ в зависимости от вероятности нахождения системы в некотором состоянии (в); изменение индекса безопасности в зависимости от вероятности нахождения системы в некотором состоянии (г)

Процесс принятия решений агентом может быть промоделирован на основе лингвистических переменных, с помощью которых формализуется качественная информация, представленная в словесной форме. По результатам контроля функционирования системы по имеющимся функциям принадлежности заполняется табл. 2. На основе этой таблицы проводятся вычисления индексов оценок уровня безопасности.

Исходная информация														
№ состояния	p_1				...	p_n			d_1		...	d_m		
	T_1	T_2	T_3	...	T_1	...	T_k	T_1	T_2	...	T_1	T_2	T_3	
1	$\mu_1(p_1)$	$\mu_2(p_1)$	$\mu_3(p_1)$	...	$\mu_1(p_n)$	...	$\mu_k(p_n)$	$\mu_1(d_1)$	$\mu_2(d_1)$	...	$\mu_1(d_m)$	$\mu_2(d_m)$	$\mu_3(d_m)$	
2	$\mu_{21}(p_1)$	$\mu_{22}(p_1)$	$\mu_{23}(p_1)$	...	$\mu_{21}(p_n)$	...	$\mu_{2k}(p_n)$	$\mu_{21}(d_1)$	$\mu_{22}(d_1)$	...	$\mu_{21}(d_m)$	$\mu_{22}(d_m)$	$\mu_{23}(d_m)$	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
ν	$\mu_{\nu 1}(p_1)$	$\mu_{\nu 2}(p_1)$	$\mu_{\nu 3}(p_1)$	...	$\mu_{\nu 1}(p_n)$	...	$\mu_{\nu k}(p_n)$	$\mu_{\nu 2}(d_1)$	$\mu_{\nu 2}(p_1)$	...	$\mu_{\nu 1}(d_m)$	$\mu_{\nu 2}(d_m)$	$\mu_{\nu 3}(d_m)$	
$p_i - i$ -й параметр системы; $d_j - j$ -й возможный ущерб														
Расчет индекса														
1 этап														
	p_1				...	p_n			d_1		...	d_m		
1	$In_1\left(\tilde{S}_{p1}\right)$				...	$In_1\left(\tilde{S}_{pn}\right)$			$In_1\left(\tilde{S}_{d1}\right)$		...	$In_1\left(\tilde{S}_{dm}\right)$		
...	...				...	...			...		...	...		
ν	$In_{\nu}\left(\tilde{S}_{p1}^{*}\right)$				...	$In_{\nu}\left(\tilde{S}_{pn}^{*}\right)$			$In_{\nu}\left(\tilde{S}_{d1}\right)$		...	$In_{\nu}\left(\tilde{S}_{dm}\right)$		
2 этап														
1	$In_1\left(\tilde{S}_{p(1-n)}\right)$							$In_1\left(\tilde{S}_{d(1-m)}\right)$						
...	...							...						
ν	$In_{\nu}\left(\tilde{S}_{p(1-n)}\right)$							$In_{\nu}\left(\tilde{S}_{d(1-m)}\right)$						
3 этап														
1	$In_{Risk1}\left(\tilde{S}\right)$													
...	...													
ν	$In_{Risk\nu}\left(\tilde{S}\right)$													

Таблица 2. Исходная информация для определения индексов безопасности и результаты вычислений

Если учитываются ущербы, которые возникают в процессе работы системы, то агент может определять и риск, который возникает при различных состояниях системы. Определение оценки индекса риска реализуются на основе того же механизма вычислений, что и оценки индекса безопасности. Риск в данном случае определяется как двойка <индекс безопасности, индекс ущерба>:

$$In_{Risk}(\tilde{S}^*) = \{In(\tilde{S}_p^*), In(\tilde{S}_d^*)\}.$$

Методика расчета оценки состояний надежности МАС топологическим методом

В рамках разработки моделей надежности МАС и ЕВИП Арктической зоны РФ проведены исследования по адаптации методов оценки надежности технологических систем к задачам оценки надежности распределенных мультиагентных информационных систем.

Для формирования показателей безопасного функционирования различных систем в качестве параметров можно использовать следующие показатели надежности: показатели безотказности (вероятность безотказной работы; средняя наработка до отказа; средняя наработка на отказ; гамма-процентная наработка до отказа; интенсивность отказов; параметр потока отказов; средняя доля безотказной наработки; плотность распределения времени безотказной работы); показатели долговечности (средний ресурс; гамма-процентный ресурс; назначенный ресурс; средний срок службы; гамма-процентный срок службы; назначенный срок службы); показатели ремонтпригодности (вероятность восстановления работоспособного состояния; среднее время восстановления работоспособного состояния; интенсивность восстановления); показатели сохраняемости (средний срок сохраняемости; гамма-процентный срок сохраняемости).

Выше перечислены все показатели, входящие в стандарт по надежности. Исследователь выбирает те показатели, которые документируются в процессе работы системы. На этой основе информации определяются функции принадлежности для показателей надежности. В список показателей могут быть включены при необходимости и ущербы, которые проявляются в результате возникновения внештатных ситуаций. Формируется табл. 2, на основе которой рассчитывается интегрированный показатель надежности. Это может быть и риск-показатель, характеризующий безопасность работы системы.

Процесс функционирования системы – это непрерывная смена состояний. Смена состояний происходит под воздействием внешних и внутренних факторов. Могут возникать различные внештатные ситуации. Процесс смены состояний системы сопровождается также непрерывной сменой значений индексов безопасности или рисков, и происходит это непрерывно во времени. В работе делается предположение, что смену состояний системы можно описать марковскими случайными процессами. Марковские случайные процессы с конечным или счетным множеством возможных состояний обычно называют цепями Маркова. Например, для системы с четырьмя состояниями (рис. 5) может быть определена следующая система уравнений Колмогорова [11]:

$$\begin{aligned}\frac{dIn_1^{Risk}}{dt} &= -\lambda_{12}In_1^{Risk} + \lambda_{31}In_3^{Risk}; \\ \frac{dIn_2^{Risk}}{dt} &= -\lambda_{23}In_2^{Risk} - \lambda_{24}In_2^{Risk} + \lambda_{12}In_1^{Risk} + \lambda_{42}In_4^{Risk}; \\ \frac{dIn_3^{Risk}}{dt} &= -\lambda_{31}In_3^{Risk} - \lambda_{34}In_3^{Risk} + \lambda_{23}In_2^{Risk}; \\ \frac{dIn_4^{Risk}}{dt} &= -\lambda_{42}In_4^{Risk} - \lambda_{24}In_2^{Risk} + \lambda_{34}In_3^{Risk}.\end{aligned}$$

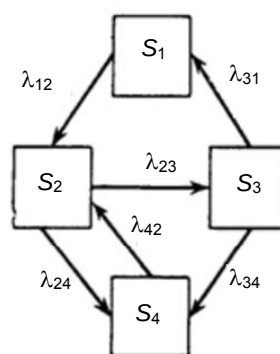


Рис. 5. Пример размеченного графа с четырьмя состояниями

Здесь $\lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{Ind_{ij}^{Risk}(\Delta t)}{\Delta t}$ – плотность перехода λ_{ij} – предел отношения значения индекса за время Δt при переходе из состояния S_i в состояние S_j к длине промежутка Δt . Левая часть уравнений

есть производная функции $Ind^{Risk}(t)$, которую можно называть интенсивностью перехода индекса оценки идеала агента.

Вывод аналогичных соотношений рассмотрен в [11]. Условия, при выполнении которых цепь Маркова обладает эргодическими свойствами и имеет единственное стационарное распределение, приведены в [12]. Используя структурное подобие вероятностного графа и нечеткого графа для моделирования процессов смены состояний мультиагентной информационной системы, рассматривается один из методов расчета марковских процессов – топологический метод [13]. Подобие графов позволяет использовать одни и те же формулы для расчета:

- вероятности нахождения системы в некотором состоянии (индексы оценки состояния надежности для нечеткой системы);
- коэффициентов готовности и простоя (индексы оценки состояния коэффициентов готовности и простоя);
- среднего времени наработки на отказ и среднего времени восстановления (индексы оценки состояний времени наработки на отказ и времени восстановления).

Если при этом учитывать свойства функций принадлежности, появляется возможность проводить диагностирование информационной системы.

Обозначим X – множество состояний системы: $X = \{x_i, i \in I, i = \overline{1, n}\}$, где x_i – i -е состояние; I – множество индексов всех возможных состояний системы; n – количество возможных состояний системы.

Разобьем множество X на два подмножества:

1. подмножество работоспособных состояний системы X_p : $X_p = \{x_i, i \in I_p \subset I\}$, где X_p – подмножество работоспособных состояний системы; I_p – множество индексов работоспособных состояний системы;

2. подмножество неработоспособных состояний системы $\bar{X}_p : \bar{X}_p = \{x_i, i \in J \subset I\}$, где $\bar{X}_p$ – подмножество неработоспособных состояний системы; J – множество индексов неработоспособных состояний системы.

Нахождение системы в том или ином состоянии обуславливает случайный процесс $X(t)$ перехода системы в пространстве ее состояний. $X(t)$ называют также траекторией системы.

Представим $X(t)$ в виде нечеткого графа состояний $G(X, W)$, где X – множество вершин графа, соответствующих множеству состояний X ; W – множество дуг, соединяющих вершины данного графа; $P_1(t), \dots, P_i(t), \dots, P_6(t)$ – вероятности нахождения системы в i -м состоянии; $d(w_{ij})$ – вес дуги w_{ij} ; α_{ij} – нечеткая интенсивность перехода из состояния i в состояние j (рис. 6).

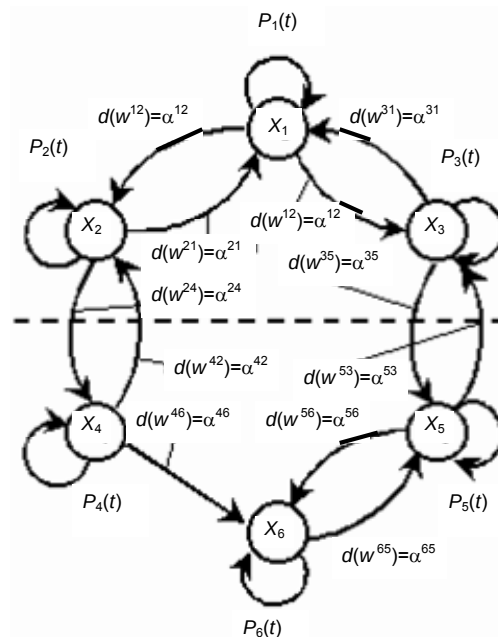


Рис. 6. Пример вероятностного графа состояний $G(X, W)$

Вернемся к модели Маркова. Если заданы интенсивности α_{ij} , то, составляя и решая систему уравнений Колмогорова, можно определить вероятности $P_i(t)$ нахождения системы в i -м состоянии, а значит, и показатели надежности. Однако составление и решение системы уравнений Колмогорова является трудоемкой операцией, поэтому для решения подобных задач применяют топологический метод. Топологический метод использует аппарат теории графов применительно к решению задач надежности.

Рассмотрим методику решения задач методом, который позволяет непосредственно по графу состояний $G(X, W)$ без составления и решения уравнений Колмогорова вычислять показатели надежности. Для этого введем некоторые определения.

Прямой путь I^j из вершины x_i в вершину x_j – цепь последовательно соединенных однонаправленных дуг, где каждая вершина имеет входящую и одну выходящую дуги, за исключением начальной и конечной, имеющих по одной дуге (рис. 7).

Вес k -го прямого пути из вершины i в вершину j рассчитывается как $d(I_k^{ij}) = \prod_{w_{ij} \in W(I_k^{ij})} d(w_{ij})$, где

$W(I_k^{ij})$ – множество дуг, которые составляют k -й прямой путь.

Замкнутый контур r – прямой путь, на котором начальная и конечная вершины совпадают (рис. 8, а). Вес замкнутого контура r равен $d(r) = \prod_{w_{ij} \in W(r)} d(w_{ij})$, где $W(r)$ – множество дуг, входящих в

замкнутый контур r . Частным случаем замкнутого контура является петля (рис. 8, б), в которой входящая и выходящие дуги сливаются в одну. Вес петли при вершине определяется как отрицательная сумма весов дуг, исходящих из этой вершины: $d(w_{ij}) = -\sum_{j \in J_n} d(w_{ij})$, где J_n – множество индексов вершин, которые

связаны с i -й вершиной выходящими из нее дугами.

Соединение графа S – это частичный граф, который образуют только замкнутые контуры. Частичный граф представляет собой все вершины, некоторые дуги и петли исходного графа, которые составляют независимые замкнутые контуры (т.е. контуры, не имеющие общих вершин). Один граф может

располагать несколькими соединениями (рис. 9). При образовании соединений следует помнить, что каждая вершина графа $G(X, W)$ имеет петлю.

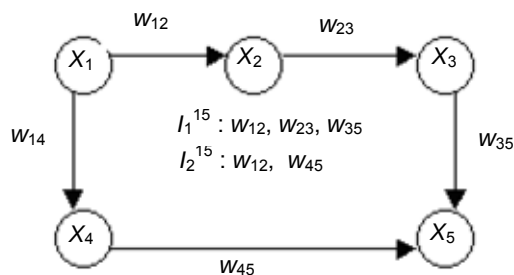


Рис. 7. Определение прямых путей на графе

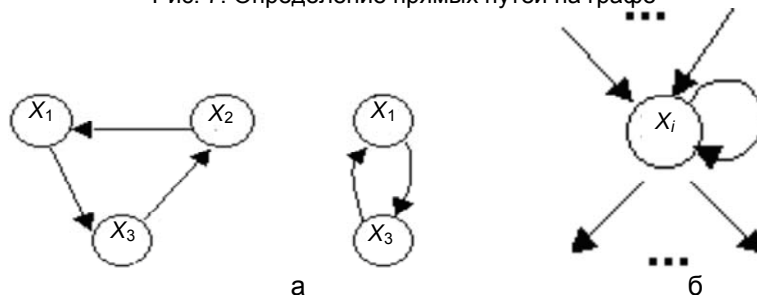


Рис. 8. Примеры замкнутых контуров: замкнутый контур (а); петля (б)

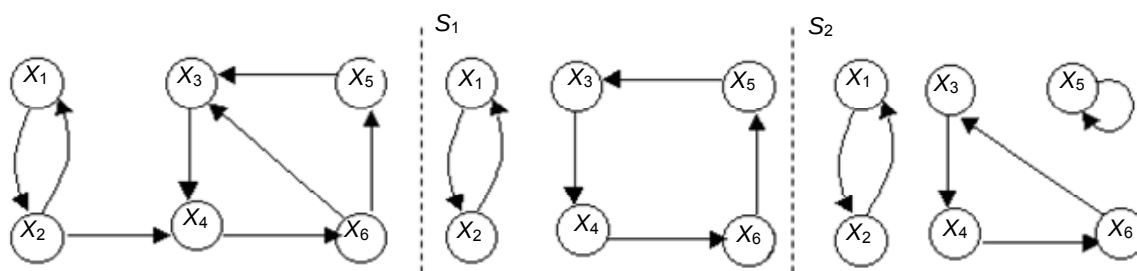


Рис. 9. Пример образования соединения графа

Вес j -го соединения равен $d(S_j) = (-1)^v \cdot \prod_{r \in R(S_j)} d(r)$, где v – число независимых замкнутых контуров, образующих соединение; $R(S_j)$ – множество независимых замкнутых контуров, образующих соединение.

Определитель графа $\Delta G = \sum_{S_j \in S} d(S_j)$, где S – множество всех возможных соединений графа.

Теперь рассмотрим методику расчета показателей надежности вероятностного графа топологическим методом в установившемся режиме, где топологические коэффициенты C_i для каждой x_i вершины графа определяются непосредственно по графу, а затем вычисляется нужный показатель по ниже приведенным топологическим формулам.

Для определения коэффициента C_i необходимо:

1. Выбрать начальную вершину графа x_q отдельно для определения каждого из коэффициентов C_i ($i = \overline{1, n}$); начальная вершина может быть выбрана произвольно, однако выбор влияет на объем вычислений, поэтому ее надо выбирать так, чтобы могли быть выделены длинные прямые пути.
2. Построить множество K прямых путей из начальной вершины x_q в вершину x_i , для которой определяется коэффициент.
3. Для каждого k -го прямого пути построить множество замкнутых контуров подграфа $G\{X_k\}$ и образовать возможные комбинации независимых замкнутых контуров (множество соединений S), где $G\{X_k\}$ – подграф графа $G(X, W)$, образованный удалением множества вершин, входящих в k -й путь и прилегающих к нему дуг.
4. Записать коэффициенты C_i по найденным составляющим по формуле $C_i = \sum_{k \in K} d(I_k^{q_i}) \cdot \Delta G\{X_k\}$, где K – множество прямых путей из произвольно выбранной вершины x_q в x_i ; X_k – множество вершин, входящих в k -й прямой путь.

Используя топологические коэффициенты, можно записать основные показатели надежности системы в установившемся режиме:

- вероятность нахождения системы в i -м состоянии (индекс оценки состояния надежности вершины x_i): $Ind_{p_i} = \frac{C_i}{\sum_{j=1}^n C_j}$, где n – число вершин графа;
- индекс оценки состояния коэффициента готовности системы $Ind_{K_g} = \frac{\sum_{i \in I_p} C_i}{\sum_{j \in I} C_j}$, где I_p – множество индексов работоспособных состояний системы;
- индекс оценки состояния коэффициента простоя системы $Ind_{K_p} = \frac{\sum_{i \in J} C_i}{\sum_{j \in I} C_j}$, где J – множество индексов неработоспособных состояний системы;
- индекс оценки состояния средней наработки на отказ $Ind_{T_o} = \frac{\sum_{i \in I_p} C_i}{\sum_{i \in I_p^+} C_i \left(\sum_{j \in J} \alpha_{ij} \right)} = \frac{\sum_{i \in I_p} C_i}{\sum_{i \in J^+} C_i \left(\sum_{j \in I_p} \alpha_{ij} \right)}$, где I_p^+ – подмножество индексов граничных состояний из X_p , из которых в неработоспособное состояние можно попасть за один переход;
- индекс оценки состояния среднего времени восстановления $Ind_{T_v} = \frac{\sum_{i \in J} C_i}{\sum_{i \in I_p^+} C_i \left(\sum_{j \in J} \alpha_{ij} \right)} = \frac{\sum_{i \in J} C_i}{\sum_{i \in J^+} C_i \left(\sum_{j \in I_p} \alpha_{ij} \right)}$,

где J^+ – подмножество индексов граничных состояний из $\bar{X}_p$, из которых в работоспособное состояние можно попасть за один переход.

Основные положения топологического метода могут быть применены для определения показателей надежности системы в неустановившемся режиме с использованием преобразований Лапласа. Необходимо отметить, что показатели надежности, вычисленные по нечеткой модели, должны совпадать с показателями надежности, вычисленными по вероятностной модели. В отличие от вероятностной марковской модели, где суммы вероятностей состояний для каждого момента времени равны единице, в нечеткой системе такое условие на индекс надежности не накладывается. В связи с этим по равенству показателей можно проводить верификацию нечеткой модели.

Алгоритм оценки состояний объекта и расчета показателей надежности по нечеткой модели включает следующие основные этапы:

1. ввод информации о реальной ситуации на объекте;
2. оценка индексов надежности состояний [14];
3. оценка нечетких интенсивностей переходов из состояния в состояние;
4. расчет показателей надежности системы.

Рассмотренные в настоящей работе модель, методика и алгоритм определения показателей надежности сложных информационных систем на основе использования нечеткой марковской модели реализованы в виде технологии оценки состояний надежности многоагентных распределенных информационных систем, ориентированных на решения задач управления региональным развитием и использование для формирования ЕВИП Арктических регионов.

Пример практической реализации

Пусть имеется распределенная информационная система, которая состоит из двух серверов. В случае аппаратной или программной реализации агентов данные серверы могут быть представленными отдельными агентами. При работоспособности одного из серверов (агентов) система работоспособна, так как каждый сервер (агент) может выполнять все функции.

Обозначим λ_1 , λ_2 – интенсивность отказа первого и второго сервера; μ_1 , μ_2 – интенсивность восстановления первого и второго сервера.

Для обслуживания серверов существует одна бригада, ремонтирующая одновременно только один сервер, который отказал первым.

Необходимо определить топологическим методом показатели надежности.

Работоспособные состояния: x_1 – оба сервера в работоспособном состоянии; x_2 – отказ первого сервера и его восстановление, второй в работоспособном состоянии; x_3 – отказ второго сервера и его восстановление, первый в работоспособном состоянии.

Неработоспособные состояния: x_4 – при восстанавливающемся первом сервере, отказал и второй; x_5 – при восстанавливающемся втором сервере, отказал первый сервер (рис. 10).

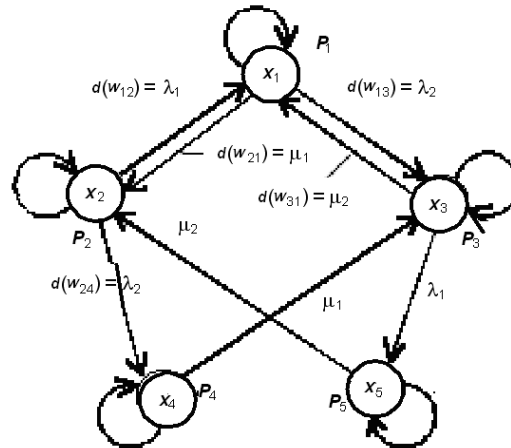


Рис. 10. Граф состояний

Вычисляем топологические коэффициенты. Для нахождения коэффициента C_1 в качестве начальной выбираем вершину x_5 .

$$C_1 = \alpha_{52}\alpha_{24}\alpha_{43}\alpha_{31} + \alpha_{52}\alpha_{21}(-1)^{v=2}d(w_{33})d(w_{44}) = \alpha_{52}\alpha_{21}[-(\alpha_{35} + \alpha_{31})] \times (-\alpha_{43}) + \alpha_{52}\alpha_{24}\alpha_{43}\alpha_{31} = \mu_2\mu_1^2(\lambda_1 + \mu_2) + \mu_2^2\lambda_2\mu_1 = \mu_2\mu_1(\mu_1\lambda_1 + \mu_1\mu_2 + \mu_2\lambda_2).$$

Для нахождения коэффициента C_2 в качестве начальной выбираем вершину x_4 .

$$C_2 = \alpha_{43}\alpha_{35}\alpha_{52}(-1)^{v=1}d(w_{11}) + \alpha_{43}\alpha_{31}\alpha_{12}(-1)^{v=1}d(w_{55}) = \mu_1\lambda_1\mu_2(\lambda_1 + \lambda_2) + \mu_1\mu_2^2\lambda_1 = \mu_1\mu_2(\lambda_1^2 + \lambda_1\lambda_2 + \lambda_1\mu_2),$$

$$C_3 = \mu_1\mu_2(\lambda_2^2 + \lambda_1\lambda_2 + \lambda_2\mu_1),$$

$$C_4 = \lambda_1\lambda_2\mu_2(\lambda_1 + \mu_2 + \lambda_2),$$

$$C_5 = \lambda_1\lambda_2\mu_1(\lambda_2 + \mu_1 + \lambda_1).$$

Теперь можно определить искомые оценки состояний показателей надежности:

$$Ind_{P_1} = \frac{C_1}{C_1 + C_2 + C_3 + C_4 + C_5}, \quad Ind_{K_s} = \frac{C_1 + C_2 + C_3}{C_1 + C_2 + C_3 + C_4 + C_5}, \quad Ind_{K_p} = \frac{C_4 + C_5}{C_1 + C_2 + C_3 + C_4 + C_5},$$

$$Ind_{T_o} = \frac{C_1 + C_2 + C_3}{C_2\lambda_2 + C_3\lambda_1}, \quad Ind_{T_v} = \frac{C_4 + C_5}{C_4\mu_1 + C_5\mu_2}.$$

Определяя интенсивности λ_1 , λ_2 , μ_1 , μ_2 , можно получить количественные оценки состояний показателей надежности системы.

Заключение

В ходе исследований разработана технология оценки состояний надежности многоагентных распределенных информационных систем, основанная на использовании нечеткой марковской модели, узлами которой являются интегральные показатели надежности. Узлы графа образуют интегрированные оценки состояния показателей надежности, которые определяются степенью нечеткого равенства некоторого текущего состояния и исправного состояния, принадлежащего области безопасного функционирования системы. Узлы графа могут быть образованы и на основе риск-показателей. В этом состоит основное отличие общепринятой цепи Маркова от рассматриваемой в работе. Технология позволяет определить распределение информационных и аппаратных ресурсов в сложных информационных системах, обеспечивающее требуемый уровень надежности.

Предложенный подход к оценке надежности распределенных МАС и ЕВИП на основе нечеткой марковской модели имеет ряд преимуществ по сравнению с использованием вероятностной модели Маркова. Основное преимущество нечетких марковских моделей состоит в том, что на основе экспертных оценок можно учесть наличие помех в радиосвязи в период геомагнитных возмущений, влияние интенсивности космических лучей, психофизиологическое состояние оператора, которое во многом зависит как от вариации геомагнитного поля, так и от интенсивности космических лучей. В данном случае для оценки состояний необходим меньший объем статистического материала, так как за основу берутся экспертные знания. Подход обеспечивает более гибкую адаптацию к конкретной задаче и позволяет выпол-

нять диагностирование объекта уже на этапе расчета интегрального показателя надежности. Гибкость достигается за счет того, что состояние надежности может оцениваться сразу по нескольким надежностным показателям. Диагностика на этапе вычисления интегрального показателя состояний достигается за счет того, что вычисления можно разделить на разные этапы. Каждый этап оценивает какой-либо из отдельных показателей, что в конечном итоге позволяет сделать вывод о состоянии соответствующего элемента.

Таким образом, можно сделать вывод о целесообразности исследования нечетких марковских моделей для целей обеспечения надежности распределенных МАС и ЕВИП.

Полученные результаты исследований в области создания и развития моделей и методов оценки надежности распределенных мультиагентных информационных систем, могут найти применение при формировании ЕВИП Арктических регионов РФ в рамках реализации «Стратегии развития Арктической зоны Российской Федерации и обеспечения национальной безопасности на период до 2020 года» на территории Мурманской области.

Работа выполнена при поддержке РФФИ (проект №12-07-00138-а «Разработка когнитивных моделей и методов формирования интегрированной информационной среды поддержки управления безопасностью Арктических регионов России»).

Литература

1. Олейник А.Г., Федоров А.М. Проблемы и задачи формирования единого информационного пространства Арктической зоны Российской Федерации // Труды Кольского научного центра РАН. Информационные технологии. – Апатиты: Изд-во КНЦ РАН, 2011. – Вып. 2. – С. 19–28.
2. Маслобоев А.В., Шишаев М.Г. Архитектура и технологии формирования интегрированной информационной среды поддержки управления безопасностью развития региона // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 6 (76). – С. 98–104.
3. Затуливетер Ю.С. Компьютерный базис сетевидного управления // Технические и программные средства систем управления, контроля и измерения: Труды конф. – М.: ИПУ РАН, 2010. – С. 492–511.
4. Путилов В.А., Шишаев М.Г., Олейник А.Г. Технологии распределенных систем информационной поддержки инновационного развития региона // Труды Института системного анализа РАН: Прикладные проблемы управления макросистемами. – 2008. – Т. 39. – С. 40–63.
5. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. – М.: URSS, 2002. – 352 с.
6. Рыбина Г.В., Паронджанов С.С. Модели, методы и программные средства поддержки взаимодействия интеллектуальных агентов // Информационные технологии и вычислительные системы. – М.: URSS, 2008. – Вып. 3. – С. 22–29.
7. Маслобоев А.В. Подходы к обеспечению информационной безопасности в открытых распределенных мультиагентных виртуальных бизнес-средах // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 6 (70). – С. 98–104.
8. Виноградов Г.П. Индивидуальное принятие решений: поведение целеустремленного агента. – Тверь: Изд-во ТГТУ, 2011. – 164 с.
9. Мелихов А.Н., Бернштейн Л.С., Коровин С.Я. Ситуационные советующие системы с нечеткой логикой. – М.: Наука, 1990. – 272 с.
10. Поспелов Д.А. Ситуационное управление: Теория и практика. – М.: Наука, 1986. – 288 с.
11. Вентцель Е.С. Исследование операций. – М.: Советское радио, 1972. – 551 с.
12. Дудин А.Н., Медведев Г.А., Меленец Ю.В. Практикум на ЭВМ по теории массового обслуживания: Учебное пособие. – Минск: Изд-во БГУ, 2000. – 109 с.
13. Палюх Б.В., Мироненко А.С. Надежность и эффективность экономических информационных систем. – Тверь: Изд-во ТГТУ, 2003. – 157 с.
14. Богатилов В.Н. и др. Приложения метода разделения состояний к управлению технологической безопасностью на основе индекса безопасности. – Тверь: Изд-во ТГТУ, 2009. – 398 с.

Маслобоев Андрей Владимирович	–	ФГБУН Институт информатики и математического моделирования технологических процессов Кольского научного центра РАН, кандидат технических наук, доцент, докторант, masloboev@iimm.kolasc.net.ru
Богатилов Валерий Николаевич	–	ФГБУН Институт информатики и математического моделирования технологических процессов Кольского научного центра РАН, доктор технических наук, профессор, vnbgtk@iimm.kolasc.net.ru

УДК 004.056

МЕТОД ПОЛУАВТОМАТИЧЕСКОГО ФОРМИРОВАНИЯ СЛОВАРЯ МОРФОЛОГИЧЕСКИХ ОПИСАНИЙ СЛОВ

С.В. Лапшин, И.С. Лебедев

Предложен метод полуавтоматического формирования словаря морфологических описаний слов. Предложенный метод позволяет существенно упростить процедуру пополнения и актуализации словарей новыми словоформами и повысить таким образом показатели точности и полноты морфологических анализаторов DLP- и IPC-систем.

Ключевые слова: полуавтоматическое формирование словаря, морфологические описания слов, DLP- и IPC-системы.

Введение

Усиление влияния информационной сферы деятельности на жизненно важные интересы общества и государства требует ее защиты от информационных воздействий. Это обуславливает возрастание роли и значения обработки и мониторинга текстовой информации в компьютерных сетях и открытых системах документооборота.

Одной из ключевых частей современных систем защиты информации классов DLP (Data loss prevention) и IPC (Information Protection and Control) является морфологический анализатор. В своей работе он использует словарь, который содержит морфологические описания словоформ того или иного языка. При работе с аналитическими языками, например, с французским, задача морфологического анализа не столь актуальна, поскольку основными передатчиками грамматического значения являются порядок слов и другие служебные слова. Но при обработке текстов на синтетическом языке, например, на русском, морфологическое описание слова является основой для построения естественно-языковой конструкции. Таким образом, возникает задача генерации словарей, которые содержат эту информацию.

Текстовые сообщения, циркулирующие в вычислительных сетях, обрабатываемые с целью мониторинга состояния информационной безопасности, имеют ряд особенностей, среди которых необходимо отметить небольшую длину и использование специфических выражений и аббревиатур [1]. Примером могут являться сообщения в интернет-мессенджерах или социальных сетях. Поскольку качество обработки текстов естественного языка напрямую зависит от полноты используемых для этого словарей, очень важно иметь достаточно полные и актуальные словари с необходимыми описаниями для каждой словоформы.

Ручное пополнение словаря новыми словами с их морфологическими описаниями является трудоемким процессом. То, что естественный язык не является статичным, особенно в разговорной речи, еще более усложняет задачу поддержания актуальности таких словарей.

Работы в этом направлении ведутся уже достаточно давно. Большой вклад в рассматриваемый вопрос внес коллектив Санкт-Петербургского экономико-математического института (ЭМИ РАН) [2], а также компании AOT, Noolab, RCO и др. [3]. В настоящей работе рассматриваются вопросы адаптации морфологического анализа для систем мониторинга, обрабатывающих текстовую информацию. В связи со спецификой анализируемых текстов в таком словаре должны быть не только корректные словоформы, но и словоформы с типичными ошибками, которые допускаются людьми при написании текстов, а также выражения и аббревиатуры, специфичные для конкретной группы людей или области знаний.

Предлагаемый в работе метод позволяет существенно упростить процедуру пополнения и актуализации словарей новыми словоформами и, таким образом, повысить показатели точности и полноты морфологических анализаторов DLP- и IPC-систем.

Постановка задачи

Основой предлагаемого предметно-ориентированного морфологического анализатора, содержащего идентификационные признаки словоформ предметной области, разработанного для русского языка, служит словарь А.А. Зализняка [4].

Формализация морфологии в словарных базах данных программы представлена следующим образом.

Пусть $S = \{ S_i \}$, $i=1, \dots, n$ – множество исходных форм слов базы данных средства защиты информации (БД СЗИ). Пусть $M = \{ M_j \}$, $j=1, \dots, k$ – множество парадигм, причем каждому элементу множества соответствует морфологический признак $M_j \rightarrow P_j$. Пусть s – словоформа. Пусть $c = \{ c_r \}$, $r=1, \dots, z$ – множество стандартных окончаний слов.

Тогда необходимо найти такие функции f и g , что

$$S \xrightarrow{f} M; \tag{1}$$

$$M \xrightarrow{g} S,$$

где f – функция соответствия элементов множества S элементам множества M ; g – функция соответствия элементов множества M элементам множества S . Нахождение этих функций позволяет говорить о решении прямой и обратной задач морфологии для СЗИ.

Предлагаемый метод решения основан на том, что любой словоформе сопоставим класс основ B и класс окончаний C , из которого состоит данная словоформа:

$$\forall s \longrightarrow \{B; C\}.$$

Тогда для каждого слова БД СЗИ можно выделить морфологический класс k его парадигм M_k , такой, что словоформа данного морфологического класса S_k является подмножеством парадигм этого класса, и выражается суммой основ и окончаний слова данного морфологического класса.

$$f_k: S_k \in M_k = B_k + C_k.$$

Это означает, что частная задача нахождения функции f прямой задачи морфологии решена. Совокупность решения частных задач даст решение прямой задачи в общем виде.

Соответствие $M_j \rightarrow P_j$, позволяет получить морфологический и идентификационный признак, содержащий информацию, используемую для обнаружения угроз информационной безопасности (морфологический шаблон).

Решение обратной задачи предполагает существование некоторой словоформы БД СЗИ.

Пусть $s_k \subset M_k$. Определим ее основу как разность между словоформой и ее окончанием:

$$\{B_k\} = S_k - \{C_r\}_k, k=1, \dots, n, r=1, \dots, z.$$

Сопоставим эту основу с множеством допустимых для нее исходных словоформ:

$$B_k \longrightarrow S_k, k=1, \dots, n.$$

Используя прямую задачу, вычислим множество парадигм этих основ:

$$\{S_k\} \xrightarrow{f} \{M_k\}.$$

Сравним исходную словоформу с этими парадигмами. В случае совпадения определяется исходная форма слова для данной парадигмы:

$$g_k: M_k \longrightarrow S_k.$$

Совокупность решения частных задач даст решение обратной задачи в общем виде для СЗИ, обрабатывающих текстовую информацию.

Метод формирования словаря

Рассмотрим словарь $Z = \{z_i\}$, полученный путем «чтения» предметно-ориентированных текстов новостных агентств, блогов и комментариев, каждая запись z_i в котором имеет структуру

$$z_i = \{s_i; S_i; P_i\},$$

т.е. состоит из словоформы s_i , исходной формы слова S_i и морфологического описания P_i .

Словарь Z на сегодняшний день содержит более 2,5 млн словоформ. Особенностью словаря является то, что в нем содержатся последовательности символов, употребляемые внутри групп пользователей, и словоформы, имеющие специфические опечатки. Задача состоит в том, чтобы, учитывая регулярность русского языка [5], поддерживать актуальность и полноту словарной базы данных в условиях его постоянного пополнения новыми словоформами с наименьшими трудозатратами. Решение поставленной задачи основывается на словаре, содержащем морфологические описания словоформ А.А. Зализняка, куда входят только базовые словоформы русского языка и множество соответствующих им окончаний.

Рассмотрим, как образованы словоформы слов «ПРЕОБРАЗОВАТЕЛЬ» и «РОЯЛЬ» (таблица).

Словоформы M_i		Морфологические описания P_i
ПРЕОБРАЗОВАТЕЛЬ	РОЯЛЬ	Сущв Муж Неодуш Им, Вин
ПРЕОБРАЗОВАТЕЛЯ	РОЯЛЯ	Сущв Муж Неодуш Род
ПРЕОБРАЗОВАТЕЛЮ	РОЯЛЮ	Сущв Муж Неодуш Дат
ПРЕОБРАЗОВАТЕЛЕМ	РОЯЛЕМ	Сущв Муж Неодуш Тв
ПРЕОБРАЗОВАТЕЛЕ	РОЯЛЕ	Сущв Муж Неодуш Пред
ПРЕОБРАЗОВАТЕЛИ	РОЯЛИ	Сущв Муж Неодуш Им, Вин
ПРЕОБРАЗОВАТЕЛЕЙ	РОЯЛЕЙ	Сущв Муж Неодуш Род
ПРЕОБРАЗОВАТЕЛЯМ	РОЯЛЯМ	Сущв Муж Неодуш Дат
ПРЕОБРАЗОВАТЕЛЯМИ	РОЯЛЯМИ	Сущв Муж Неодуш Тв
ПРЕОБРАЗОВАТЕЛЯХ	РОЯЛЯХ	Сущв Муж Неодуш Пред

Таблица. Словоформы и их морфологические описания

Из таблицы видно, что словоформы получены из базовой формы S одинаковым образом, путем добавления соответствующих окончаний C . Следовательно, достаточно иметь морфологические описания P словоформ слова «ПРЕОБРАЗОВАТЕЛЬ», чтобы построить аналогичные описания для словоформ слова «РОЯЛЬ».

На основе этой идеи разработан предлагаемый метод полуавтоматического формирования словаря. Он состоит из следующих частей:

1. разбор словаря Зализняка, генерация всех словоформ на основе исходных форм слова;
2. разбор словаря с некоторыми морфологическими описаниями описанного выше вида;
3. сопоставление словоформ из словарей, полученных на первых двух шагах, с целью выделения характерных морфологических описаний для каждого окончания;
4. на основе множества соответствий вида окончание–морфологическое описание, полученных на предыдущем шаге, словоформам из словаря Зализняка дается морфологическое описание.

Задачи на первых двух шагах являются чисто техническими, и их описание не представляет какого-либо интереса.

Выделение характерных морфологических описаний для каждого окончания, описанное на третьем шаге, осуществляется следующим образом. Каждое окончание входит в свой «класс» окончаний. Для слова «ПРЕОБРАЗОВАТЕЛЬ» это окончания «Я», «Ю», «ЕМ», «Е», «И», «ЕЙ», «ЯМ», «ЯМИ» и «ЯХ». Окончание «Ю» слова «ЗЕМЛЮ», хотя и совпадает с окончанием «Ю» слова «ПРЕОБРАЗОВАТЕЛЮ», но входит в совершенно другой «класс», и поэтому будет иметь другой набор морфологических описаний. Кроме класса, также учитывается часть речи слова и одушевленность/неодушевленность для имен существительных.

Таким образом, полный ключ в ассоциативном массиве с морфологическими описаниями состоит из «класса» окончания, части речи и признака одушевленности/неодушевленности для имен существительных. Таким образом, в случае, когда одна исходная форма относится к разным частям речи, для каждой части будет храниться свой набор морфологических описателей.

Полученные на третьем шаге соответствия применяются на четвертом шаге для составления словаря с морфологическими описаниями. Из сгенерированного на основе словаря Зализняка списка словоформ берется словоформа, а затем по ключу, описанному выше, в ассоциативном массиве находится морфологическое описание для этой словоформы. Таким образом, реализуется функция f , приведенная в формуле (1).

За счет того, что русский язык достаточно регулярен и многие слова формируются похожим способом, появляется возможность автоматически получить для исходной формы слова морфологическое описание всех его словоформ.

Важной особенностью представленной системы является то, что она не определяет автоматически класс добавляемого слова, а предполагает, что он уже задан. Задача классификации нового слова отведена специалисту-лингвисту либо может быть решена вспомогательными средствами, например, описанными в [6].

Таким образом, процедура пополнения словаря новыми словами с их морфологическими описаниями сводится к определению «морфологического класса» исходной формы слова.

Эксперимент

Для оценки качества метода введем следующие обозначения: количество правильных извлечений системы анализа DLP-фильтра – h , количество требуемых извлечений – d , общее количество извлечений – n . Тогда для полноты R и точности P справедливы следующие соотношения:

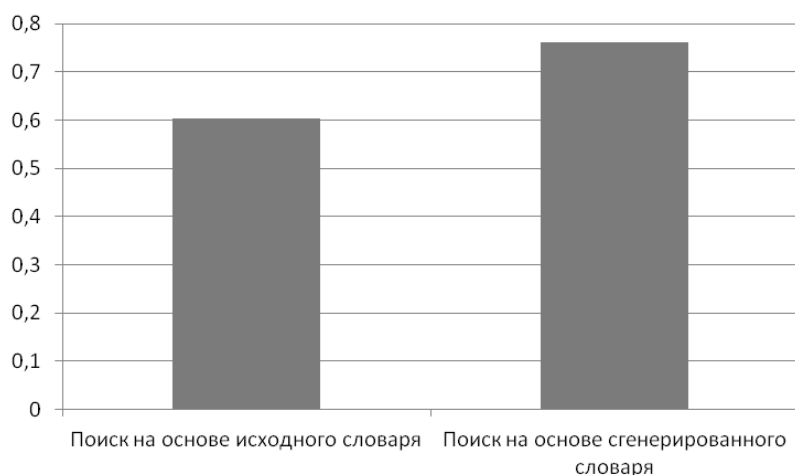
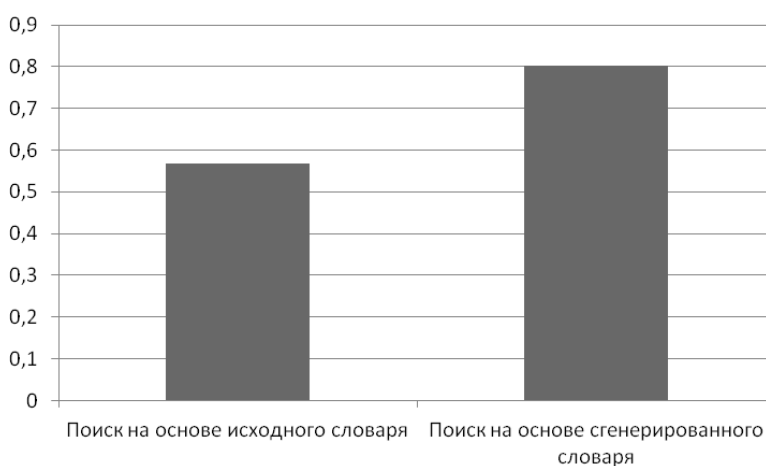
$$R_i = \frac{h_i}{d_i} \text{ и } P_i = \frac{h_i}{n_i}.$$

Эксперимент по поиску с использованием словарей проводился на основе случайной выборки предложений из национального корпуса русского языка [7]. Объем выборки – 180 000 словоупотреблений, из которых 90 000 взяты из прессы и по 30 000 – из научных текстов, художественных текстов и законодательства.

Для проведения эксперимента была разработана простая поисковая система на основе булевской модели поиска [8]. Разработанная система позволяла автоматически формировать поисковые запросы и обрабатывать результаты поиска. Таким образом, значение d числа требуемых извлечений было известно при формировании поисковых запросов, что обеспечивало правильность полученного результата. Общее количество извлечений p и количество правильных извлечений h вычислялись в ходе эксперимента после обработки каждого поискового запроса.

В первом случае поисковая система использовала словарь Зализняка и словарь с полными морфологическими описаниями для только одного слова каждого класса. Во втором случае использовался словарь, сгенерированный с помощью описанного выше метода.

В ходе эксперимента измерялись полнота R и точность P поиска на случайной выборке из национального корпуса русского языка. Результаты измерения приведены на рис. 1, 2.

Рис. 1. Результаты измерения точности поиска P Рис. 2. Результаты измерения полноты поиска R

Эксперимент показал, что при использовании сгенерированного описанным выше методом словаря точность поиска возросла на 26%, а полнота – на 42%.

Заключение

Предложенный метод полуавтоматического формирования словаря морфологических описаний слов позволяет существенно упростить и ускорить процедуру получения таких словарей. Человеку достаточно задать класс добавляемого слова, после этого все словоформы автоматически получают морфологические описания. Это позволяет быстро адаптировать систему анализа естественного языка и, таким образом, приводит к повышению показателей точности и полноты морфологических анализаторов DLP-и ИРС-систем.

Работа выполнена в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы» по государственному контракту № 07.524.12.4009.

Литература

1. Лебедев И.С., Борисов Ю.Б. Анализ текстовых сообщений в системах мониторинга информационной безопасности // Информационно-управляющие системы. – 2011. – № 2. – С. 37–43.
2. Каневский Е.А. Некоторые вопросы пополнения морфологического словаря терминами предметной области // Труды Международного семинара «Диалог'2001» по компьютерной лингвистике и ее приложениям. – М.: РосНИИ искусственного интеллекта, 2001. – Т. 2. – С. 156–160.
3. Большаков И.А., Большакова Е.И. Автоматический морфоклассификатор русских именных групп // Компьютерная лингвистика и интеллектуальные технологии. По материалам конференции «Диалог» (2012). – Т. 1. – Вып. 11. – С. 81–92.

4. Зализняк А.А. Грамматический словарь русского языка. – М.: Русский язык, 1987. – Изд. 4-е, испр. и доп. – 880 с.
5. Тузов В.А. Компьютерная семантика русского языка. – СПб: Изд-во СПбГУ, 2004. – 400 с.
6. Боярский К.К., Каневский Е.А. Проблемы пополнения семантического словаря // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 2 (72). – С. 132–137.
7. Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://ruscorpora.ru/corpora-usage.html>, свободный. Яз. рус. (дата обращения 30.05.2012).
8. Manning C.D., Raghavan P., Schütze H. An Introduction to Information Retrieval. – Cambridge University Press, Cambridge, England. – 2009. – 504 p.

Лапшин Сергей Владимирович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, sv.lapshin@gmail.com

Лебедев Илья Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, lebedev@cit.ifmo.ru

УДК 004.934.2

ПРИМЕНЕНИЕ МЕТОДОВ НЕЛИНЕЙНОЙ ДИНАМИКИ ДЛЯ РАСПОЗНАВАНИЯ ЭМОЦИИ РАДОСТИ В РЕЧИ

К.В. Сидоров, Н.Н. Филатова

Рассмотрена задача распознавания образцов речи, зарегистрированных в момент проявления испытуемыми эмоции радости, от образцов речи этих же дикторов в нейтральном состоянии. Для решения задачи использованы методы нелинейной динамики. Исследования проведены на записях, взятых из базы Emo-DB (Берлин), и фрагментах русскоязычной базы (Тверь). Сформирован модельный корпус эмоциональной речи, состоящий из базы данных двух уровней (фраз и фонем), послуживший основанием для оценки работоспособности разрабатываемых алгоритмов. Выделены устойчивые признаки нелинейной динамики – реконструкция аттрактора и рекуррентный график. Предложены новые количественные признаки для классификации образцов речи человека, испытывающего эмоцию радости, основанные на оценках максимальных векторов реконструкции аттрактора для четырех квадрантов.

Ключевые слова: эмоция, эмоциональное состояние, речь, речевой сигнал, нелинейная динамика, реконструкция аттрактора, рекуррентный график.

Введение

На современном этапе развития информационных технологий разработка методов и систем распознавания эмоционального состояния человека по речевому сигналу с помощью аппаратно-программных средств является актуальной задачей, позволяющей решить ряд проблем в области биомедицинских технологий. В последние годы наблюдается явное усиление интереса к анализу речевого сигнала как объективного показателя эмоционального состояния человека [1, 2]. Различные исследования в области акустики, психолингвистики и психофизиологии позволили собрать сведения о множестве акустических, просодических и лингвистических характеристик речи, которые можно использовать в качестве информативных признаков при распознавании эмоционального состояния, проявляющихся на уровне сегментов, фонем (звуков), слогов, целых слов и фраз. Чаще всего используются следующие признаки речевого сигнала [3]: спектрально-временные, амплитудно-частотные, вейвлет, кепстральные и характеристики (инварианты) нелинейной динамики. Судя по полученным результатам, перечисленные признаки зарекомендовали себя с положительной стороны. Однако, несмотря на большое количество проведенных в данном направлении исследований, ряд проблем все еще остается нерешенным, и многие идеи требуют дальнейшего развития. В частности, отсутствует универсальная теоретическая модель описания речевых образцов в условиях проявления разных видов эмоций, отражающая взаимосвязь вида эмоций и объективных характеристик речевого сигнала.

На текущий момент времени выделение новых информативных признаков, по возможности родственных человеческому восприятию, и поиск эффективных методик распознавания эмоций, являются важнейшей задачей. В работе рассматривается способ решения этой задачи методами нелинейной динамики, позволяющими получить количественную и качественную оценку признаков, проявляющихся в речевом сигнале человека, испытывающего эмоцию радости.

Модельный корпус эмоциональной речи

В настоящее время в Тверском государственном техническом университете активно ведутся разработки системы распознавания эмоционального состояния человека по образцам речевого сигнала. Для проведения исследований необходимо наличие модельного корпуса эмоциональной речи, т.е. базы дан-

ных, в которой хранятся образцы речи испытуемых, находящихся в различных эмоциональных состояниях. В связи с этим был сформирован модельный корпус эмоциональной речи, состоящий из двух частей (русской и немецкой). При создании русскоязычной части в качестве дикторов (испытуемых) выступили 5 человек, каждый из которых, на основе одного нейтрального образца, создал несколько клонов с различным уровнем проявления положительной эмоции радости [4], выбор которой обусловлен интересами дальнейшего применения разрабатываемой технологии. При формировании немецкоязычной части использовались записи эмоции радости и нейтрального состояния, взятые из берлинской базы данных эмоциональной речи Emo-DB (Berlin Database of Emotional Speech) [5], состоящей из 535 фраз 10 дикторов, имитирующих набор эмоциональных состояний: гнев, скука, отвращение, беспокойство/страх, печаль, радость/счастье и нейтральное состояние. В целом, модельный корпус состоит из двух уровней, связанных иерархически. Уровень 1 включает образцы фраз от разных дикторов. Используя алгоритм автоматической генерации речевых объектов [6] для каждой записи уровня 1, получены объекты уровня 2 – фонемы. Всего для проведения исследований сформированы 4 обучающие выборки (ОВ):

1. ОВ 1.1 – 18 русских записей контрольной фразы «А голос мой звучит примерно так»;
2. ОВ 1.2 – 180 гласных фонем, полученных из ОВ 1.1;
3. ОВ 2.1 – 120 немецких фраз;
4. ОВ 2.2 – 300 гласных фонем, сформированных из ОВ 2.1.

Реконструкция аттрактора

Для конструктивного решения задачи распознавания эмоций по речи необходимо количественно охарактеризовать речевой сигнал и выделить существенные параметры, отвечающие за эмоциональное состояние человека, т.е. необходимо подобрать соответствующий математический аппарат. Перспективным, по мнению авторов, в этом плане является аппарат нелинейной динамики, позволяющий реконструировать фазовый портрет аттрактора по временному ряду или по одной его координате. Для реконструкции аттрактора исследуемый временной ряд $x_n, \dots, x_{n-1}$ подвергается задержке координат [7]:

$$y_t = (x_t, x_{t+\tau}, \dots, x_{t+(m-1)\tau}), \quad t = 0, \dots, s-1, \quad s = N - (m-1)\tau, \quad (1)$$

где N – общее число элементов (точек) временного ряда; τ – задержка по времени между элементами временного ряда (временной лаг); m – размерность вложения (размерность лагового пространства).

При выборе значения временной задержки τ используется идея о том, что если точки, образующие временной ряд, независимы друг от друга, то реконструированные вектора (1) несут в себе наибольшее количество информации об исследуемом ряде. По этой причине необходимо выбирать τ таким образом, чтобы корреляция между элементами временного ряда x_t и $x_{t+\tau}$ была по возможности минимальной. Такой выбор осуществляется при вычислении автокорреляционной функции $B(\tau) = 1/k \sum_{k=0}^{k-1} (x_n - \bar{x}) \cdot (x_{n+\tau} - \bar{x})$, $k = N - \tau$, где $\bar{x}$ – математическое ожидание. Временная задержка τ выбирается равной времени первого пересечения нуля автокорреляционной функцией [8] (рис. 1, а), значение задержки составляет 13 ($\tau = 13$). Величина размерности вложения m определяется с точки зрения достаточности (насыщения) посредством вычисления корреляционного интеграла $C(\varepsilon)$ и корреляционной размерности D_2 реконструкции аттрактора [8]. Корреляционный интеграл $C(\varepsilon)$, показывающий относительное число пар точек аттрактора x_i, x_j , находящихся на расстоянии не больше ε , определяется как

$$C(\varepsilon) = \lim_{M \rightarrow \infty} 1/M(M-1) \cdot \sum_{i,j=1}^M \theta(\varepsilon - r(x_i, x_j)), \quad i, j = 1, \dots, M, \quad (2)$$

$$D_2 = \lim_{\varepsilon \rightarrow 0} \log C(\varepsilon) / \log \varepsilon, \quad (3)$$

где M – число рассматриваемых состояний x_i (количество точек x_i на аттракторе); r – расстояние между точками аттрактора; $\theta(\alpha)$ – ступенчатая функция Хевисайда. После нахождения $C(\varepsilon)$ (2) и D_2 (3), строится зависимость корреляционной размерности D_2 от размерности вложения m (1), определяется точка, при которой кривая наклонов насыщается (рис. 1, б), корреляционная размерность аттрактора составляет 3,6 ($D_2 = 3,6$), она достигается при размерности вложения, равной 5 ($m = 5$).

Проведен нелинейный анализ фраз (ОВ 1.1, ОВ 2.1) и фонем (ОВ 1.2, ОВ 2.2) (рис. 2) на основе реконструкции аттрактора (1). Выявлено, что в большинстве случаев наблюдается взаимосвязь геометрии аттрактора с состоянием эмоционального возбуждения (объектам нейтрального состояния присуща более правильная форма, стремящаяся к эллипсообразной). Установлено, что эмоция радости по сравнению с нейтральным состоянием имеет меньшую траекторию разброса реконструкции, как для фраз, так и для фонем.

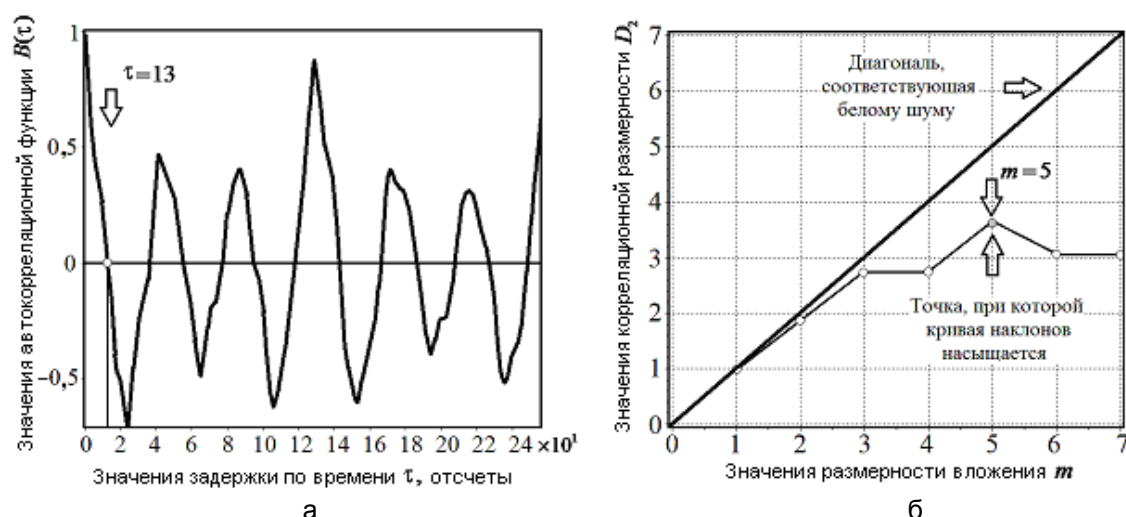


Рис. 1. Выбор оптимальных значений реконструкции: автокорреляционная функция объекта ОВ 1.1 (а); зависимость значений D_2 от m (б)

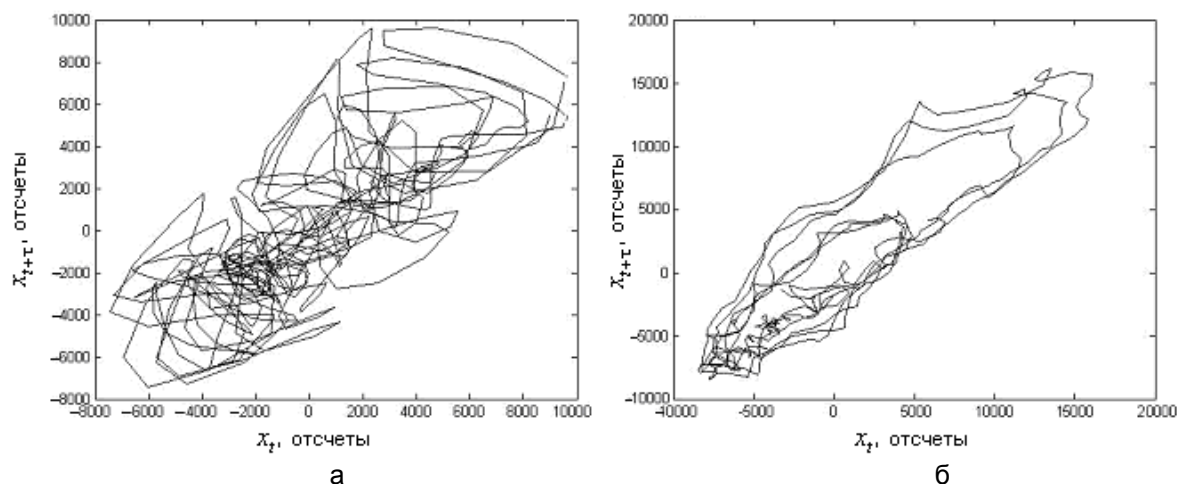


Рис. 2. Аттракторы фонемы «и» ОВ 1.2: радость (а); нейтральное состояние (б)

Предложен новый признак, определяемый по результатам реконструкции, который существенно снижает размерность описаний речевых образцов и позволяет осуществлять количественно сравнение аттракторов – усредненный максимальный вектор реконструкции аттрактора по четырем квадрантам $R_{\max}^{all}$. Вначале находится первый вектор реконструкции в первом квадранте $R_1^1 = \sqrt{x_i^2 + x_{i+\tau}^2}$, где x_i – значение временного ряда в i -й момент времени, τ – временная задержка. Далее вычисляются оставшиеся n векторов в первом квадранте, в результате получается множество значений векторов реконструкции $R^1 = \{R_1^1, R_2^1, \dots, R_n^1\}$. Из множества R^1 выбирается максимальный вектор $R_{\max}^1$. Аналогично находятся максимальные вектора реконструкции аттрактора в других квадрантах $R_{\max}^2$, $R_{\max}^3$ и $R_{\max}^4$. Далее рассчитывается усредненный максимальный вектор реконструкции аттрактора по четырем квадрантам $R_{\max}^{all}$, который является новым количественным признаком для распознавания i -го речевого образца:

$$R_{\max}^{all}(i) = 0,25 \sum_{j=1}^4 R_{\max}^j(i), \quad i = 1, \dots, M, \quad (4)$$

где j – номер квадранта, i – номер речевого образца (предложение или фонема); $M = 18$ для ОВ 1.1; $M = 180$ для ОВ 1.2; $M = 120$ для ОВ 2.1; $M = 300$ для ОВ 2.2.

Количественная оценка реконструкций аттракторов на выборках речевых образцов разной длительности (таблица) выполнена с использованием следующих характеристик:

$$\bar{R}_{\max}^{all} = M^{-1} \cdot \sum_{i=1}^M R_{\max}^{all}(i), \quad \bar{R}_{\max}^j = M^{-1} \cdot \sum_{i=1}^M R_{\max}^j(i), \quad j = 1, \dots, 4. \quad (5)$$

Установлено, что как на уровне фраз (ОВ 1.1, ОВ 2.1), так и на уровне фонем (ОВ 1.2, ОВ 2.2) эмоция радости по сравнению с нейтральным состоянием характеризуется меньшим значением $\bar{R}_{\max}^{all}$ (4),

(5). Следует отметить тот факт, что образцы русскоязычной части корпуса с эмоцией радости (на всех уровнях) имеют приблизительно в два раза меньшее значение признака $\bar{R}_{\max}^{all}$.

Объекты	Эмоциональное возбуждение	Выборка	Признаки, отсчеты				
			$\bar{R}_{\max}^1$	$\bar{R}_{\max}^2$	$\bar{R}_{\max}^3$	$\bar{R}_{\max}^4$	$\bar{R}_{\max}^{all}$
Фразы (предложения)	Радость	ОВ 1.1	19596	18786	16229	18561	18293
		ОВ 2.1	28257	34587	29716	39138	32925
	Нейтральное состояние	ОВ 1.1	37536	35547	31384	38358	35706
		ОВ 2.1	31671	33045	34846	40882	35111
Фонемы (звуки)	Радость	ОВ 1.2	13067	7969	9456	5361	8963
		ОВ 2.2	11098	10208	11800	11762	11217
	Нейтральное состояние	ОВ 1.2	28387	13795	18267	9194	17411
		ОВ 2.2	15590	11801	18777	14434	15151

Таблица. Усредненный максимальный вектор реконструкций аттракторов $\bar{R}_{\max}^{all}$

Рекуррентный график

В 1987 г. Экман и соавторы [9] разработали так называемые рекуррентные графики (диаграммы), позволяющие исследовать m -размерную траекторию лагового пространства (1) посредством двухмерного представления ее рекуррентности (повторяемости траекторий по происшествии некоторого времени в пространстве реконструкции аттрактора). Рекуррентный график представляется в виде двумерной или треугольной (так как обе стороны от главной диагонали под углом $\pi/4$ являются симметричными) матрицы размером $N \times N$, по обеим осям которой откладывается время. Матрица заполнена черными и белыми точками (единицами и нулями), где черные точки обозначают наличие рекуррентности, а белые – отсутствие [10]:

$$R_{ij} = \theta(\epsilon_i - \|x_i - x_j\|), \quad i, j = 1, \dots, N, \quad (6)$$

где N – число рассматриваемых состояний x_i ; ϵ_i – радиус выбранной окрестности (расстояние от центра окрестности x_i до ее границы); $\|\bullet\|$ – норма.

Если точка траектории реконструкции аттрактора в момент времени x_j попадает в выбранную окрестность другой точки в момент x_i , то такие точки считаются рекуррентными, вследствие чего на рекуррентном графике появляется точка черного цвета с координатами x_{ij} , соответствующая единице, и наоборот [7]. Радиус выбранной окрестности ϵ_i (6) выбирается не более 10% от максимального значения диаметра восстановленной реконструкции аттрактора [8]. На рис. 3 приведены примеры рекуррентных графиков объектов ОВ 1.1.

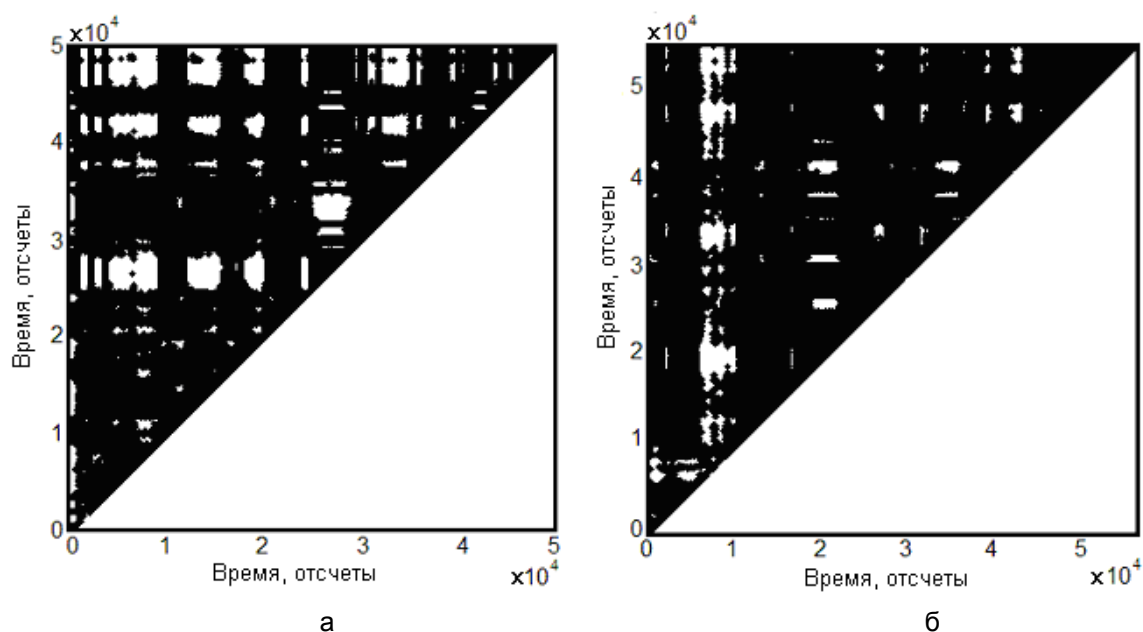


Рис. 3. Рекуррентные графики фраз: радость (а); нейтральное состояние (б)

Визуально установлено, что для объектов, выражающих эмоцию радости (рис. 3, а), характерна более контрастная топология по сравнению с нейтральным состоянием (рис. 3, б). Эмоция радости характеризуется более резкими изменениями динамики временного ряда и нестационарностью, вследствие чего в структуре рекуррентного графика появляются характерно выраженные белые зоны, указывающие на нерегулярность процесса. Текстура эмоции радости характеризуется более выраженными скоплениями горизонтальных и вертикальных линий, повторяющихся с некоторой периодичностью.

Заключение

В среде MATLAB в виде *m*-файлов реализован программный модуль распознавания эмоции радости человека по речевому сигналу, основанный на использовании двух качественных (y_i , R_{ij}) и пяти количественных ($\bar{R}_{\max}^1 - \bar{R}_{\max}^4$; $\bar{R}_{\max}^{all}$) признаков нелинейной динамики. При тестировании программного модуля на модельном корпусе эмоциональной речи точность распознавания, т.е. отнесения к одному из двух возможных классов (радость или нейтральное состояние), составила 93% для немецкоязычной и 95% для русскоязычной частей корпуса. Для сравнения отметим, что при распознавании образцов «нейтральной» и «агрессивной» речи из базы Emo-DB точность распознавания 96% получена при использовании 4 признаков, а 98% – при использовании 384 признаков [1]. В работе [2] классификатор, построенный для этой же базы Emo-DB, решал задачу разделения двух классов образцов речи (нормальное состояние и отклонение от него, возникающее у человека, испытывающего различные эмоции). Точность классификации составила 97 % при использовании 211 признаков и 87 % – при 15 признаках. Предлагаемый набор параметров аппарата нелинейной динамики после соответствующей адаптации будет использоваться для формирования динамической модели, отображающей взаимосвязь эмоционального состояния человека с характеристиками речевого сигнала.

Литература

1. Давыдов А.Г., Киселев В.В., Кочетков Д.С. Классификация эмоционального состояния диктора по голосу: проблемы и решения // Труды международной конференции «Диалог 2011». – М.: РГТУ, 2011. – С. 178–185.
2. Лукьяница А.А., Шишкин А.Г. Автоматическое определение изменений эмоционального состояния по речевому сигналу // Речевые технологии. – М.: Народное образование, 2009. – № 3. – С. 60–76.
3. Сидоров К.В., Филатова Н.Н. Анализ признаков эмоционально окрашенной речи // Вестник Тверского государственного технического университета. – Тверь: ТвГТУ, 2012. – Вып. 20. – С. 26–31.
4. Сидоров К.В., Филатова Н.Н., Калюжный М.В. Модельный русскоязычный корпус эмоциональной речи // Приоритетные направления развития науки и технологий: доклады XI всероссийской научн.-техн. конф. – Тула: Инновационные технологии, 2012. – С. 115–117.
5. Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss B. A Database of German Emotional Speech // Proc. Intern. Conf. Interspeech. – Lissabon, 2005 [Электронный ресурс]. – Режим доступа: <http://pascal.kgw.tu-berlin.de/emodb/index-1280.html>, свободный. Яз. англ. (дата обращения 10.07.2012).
6. Сидоров К.В., Филатова Н.Н. Алгоритм автоматической генерации речевых объектов // Сборник материалов I Международной научн.-практ. конф. «Технические науки – основа современной инновационной системы». – Ч. 1. – Йошкар-Ола, 2012. – С. 118–120.
7. Сидоров К.В. Диагностика эмоционального состояния диктора на основе рекуррентного анализа речевого сигнала // Междисциплинарные исследования в науке и образовании. – 2012. – № 1 Sp. – [Электронный ресурс]. – Режим доступа: <http://www.es.rae.ru/mino/157-702>, свободный. Яз. рус. (дата обращения 10.07.2012).
8. Горшков В.А., Касаткин С.А. Идентификация временных рядов авиационных событий методами и алгоритмами нелинейной динамики. – М.: Бланк Дизайн, 2008. – 208 с.
9. Eckmann J.P., Kamphorst S.O., Ruelle D. Recurrence Plots of Dynamical Systems // Europhys. Lett. 5. – 1987. – P. 973–977.
10. Киселев В.Б. Рекуррентный анализ – теория и практика // Научно-технический вестник СПбГУ ИТМО. – 2006. – № 29. – С. 118–127.

Сидоров Константин Владимирович – Тверской государственный технический университет, аспирант, bmisidorov@rambler.ru, bmisidorov@mail.ru
Филатова Наталья Николаевна – Тверской государственный технический университет, доктор технических наук, профессор, nfilatova99@mail.ru

УДК 004.85

МЕТОД ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ЭВОЛЮЦИОННЫХ АЛГОРИТМОВ С ПОМОЩЬЮ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

А.С. Буздалова, М.В. Буздалов

Предлагается метод скалярной оптимизации, основанный на применении эволюционных алгоритмов, контролируемых с помощью обучения с подкреплением. Обучение применяется для динамического выбора наиболее эффективной функции приспособленности для каждого вновь генерируемого поколения эволюционного алгоритма. Представлены результаты эксперимента по решению модельной задачи H-IFF с помощью предлагаемого метода. Проведено сравнение разработанного метода с методами многокритериальной оптимизации. Эксперименты показывают, что предлагаемый метод позволяет повысить эффективность работы эволюционных алгоритмов.

Ключевые слова: скалярная оптимизация, многокритериальная оптимизация, обучение с подкреплением, эволюционные алгоритмы, H-IFF.

Введение

Существуют различные способы повышения эффективности скалярной оптимизации. Некоторые из них основаны на использовании вспомогательных критериев. Например, задача скалярной оптимизации может быть преобразована в задачу многокритериальной оптимизации путем разработки дополнительных критериев, обладающих определенными заранее заданными свойствами, что позволяет избежать остановки поиска решения в локальном оптимуме [1]. Также в качестве источника вспомогательных критериев может выступать предметная область [2]. В этом случае свойства критериев чаще всего заранее не известны, причем они могут меняться в зависимости от того, на каком этапе находится процесс оптимизации. Важно отметить, что в настоящей работе задача оптимизации самих вспомогательных критериев не ставится. В то же время в традиционной теории многокритериальной оптимизации одинаково важны все критерии [3]. В поставленной задаче должен быть оптимизирован только один целевой критерий.

В данной работе предлагается метод повышения эффективности скалярной оптимизации с использованием вспомогательных критериев. Предполагается, что набор критериев задан заранее и об их свойствах ничего не известно. Таким образом, возникает задача скалярной оптимизации со вспомогательными критериями. Задача решается с применением эволюционного алгоритма (ЭА) (evolutionary algorithm, EA) [4], настраиваемого во время выполнения с помощью обучения с подкреплением (reinforcement learning, RL) [5, 6]. В дальнейшем предлагаемый метод будет называться EA+RL.

Метод EA+RL позволяет выбирать из заранее подготовленного набора наиболее эффективную функцию приспособленности (ФП), соответствующую критерию оптимизации, для генерации каждого последующего поколения эволюционного алгоритма. В других существующих методах настройки эволюционных алгоритмов обычно настраиваются вещественные параметры фиксированной ФП, причем настройка ФП освещена в литературе в меньшей степени, чем настройка иных параметров ЭА [7, 8].

Новизна предлагаемого подхода заключается в применении обучения с подкреплением для настройки ЭА. Обучение с подкреплением является современной развивающейся технологией, применимость которой в различных областях человеческой деятельности находится в процессе исследования [6]. Насколько известно авторам, существуют лишь две работы, в которых исследуется возможность использования обучения с подкреплением для настройки ЭА [8, 9]. В обеих работах рассматривается настройка вещественных параметров, таких как, например, вероятность мутации или размер поколения. Данная работа вносит вклад в исследование применимости обучения с подкреплением к настройке ФП.

В предыдущей работе авторов настоящей статьи [10] предложен прототип разработанного метода, предназначенный для решения конкретной модельной задачи с помощью настройки генетического алгоритма (ГА). Результаты, представленные в той работе, подтверждают, что разрабатываемый метод позволяет динамически выбирать наиболее эффективную ФП. В настоящей работе формулируется задача скалярной оптимизации со вспомогательными критериями, что позволяет дать обобщенное описание предлагаемого метода EA+RL, применимое для решения любой задачи, сводящейся к сформулированной. Эффективность предлагаемого метода протестирована на задаче оптимизации функции H-IFF, применяющейся для тестирования ГА, а также для иллюстрации методов повышения эффективности скалярной оптимизации путем сведения ее к многокритериальной оптимизации. Проведено сравнение EA+RL с упомянутыми методами.

Задача оптимизации со вспомогательными критериями

Рассмотрим формализацию задачи скалярной оптимизации со вспомогательными критериями. Обозначим как W дискретное пространство, в котором осуществляется поиск решений. Пусть $X \subset W$ – множество допустимых решений, содержащихся в пространстве поиска. Определим целевой критерий $g: W \rightarrow R$. Определим множество H , состоящее из k вспомогательных критериев: $H = \{h_i\}_{i=1}^k, h_i: W \rightarrow R$. Целью описываемой задачи является максимизация целевого критерия g с использованием вспомога-

тельных критериев H для ускорения процесса оптимизации: $g(x) \rightarrow \max_{x \in X}$. Решением задачи является $x^* \in X : g(x^*) \geq g(x), \forall x \in X$.

В общем случае характер корреляции между целевым и вспомогательными критериями неизвестен. Однако на практике часто возникает ситуация, при которой некоторые вспомогательные критерии коррелируют с целевым, по крайней мере, на некоторых этапах процесса оптимизации [2]. Предположение о том, что некоторые вспомогательные критерии обладают подобными полезными свойствами, позволяет использовать их для ускорения процесса оптимизации.

Задача обучения с подкреплением

Опишем задачу повышения эффективности ЭА, решающего задачу скалярной оптимизации со вспомогательными критериями, как задачу обучения с подкреплением [5]. Для этого достаточно задать множество действий агента A , способ определения состояний среды $s \in S$ и функцию вознаграждения $K : S \times A \rightarrow X \subseteq R$.

Будем обозначать особи, выращиваемые ЭА, как x . Пусть G_i – i -ое поколение. Множество действий A соответствует множеству функций приспособленности, состоящему из g – целевой ФП и элементов множества H – вспомогательных ФП. Применение действия реализуется как выбор некоторой ФП $f_i \in A$ в качестве функции, используемой для оценки приспособленности особей ЭА, и формирования поколения $G_i : A = H \cup g$.

Введем обозначение для лучшей особи поколения G_i , обладающей максимальным значением выбранной для этого поколения ФП $f_i : z_i = \arg \max_{x \in G_i} f_i(x)$. Также введем обозначение для нормированной разности значений некоторой ФП, вычисленной на лучших особях двух последовательных поколений: $\Delta(f, i) = \frac{f(z_i) - f(z_{i-1})}{f(z_i)}, f \in A$.

Каждому поколению ЭА поставим в соответствие состояние среды. Состояние s_i , соответствующее поколению G_i , представляет собой вектор ФП $f \in A$, упорядоченный по убыванию значений нормированных разностей $\Delta(f, i) : s_i = \langle f_1, f_2 \dots f_{k+1} \rangle, \Delta(f_1, i) \geq \Delta(f_2, i) \geq \dots \geq \Delta(f_{k+1}, i)$. В том случае, если для некоторых f_a, f_b значение $\Delta(f_a, i)$ совпадает со значением $\Delta(f_b, i)$, функции f_a, f_b располагаются в заранее установленном порядке. Например, пусть число вспомогательных ФП $k=2$ и в некотором поколении G_i выполняется неравенство $\Delta(h_2, i) = \Delta(g, i) > \Delta(h_1, i)$. Тогда соответствующее состояние среды может иметь вид $s_i = \langle h_2, g, h_1 \rangle$ или $s_i = \langle g, h_2, h_1 \rangle$ в зависимости от начальной договоренности.

В заключение определим функцию вознаграждения $K : S \times A \rightarrow R$, которая вычисляется после выбора действия f_i в состоянии s_{i-1} и генерации поколения $G_i : K(s_{i-1}, f_i) = g(z_i) - g(z_{i-1})$. Таким образом, вознаграждение зависит от разности значений целевой ФП, посчитанной на лучших особях двух последовательных поколений. Значение вознаграждения наиболее высоко, когда целевая ФП растет. Заметим, что в обучении с подкреплением целью агента является максимизация суммарной награды, причем для ряда алгоритмов обучения с подкреплением доказана их сходимость к оптимальной стратегии поведения [11]. Следовательно, задача обучения с подкреплением определена таким образом, что оптимальные действия агента будут приводить к максимизации прироста целевой ФП.

Описание алгоритма EA+RL

Предлагаемый метод позволяет управлять ходом выполнения эволюционного алгоритма путем назначения текущей ФП для каждого вновь сгенерированного поколения. Можно выделить две независимые сущности, составляющие основу метода: модуль обучения и эволюционный алгоритм. Будем называть эволюционный алгоритм средой обучения. Модулю обучения могут быть переданы награда и состояние среды. Он способен сообщать действие, которое необходимо применить к среде. В листинге представлен псевдокод разработанного алгоритма.

1. Установить номер текущего поколения: $i \leftarrow 0$
2. Сгенерировать начальное поколение G_0
3. ПОКА (условие останова ЭА не выполнено)
4. Вычислить состояние s_i и передать его модулю обучения
5. Получить ФП для следующего поколения f_{i+1} из модуля обучения
6. Сгенерировать следующее поколение G_{i+1}

Модуль обучения может быть реализован на основе произвольного алгоритма обучения с подкреплением и взаимодействовать с произвольным эволюционным алгоритмом. В ходе выполнения работы было реализовано четыре различных алгоритма обучения: Q-learning [6], Delayed Q-learning [11], Dyna [5] и R-learning [5]. Для обозначения различных реализаций предлагаемого метода EA+RL будем заменять в названии метода «EA» на название конкретного эволюционного алгоритма, «RL» – на название алгоритма обучения с подкреплением. Например, если с помощью предлагаемого метода реализуется контроль над ГА с помощью алгоритма обучения Q-learning, то соответствующая реализация метода будет называться ГА+Q-learning.

Модельная задача H-IFF

Определим задачу скалярной оптимизации функции H-IFF (Hierarchical-if-and-only-if function) [1]. Пространство поиска состоит из битовых строк $B = b_1b_2...b_l$ фиксированной длины l . Требуется максимизировать функцию H-IFF:

$$f(B) = \begin{cases} 1, & |B| = 1; \\ |B| + f(B_L) + f(B_R), & |B| > 1 \wedge (\forall i \{b_i = 0\} \vee \forall i \{b_i = 1\}); \\ f(B_L) + f(B_R), & \text{иначе.} \end{cases}$$

Функция задана таким образом, что существует два оптимальных решения: строка, полностью состоящая из единиц, и строка, полностью состоящая из нулей. Особенностью задачи является то, что поиск ее оптимального решения с помощью эволюционных алгоритмов часто останавливается в локальном оптимуме. Существует подход к решению этой проблемы, при котором скалярная задача оптимизации H-IFF заменяется многокритериальной задачей оптимизации функции MH-IFF [1]. Вместо исходной функции f вводятся функции f_0 и f_1 :

$$f_n(B) = \begin{cases} 0, & |B| = 1 \wedge b_1 \neq n; \\ 1, & |B| = 1 \wedge b_1 = n; \\ |B| + f_n(B_L) + f_n(B_R), & |B| > 1 \wedge \forall i \{b_i = n\}; \\ f_n(B_L) + f_n(B_R), & \text{иначе.} \end{cases}$$

Затем проводится максимизация предложенных функций с помощью алгоритмов многокритериальной оптимизации. Этот подход позволяет найти решения с более высокими значениями исходной функции, чем подход, основанный на скалярной оптимизации.

Задача максимизации функции H-IFF может быть представлена как задача скалярной оптимизации целевой функции $g = f$ со вспомогательными критериями $H = \{f_0, f_1\}$. Подобное представление задачи позволяет использовать предлагаемый метод для повышения эффективности эволюционных алгоритмов, применяемых для ее решения.

Описание и результаты эксперимента

В ходе эксперимента было реализовано решение задачи оптимизации H-IFF предлагаемым методом. Использовались два различных эволюционных алгоритма: генетический алгоритм (ГА) и $(1 + m)$ -эволюционная стратегия (ЭС). В ГА с вероятностью 70% применялся оператор одноточечного скрещивания и оператор мутации, инвертирующий каждый бит каждой особи с вероятностью $2 / l$. В ЭС оператор мутации инвертировал один бит каждой особи, выбранный случайным образом.

Параметры эксперимента соответствовали параметрам, примененным в работе [1], что позволяет сравнить новые результаты с результатами, полученными ее авторами. Длина особи составляла 64 бита. Соответствующее максимально возможное значение H-IFF равно 448. В табл. 1 представлены результаты оптимизации функций H-IFF и MH-IFF с помощью алгоритмов скалярной и многокритериальной оптимизации соответственно. Результаты отсортированы по среднему значению целевой ФП лучших особей, полученных в результате 30 запусков соответствующих алгоритмов. Вычисления запускались на фиксированное число поколений, равное 500000. Успешными считаются запуски, в которых была выращена особь с максимальной приспособленностью. Алгоритмы 1, 2, 4, 5, 7 реализованы с помощью предлагаемого метода с использованием различных алгоритмов обучения. Результаты 3, 6, 9, 11 получены авторами статьи, причем алгоритмы 3 и 6 (PESA и PAES) являются алгоритмами многокритериальной оптимизации. Можно видеть, что предлагаемый метод в случае использования алгоритма обучения R-learning [5] позволяет преодолеть проблему остановки в локальном оптимуме столь же эффективно, как и метод PESA, и более эффективно, чем метод PAES.

Также был проведен эксперимент, показавший, что если среди вспомогательных ФП есть мешающая ФП, оптимизация по которой ведет к убыванию целевой ФП, предлагаемый метод по-прежнему эффективен, с его помощью удастся вырастить оптимальную особь в 92% запусков. Однако алгоритмы

многокритериальной оптимизации в этом случае не позволяют выращивать особи с максимальным значением целевой ФП, так как они оптимизируют все предложенные критерии, в том числе мешающий.

№	Алгоритм	Лучшее значение	Среднее значение	σ	% успешных запусков
1	(1+10)-ЭС + R-learning	448	448,00	0,00	100
2	ГА + R-learning	448	448,00	0,00	100
3	PESA	448	448,00	0,00	100
4	ГА + Q-learning	448	435,61	32,94	87
5	ГА + Dyna	448	433,07	38,07	80
6	PAES	448	418,13	50,68	74
7	ГА + Delayed QL	448	397,18	49,16	53
8	ГА + Random	384	354,67	29,24	0
9	DCGA	448	323,93	26,54	3
10	ГА	384	304,53	27,55	0
11	SHC	336	267,47	29,46	0
12	(1+10) -ЭС	228	189,87	17,21	0

Таблица 1. Результаты оптимизации H-IFF и MH-IFF

В табл. 2 отдельно рассмотрена оптимизация H-IFF с использованием ЭС. Применяемая ЭС устроена таким образом, что решает задачу весьма неэффективно: ни в одном из запусков не удастся получить особь с максимальной приспособленностью. Однако применение предлагаемого метода позволяет добиться выращивания оптимальной особи в 73% запусков в случае использования наименее эффективной (1+1)-ЭС и в 100% запусков в остальных рассмотренных случаях.

№	Алгоритм	Лучшее значение	Среднее значение	σ	% успешных запусков
1	(1+10)-ЭС + R-learning	448	448,00	0,00	100
2	(1+10)-ЭС	228	189,87	17,21	0
3	(1 + 5)-ЭС + R-learning	448	448,00	0,00	100
4	(1 + 5)-ЭС	216	179,07	16,99	0
5	(1 + 1)-ЭС + R-learning	448	403,49	59,48	73
6	(1 + 1)-ЭС	188	167,07	11,98	0

Таблица 2. Результаты оптимизации H-IFF с помощью эволюционных стратегий.
Алгоритмы 1, 3, 5 реализованы с применением предлагаемого метода

Заключение

Предложен метод, повышающий эффективность скалярной оптимизации со вспомогательными критериями. Метод основан на выборе функции приспособленности эволюционного алгоритма с помощью обучения с подкреплением. Работа вносит вклад в исследование применимости обучения с подкреплением для настройки эволюционных алгоритмов. В ходе эксперимента подтверждена эффективность метода, а также проведено его сравнение с методами многокритериальной оптимизации. Предлагаемый метод, примененный к (1+m) эволюционным стратегиям для решения задачи оптимизации функции H-IFF, позволяет получать особи с максимальной возможной приспособленностью в 73–100% запусков, в то время как с помощью эволюционных стратегий без обучения не удастся вырастить оптимальную особь.

Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы.

Литература

1. Knowles J.D., Watson R.A., Corne D. Reducing Local Optima in Single-Objective Problems by Multi-objectivization // Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization EMO '01. – London, UK: Springer -Verlag. – 2001. – P. 269–283.

2. Буздалов М.В. Генерация тестов для олимпиадных задач по теории графов с использованием эволюционных алгоритмов. Магистерская диссертация. СПбГУ ИТМО, 2011 [Электронный ресурс]. – Режим доступа: <http://is.ifmo.ru/papers/2011-master-buzdalov/>, свободный. Яз. рус. (дата обращения 21.06.2012).
3. Лотов А.В., Поспелова И.И. Многокритериальные задачи принятия решений: Учебное пособие. – М.: МАКС Пресс, 2008. – 197 с.
4. Luke S. Essentials of Metaheuristics [Электронный ресурс]. – Режим доступа: <http://cs.gmu.edu/~sean/book/metaheuristics/>, свободный. Яз. англ. (дата обращения 21.06.2012).
5. Kaelbling L.P., Littman M.L., Moore A.W. Reinforcement Learning: A Survey // Journal of Artificial Intelligence Research. – 1996. – V. 4. – P. 237–285.
6. Gosavi A. Reinforcement Learning: A Tutorial Survey and Recent Advances // INFORMS Journal on Computing. – 2009. – V. 21. – № 2. – P. 178–192.
7. Eiben A.E., Michalewicz Z., Schoenauer M., Smith J.E. Parameter Control in Evolutionary Algorithms // In Parameter Setting in Evolutionary Algorithms. – 2007. – P. 19–46.
8. Müller S., Schraudolph N.N., Koumoutsakos P.D. Step Size Adaptation in Evolution Strategies using Reinforcement Learning // Proceedings of the Congress on Evolutionary Computation, IEEE. – 2002. – P. 151–156.
9. Eiben A.E., Horvath M., Kowalczyk W., Schut M.C. Reinforcement Learning For Online Control Of Evolutionary Algorithms // Proceedings of the 4th International Conference On Engineering Self-Organising Systems ESOA'06. – Springer -Verlag, Berlin, Heidelberg, 2006. – P. 151–160.
10. Афанасьева А.С., Буздалов М.В. Выбор функции приспособленности особей генетического алгоритма с помощью обучения с подкреплением // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 1 (77). – С. 77–81.
11. Strehl A.L., Li L., Wiewora E., Langford J., Littman M.L. PAC Model-Free Reinforcement Learning // ICML'06: Proceedings of the 23rd International Conference On Machine Learning. – 2006. – P. 881–888.

Буздалова Арина Сергеевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, aduzdalova@gmail.com

Буздалов Максим Викторович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, mbuzdalov@gmail.com

УДК 537.29

**ИССЛЕДОВАНИЕ ЭФФЕКТА ПЕРЕМЕЩЕНИЯ НАНОЧАСТИЦ
ПОД ДЕЙСТВИЕМ ПОЛЕЙ И ЗАРЯДОВ, СОЗДАВАЕМЫХ
СФОКУСИРОВАННЫМ ЭЛЕКТРОННЫМ ПУЧКОМ****Ф.Э. Комиссаренко, А.И. Денисюк**

Исследуется эффект перемещения диэлектрических наночастиц, наблюдаемый в электронном микроскопе. В проведенных экспериментах наблюдалось перемещение наночастиц полистирола размером 120 нм по подложке и их фиксация на острие подведенной металлической иглы, заряжающейся под электронным пучком. Проведен анализ механизмов взаимодействия в системе игла—частица—подложка под электронным пучком. Результаты моделирования и расчетов коррелируют с экспериментально наблюдаемыми явлениями. Наблюдаемый эффект может служить основой нового метода прецизионного перемещения нанообъектов.

Ключевые слова: электронная микроскопия, методы прецизионного перемещения наночастиц.

Введение

Наряду с развитием методов создания наноструктур, важной задачей нанотехнологий является прецизионное перемещение и фиксация нанообъектов. Метод оптического (лазерного) пинцета является исторически первой реализацией прецизионного перемещения микро- и наночастиц. В настоящее время этот метод [1, 2] достаточно развит и с его помощью можно создавать различные устройства опто- и наноэлектроники, биологические и химические сенсоры. В основе метода лежит взаимодействие электрического поля световой волны с индуцированным дипольным моментом микро- или наночастиц, в результате этого частица смещается вдоль градиента электрического поля волны. Данный метод весьма успешно работает при прецизионном перемещении микрочастиц, однако он не вполне пригоден для манипулирования наночастицами с размерами меньше длины волны управляющего лазера.

Прецизионное перемещение наночастиц возможно с помощью различных типов сканирующих зондовых микроскопов. В данных приборах для перемещения наночастиц могут быть задействованы различные механизмы взаимодействия зонда с частицей — механическое [3], химическое [4–6], электростатическое [7]. Более того, в зондовых микроскопах можно не только осуществлять прецизионное перемещение наночастиц по подложке, но и фиксировать наночастицу на острие зонда микроскопа. Тем самым можно создавать различные функциональные наноструктуры, такие как специализированные зонды для сканирующих зондовых микроскопов, обладающие уникальными характеристиками [4–6]. Существенным недостатком всех методов манипулирования наночастицами в зондовых микроскопах является отсутствие возможности непосредственной визуализации процесса и результата перемещения и фиксации наночастицы. Такая возможность реализуется в электронном микроскопе, снабженном микроманипулятором. Однако существующие в настоящее время работы по манипулированию частицами в электронном микроскопе [8, 9] описывают, в основном, возможность перемещения и фиксации крупных частиц микронного и субмикронного размера и основываются на механическом взаимодействии микроманипулятора с образцом. Также необходимо отметить работу [10], в которой описан эффект перемещения наночастиц под действием электромагнитного поля, создаваемого сфокусированным электронным пучком. Однако описанное явление наблюдалось лишь для наночастиц, находящихся в расплавах.

В настоящей работе экспериментально и теоретически исследуется эффект перемещения наночастиц по подложке под действием электронного пучка. Наблюдаемый эффект может служить основой нового метода прецизионного перемещения нанообъектов, который важен как при изучении их свойств, так и при создании различных функциональных элементов на основе этих объектов [11]. Преимуществом данного метода является то, что весь процесс и результат перемещения визуализируются на электронных изображениях в микроскопе в реальном времени.

Эксперимент

Экспериментальные исследования проводились со сферическими наночастицами полистирола диаметром 120 нм. Частицы были нанесены из водного раствора на кремниевую подложку, которая затем была помещена в камеру электронного микроскопа Neon 40 EsB (фирмы Carl Zeiss). Микроскоп дополнительно оборудован механическим микроманипулятором. Для проведения экспериментов по перемещению наночастиц к подвижной части микроманипулятора была приклеена заостренная вольфрамовая игла, полученная с помощью электрохимической заточки вольфрамовой проволоки до радиуса скругления острия менее 100 нм. Важно, что вольфрамовая игла не была заземлена. Движение вольфрамовой иглы над подложкой с частицами полистирола можно было непосредственно отслеживать на электрон-

ном изображении (ускоряющее напряжение электронного пучка 5 кВ, ток 60 пА). Оказалось, что при подведении вольфрамовой иглы к наночастицам на подложке на расстояние порядка нескольких сотен нанометров одна или несколько частиц притягивались к игле и фиксировались вблизи ее острия (рис. 1).

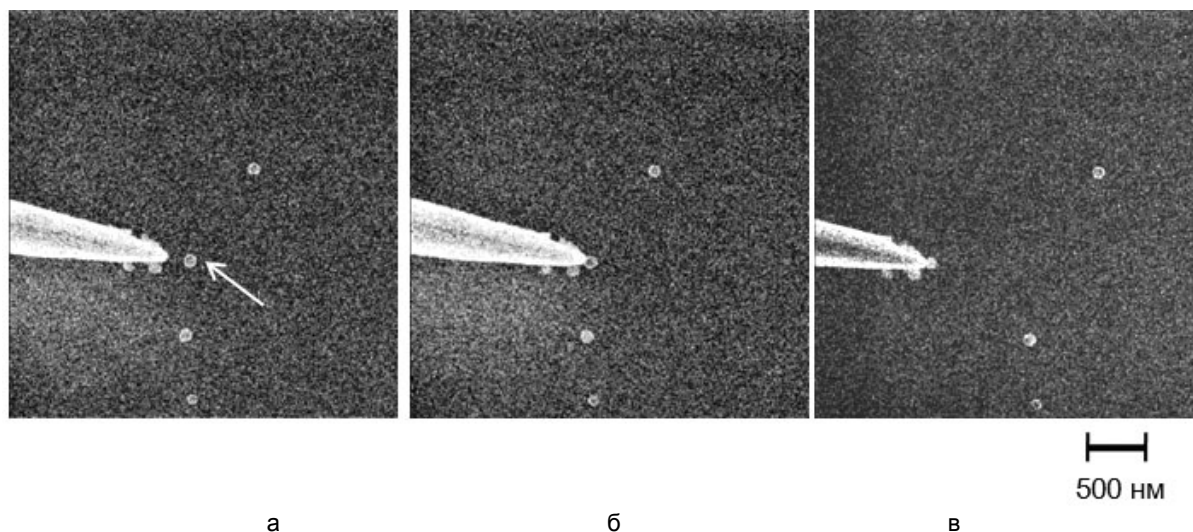


Рис. 1. Электронные изображения, иллюстрирующие фиксацию одиночной наночастицы полистирола на острие вольфрамовой иглы: подведение иглы к частице (а); фиксация частицы на острие иглы (б); отвод иглы (в)

Расчеты и моделирование

Обнаруженный эффект фиксации диэлектрической наночастицы на острие металлической иглы, по-видимому, объясняется электростатическим взаимодействием в системе игла–частица–подложка под электронным пучком. Диэлектрические наночастицы на подложке удерживаются силами Ван-дер-Ваальса. Незаземленная металлическая игла заряжается отрицательно под электронным пучком. Таким образом, вблизи острия иглы будет создано сильное градиентное электростатическое поле. Это поле поляризует диэлектрическую частицу и притягивает ее (движение диэлектрических частиц в градиентных электрических полях известно как диэлектрофорез [11]) (рис. 2, а). Однако нельзя не учесть тот факт, что диэлектрическая частица сама заряжается под электронным пучком отрицательно. Тогда будет иметь место отталкивание частицы от иглы при взаимодействии двух отрицательных зарядов (рис. 2, б). Наконец, еще один механизм связан с тем, что отрицательный заряд частицы создает свое изображение в металлической игле. Таким образом, заряженная частица будет притягиваться к своему изображению в металлической игле (рис. 2, в).

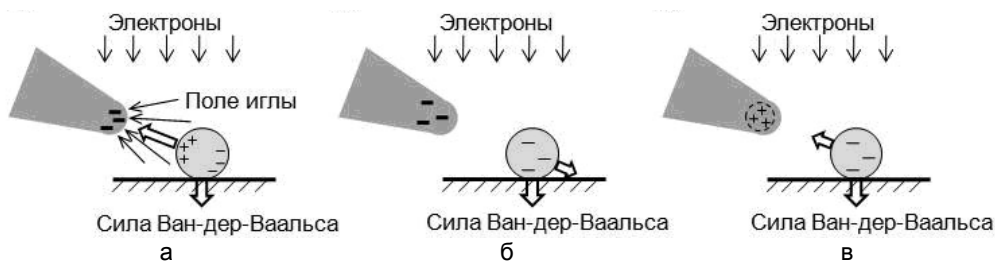


Рис. 2. Различные механизмы электростатического взаимодействия в системе игла–частица–подложка под электронным пучком: притяжение частицы градиентным полем заряженной иглы (а); отталкивание частицы от иглы при взаимодействии двух отрицательных зарядов (б); притяжение частицы к своему изображению в металлической игле (в)

Проведем оценку сил, действующих в системе игла–частица–подложка, согласно механизму, представленному на рис. 2, а. Исходные данные для расчета взяты из экспериментальных результатов. Для расчета силы Ван-дер-Ваальса, удерживающей частицу на подложке, воспользуемся формулой, представленной в работе [12]. Расчет показал, что сила Ван-дер-Ваальса в данном случае составляет около 6 нН.

Для оценки диэлектрофоретической силы рассчитаем заряд, накопленный вольфрамовой иглой, исходя из следующих данных: ток электронного пучка 50 пА, время облучения иглы 0,6 с (получено исходя из времени проведения эксперимента с учетом площади иглы), коэффициент поглощения электронов 0,6 (получено, исходя из результатов моделирования методом Монте-Карло рассеяния электронов с энергией 5 кэВ в вольфраме). Расчет показал, что заряд, накопленный вольфрамовой иглой, составляет

около 18 пКл. Следующим шагом является моделирование электростатического поля вблизи острия иглы. Эта задача решена методом конечных элементов, форма острия принята как круговой конус длиной 1 мм и радиусом основания 200 мкм, вершина конуса (острие иглы) имеет радиус скругления 50 нм. Результаты моделирования представлены на рис. 3, а (кривая значения напряженности в зависимости от расстояния от вершины иглы по ее оси), и рис. 3, б (картина поля вблизи острия иглы). На рис. 3, а, также представлены кривые, соответствующие радиусам скругления острия иглы 30 и 70 нм, а также двойному заряду, накопленному на игле (36 пКл) при радиусе скругления острия 50 нм.

При расчете диэлектростатической силы использована формула из обзора [12]. Согласно этой оценке диэлектростатическая сила, создаваемая иглой с зарядом 18 пКл и радиусом скругления острия 50 нм на расстоянии 100 нм от нее, составит порядка 7 нН, т.е. станет равной силе Ван-дер-Ваальса, удерживающей частицу на подложке. Данный расчет подтверждает возможность притяжения частицы к острию заряженной иглы. Расчеты также показывают, что диэлектростатическая сила, создаваемая заряженной иглой, слабо меняется при изменении радиуса острия иглы в диапазоне 30–70 нм, однако эта сила возрастает в два раза при двукратном увеличении накопленного заряда.

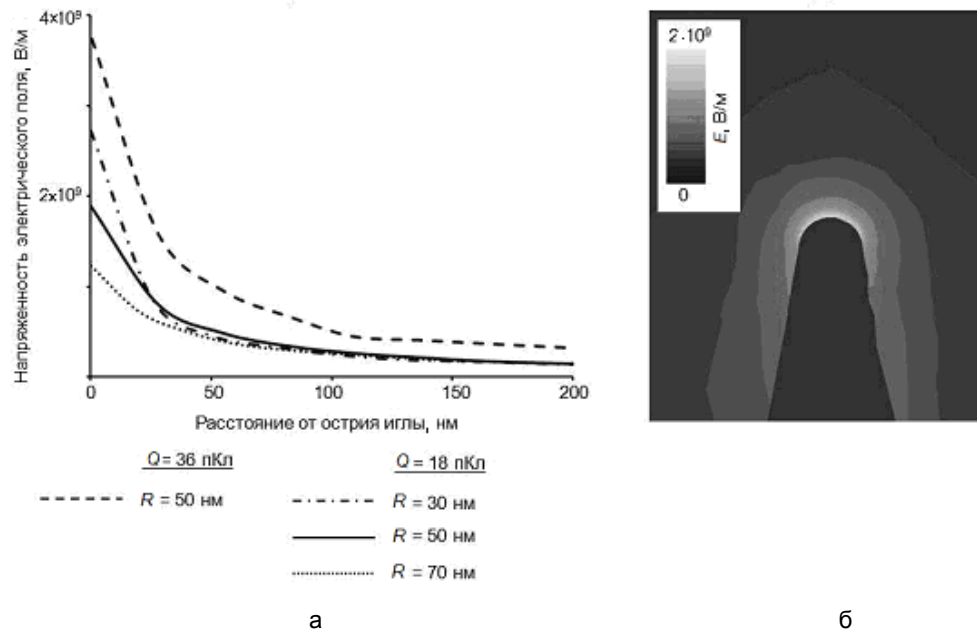


Рис. 3. Результаты расчетов электростатического поля вблизи острия заряженной металлической иглы: зависимость поля от расстояния от острия иглы (а); картина распределения электрического поля вблизи острия иглы (б)

Заключение

В работе теоретически и экспериментально изучен эффект перемещения наночастиц под действием полей и зарядов, создаваемых электронным пучком. Проведенные эксперименты показали возможность перемещения сферических наночастиц полистирола диаметром 120 нм и их фиксации на острие подведенной металлической иглы, заряжающейся под электронным пучком. Расчеты взаимодействия в системе игла–частица–подложка под электронным пучком подтвердили экспериментально наблюдаемые результаты. Описанный эффект может являться основой нового метода прецизионного перемещения нанообъектов, который будет полезен как для изучения их свойств, так и для создания различных функциональных наноматериалов.

Работа выполнена в рамках реализации и при финансовой поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (ГК № П1285).

Литература

1. Neuman K.C., Block S.M. Optical trapping // Review of scientific instruments. – 2004. – V. 75. – P. 2787–2809.
2. Dienerowitz M., Mazilu M., Dholakia K. Optical manipulation of nanoparticles: a review // Journal of Nanophotonics. – 2008. – V. 2. – P. 021875.
3. Sitti M., Hashimoto H. Controlled Pushing of Nanoparticles Modeling and Experiments // IEEE/ASME transaction on mechatronics. – 2000. – V. 5. – P. 199–211.

4. Ducker W.A., Senden T.J., Pashley R.M. Direct Measurement Of Colloidal Forces Using An Atomic Force Microscope // Nature. – 1991. – V. 353. – P. 239–241.
5. Лебедев Д.В., Чуکلанов А.П., Бухараев А.А., Дружинина О.С. Измерение модуля Юнга биологических объектов в жидкой среде с помощью специального зонда атомно-силового микроскопа // Письма в ЖТФ. – 2009. – С. 54–61.
6. Höppener C., Novotny L. Imaging of membrane proteins using antenna-based optical microscopy // Nanotechnology. – 2008. – V. 19. – P. 384012.
7. Grobelny J., Tsai D.-H., Kim D.-I., Pradeep N., Cook R.F., Zachariah M.R. Mechanism of nanoparticle manipulation by scanning tunnelling microscopy // Nanotechnology. – 2006. – V. 17. – P. 5519–5524.
8. Meyer E., Braun H.-G. Micro- and nanomanipulation inside the SEM // Journal of Physics: Conference Series. – 2008. – V. 126. – P. 012074.
9. Nakazato Y., Yuasa T., Sekine G., Miyazawa H., Jin M., Takeuchi S., Ariga Y., Murakawa M. Micromanipulation system using scanning electron microscope // Microsyst Technol. – 2009. – V. 15. – P. 859–864.
10. Oleshko V.P., Howe J.M. Are electron tweezers possible? // Ultramicroscopy. – 2011. – V. 111. – P. 1599–1606.
11. Kadaksham A.T.J., Singh P., Aubry N. Dielectrophoresis of nano-particles // Electrophoresis. – 2004. – V. 25. – P. 3625–3632.
12. Hamaker H.C. The London – van der Waals attraction between spherical particles // Physica. – 1937. – V. 4. – № 10. – P. 1058–1072.

Комиссаренко Филипп Эдуардович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, malkav86@mail.ru

Денисюк Андрей Игоревич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, доцент, aidenisyuk@gmail.com

УДК 54-185, 539.51

НЕОДНОРОДНОЕ УШИРЕНИЕ ЭЛЕКТРОННЫХ СПЕКТРОВ МОЛЕКУЛ КРАСИТЕЛЕЙ В НАНОПОРИСТОМ СТЕКЛЕ

Ю.Л. Колесников, А.Ф. Новиков

Исследовано влияние частоты возбуждающего излучения на параметры спектров флуоресценции адсорбированных молекул. Приведена трактовка наблюдаемых эффектов, исходя из того, что электронные спектры молекул красителей в нанопористом стекле неоднородно уширены. Обнаруженные зависимости в ряде случаев имеют немонотонный характер, поэтому неоднородное уширение не всегда может быть описано простой функцией Гаусса. Показано, что природа неоднородного уширения в этом случае связана прежде всего с вариациями потенциала межмолекулярных взаимодействий адсорбата с адсорбентом. Эти вариации определяются переменной локальной кривизной поверхности пор и сосуществованием различных устойчивых ориентаций молекул на поверхности.

Ключевые слова: спектроскопия молекулярной флуоресценции и поглощения; молекулы органических красителей; межмолекулярные взаимодействия; нанопористые силикатные стекла.

Введение

Методы осаждения различных веществ в порах матриц, изготовленных из силикатного стекла, известны достаточно давно [1–3]. Однако интерес к подобным системам не ослабевает, главным образом, в связи с интенсивным развитием наноконпозиционных технологий. Предприняты активные исследования органических и неорганических веществ в виде молекулярных кластеров и даже отдельных молекул, распределенных в объеме матриц, содержащих в себе систему пор с размерами вплоть до единиц нанометров [4–7].

Нанопористые стекла не всегда играют лишь роль пассивного носителя для внедренного в поры активного вещества. Как правило, состояние и свойства такого рода стекол оказывают существенное влияние на свойства системы в целом. К настоящему времени структура и свойства нанопористых стекол достаточно подробно изучены во многих работах [3–9]. Особый интерес представляют нанопористые стекла с удельной поверхностью до $100 \text{ м}^2/\text{см}^3$ и более: они обеспечивают диспергирование вещества внутри матрицы вплоть до отдельных молекул, так что появляется возможность исследовать свойства изолированных молекул.

Среди прочих методов исследования состояния сложных молекул на внутренней поверхности пор матрицы, в частности, молекул органических красителей, особое место принадлежит спектрально-оптическим методам, включая интерпретацию получаемых спектров люминесценции, поглощения и флуоресценции.

В отличие от состояния молекул в жидких растворах, молекулы в порах матрицы адсорбированы на поверхности диэлектрика при отсутствии растворителя. Гетерогенный характер таких систем оказывает существенное влияние на флуоресцентные свойства молекул красителей, закрепленных на поверхности пор, так что спектральное поведение этих молекул отличается от того, что имеет место в растворе. Одним из проявлений таких отличий является наблюдаемое уширение спектров флуоресценции.

Среди сравнительно недавних исследований, проведенных в этом направлении независимо от данной работы и с другими красителями, следует отметить [10–12]. Так, в [10,12] сообщается о наблюдаемой спектральной неоднородности, вызываемой взаимодействием адсорбированных молекул (эозина, антрацена, родамина 6Ж). Отмечается неоднородное уширение энергетических уровней электронов, приводящее к дисперсии спектральных полос и линий. Высказывается предположение о возможных вариациях угла между молекулами, закрепленными на поверхности пор стекла.

Целью данной работы являются наблюдение и интерпретация неоднородного уширения спектров флуоресценции молекул ряда красителей, что позволит оценить состояние и ориентацию молекул на поверхности неорганического диэлектрика, притом не плоской, а искривленной.

Методика эксперимента

Образцы пористого натриевоборосиликатного стекла ДВ-1М изготавливались с использованием поэтапной термической и химической обработки исходного монолитного натриевоборосиликатного стекла ДВ-1, следуя известной технологии [4, 5]. Исходное стекло представляло собой ликвидировавшее стекло (оксид натрия – 6,8%; оксид бора – 26,7%; диоксид кремния – 66,0%; остальное – прочие компоненты). В объеме стекла было обеспечено разделение щелочной и силикатной фаз. Затем образцы выдерживались в водном трехнормальном растворе соляной кислоты, в результате чего натриевоборатная фаза выщелачивалась и удалялась из объема стекла. Затем образцы достаточное время промывались в дистиллированной воде, после чего высушивались и отжигались на воздухе при температуре +550 °С. В результате в стекле создавалась сквозная система сообщающихся пор внутри силикатного по преимуществу каркаса.

Полученная пористая матрица пропитывалась раствором выбранного флуоресцентного красителя до достижения сорбционного равновесия (время пропитки – до 24 часов). Молекулы красителей вводились в поры матрицы из этанольного раствора, после чего образцы извлекались из раствора и нагревались на воздухе при +50 °С для испарения растворителя. Дополнительно проводилась вакуумная термообработка образцов в условиях +150 °С (10^{-4} Торр). В результате такой обработки молекулы флуоресцентных красителей оказывались закрепленными на поверхности нанопор и практически изолированными друг от друга (объемная концентрация – около 10^{-17} см⁻³).

Спектры поглощения нанопористых образцов и пропитывающих растворов снимались на спектрофотометрах U-3200 (Hitachi, Japan), Lambda19 (Perkin-Elmer, USA) и СФ-26 (ЛОМО, Россия) в спектральном диапазоне 300–800 нм. Спектры флуоресценции снимались на спектрофотометрах Fluorolog (USA) и ДФС-24 (ЛОМО, Россия).

Экспериментальные результаты

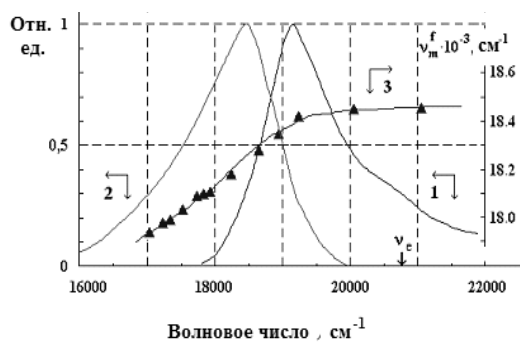
Предпринятые нами рентгеновские исследования в области малых углов рассеяния, а также рутинная обработка изотерм адсорбции в области капиллярной конденсации однозначно продемонстрировали практически моносферное распределение размеров пор внутри нанопористого стекла с максимумом в районе 8–9 нм [8, 9]. Рассчитанный фактор объемной пористости составлял примерно 28%, а удельная поверхность – 106 м²/см³. Исследуемые нанопористые стекла характеризуются изотермой IV типа (подтип А) по классификации Грегга [13]. Этот факт свидетельствует о присутствии в объекте пористой структуры в виде вытянутых сообщающихся нанокapилляров с открытыми концами. Такая структура обеспечивает, во-первых, проницаемость для молекул из окружающей жидкой и газообразной среды, во-вторых, оптическую прозрачность образцов в видимом диапазоне спектра.

Предварительно нами были исследованы и проанализированы спектры флуоресценции многих красителей, сорбированных нанопористыми матрицами. Было показано [14], что энергии связи между молекулами красителя и поверхностью значительно слабее, чем энергии химической связи. Учитывая это, а также обнаруженную нами обратимость сорбционных процессов [9], мы делаем вывод именно о физическом (не химическом) характере адсорбции молекул внутри пористого стекла. Об этом же свидетельствует и подобие формы спектральных кривых красителей в матрице и в растворе.

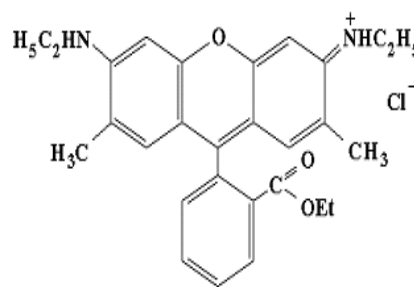
На рис. 1, а, приведены спектры поглощения (кривая 1 – длинноволновая полоса) и флуоресценции (кривая 2) красителя родамина 6Ж (Rhodamine 6G) в пористом стекле. Структурная формула молекулы представлена на рис. 1, б. Отметим, что положение максимумов полос смещается по оси частот относительно таковых для этанольных растворов, так что величина стоксовского сдвига уменьшается. На рис. 1, а (кривая 3), показана зависимость частоты ν_m^f , соответствующая спектральному максимуму флуоресценции, от частоты возбуждающего света $\nu_{вх}$. Как видно из рисунка, по мере перехода типа возбуждения

от стоксовского к антистоксовскому полюса флуоресценции монотонно сдвигается в длинноволновую сторону. С понижением возбуждающей частоты $\nu_{\text{вх}}$ при антистоксовском возбуждении интенсивность флуоресценции резко падает. Однако даже в этом случае сохраняется флуоресценция именно Rhodamine 6G, а не примесей, как это показано в ходе последующих экспериментов.

Нагрев образцов приводит к тушению флуоресценции в стоксовской полосе и к экспоненциальному росту флуоресценции – в антистоксовской. Возрастание флуоресценции во втором случае определяется экспоненциальным ростом заселенности на верхних колебательных уровнях основного синглетного состояния молекулы. Если бы флуоресцировали примеси, то имело бы место только тушение флуоресценции. Помимо этого, на характеристики флуоресценции закрепленных молекул оказывает влияние и сама поверхность пористой матрицы. Особенно отчетливо это влияние проявляется в спектрах, имеющих колебательную структуру. Например, на рис. 2 (кривые 1 и 2) представлены спектры флуоресценции и возбуждения другого красителя, Oxazine 1, в пористом стекле. Структурная формула молекулы представлена на рис. 2 справа. Эти спектры имеют выраженную колебательную структуру при низких температурах. Спектры возбуждения Oxazine 1, полученные при длинноволновой регистрации (например, на частотах ν_1 и ниже), идентичны друг другу и соответствуют спектру поглощения молекул этого красителя. Как известно [15], спектры поглощения и флуоресценции зеркально симметричны друг другу относительно частоты ν_{00} для чисто электронного перехода. На рис. 2, однако, эта симметрия нарушена из-за присутствия дополнительной, более коротковолновой полосы флуоресценции ν_3 . С другой стороны, при регистрации флуоресценции на частоте ν_2 спектр возбуждения сдвигается как целое на 500 см^{-1} в коротковолновую сторону. Разумеется, здесь мог быть снят спектр возбуждения только в его коротковолновой части. Как будет показано ниже, этот эффект, вероятнее всего, связан с наличием в пористом стекле двух подсистем адсорбированных молекул Oxazine 1.



а

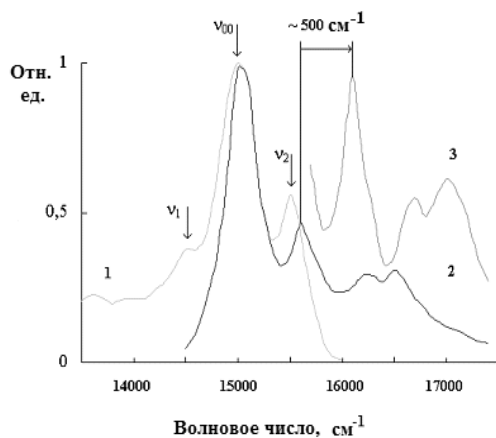


б

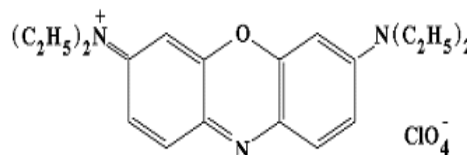
Рис. 1, а. Нормализованные спектры поглощения (1), флуоресценции (2) и зависимость частоты флуоресценции в максимуме ν_m^f от частоты возбуждения $\nu_{\text{вх}}$ (3) для Rhodamine 6G в нанопористом стекле при температуре 296 К. Спектр флуоресценции (2) снят при частоте возбуждения $\nu_{\text{вх}} = 20800 \text{ см}^{-1}$.

Концентрация красителя $N \approx 6 \cdot 10^{16} \text{ см}^{-3}$. Погрешность измерения частоты ν_m^f меньше 40 см^{-1} .

б. Структурная формула молекулы Rhodamine 6G



а



б

Рис. 2, а. Спектры флуоресценции (1) и возбуждения (2, 3) для Oxazine 1 в нанопористом стекле при температуре 80 К. Спектры возбуждения флуоресценции снимались на частотах $\nu_1 = 14500 \text{ см}^{-1}$ (2), $\nu_2 = 15500 \text{ см}^{-1}$ (3). Частота ν_{00} соответствует чисто электронному переходу.

б. Структурная формула молекулы Oxazine 1.

Обсуждение результатов

Для обсуждения полученных результатов была использована модель физического взаимодействия многоатомной молекулы с неплоской поверхностью твердотельного диэлектрика. При этом безразлично, является ли адсорбированная молекула ионом (как Rhodamine 6G и Oxazine 1) или нейтральной частицей (Coumarin 30 и Coumarin 7). Для расчета энергии взаимодействия в работе [14] использованы представления о парных потенциалах межмолекулярных сил с последующим их суммированием по частицам адсорбата и адсорбента. При этом молекула адсорбата аппроксимируется однородным трехосным эллипсоидом (см. рис. 3), а адсорбент рассматривается как однородный твердый диэлектрик с искривленной поверхностью. Эллипсоид наделяется электрическими характеристиками молекулы – тензором поляризуемости, вектором дипольного момента, эффективным электрическим зарядом (для ионов) и другими, ответственными за ван-дер-ваальсовское и кулоновское взаимодействия адсорбата и адсорбента.

Оказалось, что энергия взаимодействия многоатомной молекулы с твердой поверхностью как функция расстояния взаимодействия z имеет вид

$$\Phi(z) = - \left\{ C_1 \frac{\alpha}{(z+a_0)^3} \left[1 - \left(\frac{z}{z+a_0} \right)^2 \right]^{-2} + C_2 \frac{\mu}{z^3} + \frac{C_3}{z} \right\} \cdot \left(1 - \frac{z}{2R} \right)^{-3}, \quad (1)$$

где a_0 – онзагеровский радиус поверхностных частиц адсорбента, α и μ – поляризуемость и дипольный момент молекулы адсорбата; C_1 , C_2 и C_3 – константы, отвечающие за действие дисперсионных, индукционно-ориентационных и кулоновских сил соответственно. Множитель в квадратных скобках – поправка на неоднородную поляризуемость молекулы в поле ван-дер-ваальсовых сил; множитель в круглых скобках – поправка на искривленность поверхности радиуса R . Учет сил отталкивания производился введением идеализации «твердой стенки» с привлечением данных о ван-дер-ваальсовых радиусах атомов.

Расчет по формуле (1) позволяет определить форму потенциальных поверхностей для адсорбата. Для этого по структурной формуле молекулы изучаемого адсорбата нужно определить полуоси эллипсоида формы, а также эллипсоида поляризуемости по аддитивной схеме.

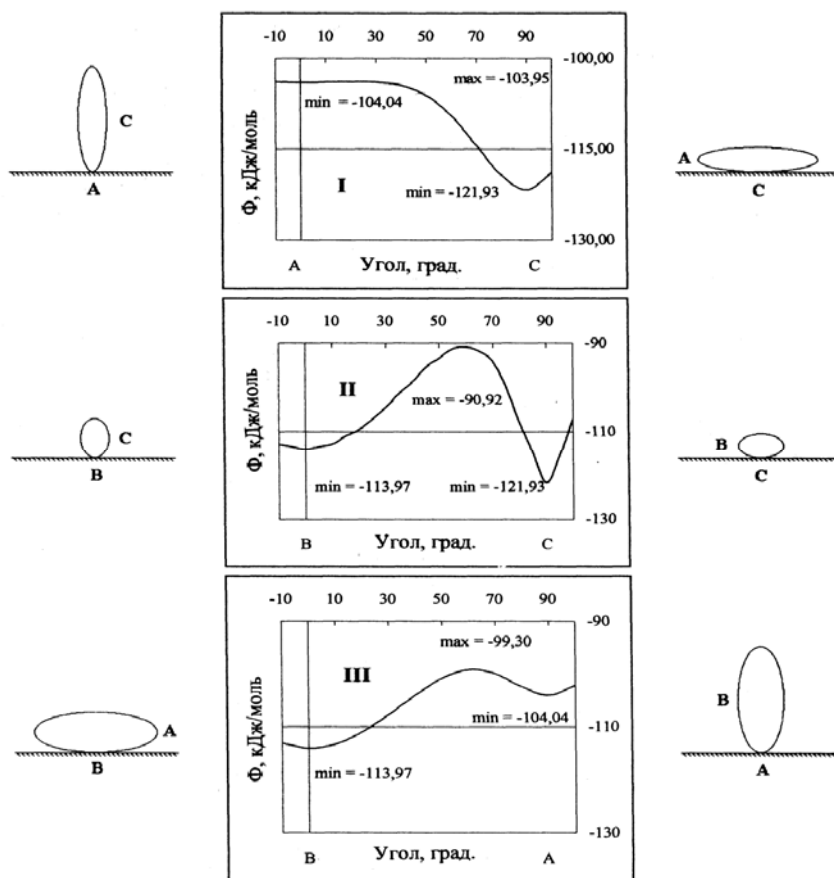


Рис. 3. Рассчитанные поперечные сечения потенциальных поверхностей для молекулы Oxazine 1. Ориентации эллипсоидов относительно поверхности поры представлены схематически как их сечения: I – ac; II – bc; III – ab; a – длинная ось эллипсоида, b – средняя, c – короткая

Три сечения потенциальной поверхности для молекулы Oxazine 1, рассчитанные по формуле (1), показаны на рис. 3 (I, II и III). Особенностью его взаимодействия с поверхностью адсорбента является наличие трех минимумов энергии, соответствующих точкам контакта эллипсоида с поверхностью. Более или менее глубокие минимумы потенциальных кривых появляются в случае контактирования в точках С (рис. 3 – I и II, справа) и В (рис. 3 – II и III, слева). В то же время минимум, соответствующий точке контакта А (рис. 3 – I, слева, и III, справа), слабо выражен при переходе молекулы в положение В (рис. 3 – III) и вовсе отсутствует при переходе к положению С (рис. 3 – I).

Поэтому представляется, что молекулы красителя Oxazine1 преимущественно адсорбируются в порах стекла в положениях В и С (рис. 3 – II). Это и формирует две подсистемы молекул, для которых частоты чисто электронного перехода отличаются на $\Delta\nu = 500 \text{ см}^{-1}$. Таким образом, пик при частоте ν_3 на высокочастотном крыле спектра флуоресценции (рис. 2) обусловлен флуоресценцией «синих» молекул (обладающих большей частотой чисто электронного перехода). Более длинноволновая часть спектра этих молекул маскируется более интенсивным спектром флуоресценции «красных» молекул (с меньшей частотой чисто электронного перехода).

Согласно формуле (1), разность энергий адсорбции Oxazine 1 в положениях С и В равна 670 см^{-1} (рис. 3 – II). Учитывая, что модельные расчеты проведены только для основного состояния молекулы, это неплохо согласуется с наблюдаемой в эксперименте величиной «скачкообразного» сдвига спектра возбуждения Oxazine1 в пористом стекле (см. рис. 2). С другой стороны, учитывая более сложную форму молекулы Rhodamine 6G, следует ожидать большего разнообразия ориентаций этой молекулы относительно поверхности пористого стекла. Вследствие этого имеется целый набор молекул Rhodamine 6G с разными частотами чисто электронного перехода. Поэтому при коротковолновом возбуждении флуоресценции ($\nu_e > 20000 \text{ см}^{-1}$) положение ее спектра не зависит от частоты возбуждения, так как возбуждаются все молекулы. Но по мере уменьшения частоты возбуждения начинает преобладать свечение «красных» центров, и наблюдается длинноволновый сдвиг полосы флуоресценции (рис. 1, кривая 3).

Заключение

Выполненные в работе исследования показывают, что электронные спектры адсорбатов испытывают значительное неоднородное уширение. Оно имеет сложную природу и вызвано одновременным действием ряда факторов. Наиболее важные среди них – неоднородность поверхности адсорбента (в частности, непостоянство локальной кривизны поверхности пор) и сосуществование различных устойчивых ориентаций молекул на поверхности.

Исследования такого рода молекулярных систем дают основу для более глубокого понимания физики процессов, оно же, в свою очередь, открывает пути для разработки все более совершенных активных элементов лазеров на красителях [16, 17] и сенсоров состава окружающей среды [9, 18].

Литература

1. Elmer T.H., Nordberg M.E. Survey on the Porous Glasses // Journal of American Ceramic Society. – 1970. – V. 53. – P.171.
2. Macedo P.B., Litovitz T.A. Method of Precipitation of Dopants in a Porous Silicate Glass. US Patent No. 4.110.096. Aug. 29, 1978.
3. Yanowski F., Heyer W. Poröse Gläser: Herstellung, Eigenschaften, Anwendung. –Leipzig: Springer, 1981. – 307 p.
4. Двухфазные стекла: структура, свойства, применения. / Под ред. Варшала В.Г. – Л.: Наука, 1991. – 276 с.
5. Мешковский И.К., Композиционные оптические материалы на основе пористых матриц (Монография). – СПб: СПбГУ ИТМО, 1998. – 332 с.
6. Enke D., Yanowski F., Schwieger W. Porous glass in 21st century. A short review // Microporous and Mesoporous Materials. – 2003. – V. 60. – № 1. – P. 19–30.
7. Алексашкина М.А., Вензель Б.И., Сватовская Л.Г. Пористые стекла как матрица для получения нанокомпозитов. // Физика и химия стекла. – 2005. – Т. 31. – № 3. – P. 361–368.
8. Novikov A.F. Characterization of the inner structure and surface of nanoporous sodium-borate-silicate glasses. // Optica Applicata. – 2005. – Vol. XXXV. – № 4. – P. 702–708.
9. Новиков А.Ф. Цвет молекул. Время. Цвет. (Сер. «Выдающиеся ученые НИУ ИТМО», вып. 16). – СПб: НИУ ИТМО, 2011. – 308 с.
10. Гордеева Ю.А. Спектроскопические исследования фотофизических процессов в структурах нанопористое стекло–адсорбированные сложные молекулы. Автореф. дисс. канд. физ.-мат. наук. – МГУ. – 2007.
11. Гордеева Ю.А., Салеский А.М. Влияние размеров пор силикатного стекла на степень ассоциации молекул красителей. // ЖПС. – 2006. – Т. 73. – № 6. – P. 824–827.

12. Антропова Т.В., Гордеева Ю.А., Рыжиков Б.Д., Салецкий А.М. Перенос энергии электронного возбуждения между разнотипными молекулами красителей в матрице пористого стекла. // ЖПС. – 2005. – Т. 72. – № 4. – Р. 446–449.
13. Gregg S.J., Sing K.S.W., Adsorption, Surface Area & Porosity. – L.– N.Y.: Academic Press, 1967. – 410 p.
14. Сечкарев А.В., Бегер В.Н., Земский В.И. Конфигурационные переходы многоатомных молекул, адсорбированных неоднородной поверхностью диэлектрика // Журнал физической химии. – 1993. – Т. 67. – № 2. – С. 400–407.
15. Lacovicz J.R., Principles of Fluorescence Spectroscopy. – N.Y.: Klew Academic Plenum Publishers, 1995. – 682 p.
16. Земский В.И., Колесников Ю.Л., Мешковский И.К. Физика и техника импульсных лазеров на красителях. (Сер. «Выдающиеся ученые СПбГУ ИТМО», вып. 11). – СПб: СПбГУ ИТМО, 2005. – 176 с.
17. Meshkovsky I.K., Zemskii V.I., Kolesnikov Yu.L. Active Dye Laser Microporous Componenets // Proc. SPIE. – 1995. – V. 2380. – P. 298.
18. Кабакова М.М., Колесников Ю.Л., Новиков А.Ф., Тиль Э.Р. Спектрально-оптические наноконпозиционные материалы для применения в химических сенсорах на основе моно- и полимолекулярных индикаторов // Труды Международной конференции «Фундаментальные проблемы оптики–2004». – СПб: СПб ГУ ИТМО, 2005. – С. 105–108.

Колесников Юрий Леонидович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, проректор, kolesnikov@mail.ifmo.ru

Новиков Александр Федорович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, alexfednov@gmail.com

УДК 621.315.592; 621.382.002

ПРИМЕНЕНИЕ ИМПУЛЬСНОГО ВОЛОКОННОГО ЛАЗЕРА ДЛЯ МИКРОСТРУКТУРИРОВАНИЯ СИСТЕМЫ SiO₂/Si

А.М. Скворцов, В.П. Вейко, Хуинь Конг Ту

Исследовано влияние облучения системы SiO₂/Si импульсным иттербиевым волоконным лазером ИЛИ-1-50 на морфологию и структурные свойства системы. Показано, что при лазерном облучении поверхности монокристаллической кремниевой пластины, покрытой тонким термически выращенным слоем SiO₂, происходят существенные структурные изменения, связанные с локализацией на поверхности кремния полос линий скольжения и сеток линий скольжения, образующихся в результате пересечения этих полос. Рассмотрена фоновая теория появления микроструктурного рельефа.

Ключевые слова: лазерное облучение, дислокации, линии скольжения, сетка линий скольжения, механические напряжения, пластическая деформация, рекристаллизация, наноконкомпозит.

Введение

В последнее время интенсивно развиваются технологии создания наноконкомпозитных материалов. Эти материалы проявляют уникальные свойства, существенно отличающиеся от свойства вещества в макроскопическом (объемном) состоянии. Особое место занимают наноконкомпозиты, содержащие наноразмерные частицы полупроводников – нанокластеры и «квантовые точки». К таким наноконкомпозитам, в первую очередь, относится материал, представляющий собой широкозонную диэлектрическую матрицу SiO₂ с внедренными нанокластерами кремния.

Имеется большое количество экспериментальных и теоретических работ, связанных с исследованиями различных методов микроструктурирования. Существенно меньше число работ по изучению влияния лазерного излучения на микроструктурирование кремния и систему диоксид кремния–кремний (SiO₂/Si). Краткий обзор методов микроструктурирования поверхности кремния приведен в [1]. За период, истекший после этой публикации, расширился диапазон исследований взаимодействия лазерного излучения с системой SiO₂/Si [2–5]. С 2008 г. в СПб НИУ ИТМО проводятся исследования по разработке методов формирования наноконкомпозитов с кластерами кремния в тонких пленках SiO₂. Эти пленки являются составной частью системы SiO₂/Si (кремниевая монокристаллическая пластина с пленкой диоксида кремния, полученной с помощью высокотемпературного окисления). Для формирования таких наноконкомпозитов использовался метод лазерного облучения системы SiO₂/Si, который, с точки зрения авторов, является наиболее перспективным и экономичным.

Экспериментальные образцы представляли собой термически окисленные пластины кремния марки КЭФ-4,5, рабочая поверхность которых совпадала с кристаллографической плоскостью (100). Толщина окисной пленки составляла около 100 нм. Облучение образцов производилось импульсным волоконным YLP-лазером: длина волны $\lambda = 1,06$ мкм; длительность импульса – 250 нс; частота следования импульсов –

200 Гц. Облучение экспериментальных образцов системы SiO_2/Si производилось при значениях мощности меньше критической, т.е. такой, при которой начиналось плавление кремния в отдельных микроскопических точках подложки (подплавление кремния). Плотность мощности составляла $(2-4) \cdot 10^3 \text{ Вт/см}^2$. Выбор режима облучения обуславливался желанием получить надежный и однозначный ответ на каждый из двух вопросов – как изменяются электрофизические параметры межфазной границы системы SiO_2/Si непосредственно в зоне лазерного воздействия и насколько локальным является это воздействие.

На облученных лазером экспериментальных образцах проводились измерения вольт-фарадных характеристик и фиксировались структурные особенности облученных и необлученных участков подложек. Структурные исследования проводились методом катодолюминесценции и просвечивающей электронной микроскопии. Полученные результаты свидетельствуют о наличии в пленках SiO_2 нанокластеров кремния [6–9].

В настоящей работе приведены результаты исследования микроstructuring системы SiO_2/Si импульсным иттербиевым волоконным лазером типа ИЛИ-1-50 с той же длиной волны, что и в предыдущих исследованиях ($\lambda = 1,06 \text{ мкм}$), при этом значительно увеличены энергия импульсов (до $1,0 \text{ мДж}$) и частота их следования (50 кГц). Сделана попытка объяснения природы микроstructuring.

Методика эксперимента

В качестве экспериментальных образцов использовались пластины монокристаллического кремния марок КЭФ-4,5, ориентированные в кристаллографической плоскости (100), и КДБ-10, ориентированные в кристаллографической плоскости (111). На пластинах методом термического окисления был выращен слой SiO_2 толщиной 150 нм .

Для облучения структур SiO_2/Si использовался импульсный иттербиевый волоконный лазер типа ИЛИ-1-50 с длиной волны $\lambda = 1062 \text{ нм}$ и случайной поляризацией. Номинальная выходная мощность лазера – 50 Вт , длительность импульсов – 120 нс . Лазер имеет две частоты следования импульсов – 50 и 100 кГц (в эксперименте – 50 кГц). Номинальная энергия в импульсе при максимальной выходной мощности 50 Вт составляет $1,0 \text{ мДж}$. Контроль облученных образцов проводился на оптическом микроскопе Axio Cam 1ss3 фирмы Carl Zeiss, укомплектованном цифровой видеокамерой высокого разрешения.

Результаты экспериментальных исследований

При проведении исследований влияния мощности сфокусированного лазерного пучка на модификацию системы SiO_2/Si выяснилось, что лазерное излучение мощностью $(2-4) \cdot 10^6 \text{ Вт/см}^2$, примененное в предыдущих экспериментах и позволявшее без видимого разрушения пленки кремния модифицировать структуру SiO_2 , в новом эксперименте приводит к разрушению пленки, а также к плавлению и испарению кремния (рис. 1, а), причем процесс разрушения происходит скачком, а диаметр разрушенной области (139 мкм) существенно превышает диаметр пятна в фокусе (84 мкм). По этой причине при исследовании режимов облучения структур авторами был выбран метод облучения в сходящихся лучах [1]. На рис. 1, б, приведена фотография области, облученной по этому методу при тех же режимах. Подложка располагалась на 54 мм выше фокуса. Центральная часть (темная область), расположенная в середине облученного пятна, – область, в которой разрушилась пленка SiO_2 (диаметр 333 мкм). Цветные кольца вокруг разрушенной области представляют собой пленку SiO_2 , повторяющую профиль поверхности кремния, рекристаллизованного после плавления. Диаметр области плавления – рекристаллизации кремния – составляет 590 мкм .

Следует заметить, что на рис. 1, б, от центральной темной части рекристаллизованной области кремния, покрытой частицами разрушенной пленки SiO_2 , наблюдается появление микротрещин под углом 60° к касательной, характерных для плоскостей скольжения в монокристаллах кремния с кристаллографической ориентацией поверхностей $\{111\}$. В результате экспериментов были выбраны режимы облучения, при которых сохранялась целостность пленки SiO_2 , а поверхность кремния претерпевала минимальные структурные изменения (точечное подплавление кремния в центре области облучения). Начало микроstructuring поверхности кремния (появление линий скольжения (ЛС)) обнаруживалось при мощности $12,9 \cdot 10^3 \text{ Вт/см}^2$. Дальнейшее повышение дозы лазерного излучения приводит к резкому (скачкообразному) возрастанию плотности линий, т.е. формированию полос ЛС.

Следом за одной появляются полосы ЛС в других плоскостях скольжения. Полосы ЛС пересекаются под углами, соответствующими плоскостям скольжения для разных кристаллографических ориентаций поверхности кремниевых пластин. При увеличении времени облучения и постоянной мощности рост ЛС и полос ЛС происходит более равномерно, чем в случае набора дозы при постоянном времени облучения и увеличении мощности. На рис. 2 приведены микрофотографии полос скольжения облученных участков, полученных в разных режимах облучения на подложках с кристаллографической ориентацией (100) и (111).

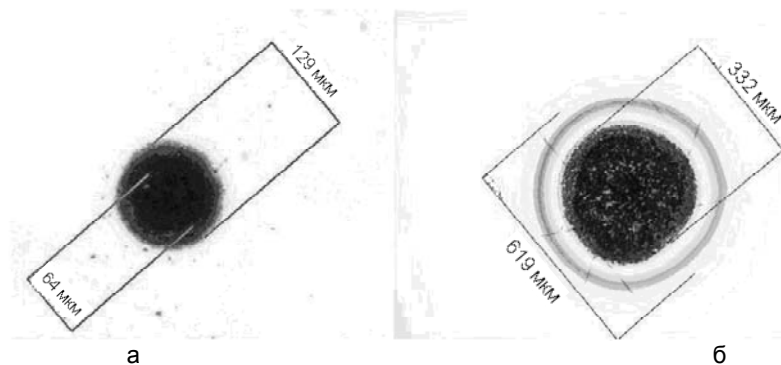


Рис. 1. Фотографии облученных областей на окисленной подложке кремния: облучение пучком, сфокусированным на подложке (а); облучение сходящимся пучком (б)

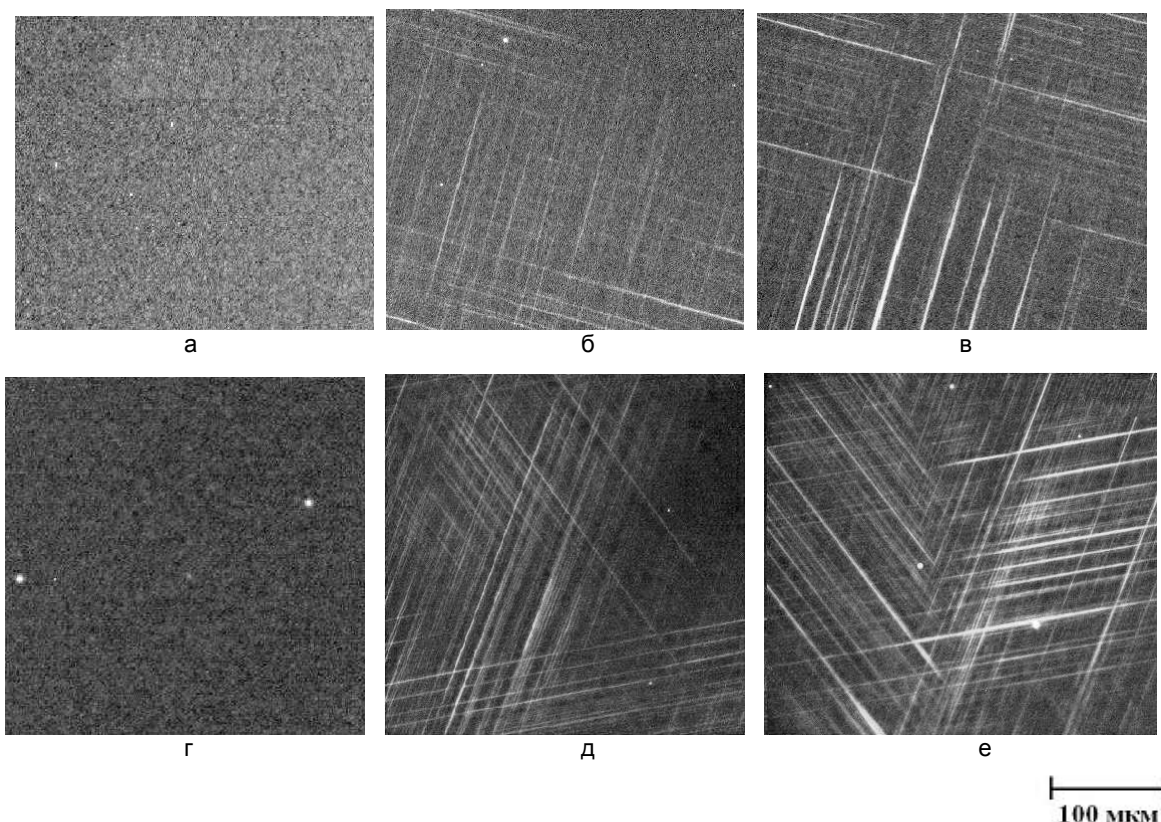


Рис. 2. Микрофотографии сеток ЛС в средней части облученных областей структур SiO_2/Si при разных дозах облучения: структура с ориентацией подложки (100) до облучения (а), 28,2 Дж/см² (б), 77,4 Дж/см² (в); структура с ориентацией подложки (111) до облучения (г), 28,2 Дж/см² (д), 77,4 Дж/см² (е). Во всех случаях время облучения равнялось 5 с

Обсуждение полученных результатов

Из приведенных результатов следует, что в результате облучения кремниевой монокристаллической пластины со слоем SiO_2 на поверхности (системы SiO_2/Si) волоконным лазером с длиной волны $\lambda = 1062$ нм происходит пластическая деформация поверхности кремния. Область пластической деформации локализована площадью облучения и представляет собой сетку ЛС, образованную пересечением полос ЛС.

При изготовлении полупроводниковых приборов и интегральных микросхем основными причинами возникновения ЛС на кремниевых подложках являются термические или термомеханические напряжения, возникающие при высокотемпературных обработках подложек (например, термическое окисление, диффузия и др.). Центрами зарождения ЛС являются крупные структурные дефекты, такие как сколы на краях пластины, царапины на поверхности, свирл-дефекты, кластеры преципитатов кислорода. Как правило, эти ЛС пересекают пластину от края до края и при очень больших напряжениях образуют

сетки из полос ЛС. Например, на кремниевых пластинах, ориентированных в главной кристаллографической плоскости (111), они образуют шестиугольную звезду, получившую название «звезда Давида».

В данном случае основной причиной пластической деформации поверхности кремния также являются термомеханические напряжения, однако природа их другая. Исходная система SiO_2/Si является изначально напряженной. Упругие механические напряжения, обусловленные структурными особенностями SiO_2 и Si , являются касательными к поверхности кремния и растягивающими. Их максимальные значения приложены к поверхности [10]. Кроме того, следует учитывать, что особыми генерационными свойствами обладают поверхностные слои монокристаллов. Здесь согласно модели, предложенной Алейхиным для межфазной границы Si-SiO_2 , находятся легко действующие поверхностные источники, генерирующие дислокации при низких напряжениях сдвига вследствие разности модулей упругости кремния и окисла.

Пленка SiO_2 практически прозрачна для излучения с длиной волны $\lambda = 1062$ нм, а коэффициент поглощения кремния на этой же длине волны $\alpha \approx 9,35 \text{ см}^{-1}$, поэтому энергия фотонов полностью передается кристаллической решетке кремния. В результате поглощения энергии квантов фононы кристаллической решетки увеличивают энергию, часть которой передается соседним фононам, а часть выделяется в виде тепла. Происходит разогрев области облучения. Приповерхностный слой кремния разогревается сильнее, так как в нем фононы получают большее число квантов энергии. В результате кристаллическая решетка в приповерхностных слоях кремния существенно деформируется, изменяется ее постоянная, химические связи между атомами ослабевают. Разогрев системы SiO_2/Si приводит к тому, что к структурным напряжениям, действующим в системе SiO_2/Si до облучения, добавляются термические напряжения, обусловленные разницей коэффициентов термического линейного расширения SiO_2 и кремния (для SiO_2 $\alpha = 0,5 \cdot 10^{-6} \text{ K}^{-1}$, для Si $\alpha = (2,55-4,15) \cdot 10^{-6} \text{ K}^{-1}$).

Учитывая вышеизложенное энергетическое состояние системы SiO_2/Si , механизм формирования на поверхности кремния сетки ЛС может быть объяснен следующим образом. В результате постоянной накачки энергии лазерным пучком с высокой частотой следования импульсов (50 кГц) происходит «перегрев» фононов, и в приповерхностных слоях кремния начинается интенсивная генерация собственных точечных дефектов (междоузельных атомов кремния и вакансий). Увеличение плотности точечных дефектов, в свою очередь, приводит к появлению и возрастанию плотности дефектов упаковки и дислокаций. Под действием касательных напряжений на границе SiO_2/Si дислокации по механизму скольжения передвигаются к поверхности кристалла. Согласно этому механизму, движение дислокации происходит следующим образом [11]. Дислокация легко движется в плоскости скольжения, т.е. в той плоскости, в которой находится дислокационная линия и ее вектор Бюргерса. Перемещение краевой дислокации на одно межатомное расстояние представляет собой согласованную перегруппировку атомов возле дислокации и не сопровождается диффузионным переносом массы. Под действием касательного напряжения ряд атомов, образующих дислокационную линию, вытесняет ряд атомов в соседней плоскости, образуя полную плоскость. Вытесненный же ряд атомов вызывает перестройку связей, и образуется новая плоскость с дислокацией на конце – дислокация переместилась на величину вектора Бюргерса, т.е. на одно межатомное расстояние. Так, перемещаясь, дислокация, в конце концов, выйдет на поверхность кристалла, и здесь появится ступенька, равная межатомному расстоянию. Эта ступенька и является выходом ЛС на поверхность монокристалла, хотя она не может быть обнаружена визуально. Так как в плоскости скольжения движутся десятки и сотни дислокаций, то в результате их выхода на поверхность высота ступеньки будет увеличиваться. Под микроскопом эта «высокая» ступенька наблюдается как ЛС.

Деформация развивается неоднородно и начинается от центра облучаемой области. По этой причине ЛС располагаются на разных расстояниях друг от друга, образуя полосу ЛС. При увеличении дозы облучения скольжение дислокаций распространяется на другие пересекающиеся плоскости скольжения системы. В результате получается сетка дислокаций, образующаяся при пересечении полос ЛС, сформированных на пересекающихся плоскостях скольжения (рис. 2, б, д). В начале пластической деформации с ростом дозы облучения плотность ЛС в полосе растет, затем начинает уменьшаться, сопровождаясь увеличением высоты ступенек ЛС. Наконец, в результате появления большого числа дислокаций, движущихся по плоскостям скольжения, и сетки ЛС возможно появление новых источников для размножения дислокаций – источников Франка–Рида [12]. Таким источником дислокаций в данном случае может быть отрезок движущейся дислокации, застопоренный точками пересечения ЛС, расположенными в разных полосах. Как известно, «взрыв» одного источника Франка–Рида способен образовать сотни новых дислокаций.

Начиная с определенных величин доз облучения подложек, часть ЛС превращается (перерастает) в микротрещины, формирование которых начинается в центральной части облученных областей в местах пересечения и скопления ЛС (рис. 2, в, е). При мощностях, которые вызывают появление микротрещин в облученном участке подложки, на обратной полированной стороне пластины возникают микротрещины в виде тонких лучей, исходящих из одной точки, совпадающей с центром облученной области. Это сви-

детельствует о том, что начавшаяся на поверхности кристалла пластическая деформация с увеличением дозы облучения распространяется на всю толщину кремниевой пластины.

О фононной природе возникновения локальной пластической деформации свидетельствует также следующий эксперимент. На стандартном экспериментальном образце с помощью алмазного резца был нанесен скрайб. Именно механические нарушения поверхности кремниевых пластин при термических обработках являются основной причиной возникновения ЛС. В данном случае при лазерном облучении скрайб не инициирует появление ЛС. Более того, скрайб является барьером при развитии ЛС даже в том случае, когда область облучения частично (краем) заходит на скрайб. Необходимо также отметить, что для формирования сеток ЛС вблизи края пластины SiO_2/Si требуется доза облучения в два–три раза меньше, чем в середине пластины. Это можно объяснить тем, что край пластины ограничивает движение «фононного ветра». Иначе говоря, край пластины задерживает передачу энергии фононами в этом направлении. Энергия фононов растет, и повышается температура в области облучения до порога пластичности при более низких дозах. В центре же пластины энергия лазерного излучения передается кристаллической решетке во все стороны на большие расстояния от облученной области. Следовательно, для разогрева кристаллической решетки до появления пластической деформации в средней части пластины требуется большая доза облучения, чем на краю пластины.

Заключение

В работе продемонстрирована возможность микроструктурирования поверхности монокристалла кремния в системе SiO_2/Si импульсным иттербиевым волоконным лазером типа ИЛИ-1-50 с энергией импульса 1 мДж и частотой следования импульсов 50 кГц.

Показано, что на поверхности кремния в области облучения возникает пластическая деформация, что обусловлено результатом воздействия на систему SiO_2/Si нескольких энергетических факторов:

- энергии фононов лазерного излучения (большая энергия и высокая частота следования импульсов);
- упругих механических напряжений, вызванных различием структур сочлененных решеток кремния и диоксида кремния;
- термических напряжений, возникающих при лазерном облучении системы SiO_2/Si за счет разницы температурных коэффициентов линейного расширения кремния и диоксида кремния;
- наличия низкоэнергетических центров зарождения дислокаций на границе раздела SiO_2/Si .

Пластическая деформация проявляется в виде полос, каждая из которых состоит из большого числа параллельных друг другу линий скольжения, расположенных на соответствующих плоскостях скольжения. Пересечение полос приводит к появлению в кристаллической решетке кремния сложно структурированной поверхности в виде сетки линий скольжения. Следует особо отметить, что при таком микроструктурировании поверхности кремния не нарушается целостность пленки SiO_2 и сохраняется величина ее пробивного напряжения.

Предложен возможный фононно-дислокационный механизм формирования морфологии микроструктурированной поверхности кремниевой подложки. Предварительные исследования вольт-амперных и вольт-фарадных характеристик первых экспериментальных образцов позволяют надеяться на возможность практического применения этого метода микроструктурирования.

Литература

1. Вейко В.П., Дышлошенко С.С., Скворцов А.М. Лазерное микроструктурирование поверхности кремния // Научно-технический сборник «Диагностика и функциональный контроль качества оптических материалов». – СПб: СПбГУ ИТМО, 2004. – Ч. 2. – С. 138–153.
2. Serano Justin R. and David G. Cahil. Micron-scale buckling of SiO_2 on Si // Appl. Phys. – 2002. – V. 92. – P. 7606–7610.
3. Скворцов А.М., Плотников В.В., Соколов В.И. Формирование нанокластеров кремния в структуре кремний/диоксид кремния // Изв. вузов. Приборостроение. – 2006. – Т. 48. – № 3. – С. 62–67.
4. Medvid Arthur, Dmitruk Igor, Onufrijevs Pavels, Pundyk Iryna. Properties of Nanostructure Formed on SiO_2/Si Interface by Laser Radiation // Solid State Phenomena. – 2007. – V. 131–133. – P. 559–562.
5. Medvid A., Onufrijevs P., Kropman D., Mellikov E., Mukapavela F., Bakradze F. Low-K factor of SiO_2 layer on Si irradiated by YAG:Nd laser // Journal of Non-Crystalline Solids. – 2007. – V. 353. – № 5–7. – P. 703–707.
6. Скворцов А.М., Лэ Зуи Туан, Чуйко В.А., Фам Куанг Тунг. Формирование регулярного массива наноструктур кремния в пленке SiO_2 на кремниевой подложке методом проекционного лазерного облучения // Изв. вузов. Приборостроение. – 2009. – Т. 52. – № 5. – С. 69–75.
7. Skvortsov A.M., Veiko V.P., Sokolov V.I., Pham Qung Tung, Khaletsky R.A. Laser modification of thermal oxide films on silicon // International conference «Fundamentals of laser assisted micro- and nanotechnologies» (FLAMN-10): Abstracts. – St. Petersburg: Russia, 2010. – P. 103.

8. Khaletsky R.A., Zamoraynskaya M.V., Kolesnikova E.V., Skvortsov A.M., Sokolov V.I., Pham Qung Tung, Veiko V.P. «Long-range action» effect under laser irradiation of SiO₂-Si system // International conference «Fundamentals of laser assisted micro- and nanotechnologies» (FLAMN-10): Abstracts. – St. Petersburg: Russia, 2010. – P. 105.
9. Veiko V.P., Skvortsov A.M., Sokolov V.I., Pham Qung Tung, Khalecki R.A., Efimov E.I., Vartanyan Tigran A. Effect of laser irradiation on the structures properties such as SiO₂/Si // Fundamentals of Laser-Assistant Micro- and Nanotechnologies 2010. Proceeding of SPIE. – 2011. – V. 7996. – P. 79960S-79960S-5.
10. Скворцов А.М., Фролков Е.Г. Дефектообразование и надежность больших интегральных схем: Учебное пособие. – СПб: СПб ГИТМО (ТУ), 2003. – 139 с.
11. Хирт Д., Лотте Н. Теория дислокаций: Пер. с англ. / Под ред. Э.М. Нагорного и Ю.А. Осипьяна. – М.: Атомиздат, 1972. – 600 с.
12. Бенгус В.З. Скорость размножения подвижных дислокаций и источники подвижных дислокаций. – Киев: Наукова думка, 1975. – С. 315–333.

- | | |
|--|---|
| <i>Скворцов Альберт Матвеевич</i> | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, a-skvortsov@yandex.ru |
| <i>Вейко Вадим Павлович</i> | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, veiko@lastech.ifmo.ru |
| <i>Хуинь Конг Ту</i> | – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, Picochip912@yahoo.com |

УДК 536.629.7

**ИССЛЕДОВАНИЕ ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ
БАТАРЕЙНЫХ ПРИЕМНИКОВ ТЕПЛОВОГО ПОТОКА
В НЕСТАЦИОНАРНОЙ ТЕПЛОМЕТРИИ****И.А. Сиваков**

Приведены результаты эксперимента, показывающие возможность применения батарейного приемника теплового потока (датчик Геращенко) в нестационарной теплометрии с использованием метода параметрической идентификации модели теплопереноса в датчике.

Ключевые слова: нестационарная теплометрия, приемники теплового потока, датчик Геращенко, обратная задача теплопроводности, параметрическая идентификация, фильтр Калмана, дифференциально-разностная модель.

Введение

В настоящее время разрабатываются и широко используются для прикладной теплометрии различного типа приемники тепловых потоков (ПТП), которые, как правило, представляют собой автономные, достаточно миниатюрные устройства с одномерным теплопереносом. Во многих практических важных случаях наибольший интерес представляют исследования переходных процессов в изучаемых объектах, когда измерения постоянных или переменных во времени плотностей тепловых потоков выполняются в нестационарных режимах работы ПТП. Таким образом, нестационарная теплометрия является одной из наиболее проблемных задач прикладной теплометрии при исследовании промышленных объектов и технологических процессов.

В настоящее время при исследовании процессов теплообмена для определения тепловых потоков используются два общих подхода. Первый, традиционный, состоит в расчете по простым, часто алгебраическим формулам, где основные проблемы заключаются в конструировании различных узлов прибора, чтобы эти формулы можно было применить. В основе метода лежит прямая градуировка ПТП, в ряде случаев – на специальных стендах, создание которых требует значительных усилий.

Второй подход, интенсивно развивающийся последние 30–40 лет, составляют методы и принципы решения обратных задач теплопроводности (ОЗТ) и различные алгоритмы оптимизации. Данный подход заключается в расчетном определении (восстановлении) плотности входящего в ПТП теплового потока по измеряемым температурам в отдельных точках ПТП. Такие задачи относятся к нестационарным граничным ОЗТ, а в аспекте измерительной техники – к косвенным методам измерений. Этот подход является прямо противоположным первому: при значительной простоте конструкции приходится обрабатывать большие объемы данных по специальным алгоритмам. С помощью решения ОЗТ можно определять температуры, которые по каким-либо причинам нельзя непосредственно измерить, глубину заделки термомпар, начальное распределение температур и многое другое. В математической физике такие задачи называются некорректно поставленными и могут иметь неустойчивое решение, очень чувствительное к погрешностям измерения температур. Существуют различные методы решения обратных задач. Выбор наиболее универсального, помехозащищенного и вычислительно эффективного метода для определенного класса задач является достаточно сложной проблемой.

В работах Д.Ф. Симбирского, А.В. Олейника, Н.В. Пилипенко, Дж. Бека и др. предложено использовать рекуррентные (последовательные) методы параметрической идентификации, в частности, базирующиеся на модифицированном алгоритме цифрового фильтра Калмана (ФК). Изначально разработанные для использования в измерительных системах реального времени, данные методы доказали свою эффективность при решении ряда граничных задач для различных типов ПТП [1–4]. Для решения таких задач удобно использовать метод параметрической идентификации дифференциально-разностной модели (ДРМ) теплопереноса в исследуемом объекте с использованием рекуррентного цифрового ФК по искомым параметрам [3, 4]. Оценка возможностей такого подхода показала [3–6], что его применение позволяет получить приемлемые для практики результаты.

Исторически наибольшее распространение и популярность на территории Российской Федерации и СНГ получили малогабаритные градиентные батарейные ПТП, разработанные в начале 1970-х гг. под руководством О.А. Геращенко и в дальнейшем развиваемые его последователями. Данный вид датчиков серийно выпускается на многих предприятиях и находит широкое применение в прикладной стационарной теплометрии. Тепломеры Геращенко обладают рядом положительных особенностей, однако, в силу своей тепловой инерционности они ограничены в части измерения быстро меняющихся или постоянных тепловых потоков, но в динамических условиях (при ограничениях на время единичного измерения).

В настоящей работе предлагается метод восстановления нестационарного теплового потока, на основе параметрической идентификации модели теплопереноса в батарейном ПТП и результаты его экспериментального исследования. Кратко остановимся на способе получения оптимальных оценок теплового потока.

Параметрическая идентификация процесса теплопереноса

Как известно, динамика тепловых процессов в общем случае описывается одним или несколькими параболическими уравнениями в частных производных с соответствующими граничными условиями и представляется в виде температурного поля в некоторой многосвязной области. Так как температурное поле обычно кусочно-непрерывно (имеет конечное число разрывов), то в каждый момент времени с некоторой погрешностью δ оно может быть описано температурами счетного числа точек $n = n(\delta)$, зависящего от δ (при $\delta \rightarrow 0, n \rightarrow \infty$) [3]. В этом и заключается смысл дифференциально-разностной аппроксимации.

Для получения оптимальных оценок теплового потока вначале рассматривается прямая задача теплопроводности, метод решения которой основан на использовании ДРМ, которые представляют собой системы обыкновенных дифференциальных уравнений первого порядка относительно вектора температурного состояния $\mathbf{T}(\tau) = [t_i(\tau)]_{i=1}^n$, где $t_i(\tau)$ – температуры элементарных участков (блоков размером Δ), отнесенные к их центрам (для граничных блоков – к их торцевым поверхностям). ДРМ позволяет учесть практически все особенности тепловых схем для ПТП различных видов и в общем случае имеет вид [1, 2, 5, 6]

$$\frac{d}{d\tau} \mathbf{T}(\tau) = F\mathbf{T}(\tau) + G\mathbf{U}(\tau),$$

где F – $(n \times n)$ -матрица обратных связей; G – $(n \times 2)$ -матрица управления. Вектор управления $\mathbf{U}(\tau)$, в частности, для условий 2-го рода на торцах ПТП, имеет вид $\mathbf{U}(\tau) = [q_1(\tau) \quad q_2(\tau)]^T$.

Тепловая схема, подробный вывод и топология ДРМ батарейного ПТП, а также структура матриц управления и обратных связей приведены в работе [7].

В ПТП подлежат измерению либо температуры в отдельных точках, либо их разности, либо средние объемные температуры, что отражено в $(m \times n)$ -матрице измерений H универсальной модели измерений

$$\mathbf{Y}_k = H\mathbf{T}_k + \boldsymbol{\varepsilon}_k,$$

где $\mathbf{Y}_k$ – $(m \times 1)$ -вектор измерений; $\boldsymbol{\varepsilon}_k$ – $(m \times 1)$ -вектор случайных погрешностей.

В дальнейшем решается ОЗТ. При этом принимается допущение о том, что известен характер изменения $q(\tau)$, который позволяет с требуемой точностью выполнить кусочно-линейную аппроксимацию на всем интервале его изменения $\overline{0, \tau_N}$. Искомый тепловой поток задается в виде [3, 5]:

$$q(\tau) = \sum_{j=1}^r q_j \varphi_j(\tau),$$

где $\varphi_j(\tau)$ – система базисных функций времени, а q_j – априори неизвестные коэффициенты, которые объединяются в $(r \times 1)$ -вектор искомых параметров

$$\mathbf{Q} = [q_1 \quad q_2 \quad \dots \quad q_r]^T.$$

В качестве базисной функции используются В-сплайны 1-го порядка. При этом интервал $\overline{0, \tau_N}$ разбивается на одинаковые участки сплайн-аппроксимации Δ_z ($z = 1, 2, 3, \dots, r-1$); каждый из них включает l моментов времени измерений $\mathbf{Y}_k$, т.е. имеет протяженность $\Delta_z = l \cdot \Delta\tau$. Тогда на каждом участке аппроксимация $q(\tau)$ имеет вид $q_z = q_{a,z} \cdot Sp_z^{(1)} + q_{b,z} \cdot Sp_{z+1}^{(1)}$, где $q_{a,z}$ и $q_{b,z}$ – значения q на левой и правой границах участка соответственно; $Sp_z^{(1)}$ и $Sp_{z+1}^{(1)}$ ($z = 1, 2, 3, \dots, r-1$) – В-сплайны.

Величины $q_{a,z}$ и $q_{b,z}$ на каждом z -ом участке объединяются в (2×1) -вектор искомых параметров $\mathbf{Q}_z = [q_{a,z} \quad q_{b,z}]^T$. Благодаря сквозной аппроксимации В-сплайнами на всем интервале измерения $\overline{0, \tau_N}$ на границах выполняются равенства $q_{a,z+1} = q_{b,z}$; $q_{a,z+2} = q_{b,z+1}$; ... По мере поступления измерений $\mathbf{Y}_k$ вычисления перемещаются от участка $z = 1$ к участку $z = 2$ и т. д., тем самым последовательно восстанавливается весь набор искомых значений теплового потока $q_1, q_2, \dots, q_r$.

Такую аппроксимацию $q(\tau)$ называют параметризацией ОЗТ. Тогда задача восстановления $q(\tau)$ сводится к параметрической идентификации ДРМ теплопереноса в ПТП – последовательному получению оптимальных оценок $\hat{\mathbf{Q}}_{z,l}$ вектора искомых параметров $\mathbf{Q}_z$ на каждом участке Δ_z путем минимизации по $\mathbf{Q}_z$ квадратичной функции невязки

$$\Phi(\mathbf{Q}_z) = \sum_{k=1}^l (\mathbf{Y}_k - \hat{\mathbf{Y}}_k(\mathbf{Q}_z))^T \cdot \mathbf{R}^{-1} \cdot (\mathbf{Y}_k - \hat{\mathbf{Y}}_k(\mathbf{Q}_z)),$$

где $\hat{\mathbf{Y}}_k(\mathbf{Q}_z)$ – аналог вектора измерений $\mathbf{Y}_k$, рассчитываемый по ДРМ теплопереноса в ПТП для различных значений искомых параметров $\mathbf{Q}_z$, который будем называть модельным вектором измерений; $\mathbf{R}$ – ковариационная $(m \times m)$ -матрица вектора $\mathbf{\epsilon}_k$ случайных погрешностей в измерениях температур ПТП.

С учетом требований к теплотемрическим измерительным системам реального времени при выборе алгоритма минимизации функции невязки $\Phi(\mathbf{Q}_z)$ предпочтение отдано рекуррентным алгоритмам, в частности, алгоритмам цифрового ФК [3, 4]. Они отличаются высокой вычислительной эффективностью и изначально предназначены для работы в измерительно-вычислительных системах реального времени. Для получения оптимальных оценок $\mathbf{Q}_{k+1}$ вектора $\mathbf{Q}$ в $(k+1)$ -й момент времени ФК по искомым параметрам имеет вид [2, 4–6]

$$\mathbf{Q}_{k+1} = \hat{\mathbf{Q}}_k + \mathbf{K}_{k+1}[\mathbf{Y}_{k+1} - \hat{\mathbf{Y}}_{k+1}(\hat{\mathbf{Q}}_k)],$$

$$\mathbf{K}_{k+1} = \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \mathbf{R})^{-1},$$

$$\mathbf{P}_{k+1} = \mathbf{P}_k - \mathbf{K}_{k+1} \mathbf{H}_k \mathbf{P}_k,$$

где $\mathbf{P}_k, \mathbf{P}_{k+1}$ – ковариационные матрицы ошибок оценок параметров для моментов времени $\tau_k = k \cdot \Delta\tau$ и $\tau_{k+1} = (k+1)\Delta\tau$; $\mathbf{H}_k$ – матрица коэффициентов чувствительности измеряемой температуры ПТП к изменению искомых параметров в момент времени τ_{k+1} ; $\mathbf{K}_k$ – весовая матрица.

В качестве иллюстрации возможностей предлагаемого метода на рис. 1 представлены результаты имитационного моделирования восстановления теплового потока для батарейного ПТП при различных внешних воздействиях (различных законах изменения теплового потока). Параметры датчика приведены в таблице.

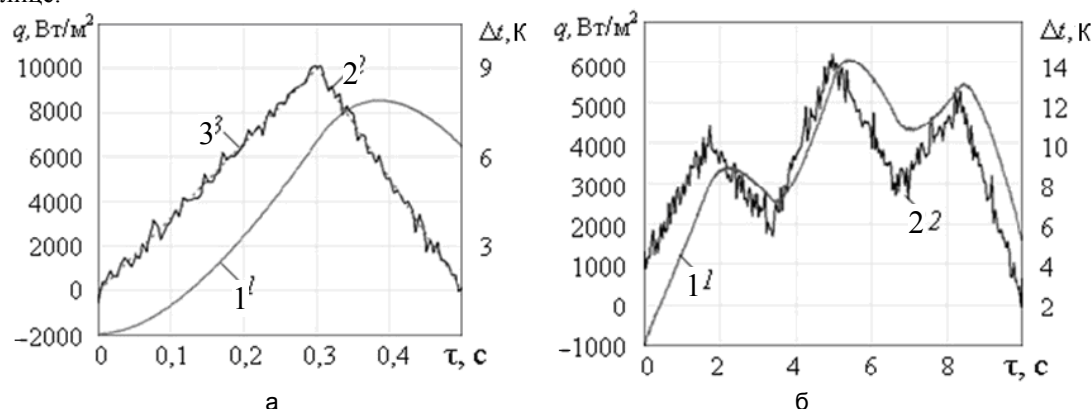


Рис. 1. Результаты имитационного моделирования восстановления теплового потока для батарейного ПТП при различных законах изменения теплового потока: импульсное воздействие (а); периодический закон (б). Обозначения: 1 – измеренная разность температур; 2 – восстановленный тепловой поток; 3 – истинный тепловой поток

Толщина $\delta \cdot 10^3$, м	Теплопроводность λ , Вт/(м·К)	Температуропроводность $a \cdot 10^6$, м²/с
$\delta_1 = 0,1$	$\lambda_1 = 0,2$	$a_1 = 0,11$
$\delta_2 = 1,3$	$\lambda_2 = 0,7$	$a_2 = 0,22$
$\delta_3 = 0,1$	$\lambda_3 = 0,2$	$a_3 = 0,11$

Таблица. Основные характеристики батарейного ПТП

Как показывают результаты имитационного моделирования, предложенный метод восстановления теплового потока путем параметрической идентификации позволяет с высокой точностью восстанавливать нестационарный тепловой поток с использованием батарейных ПТП.

Экспериментальная проверка

Для дальнейшей проверки возможности применения батарейных ПТП в нестационарной теплотметрии проведена серия экспериментов на установке для задания нестационарного теплового потока и определения динамических характеристик ПТП.

В качестве источника теплового потока используется галогенная лампа типа КГЛ-100-12В с оптической системой в виде эллиптического отражателя. Использование эллиптического отражателя позволяет собрать во втором его фокусе значительно большую часть излучаемой лампой мощности по сравнению с линзовыми конденсорами. Между оптической системой и испытуемым датчиком располагается диск модулятора, который служит для задания нестационарных тепловых потоков различной частоты. Сигнал ПТП регистрируется с помощью компьютерного цифрового осциллографа фирмы Velleman типа PCS500.

Величина теплового потока, излучаемого лампой, определена экспериментально и составляет $q_0 = 6 \cdot 10^4 \text{ Вт/м}^2$. Так как форма модуляции близка к синусоидальной, можно считать, что при включенном модуляторе действующий тепловой поток определяется соотношением $q(\tau) = 0,5q_0(1 + \sin \omega\tau)$, где q_0 – среднее значение теплового потока; ω – частота изменения потока; τ – время.

Используемый в опыте датчик, имеющий постоянную времени порядка 5 с, подвергался воздействию модулированного лучистого потока при различных частотах модуляции. Между экспериментами делалась пауза, достаточная для остывания датчика и конструктивных элементов установки. Полученные экспериментальные результаты сохранялись на компьютере, после чего производилась их обработка в программном комплексе «Heat Flow», разработанном на кафедре компьютерной теплофизики и энергофизического мониторинга СПб НИУ ИТМО. Обработка осуществлялась в два этапа:

- фильтрация и сглаживание зашумленного входного сигнала;
- восстановление плотности теплового потока по результатам термометрии.

Современное вычислительное оборудование позволяет проводить обе эти операции в реальном времени. На рис. 2 приведены результаты эксперимента по восстановлению плотности теплового потока, изменяющегося с частотой 2 Гц. Границы применимости предложенного метода для данного датчика (постоянная времени 5 с) – нестационарные тепловые потоки с частотой до 2–2,5 Гц.

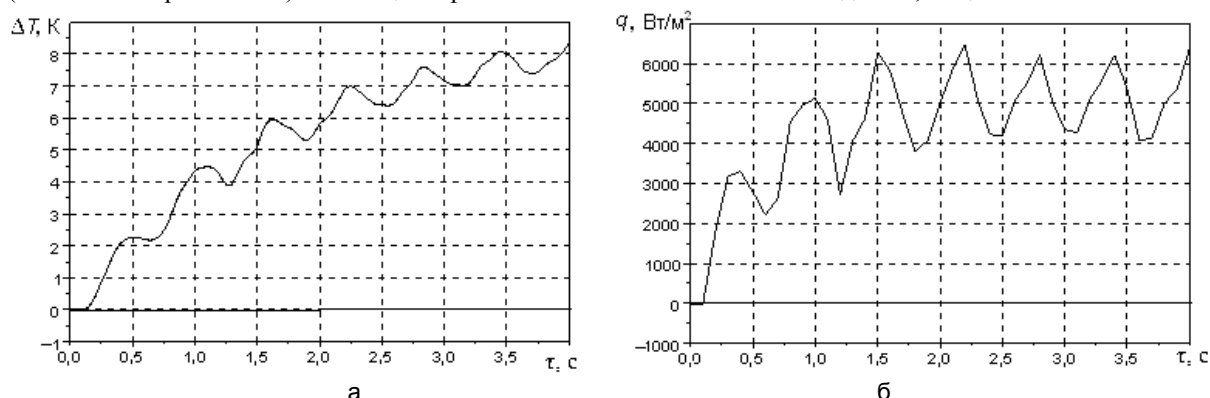


Рис. 2. Результат восстановления плотности теплового потока методом параметрической идентификации по результатам экспериментальной термометрии (после предварительной фильтрации) с использованием батарейного ПТП при воздействии нестационарного теплового потока с частотой 2 Гц: измеренный перепад температур по толщине датчика (а); восстановленный тепловой поток (б)

Заключение

Представлен и обоснован метод решения граничной обратной задачи теплопроводности – восстановления плотности теплового потока по изменению перепада температур по толщине батарейного ПТП. Для решения задачи использован метод параметрической идентификации дифференциально-разностной модели теплопереноса в ПТП.

В качестве иллюстрации возможностей предложенного метода приведены результаты имитационного моделирования по восстановлению плотности теплового потока при различных видах воздействия с использованием модели батарейного ПТП.

Проведены экспериментальные исследования батарейного датчика, которые показали возможность его использования для целей нестационарной теплотметрии и позволили установить границы применимости предложенного метода. Для исследованного батарейного датчика с постоянной времени 5 с максимальная частота изменения плотности теплового потока составляет 2–2,5 Гц. При увеличении частоты изменения плотности теплового потока скомпенсировать высокую инерционность рассматриваемого датчика расчетными методами не представляется возможным.

Полученные результаты показывают, что использование метода параметрической идентификации при восстановлении плотности теплового потока позволяет существенно расширить возможности некоторых видов датчиков, в том числе дает возможность использования стационарных датчиков для целей нестационарной теплотометрии.

Литература

1. Пилипенко Н.В. Методы параметрической идентификации в нестационарной теплотометрии. Ч. 1 // Изв. вузов. Приборостроение. – 2003. – Т. 46. – № 8. – С. 50–54.
2. Пилипенко Н.В. Методы параметрической идентификации в нестационарной теплотометрии. Ч. 2 // Изв. вузов. Приборостроение. – 2003. – Т. 46. – № 10. – С. 67–71.
3. Пилипенко Н.В. Нестационарная теплотометрия на основе параметрической идентификации дифференциально-разностных моделей теплопереноса в одномерных приемниках: Автореф. дис. ... докт. техн. наук. – СПб, 2008. – 35 с.
4. Кириллов К.В., Пилипенко Н.В. Алгоритмы программ для решения прямых и обратных задач теплопроводности при использовании дифференциально-разностных моделей // Научно-технический вестник СПб ГУИТМО. – 2010. – № 5 (69). – С. 106–110.
5. Пилипенко Н.В., Сиваков И.А. Метод определения нестационарного теплового потока и теплопроводности путем параметрической идентификации // Измерительная техника. – 2011. – № 3. – С. 48–51.
6. Pilipenko N.V., Sivakov I.A. A method of determining nonstationary heat flux and heat conduction using parametric identification // Measurement Techniques. – 2011. – V. 54. – № 3. – P. 318–323.
7. Пилипенко Н.В., Лазуренко Н.В., Лебедев П.В. Параметрическая идентификация нестационарных тепловых потоков с помощью тепломера типа «вспомогательная стенка» // Изв. вузов. Приборостроение. – 2005. – Т. 48. – № 9. – С. 47–51.

Сиваков Иван Анатольевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, i.a.sivakov@gmail.com

8

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ

УДК 004.056.52

**МНОГОМЕРНАЯ МОДЕЛЬ РАЗГРАНИЧЕНИЯ ДОСТУПА К ОБЪЕКТАМ
В СИСТЕМЕ КОНТРОЛЯ ВЕРСИЙ**

А.И. Спивак, А.В. Разумовский, И.А. Зикратов

Предложена новая модель разграничения доступа для систем контроля версий файлов. Рассматривается трехмерная модель разграничения доступа на основе дискреционной модели Харрисона–Руззо–Ульмана. Для решения задачи управления и контроля базовые операции модели Харрисона–Руззо–Ульмана дополнены новыми операциями, учитывающими доступ субъектов к версиям объектов.

Ключевые слова: модель, безопасность, разграничение доступа, контроль версий.

Введение

Разработка и эксплуатация сложных ИТ-систем предполагает участие в этих процессах коллектива разработчиков и пользователей. В частности, при разработке программного обеспечения 3D-анимации объектами являются исходные тексты программ, отдельные графические сцены и т.д. Для координации усилий по разработке одних и тех же объектов со стороны множества разработчиков применяются системы контроля версий. Известно, что разграничение доступа в таких системах, как Subversion [1], CVS [2] обеспечивается на уровне объекта доступа и подразумевает все возможные его версии, которые были получены при совместной работе разработчиков. Однако реализовать разграничение доступа субъектов на уровне версии не представляется возможным, так как такой функционал не реализован в существующих системах контроля версий. В этом случае особую актуальность приобретает задача обеспечения совместной работы различных субъектов над одними и теми же объектами, что подразумевает независимое появление различных версий исходного объекта. Для решения этой задачи авторами взята за основу известная из теории информационной безопасности модель дискреционного разграничения доступа Харрисона–Руззо–Ульмана (HRU) [3, 4]. Доработка этой модели путем введения в ее состав совокупности матриц доступа, описывающих разграничение субъектов к версиям объектов, позволила реализовать механизм контроля версий при производстве мультимедийного контента в системе 3D-анимации.

Дискреционная модель разграничения доступа в системах контроля версий

При разработке мультимедийного контента особое внимание уделяется наличию отдельных версий одних и тех же сцен, а также возможность существования отличающихся прав доступа к отдельным версиям сцен со стороны субъектов-разработчиков. Традиционные системы контроля версий обеспечивают распределение доступа к версиям бинарных файлов посредством механизма блокировок. Механизм блокировки позволяет одному из разработчиков захватить в монопольное использование файл или группу файлов для внесения в них изменений. На то время, пока файл заблокирован, все права у остальных пользователей отбираются, он остается доступным всем остальным разработчикам только на чтение, и любая попытка внести в него изменения отвергается сервером. Недостатки использования блокировок очевидны:

- блокировки мешают продуктивной работе, поскольку вынуждают ожидать освобождения заблокированных файлов;
- блокировки создают административные проблемы, когда разработчик может забыть снять блокировку с занятых им файлов.

Для разрешения подобных проблем приходится применять административные меры, в том числе включать в систему технические средства для сброса неверных блокировок, но и при их наличии на приведение системы в порядок расходуется время.

Основным отличием предлагаемой модели разграничения доступа является представление версий объектов не в виде древовидной структуры, количество узлов и ветвей в которой является случайными и трудно прогнозируемыми величинами, а виде совокупности матриц доступа. Согласно модели HRU, матрица доступа описывает права доступа именованных субъектов к именованным объектам, которые записаны на пересечении соответствующих строк и столбцов.

В качестве субъектов доступа в предлагаемой модели рассматриваются разработчики, работающие с различными версиями сцен. Объектами являются результаты работы – файлы-сцены. Тогда матрица прав доступа имеет в качестве столбцов объекты, которыми являются файлы-сцены, а строками – субъекты-разработчики. В ячейках-пересечениях находятся права доступа данного субъекта к данному объекту. При появлении i -й версии объекта создается дополнительная i -я матрица доступа, которая отражает

права к этой версии объектов со стороны субъектов. В случае отсутствия версий для какого-либо объекта ячейки, соответствующие его версиям, будут пустыми.

В соответствии с формальным описанием модели HRU обозначим: O – множество объектов; S – множество субъектов; R – множество прав доступа субъектов к объектам; V – множество версий объектов O . Для описания отношений между субъектами в модели принимается следующее отношение принадлежности множества субъектов к множеству объектов: $S \in O$.

Пространство состояний такой системы представляется в виде $S \times O \times R \times V$.

Матрица прав доступа M , где столбцами являются объекты, а строками – субъекты, содержит права доступа субъектов к объектам. При создании новой версии объекта (V_i) создается новая матрица $W_{vi}[s, o]$, где содержатся права доступа субъектов к новой версии объекта. Для упрощения получения информации о последнем изменении объекта применяется сквозная нумерация версий объектов, при этом не имеет значения, какой именно субъект создал новую версию объекта. Суть модели поясняется рисунком.

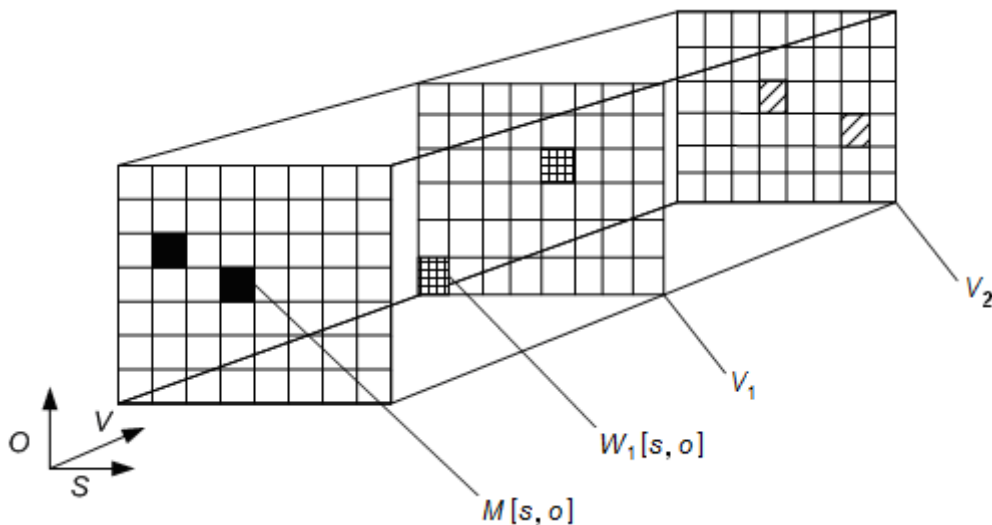


Рисунок. Совокупность матриц доступа субъектов к версиям объектов

На рисунке изображена трехмерная матрица прав доступа в системе контроля версий, в которой присутствуют различные права доступа к разным версиям. Отмеченные ячейки в матрице показывают наличие определенных прав субъекта к данному объекту для версии объекта $W_i[s, o]$.

Формальное описание операций управления системой контроля версий

Положив в основу описания системы формальный аппарат модели HRU, дополним его множеством версий объектов. Тогда состояние системы, добавление и удаление прав доступа, субъектов, объектов и их версий может быть описано следующим образом.

Пусть, учитывая введенные выше обозначения, состояние системы представлено в следующем виде:

$$Q = (S, O, V, M, W),$$

где $M[s, o]$ – ячейка, содержащая элементы из множества R , строки – субъекты, столбцы – объекты; $W_v[s, o]$ – ячейка, содержащая элементы из множества R , строки – субъекты, столбцы – версии объектов.

Изменения в состояние системы могут быть внесены посредством команд $a(x_1, \dots, x_n)$, которые содержат условия выполнения команды и базовые операторы:

$$\text{if } r_{1n} \in M[x_{s1}, x_{o1}] \text{ and } \dots \text{ and } r_{nm} \in M[x_{sn}, x_{on}] \dots \text{ and } r_{1m} \in W_{v1}[x_{s1}, x_{o1}] \text{ and } \dots \text{ and } r_{nm} \in W_{vm}[x_{sm}, x_{om}] \text{ and } \dots$$

then

$$c_1(\dots)$$

$$c_2(\dots)$$

.....

$$c_p(\dots)$$

$r_1, \dots, r_n \in R$ – права доступа;

$c_1, \dots, c_p$ – набор операторов, в которых в качестве параметров принимаются $x_1, \dots, x_n$.

При выполнении команды $a()$ система переходит из состояния Q в состояние Q' .

Элементарные операторы модели Харрисона-Руззо-Ульмана с учетом расширения модели приобретают следующий вид:

1. Добавление права субъекту по отношению к объекту **Enter** $r M[s, o]$.

- Начальное состояние $q=(S, O, V, M, W)$: $s \in S, o \in O, r \in R, v \in V$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S, O'=O, V'=V, W'[s, o]=W[s, o], M'[s, o]=M[s, o] \cup \{r\}$,
 если $(x_s, x_o) \neq (s, o) \Rightarrow M'[x_s, x_o]=M[x_s, x_o]$.
2. Удаление права у субъекта по отношению к объекту **Delete r M[s, o]**.
 Начальное состояние $q=(S, O, V, M, W)$: $s \in S, o \in O, r \in R, v \in V$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S, O'=O, V'=V, W'[s, o]=W[s, o], M'[s, o]=M[s, o] \setminus \{r\}$,
 если $(x_s, x_o) \neq (s, o) \Rightarrow M'[x_s, x_o]=M[x_s, x_o]$.
3. Создание субъекта **Create s'**.
 Начальное состояние $q=(S, O, V, M, W)$: $s \notin S$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S \cup \{s'\}, O'=O \cup \{s'\}, V'=V$, для $\forall (x_s, x_o) \in S \times O \Rightarrow M'[x_s, x_o]=M[x_s, x_o], W'[x_s, x_o]=W[x_s, x_o], M'[s', x_o]=\emptyset$ для $\forall x_o \in O', M'[s', x_s]=\emptyset$ для $\forall x_s \in S'$, для $\forall v \in V', \forall o \in O' \Rightarrow W_v'[s', o]=\emptyset$.
4. Удаление субъекта **Destroy s'**.
 Начальное состояние $q=(S, O, V, M, W)$: $s' \in S$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S \setminus \{s'\}, O'=O \setminus \{s'\}$, для $\forall (x_s, x_o) \in S' \times O' \Rightarrow M'[x_s, x_o]=M[x_s, x_o], W'[x_s, x_o]=W[x_s, x_o]$.
5. Создание объекта **Create o'**.
 Начальное состояние $q=(S, O, V, M, W)$: $o \notin O$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S, O'=O \cup \{o'\}$, для $\forall (x_s, x_o) \in S \times O \Rightarrow M'[x_s, x_o]=M[x_s, x_o], W'[x_s, x_o]=W[x_s, x_o]$, для $\forall x_s \in S' \Rightarrow M'[x_s, o']=\emptyset$, для $\forall v \in V'$ и $\forall x_s \in S' \Rightarrow W_v'[x_s, o']=\emptyset$.
6. Удаление объекта **Destroy o'**.
 Начальное состояние $q=(S, O, V, M, W)$: $o' \in O, o' \notin S$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S, O'=O \setminus \{o'\}$, для $\forall (x_s, x_o) \in S' \times O' \Rightarrow M'[x_s, x_o]=M[x_s, x_o], W'[x_s, x_o]=W[x_s, x_o]$.
 Наличие в новой модели версий объектов приводит к следующим дополнительным операциям:
1. Добавление права субъекту по отношению к версии объекта **Enter r W[s, o]**.
 Начальное состояние $q=(S, O, V, M, W)$: $s \in S, o \in O, r \in R, v \in V$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S, O'=O, V'=V, W_v'[s, o]=W_v[s, o] \cup \{r\}$, если $(x_s, x_o) \neq (s, o) \Rightarrow W_v'[x_s, x_o]=W_v[x_s, x_o], M'[s, o]=M[s, o]$.
2. Удаление права у субъекта по отношению к версии объекта **Delete r W[s, o]**.
 Начальное состояние $q=(S, O, V, M, W)$: $s \in S, o \in O, r \in R, v \in V$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S, O'=O, V'=V, W_v'[s, o]=W_v[s, o] \setminus \{r\}$, если $(x_s, x_o) \neq (s, o) \Rightarrow W_v'[x_s, x_o]=W_v[x_s, x_o], M'[s, o]=M[s, o]$.
3. Создание версии объекта **Create v'**.
 Начальное состояние $q=(S, O, V, M, W)$: $v \notin V$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S, V'=V \cup \{v'\}$, для $\forall (x_s, x_o) \in S \times O \Rightarrow M'[s, o]=M[s, o], W_v'[x_s, x_o]=\emptyset$.
4. Удаление версии объекта **Delete v'**.
 Начальное состояние $q=(S, O, V, M, W)$: $v' \in V$.
 Конечное состояние $q=(S', O', V', M', W')$: $S'=S, O'=O, V'=V \setminus \{v'\}$ для $\forall (x_s, x_o) \in S' \times O' \Rightarrow M'[s, o]=M[s, o], W_v'[x_s, x_o]=W_v[x_s, x_o]$.

Таким образом, наличие апробированного формального описания дискреционной системы контроля версий файлов позволяет устранить недостатки, присущие традиционным механизмам блокировки файлов, обеспечить ясную физическую интерпретацию и контролируемое масштабирование процесса создания версий файлов-сцен, обеспечивает простоту разработки программного кода.

Реализация модели разграничения доступа

Описанная в работе дискреционная модель разграничения доступа реализована в программном модуле системы хранения и контроля версий (СКВ) больших бинарных файлов инновационной системы 3D-анимации. Модуль включает в себя несколько обособленных программных блоков.

Функционал модели реализуется компонентом контроля доступа. В его составе выделяются блок управления субъектами и объектами, блок управления правами доступа, блок проверки прав доступа.

В задачи блока управления субъектами и объектами входит внесение изменений в принятую трехмерную матрицу доступа в соответствии с описанными переопределенными операторами, а именно, добавление субъектов, объектов, новых версий и прав к ним.

Блок управления правами доступа ответственен за редактирование прав доступа между субъектами и объектами, в данном случае происходит внесение в ячейки матриц $M[s, o]$ и $W_v[s, o]$ прав r .

Задачи по предоставлению прав субъектов к объектам решаются блоком проверки прав доступа. Благодаря реализованной модели, у пользователей появилась возможность более гибко настраивать права на версии объектов (файлов). Это, в частности, позволяет делать «срез» по правам объектов на определенный момент времени и тем самым избежать использования устаревших версий объектов, что очень важно при работе над сложной сценой в большом фильме с участием множества разработчиков.

Архитектура СКВ имеет распределенный характер. В СКВ выделены три компонента: ядро СКВ, агенты СКВ и клиент СКВ. Благодаря тому, что данные (файлы) передаются только между агентом и клиентом, такая архитектура позволяет снизить нагрузку на ядро СКВ.

Интерфейсы компонентов для общения между собой и программными модулями, использующими СКВ для получения файлов и их версий, унифицированы и представляют собой REST-интерфейс [5].

Все компоненты реализованы на языке Java. Это позволяет развернуть компоненты на базе любой операционной системы, имеющей JVM.

Разработка модели выполнена в рамках ФЦП ГК № 07.524.11.4009 «Разработка инновационной системы 3D-анимации».

Заключение

Разработанная многомерная модель позволяет использующим ее системам контроля версии обеспечивать разграничение доступа на уровне отдельных версий. Такой функционал является конкурентным преимуществом по сравнению с существующими системами контроля версий и дает возможность более гибкого доступа к объектам по сравнению с механизмом блокировки управления.

Литература

1. Система управления версиями Subversion [Электронный ресурс]. – Режим доступа: <http://subversion.apache.org/>, свободный. Яз. англ. (дата обращения 01.03.2012).
2. Система управления версиями CVS [Электронный ресурс]. – Режим доступа: <http://cvs.nongnu.org/>, свободный. Яз. англ. (дата обращения 01.03.2012).
3. Девянин П.Н. Теоретические основы компьютерной безопасности. – М.: Радио и связь, 2001. – 192 с.
4. Малюк А.А. Информационная безопасность: концептуальные и методологические основы защиты информации. – М.: Горячая линия – Телеком, 2004. – 280 с.
5. Roy Thomas Fielding. Architectural Styles and the Design of Network-based Software Architectures [Электронный ресурс]. – Режим доступа: <http://www.ics.uci.edu/Efielding/pubs/dissertation/top.htm>, свободный. Яз. англ. (дата обращения 01.03.2012).

Спивак Антон Игоревич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, anton.spivak@gmail.com

Разумовский Андрей Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, ассистент, xrew@yandex.ru

Зикратов Игорь Алексеевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, доцент, зав. кафедрой, zikratov@cit.ifmo.ru

УДК 519.688

**МЕТОДЫ РАСЧЕТА ЦЕН ОПЦИОНОВ АЗИАТСКОГО ТИПА
С ДИСКРЕТНЫМИ ДИВИДЕНДАМИ ВНЕ ПЕРИОДА УСРЕДНЕНИЯ****М.С. Косяков, М.В. Пономарев, Д.В. Иванов, Ю.А. Шполянский**

Представлены два метода расчета цен опционов азиатского типа с арифметическим усреднением с учетом выплаты многократных дискретных дивидендов вне периода усреднения: метод прямого моделирования Монте-Карло и предложенный авторами статьи оригинальный подход, основанный на конечно-разностной схеме Кранка–Николсона с использованием аналитической аппроксимации Куррана в качестве терминального условия. Результаты сравнительного тестирования показали, что точность оригинального метода является приемлемой для практических задач при использовании рекомендованных в работе размеров сеток, зависящих от входных параметров. При этом время расчета составляет единицы (доли) миллисекунд на современном персональном компьютере в сравнении с единицами секунд для метода Монте-Карло.

Ключевые слова: азиатский опцион, дискретные дивиденды, метод Монте-Карло, конечно-разностные методы, схема Кранка–Николсона, аппроксимация Куррана.

Введение

Опционом называется производный финансовый инструмент – контракт, согласно которому потенциальный покупатель или продавец получает право, но не обязательство, совершить покупку (опцион типа call) или продажу (опцион типа put) некоторого базового актива по заранее оговоренной цене исполнения (strike price) в определенный момент в будущем, называемый датой исполнения опциона [1].

Выплата по классическому европейскому опциону определяется ценой базового актива на дату исполнения и известной ценой исполнения. Данное условие открывает возможность манипуляций над ценой базового актива непосредственно перед исполнением опциона, что, в свою очередь, ведет к увеличению или уменьшению выплаты. Чтобы ослабить эффект возможной манипуляции, были предложены так называемые азиатские опционы. Их отличительной особенностью является зависимость выплаты от усредненного значения цены базового актива за некоторый промежуток времени – период усреднения. Длительность этого периода и число дат, в которые регистрируются цены для усреднения (даты усреднения), бывают различными. Широко распространены, например, азиатские опционы, в которых даты усреднения – это последние дни срока контракта (часто 10 дней), включая дату исполнения. Такие опционы называют азиатскими «хвостами» (tails).

Одной из важнейших задач систем алгоритмической торговли, позволяющих совершать торговые операции на электронных финансовых рынках с помощью специализированных компьютерных систем, является расчет теоретических цен опционов в реальном времени с учетом постоянного изменения их параметров. Важно отметить, что теоретическое ценообразование азиатских опционов из-за наличия усреднения цен значительно усложняется по сравнению с европейскими опционами. Зависимость цены азиатского опциона от текущей цены базового актива, арифметического среднего и времени описывается дифференциальным уравнением в частных производных. Среднее входит в уравнение как параметр, но при решении играет роль еще одной независимой переменной, поэтому задача определения цены фактически является трехмерной [1]. Точные аналитические решения этой задачи неизвестны [1, 2]. При условии отсутствия выплат дискретных дивидендов найдены приближенные аналитические решения [2]. Хорошее приближение обеспечивается аппроксимацией Куррана [2, 3].

Прямые конечно-разностные методы не позволяют получать решение в реальном времени из-за размерности задачи. В случае, когда дивидендные выплаты пропорциональны цене базового актива, задачу можно переписать в виде двумерного уравнения и эффективно решать численно ([4, 5] и ссылки в них). Однако, если выплаты дискретных дивидендов произвольны, понизить размерность не удастся.

В настоящей работе представлены подходы для решения задачи ценообразования азиатских опционов с произвольными дискретными дивидендами. Рассмотрен метод Монте-Карло прямого моделирования стохастического процесса изменения цены базового актива. Метод сравнительно прост в реализации, но его главным недостатком является низкая скорость сходимости и, как следствие, большое время вычислений и (или) высокие требования к производительности используемого оборудования. Это сильно ограничивает возможность его применения в современных системах алгоритмической торговли. В связи с этим в работе авторами предложен и проанализирован альтернативный оригинальный подход, основанный на конечно-разностной схеме Кранка–Николсона [1, 6] с использованием аналитической аппроксимации Куррана [3] в качестве терминального условия, что позволило осуществлять расчеты в реальном времени. Проведены сравнение результатов вычислений, анализ погрешностей, сформулированы рекомендации по настройке параметров численных алгоритмов для повышения точности результатов.

Ценовые модели опционов

Для расчета цен опционов используются стохастические модели, называемые ценовыми моделями опционов. В их основе лежит постулирование стохастического процесса, моделирующего поведение цены базового актива. Одна из наиболее популярных моделей расчета цен опционов – модель Блэка–Шоулза [1]. Модель Блэка–Шоулза отражает изменение цены опциона в зависимости от ряда параметров и формализуется в виде дифференциального уравнения

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0, \quad (1)$$

где $V(S, t)$ – цена опциона; S – цена базового актива; t – время (изменяется в сторону уменьшения от даты исполнения опциона T до 0); r – безрисковая процентная ставка; σ – волатильность.

Выплата P по европейскому колл (call) опциону определяется формулой:

$$P(S) = V(S, T) = \max[(S(T) - K), 0],$$

где T – дата исполнения; K – цена исполнения. Выплата по пут (put)-опциону равна

$$P(S) = V(S, T) = \max[(K - S(T)), 0].$$

При выплате дивиденда по базовому активу его цена S меняется скачкообразно в момент выплаты, однако цена опциона V остается непрерывной [1]:

$$V(S, t_d^-) = V(S_d(S, D), t_d^+), \quad (2)$$

где t_d^- – момент времени непосредственно перед, а t_d^+ – непосредственно после выплаты дивиденда; $S_d(S, D)$ – теоретическая модель изменения цены S при выплате дивиденда величиной D . В данной работе используется модель [7]

$$S_d(S, D) = \max(S - D, 0). \quad (3)$$

В работе рассматриваются азиатские опционы с дискретным арифметическим усреднением значения цены базового актива. Пусть t_i , $i = 1, \dots, m$ – даты усреднения, где m – число дат. Тогда среднее арифметическое в дату t_i определяется суммой

$$A_i = \frac{1}{i} \sum_{k=1}^i S(t_k).$$

Каждое последующее значение A_i , кроме того, можно выразить через предыдущее:

$$A_i = \frac{i-1}{i} A_{i-1} + \frac{1}{i} S(t_i).$$

В очередную дату усреднения t_i регистрируется цена актива $S(t_i)$ и значение усредненной цены A пересчитывается, поэтому изменяется и цена самого опциона V . Значения V непосредственно перед и после пересчета связаны соотношением [1]

$$V(S, A, t_i^-) = V\left(S, \frac{i-1}{i} A + \frac{1}{i} S, t_i^+\right),$$

где t_i^- – момент непосредственно перед, а t_i^+ – непосредственно после усреднения.

Выплаты по азиатским колл- и пут-опционам равны соответственно

$$P(S, A) = V(S, A, T) = \max[(A - K), 0], \quad (4)$$

$$P(S, A) = V(S, A, T) = \max[(K - A), 0]. \quad (5)$$

Метод Монте-Карло расчета теоретических цен азиатских опционов

Метод Монте-Карло основан на проведении большого числа испытаний согласно стохастическому процессу с параметрами, отвечающими реальному процессу. В нашем случае воспользуемся стохастической моделью изменения цены актива [8]:

$$S(t + \Delta t) = S(t) \exp\left((r - 0,5\sigma^2)\Delta t + \sigma\sqrt{\Delta t}X\right), \quad (6)$$

где $S(t)$ – цена актива в момент времени t ; X – случайная величина, имеющая стандартное нормальное распределение. Теоретическая цена азиатского опциона в настоящий момент времени t_0 определяется как

$$V(S, t_0) = \exp[-r(T - t_0)] E\{P(S, A)\}, \quad (7)$$

где $E\{x\}$ – математическое ожидание случайной величины x . В дальнейшем будем считать $t_0 = 0$.

По формуле (6) для каждой реализации стохастического процесса можно определить значение цены актива $S(t_d)$ в момент выплаты дивиденда t_d , а также в каждый из моментов усреднения t_i . Необходимо

отметить, что в момент выплаты дивиденда t_d цена актива подлежит изменению согласно выражению (3). Выплаты по азиатским колл- и пут-опционам P_k для k -го испытания определяются выражениями (4) и (5). Оценки математического ожидания и дисперсии для N реализаций равны соответственно

$$\hat{P} = \frac{1}{N} \sum_{k=1}^N P_k, \quad s^2 = \frac{1}{N-1} \sum_{k=1}^N (P_k - \hat{P})^2,$$

а оценка теоретической цены азиатского опциона определяется согласно (7). В работе рассматривался 95%-ный доверительный интервал получаемой оценки выплаты по азиатскому опциону, определяемый диапазоном $[\hat{P} - 1,96s/\sqrt{N}; \hat{P} + 1,96s/\sqrt{N}]$.

Как уже было отмечено, основным недостатком прямого метода Монте-Карло является низкая скорость сходимости: ошибка вычислений пропорциональна $s/\sqrt{N}$. Это сильно ограничивает использование прямого метода Монте-Карло в задачах алгоритмической торговли и чаще всего требует применения специализированных вычислительных средств, если расчет цен другими методами невозможен.

Анализ оригинального сеточного метода: комбинация схемы Кранка–Николсона и аппроксимации Куррана

Низкая скорость сходимости метода Монте-Карло и, как следствие, большое время вычислений и (или) высокие требования к производительности используемого оборудования вынуждают искать альтернативные подходы к расчету цен азиатских опционов с учетом выплаты дискретных дивидендов. В этой связи авторами предложен оригинальный численный метод, схема которого показана на рис. 1. Временной промежуток от даты исполнения T до даты выплаты последнего дивиденда t_d (в сторону уменьшения t), включающий весь период усреднения, покрывается приближенной аналитической аппроксимацией Куррана для цен азиатских опционов без дивидендов [3]. Так с учетом (2) формируется распределение цены $V(S, t_d^-)$ для всех точек сетки по S . Это распределение используется в качестве терминального условия разностной схемы Кранка–Николсона численного решения дифференциального уравнения (1), в общем случае с учетом многократных выплат дивидендов согласно (2) [1]. Таким образом, схема Кранка–Николсона используется на участке от даты последнего дивиденда t_d до текущего момента $t_0 = 0$ (также в сторону уменьшения t).

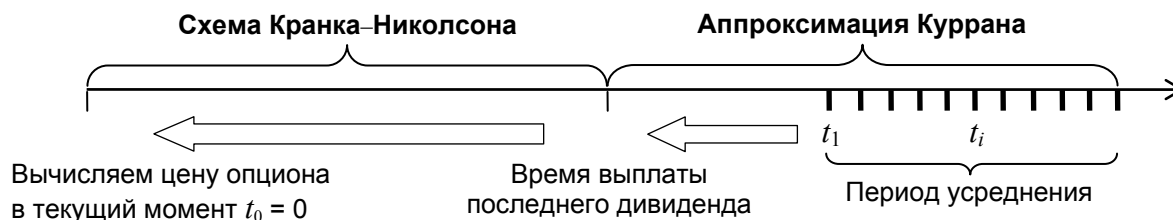


Рис. 1. Аппроксимация Куррана как терминальное условие для схемы Кранка-Николсона

Для проверки применимости предлагаемого подхода и определения параметров, влияющих на точность расчетов, было проведено тестирование оригинального численного метода с помощью метода Монте-Карло. В ходе тестирования теоретическая цена опциона рассчитывалась обоими методами при различных наборах параметров опциона. При моделировании методом Монте-Карло значения нормально распределенных случайных величин формировались из равномерно распределенных величин с помощью преобразования Бокса-Мюллера [9, 10].

На рис. 2, а, б, показаны погрешность Δ и относительная погрешность δ определения цены азиатского колл опциона V сеточным методом в зависимости от цены базового актива S при различных значениях волатильности при условии выплаты одного дивиденда. Зависимости построены для следующих значений параметров: цена исполнения $K = 166$ у.е.; дата исполнения $T = 1$ год; даты усреднения – ежедневно в течение последних 10 дней ($m = 10$); безрисковая процентная ставка $r = 0,10$ в год; момент выплаты дивиденда $t_d = 0,75$ года; размер дивиденда $D = 12,7$ у.е. Для метода Монте-Карло число испытаний составляло $N = 10^7$, а ширина доверительного интервала не превышала 0,1% от получаемого значения V . Параметры сетки в численном методе решения дифференциального уравнения (1) определялись значениями максимальной цены $S_{\max} = 3K$, минимальной цены $S_{\min} = 0$ и количеством шагов равномерной сетки по цене актива $n_S = 150$ и времени $n_t = 200$ шагов/год. Как видно из рис. 2, а, б, погрешность предложенного метода расчета цен азиатских опционов тем выше, чем выше значение волатильности.

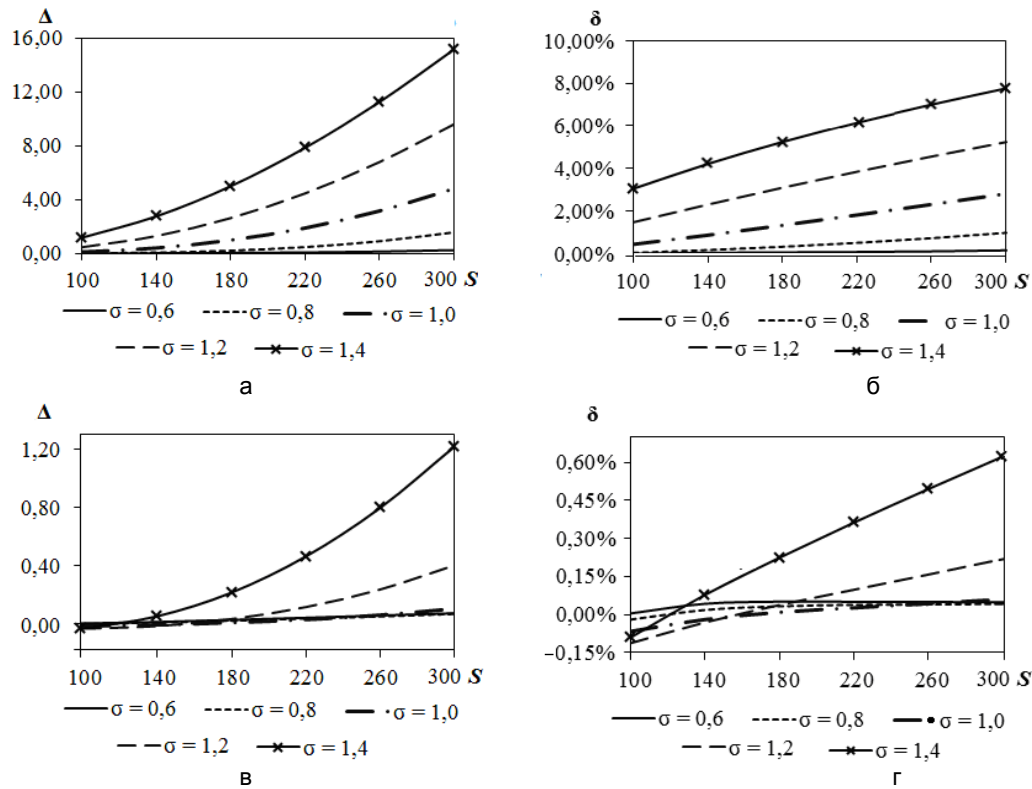


Рис. 2. Погрешность Δ и относительная погрешность δ сеточного метода при условии выплаты одного дивиденда для параметров сетки $S_{\max} = 3K$ и $n_S = 150$ (а)–(б) и для $S_{\max} = 6K$ и $n_S = 300$ (в)–(г)

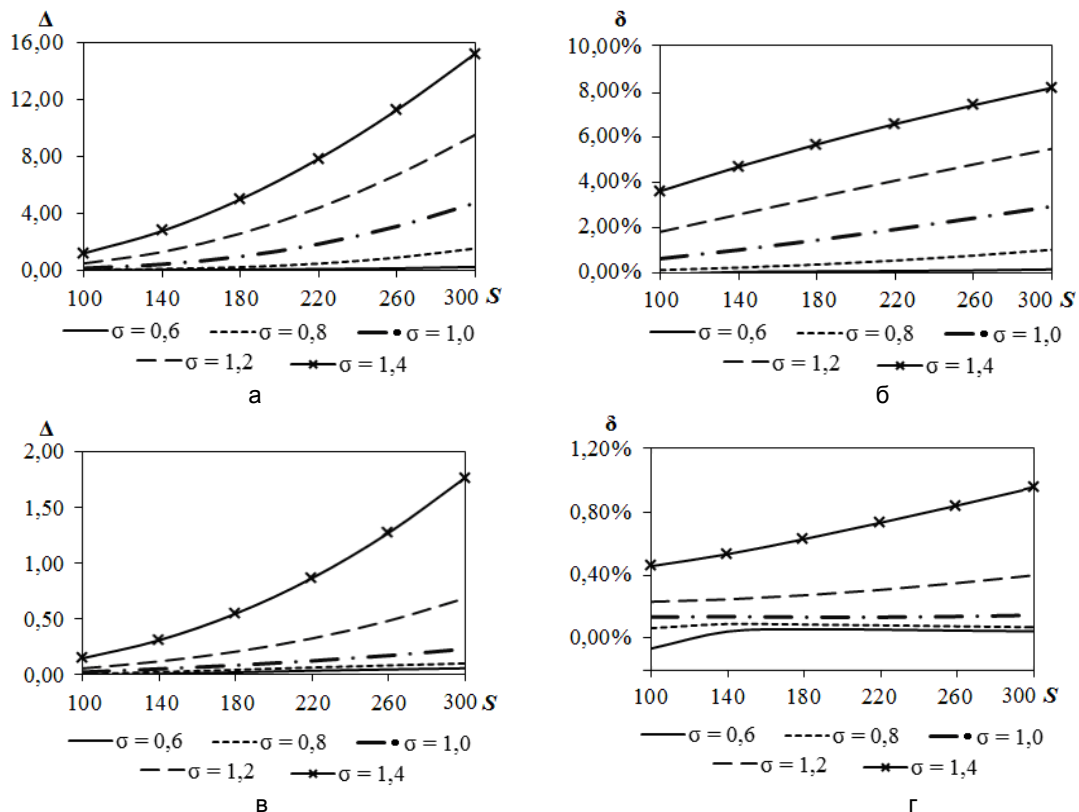


Рис. 3. Погрешность Δ и относительная погрешность δ сеточного метода при условии выплаты трех дивидендов для параметров сетки $S_{\max} = 3K$ и $n_S = 150$ (а)–(б) и для $S_{\max} = 6K$ и $n_S = 300$ (в)–(г)

Для уменьшения погрешности численного метода при значениях $\sigma \geq 1$ было предложено увеличивать количество точек в сетке, а также ее верхний предел в два раза, сохранив шаг сетки по цене неизменным. Из рис. 2, в, г, на которых представлены погрешности предложенного сеточного метода для

$S_{\max} = 6K$ и $n_S = 300$, следует, что при достаточном количестве узлов сетки относительная погрешность расчета предложенным численным методом не превышает 1%. Подобное поведение наблюдалось и при других значениях параметров вычисления. Так, на рис. 3, а–г, показаны соответствующие зависимости при условии выплаты трех дивидендов: моменты выплаты $t_d^1 = 0,11$; $t_d^2 = 0,41$; $t_d^3 = 0,75$ года; соответствующие размеры дивидендов: $D_1 = 5,3$; $D_2 = 9,2$; $D_3 = 12,7$ у.е.; параметры сеток совпадают с параметрами для результатов, представленных на рис. 2.

Таким образом, предложенный сеточный метод при правильно выбранных параметрах сетки является достаточно точным решением задачи ценообразования азиатских опционов с учетом выплаты дискретных дивидендов. Время расчета составляет единицы (доли) миллисекунд на стандартном персональном компьютере по сравнению с единицами секунд при использовании метода Монте-Карло.

Заключение

В работе рассмотрена задача ценообразования опционов азиатского типа с учетом выплаты дискретных дивидендов и арифметическим усреднением, не имеющая точного аналитического решения. Описан алгоритм решения данной задачи методом Монте-Карло. Расчет цены опциона с помощью метода Монте-Карло является прямым многократным моделированием стохастического процесса изменения цены базового актива и имеет вероятностную оценку погрешности, определяемую шириной получаемого доверительного интервала. Главным недостатком данного метода является низкая скорость сходимости и, как следствие, большое время вычислений и (или) высокие требования к производительности используемого оборудования. Несмотря на возможности использования модификаций метода Монте-Карло, направленных на уменьшение дисперсии случайной величины для ускорения его сходимости, применимость данного метода в современных системах алгоритмической торговли весьма ограничена.

В этой связи в работе предложен и проанализирован альтернативный оригинальный подход, основанный на конечно-разностной схеме Кранка–Николсона с использованием аналитической аппроксимации Куррана в качестве терминального условия, что позволяет удовлетворить требованию расчета в реальном времени. Для проверки применимости предлагаемого подхода и определения параметров, влияющих на точность вычислений, было проведено сравнительное тестирование оригинального сеточного метода с помощью метода Монте-Карло. Результаты тестирования показали, что точность оригинального метода является приемлемой для практических задач при использовании рекомендованных в работе размеров сеток, зависящих от входных параметров расчета, в первую очередь от волатильности. При этом время вычислений составляет единицы (доли) миллисекунд на современном персональном компьютере в сравнении с единицами секунд для метода Монте-Карло.

Литература

1. Willmott P. On quantitative finance. – 2-d ed. – John Wiley & Sons, Ltd., 2006. – V. 1–3. – 1500 p.
2. Haug E.G. The complete guide to option pricing formulas. – 2-d ed. – McGraw-Hill, 2007. – 530 p.
3. Curran M. Beyond average intelligence // Risk. – 1992. – V. 5. – № 10. – P. 60–63.
4. Vecer J. Unified pricing of Asian options // Risk. – 2002. – V. 15. – № 6. – P. 113–116.
5. Lord R. Partially exact and bounded approximations for arithmetic Asian options // Journal of Computational Finance. – 2006. – V. 10. – № 2. – P. 1–52.
6. Формалев В.Ф., Ревизников Д.Л. Численные методы. – М.: ФИЗМАТЛИТ, 2004. – 400 с.
7. Haug E.G., Haug J., Lewis A. Back to Basics: a new approach to the discrete dividend problem // Wilmott magazine. – September, 2003. – P. 37–47.
8. Hongbin Zhang. Pricing Asian Options using Monte Carlo Methods. – Department of Mathematics Uppsala University, 2009. – 36 p.
9. Reuven Y. Rubinstein, Dirk P. Kroese. Simulation and the Monte Carlo Method. – John Wiley & Sons, Ltd, 2008. – 372 p.
10. Joshi. M.S. C++ Design Patterns and Derivatives Pricing. – University of Melbourne, 2008. – 292 p.

Косяков Михаил Сергеевич	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, mkosyakov@gmail.com
Пonomarev Максим Васильевич	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, maxim.v.ponomarev@gmail.com
Иванов Дмитрий Владимирович	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант; Тбрикс АБ, математик; dm.vl.ivanov@gmail.com
Шполянский Юрий Александрович	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, доцент; Тбрикс АБ, ведущий математик, shpolyan@mail.ru

КРАТКИЕ СООБЩЕНИЯ

УДК 621.865.8

К ВОПРОСУ О ТЕРМИНОЛОГИИ В ОБЛАСТИ МЕХАТРОНИКИ

Е.В. Шалобаев, Р.Т. Толочка

Рассматриваются актуальные вопросы терминологии в области мехатроники в связи с 25-летием появления в России термина «мехатроника», а также – распространением терминов, рекомендованных постоянным комитетом по стандартизации терминологии Международной организации по теории машин и механизмов.

Ключевые слова: мехатроника, микро- и наномехатроника, терминология, международный транслятор, микросистемная техника, уровневый подход, триада «сенсоры–контроллер–актюатор».

В канун 25-летия выхода в свет русского издания монографии «Мехатроника» японских ученых под редакцией Т. Исии [1] необходимо вернуться к проблемам терминологии в области мехатроники. Можно с полной уверенностью констатировать, что становление мехатроники как науки состоялось, что показывают публикации [2–5], отражающие положение дел как в науке и технике, так и в деле подготовки специалистов в рассматриваемой области [6–10]. При этом, несмотря на две волны интереса к рассматриваемой проблеме [11–22], общепризнанная терминология до сих пор отсутствует.

В России действует государственный образовательный стандарт для подготовки специалистов в области мехатроники, который является лишь отраслевым нормативным документом. В первой (1995 г.) и второй (2000 г.) редакциях даны определения термина «мехатроника». В последней редакции ФГОС-2009 определен как сам термин «мехатроника», так и термин «мехатронная система» [10].

Мехатроника – область науки и техники, основанная на системном объединении узлов точной механики, датчиков состояния внешней среды и самого объекта, источников энергии, исполнительных механизмов, усилителей, вычислительных устройств.

Мехатронная система – единый комплекс электромеханических, электрогидравлических, электронных элементов и средств вычислительной техники, между которыми осуществляется постоянный динамически меняющийся обмен энергией и информацией, объединенный общей системой автоматического управления, обладающей элементами искусственного интеллекта.

Постоянный комитет по стандартизации терминологии Международной организации по теории механизмов и машин (ТММ) (РК «А» IFToMM) в 2002 г. разослал проект, в котором предлагались трактовки терминов для использования в области мехатроники. После ознакомления научной общественности с указанным проектом и заключениями ряда экспертов [14] РК «А» IFToMM были подготовлены рекомендации по терминологии в области мехатроники, принятые IFToMM [15]. В международном трансляторе по ТММ даны термины и их толкование как на английском, так и на русском языке.

Несмотря на наличие рекомендаций IFToMM, их использование не стало общепринятым фактом. Так даже в широко известных и используемых книгах проф. Ю.В. Подураева и его коллег [3, 5, 6] есть упоминания о рекомендациях IFToMM, которые, однако, фактически не используются. Интерес к проблемам терминологии в области мехатроники периодически проявляется, как это следует из работ [12–17], а также [18–22].

Данная проблема была обсуждена на 23-й и 24-й рабочих совещаниях комитета «А» IFToMM (Беларусь, Гомель, 2010; Германия, Ильменау, 2012) [20], и данная публикация, принадлежащая членам РК «А», является реализацией одного из направлений принятых на них решений по ознакомлению широкой научной общественности с рекомендациями международного транслятора в области мехатроники. В качестве иллюстрации приведем ряд следующих терминов на русском языке.

Мехатроника – связующая комбинация механики, электротехники, электроники и информационных технологий для создания технических систем с искусственным интеллектом, в особенности механизмов и машин.

Мехатронная система – система, созданная на основе принципов мехатроники.

Мехатронный подход – интегральный междисциплинарный подход к разработке мехатронных систем.

Микромехатроника – подобласть мехатроники, относящаяся к устройствам и системам с размерами в несколько миллиметров и меньше.

Наномехатроника – подобласть мехатроники, относящаяся к устройствам и их системам, соизмеримым с молекулами.

Адаптивная механика – раздел мехатроники, в котором рассматриваются интеллектуальные устройства и адаптивные структуры, способные к изменению их поведения в соответствии с изменениями в окружающей среде и их внутреннего состояния.

Актуатор – устройство, оказывающее непосредственное воздействие на процесс, другие механические устройства или окружающую среду для выполнения какого-либо полезного действия.

Интеллектуальный актуатор – автономно действующий актуатор, состоящий из сервопривода и вычислительного устройства с соответствующим программным управлением, служащим для управления, восприятия и обмена информацией.

Интеллектуальная машина – машина, обладающая искусственным интеллектом.

Интеллектуальный материал – композиционный материал (структура), обладающий свойствами, соответствующими примитивному интеллекту.

«Умный» материал (система, продукт) – материал (система, продукт), способный к достаточно сложному поведению за счет присущих ему свойств искусственного интеллекта.

Адаптивная машина – интеллектуальная машина, обладающая интеллектуальным свойством адаптивности.

Активный материал (структура) – материал (структура), способный реагировать на возмущения, благодаря внутреннему преобразованию энергии или изменению собственных параметров.

Механизм переменной структуры (с переменными параметрами) – регулируемый механизм, структура (параметры) которого могут быть изменяемы.

Регулируемая структура – структура, параметры или компоновка составляющих которой могут изменяться с целью улучшения функционирования.

Интеллектуальный датчик – автономная единица, интегрирующая в себе функции ощущения, восприятия, переработки аналогового и дискретного сигнала, автоматической и самокалибровки и компенсации.

Система датчиков – система, состоящая из нескольких датчиков, используемая с целью дополнения данных одного датчика данными, полученными от других.

Архитектура мехатронной системы – иерархия или устройство компонентов мехатронной системы.

Решающая система – часть мехатронной системы, которая оценивает воспринятую информацию и планирует последующие действия.

Автоматизированная система управления – система для автоматического управления, обычно основанная на использовании компьютеров.

Исполнительная система – часть мехатронной системы, которая управляет работой машины на основе данных, полученных от решающей системы или непосредственно от системы восприятия (от датчиков).

Система восприятия – часть мехатронной системы, которая выполняет сбор, хранение, переработку и распределение информации о состоянии машины и окружающей среды.

Самоорганизующая система – система, которая полностью или частично формирует свою структуру путем самоорганизации.

Самоорганизация – способность создавать структуру без каких-либо внешних воздействий; свойство системы, заключающееся в ее способности к возникновению.

Саморегулирование – способность машины достигать и поддерживать желаемое поведение при работе в окружающей среде, которая претерпевает конечные изменения во времени.

Самовосстановление – способность машины к восстановлению работоспособности.

Самодиагностика – способность машины контролировать и оценивать рабочее состояние.

Указанные выше термины в англоязычной литературе используются около 10 лет, их толкование не является догмой и, естественно, должно корректироваться в процессе использования. Так, термин «мехатронный подход» должен, по мнению авторов данной публикации, рассматриваться как часть системного подхода. Опыт применения термина «актуатор» показал, что в русской транскрипции лучше использовать написание «актюатор». Синонимом термина «решающая система» может быть использован термин «контроллер», широко применяемый в зарубежной и отечественной литературе. Термин «система восприятия» логично трактовать как «сенсорную систему». Отсюда мехатронная система может трактоваться как триада «сенсоры–контроллер–актюатор» [11]. Необходимо зафиксировать и уровневый подход к мехатронике, учитывая масштабный фактор. Иначе говоря, мехатроника относится к макромиру, микромехатроника – к микромиру, наномехатроника – к наномиру [11]. Требуется установить взаимосвязь между микро- и наносистемной техникой и мехатроникой, поскольку имеются отечественные журналы «Мехатроника» (ныне «Мехатроника, автоматизация, управление») и «Микросистемная техника» (ныне «Нано- и микросистемная техника»), в которых взаимосвязь между их предметными областями не оговаривается. В работах одного из авторов данной публикации [2, 11, 14, 23] такой уровневый подход к мехатронике был предложен, что нашло фактическое отражение в рассматриваемом международном трансляторе, но формально зафиксировано не был и требует такого закрепления. Также необходимо и закрепление тождества терминов «микросистемная техника» и «микромехатроника» [2]. Кстати, нужно зафиксировать различие терминов «сенсор» и «датчик», которое должно отражать физические и конст-

руктивные аспекты [24]. Можно уточнить, что систему, являющуюся частью другой системы, корректнее называть подсистемой. Есть и другие замечания.

Данная публикация является предложением для дискуссии, которая и должна учесть разные мнения и создать предпосылки для создания общепризнанной терминологии. Предстоящее в этом году вступление России в ВТО ставит вопрос об использовании международных стандартов, норм и рекомендаций для поддержания конкурентоспособности в области не только собственно торговли, но и различных форм отечественной продукции, в том числе и научной, что еще раз подчеркивает актуальность данной публикации. В ряде других научно-технических областей подобные международные трансляторы уже не только разработаны с участием постоянного комитета по стандартизации терминологии IFToMM [25], но и нашли широкое применение, о чем говорят многочисленные переиздания словаря-справочника [26].

1. Мехатроника / Под ред. Т. Исии. – Пер. с яп. – М.: Мир, 1988. – 318 с.
2. Шалобаев Е.В. Фундаментальные и прикладные проблемы развития мехатроники // Сборник: Современные технологии / Под ред. С.А. Козлова. – СПб: СПб ГИТМО (ТУ), 2001. – С. 46–66.
3. Подураев Ю.В. Мехатроника: основы, методы, применение. – 2-е изд. – М.: Машиностроение, 2007. – 256 с.
4. Toločka R.T. Adaptive mechanics for mechatronics // The 22th Working Meeting of the IFToMM PC for Standardization of Terminology, June 29–July, 2008. – Villeurbanne, France, Lyon : IFToMM, 2008. – P. 25–29.
5. Подураев Ю.В. Основы мехатроники. – М.: Изд-во МГТУ СТАНКИН, 2000. – 80 с.
6. Егоров О.Д., Подураев Ю.В. Конструирование мехатронных модулей: Учебник. – М.: Изд-во МГТУ СТАНКИН, 2004. – 360 с.
7. Toločka R.T. Engineering, mechatronics and its educational programs // Global Cooperation in Engineering Education: Innovative Technologies, Studies and Professional Development : the 3rd international conference proceedings, October 1–3, 2009. – Kaunas: Technologija, 2009. – P. 32–35.
8. Bansevicius R.P.; Toločka R.T., Macha Ewald, Pawliczek Roland. Adaptive mechanics: concept and course for mechatronics study programme // Mechatronic Systems and Materials: selected papers. – Opole University of Technology, 2007. – P. 7–14.
9. Toločka R.T., Macha E., Pawliczek R. Mechanical engineering and mechatronics: development of study programmes // Mechatronic Systems and Materials 2010. – Opole University of Technology, 2011. – P. 249–254.
10. ФГОС ВПО по направлению подготовки 221000 «Мехатроника и робототехника» (квалификация (степень) «Магистр»). – М., 2009. – 23 с.
11. Шалобаев Е.В. Микросистемная техника и мехатроника: особенности соотношения макро- и микроуровней // Микросистемная техника. – 2000. – № 4. – С. 5–10.
12. Аршанский М.М., Шалобаев Е.В. Мехатроника: основы глоссария // Мехатроника, 2001. – № 2. – С. 47–48.
13. Шалобаев Е.В. К вопросу об определении мехатроники и иерархии мехатронных объектов // Датчики и системы. – 2001. – № 7. – С. 62–65.
14. Шалобаев Е.В. К вопросу о международном трансляторе по мехатронике // Мехатроника. – 2002. – № 4. – С. 6–11.
15. Международный транслятор по науке механизмов и машин в редакции 2003, MMS-terms-2003 // [Электронный ресурс]. – Режим доступа: <http://www.iftomm.3me.tudelft.nl>, свободный. Яз. рус. (дата обращения 09.07.2012).
16. Шалобаев Е.В. Проблемы и тенденции развития терминологии в современных условиях // Микросистемная техника. – 2004. – № 4. – С. 29–32.
17. Шалобаев Е.В. Определение неологизма «мехатроника» и его развитие // В кн.: Проблемы интеллектуального управления в авионике. – СПб: СПбГУ ИТМО, 2005. – С. 12–18.
18. Осипов Ю.М. К вопросу о развитии понятия «мехатроника» // Доклады ТУСУРа, июнь 2010. – № 1 (21). – Ч. 2. – С. 193–198.
19. Кориков А.М. О развитии понятия «мехатроника» // Доклады ТУСУРа, июнь 2010. – № 1 (21). – Ч. 2. – С. 199–202.
20. Shalobaev E.V. Mechatronics: Today Problems and Development trends of Terminology // Proceeding 23rd Working Meeting of the IFToMM Permanent Commission for Standardization of Terminology on MMS Minsk. – Gomel, Belarus, June 21–26, 2010. – P. 111–118.
21. Кориков А.М. Еще раз о мехатронике как о науке // Мехатроника, автоматизация, управление. – 2011. – № 1. – С. 2–8.
22. Толочка Р.А., Шалобаев Е.В. Терминология в области мехатроники // Материалы Международной научной школы «Фридендеровские чтения», ВУР-2011. – СПб: ИПМаш РАН, 2011. – С. 38–44.
23. Шалобаев Е.В., Старжинский В.Е., Шилько С.В. Тенденции развития современной трибологии на микро- и наноуровнях // Труды Международного симпозиума: Гидродинамическая теория смазки. В 2-х томах. – М.: Машиностроение, 2005. – Т. 2. – С. 289–295.

24. Шалобаев Е.В. Сенсорика и 21 век // Датчики и системы. – 2001. – № 1. – С. 63–65.
25. Starghisky V.E., Shalobaev E.V., Sherbakov S.V. On compiling a terminological Reference-Dictionary on gearing // Proceedings of International conference «Power Transmissions'03», 11–12 September, 2003, Section I. – Sofia, Varna : БолгАН, 2003. – Р. 180–186.
26. Гольдфарб В.И., Старжинский В.Е., Шалобаев Е.В. и др. Словарь-справочник по зубчатым передачам: русско-англо-немецко-французский / Под ред. В.Е. Старжинского. – Изд. 5-е. – Гомель: ИММС НАН Б, 2011. – 220 с.

Шалобаев Евгений Васильевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, профессор, shalobaev47@yandex.ru
Толочка Римантас Тадас – Каунасский технологический университет, доктор технических наук, профессор, tadas.tolocka@ktu.lt

УДК 629.1

ОБ ИСПОЛЬЗОВАНИИ МЕТОДА СПРЯМЛЕННЫХ ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК В ЗАДАЧАХ СГЛАЖИВАНИЯ

А.В. Лопарев, О.А. Степанов, О.М. Яшников

Предлагается обобщение приближенного метода спрямленных логарифмических характеристик в случае решения стационарных задач сглаживания.

Ключевые слова: фильтрация, сглаживание, метод логарифмических характеристик, стационарные процессы.

При решении задач обработки измерительной информации широкое применение получили алгоритмы, разрабатываемые в рамках калмановского подхода. Вместе с тем для задач обработки стационарных сигналов сохраняет свою актуальность винеровский подход. Суть подхода заключается в нахождении передаточной (или весовой) функции оптимального фильтра, минимизирующего среднеквадратическую ошибку оценивания в установившемся режиме. Одно из достоинств винеровского подхода заключается в том, что для построения алгоритмов разработаны различные упрощенные методы. В частности, применительно к навигационным приложениям наибольшее применение получил так называемый метод спрямленных логарифмических характеристик [1, 2]. В работе предлагается обобщение этого метода для решения задач сглаживания.

Рассмотрим классическую задачу оптимального оценивания полезного сигнала на фоне случайных ошибок измерений. Пусть скалярные измерения

$$y(t) = x(t) + n(t)$$

представляют собой аддитивную смесь полезного сигнала $x(t)$ и помехи $n(t)$, которые полагаются центрированными, некоррелированными и стационарными процессами с заданными спектральными плотностями $S_x(\omega)$, $S_n(\omega)$. Рассмотрим задачу сглаживания, особенность которой заключается в том, что при получении оценки в текущий момент времени могут быть использованы не только прошлые, как в случае задачи фильтрации, но и будущие (по отношению к этому моменту времени) измерения. В этом случае выражение для передаточной функции оптимального сглаживающего фильтра будет определяться как [1]

$$W_0^s(j\omega) = \frac{S_x(\omega)}{S_x(\omega) + S_n(\omega)}, \quad (1)$$

а для спектральной плотности ошибки оптимальной нереализуемой оценки и ее дисперсии будут в этой ситуации справедливы следующие выражения:

$$S_e(j\omega) = \frac{S_x(\omega)S_n(\omega)}{S_x(\omega) + S_n(\omega)}, \quad D_0^s = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_x(\omega)S_n(\omega)}{S_x(\omega) + S_n(\omega)} d\omega. \quad (2)$$

Метод спрямленных логарифмических характеристик основан на предположении о том, что свойства фильтра определяются характером поведения спектральных плотностей полезного сигнала и помехи в точке их пересечения. Для приближенного их описания в логарифмическом масштабе используются аппроксимирующие прямые, соответствующие так называемым условным спектральным плотностям для

полезного сигнала $x(t)$ и помехи $n(t)$, определяемым в виде $S_x(\omega) \approx a^2 \left(\frac{\rho}{\omega} \right)^{2p}$, $S_n(\omega) \approx a^2 \left(\frac{\rho}{\omega} \right)^{2q}$. Используя

аналогичный подход к задачам сглаживания, нетрудно получить общие решения для различных комбинаций простейших аппроксимаций спектральных плотностей сигнала и помехи. В таблице приведены выражения для параметров C_{if} и C_{is} , характеризующих дисперсии установившихся ошибок фильтрации

(σ^f)² = $C_{if} a^2 \rho$ и сглаживания (σ^s)² = $C_{is} a^2 \rho$, и величина, позволяющая оценить потенциальный выигрыш в точности в виде отношения среднеквадратических ошибок оценивания при решении задач фильтрации и сглаживания $\sigma^f / \sigma^s = \sqrt{C_{if} / C_{is}}$.

$S_x(\omega)$	$S_n(\omega)$	C_{if}	C_{is}	$\sqrt{C_{if}/C_{is}}$
$a^2 \left(\frac{\rho}{\omega}\right)^{2n}$	a^2	$\frac{1}{\sin \frac{\pi}{2n}}$	$\frac{1}{2n \sin \frac{\pi}{2n}}$	$\sqrt{2n}$
$a^2 \left(\frac{\rho}{\omega}\right)^{2n-2}$	$a^2 \left(\frac{\omega}{\rho}\right)^2$	$\frac{\operatorname{ctg}^2 \frac{\pi}{2n}}{\sin \frac{3\pi}{2n}}$	$\frac{1}{2n \sin \frac{3\pi}{2n}}$	$\sqrt{2n \operatorname{ctg}^2 \frac{\pi}{2n}}$
$a^2 \left(\frac{\rho}{\omega}\right)^{2n-4}$	$a^2 \left(\frac{\omega}{\rho}\right)^4$	$\frac{\cos^2 \frac{\pi}{n}}{4 \sin^4 \frac{\pi}{2n} \sin \frac{5\pi}{2n}}$	$\frac{1}{2n \sin \frac{5\pi}{2n}}$	$\sqrt{\frac{n \cos^2 \frac{\pi}{n}}{2 \sin^4 \frac{\pi}{2n}}}$
$a^2 \left(\frac{\rho}{\omega}\right)^{2n-6}$	$a^2 \left(\frac{\omega}{\rho}\right)^6$	$\frac{\cos^2 \frac{\pi}{n} \operatorname{ctg}^2 \frac{3\pi}{2n}}{4 \sin^4 \frac{\pi}{2n} \sin \frac{7\pi}{2n}}$	$\frac{1}{2n \sin \frac{7\pi}{2n}}$	$\sqrt{\frac{n \cos^2 \frac{\pi}{n} \operatorname{ctg}^2 \frac{3\pi}{2n}}{2 \sin^4 \frac{\pi}{2n}}}$

Таблица. Показатели точности фильтрации и сглаживания в установившемся режиме

Проиллюстрируем, как получить представленные в таблице выражения для спектральных плотностей полезного сигнала и помехи, определяемых как

$$S_x(\omega) \approx a^2 \left(\frac{\rho}{\omega}\right)^4, S_n(\omega) \approx a^2, \quad (3)$$

что соответствует первой строке таблицы при $n = 2$. Подставляя (3) в выражения (1) и (2) нетрудно убедиться, что выражения для передаточной функции оптимального сглаживающего фильтра, спектральной плотности ошибки оптимальной нереализуемой оценки и ее дисперсии примут следующий вид:

$$W_0^s(\omega) = \frac{\rho^4}{\omega^4 + \rho^4}, S_e^s(\omega) = \frac{a^2 \rho^4}{\omega^4 + \rho^4}, W_0^s(\omega) = \frac{1}{2\sqrt{2}} a^2 \rho.$$

В этом случае среднеквадратическая ошибка оценивания для задачи сглаживания в два раза меньше ошибки фильтрации $\sqrt{C_{if}/C_{is}} = \sqrt{2n} = 2$. Таким образом, в тех случаях, когда нахождение оценки в реальном времени не является обязательным условием, целесообразно вместо задачи фильтрации решать задачу сглаживания.

Работа выполнена при поддержке гранта РФФИ по проекту 12-08-00835-а.

1. Челпанов И.Б., Несенюк Л.П., Брагинский М.В. Расчет характеристик навигационных гироскопов. – Л.: Судостроение, 1978. – 264 с.
2. Лопарев А.В., Степанов О.А., Челпанов И.Б. Использование частотного подхода при синтезе нестационарных алгоритмов обработки навигационной информации // Гироскопия и навигация. – 2011. – № 3. – С. 115–132.

Лопарев Алексей Валерьевич – ОАО «Концерн «ЦНИИ «Электроприбор», кандидат технических наук, доцент, начальник сектора, loparev@mail15.com

Степанов Олег Андреевич – ОАО «Концерн «ЦНИИ «Электроприбор», доктор технических наук, начальник научно-образовательного центра, soalax@mail.ru

Яшников Ольга Михайловна – ОАО «Концерн «ЦНИИ «Электроприбор», аспирант, olga_evstifeeva@mail.ru

УДК 629.4.066

ДВИЖУЩИЙСЯ ПОЕЗД КАК ИСТОЧНИК ЗВУКОВЫХ ВОЛН, РАСПРОСТРАНЯЮЩИХСЯ ПО РЕЛЬСОВОМУ ПУТИ

С.В. Бибилов, А.В. Шапарь

На основании экспериментальных данных шум качения пары «колесо–рельс» выбран в качестве доминирующего источника звуковых волн для обнаружителя приближающегося поезда.

Ключевые слова: уровень шума, шум качения, приближение поезда.

Разрабатывается устройство для оповещения о приближении поезда лиц, работающих на железнодорожных путях. Один из принципов действия – анализ виброакустических колебаний, возникающих в рельсе при движении поезда. Рассматривается поезд как источник некоего сообщения о своем приближении, передаваемого в рельс, и необходимо выявить характерные особенности этого сообщения.

Исследование и классификация излучения шума от движущихся поездов в окружающую среду – предмет промышленной акустики. Источники шума движущегося поезда известны [1], и их можно разделить на две группы:

- регулярные – шум качения пары «колесо–рельс», аэродинамический шум;
- нерегулярные – определяются типом пути, типом и состоянием подвижного состава. К ним относятся стук колесной пары о стыки, удары и шум, вызванные дефектами окружности обода колеса и рельса, скрежет при прохождении поездами кривых малого радиуса, шумы двигателей, передач, грохот элементов конструкции вагонов.

Уровень шума, излучаемого поездом, обычно измеряется в воздухе на расстоянии 25 м от оси пути и на высоте 3,5 м с помощью системы микрофонов [1]. Установлено, что эквивалентный уровень шума, создаваемый зоной «колесо–рельс» грузового вагона, как минимум, на 12 дБ выше, чем шум от кузова. В разных диапазонах частот преобладают шумы разных компонентов системы «колесная пара–путь».

Есть отличия шума, излучаемого в воздух, от колебаний, распространяющихся в рельсе. С целью изучения этих отличий была проведена запись фонограмм приближения поездов на различных участках Октябрьской железной дороги с помощью виброакустического преобразователя, закрепленного непосредственно на рельсе. Было выявлено следующее.

1. Все вышеуказанные источники шума прослушиваются на различных фонограммах, полученных с рельса. Спектр шума, распространяющегося по рельсу на бесстыковом пути, на расстоянии до 200 м от приближающегося поезда близок к тому, что приведен в [2, рис. 3] для измерений шума от рельса и колес на расстоянии 3,7 м, проведенных на воздухе.
2. Нарастающий шум качения является доминирующим в большинстве ситуаций. Спектр шума качения зависит от скорости поезда и перекрывает диапазон 450–22000 Гц. Шумовые компоненты плавно нарастают при приближении поезда по бесстыковому участку пути, либо резко появляются, когда поезд минует неоднородность пути (стрелочный перевод либо искусственное сооружение). Никаких регулярных отличительных признаков в шуме качения не выявлено. Именно шум качения целесообразно использовать в качестве основного источника информации о приближении поезда.
3. В исследуемом шуме можно выделить удары колесных пар о стыки. На бесстыковых путях хорошо обнаруживаются только удары 2–4 первых колесных пар поезда об одиночные стыки. Удары о стык последующих колесных пар ослабевают в 10–50 раз. Парные удары, вызванные последовательными ударами о стык колесных пар двухосной тележки пассажирского вагона, при скоростях поезда более 120 км/ч сливаются в одиночные.
4. На звеньевых путях удары о стыки и шум качения сливаются в широкополосный шум, имеющий максимум спектра в области 800–2500 Гц. Отдельные удары колес о стыки при скорости поезда 50 км/ч и более практически неразличимы. Использование звука ударов о стыки для идентификации приближающегося поезда возможно, но трудно реализуемо в силу большой зависимости от скорости поезда и качества пути.

Эти выводы были использованы при разработке прототипа устройства «Сигнализатор П». В процессе опытной эксплуатации его в 2011 г. на Октябрьской железной дороге был сделан анализ еще более 100 фонограмм приближения поездов. Первоначальные экспериментальные данные и сделанные выводы подтверждены.

1. Определение источников шума на подвижном составе // Железные дороги мира. – 1998. – № 10 [Электронный ресурс]. – Режим доступа: <http://www.css-rzd.ru/ZDM/10-1998/7136.htm>, свободный. Яз. рус. (дата обращения 09.04.2012).
2. Шум качения и методы борьбы с ним // Железные дороги мира. – 2003. – № 12 [Электронный ресурс]. – Режим доступа: <http://www.css-rzd.ru/zdm/12-2003/03069-1.htm>, свободный. Яз. рус. (дата обращения 09.04.2012).

Бибиков Сергей Викторович – ООО «Центр речевых технологий», зам. технического директора, bibikov@speechpro.com

Шапарь Александр Владимирович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, shapar.av@gmail.com

УДК 004.9, 535.6, 681.78

НАСТРОЙКА ЦВЕТОПЕРЕДАЧИ ОПТИКО-ЭЛЕКТРОННЫХ СИСТЕМ ТЕХНИЧЕСКОГО ЗРЕНИЯ ПО МЕТОДУ «ОПОРНЫХ ЦВЕТОВ»

Е.В. Горбунова, Д.Д. Шитов

Предложена методика настройки цветопередачи оптико-электронных систем технического зрения на базе использования тестовой таблицы, позволяющая осуществлять цветовой анализ наблюдаемых объектов с заданной точностью.

Ключевые слова: анализ цвета, цветопередача, тестовая таблица, преобразование цветового пространства.

На сегодняшний день оптико-электронные системы технического зрения (ОЭС ТЗ) широко применяются для решения различных задач наблюдения и контроля в научно-технической и промышленной

сфере. Как правило, цветопередача таких систем вполне достаточна для визуального наблюдения картины на экране. Однако при контроле цветовых характеристик объектов наблюдения и измерении точных цветовых координат объекта в каждой точке изображения его поверхности в автоматическом режиме [Л] возникает проблема настройки цветопередачи используемой ОЭС ТЗ. Для решения указанной проблемы существует ряд методик по настройке и улучшению цветопередачи видеоинформационных систем, обладающих, как правило, общими недостатками – взаимозависимостью двух или более функций настройки цветопередачи, а также отсутствием учета условий внешнего освещения и его возможного изменения.

Таким образом, актуальным является создание специализированной методики для оперативной подстройки параметров цветопередачи ОЭС ТЗ под особенности конкретной задачи и условий наблюдения, обнаружения и (или) анализа объектов или явлений.

Концепция предлагаемой методики настройки цветопередачи заключается в формировании преобразованного цветового пространства ОЭС ТЗ по показаниям «опорных цветов», которыми являются цвета аттестованной тестовой таблицы. Зная погрешности определения цветовых координат хотя бы для четырех цветов тестовой таблицы, можно преобразовать их и все остальные точки, заключенные между ними, в цветовом пространстве, для минимизации общей погрешности определения цвета с помощью ОЭС ТЗ.

На рисунке представлен пример решения проблемы повышения качества цветопередачи методом выделения цветового пространства ABCD ОЭС ТЗ и последующего преобразования его в пространство A'BCD, соответствующего цветовым координатам тестовой таблицы. Таким образом, поочередно корректируя каждую из «опорных» точек, можно получить ОЭС ТЗ с цветопередачей, полностью согласующейся с цветовыми координатами тестовой таблицы. Точность этого метода напрямую зависит от «удаленности» «опорных цветов» друг от друга в используемом цветовом пространстве. По этой причине, имея тестовую таблицу с достаточно большим набором цветов, можно настроить цветовые характеристики ОЭС ТЗ для проведения точных цветовых измерений.

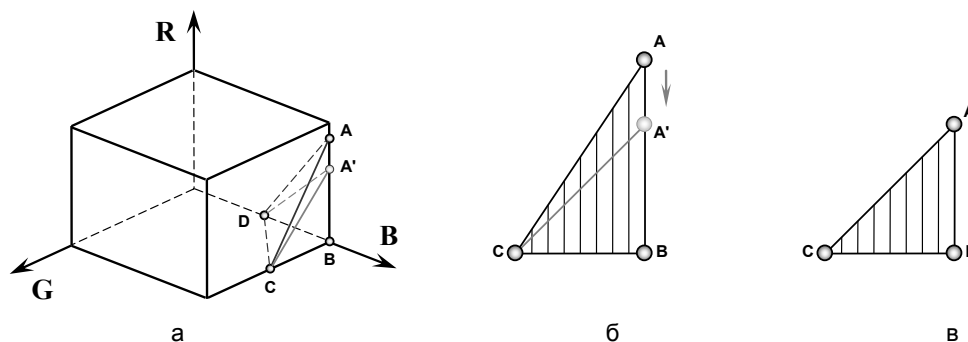


Рисунок. Преобразование цветового пространства при помощи «опорных» точек: реализация преобразования в трехмерном пространстве (а); реализация преобразования цветового пространства в треугольнике ACB (б); результат преобразования цветового пространства в треугольнике ACB (в)

Работа проводилась при финансовой поддержке Министерства образования и науки Российской Федерации в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы».

[Л]. Горбунова Е.В., Коротаев В.В., Чертов А.Н. Возможности сортировки алмазного сырья опико-электронными методами // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 4 (80). – С. 13–17.

Горбунова Елена Васильевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, ст. научный сотрудник, vredina_ia@mail.ru

Шитов Денис Дмитриевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, qdenisq@gmail.com

УДК 535.6, 628.9

МНОГОЭЛЕМЕНТНЫЙ ЦВЕТНОЙ ИСТОЧНИК ИЗЛУЧЕНИЯ

Е.В. Горбунова, О.Ю. Лашманов, В.С. Перетягин

Рассмотрены результаты разработки экспериментального образца многоэлементного цветного источника излучения в виде матрицы светодиодов, предназначенного для использования в опико-электронных системах, реализующих цветовой анализ объектов контроля.

Ключевые слова: анализ цвета, светодиод, RGB-источник излучения, равномерное освещение.

Цвет, а также процессы его восприятия и анализа являются весьма сложными объектами для изучения. В современной терминологии цветом называют характеристику цветового стимула, благодаря

которой наблюдатель может проводить различие между двумя полями зрения, одинаковыми по форме, размеру, структуре и отличающимися лишь спектральным составом. Иначе говоря, всегда, когда речь идет о «цвете», необходимо учитывать не только физические и математические закономерности его описания, но и физиологические и психологические аспекты цветового зрения. При этом всегда следует помнить о том, что цвет – это результат взаимодействия трех «участников»: источника излучения, объекта контроля и анализатора цветовой картины.

Оценки по цвету производятся во многих отраслях промышленности, например, пищевой (сортировка продуктов), текстильной (контроль качества окраски тканей), горнодобывающей (стоимостная оценка драгоценных и полудрагоценных камней), строительной (оценка сортности строительного камня, песка, гравия, пиломатериалов) и т.д. Существует большое количество разнообразных приборов и технических средств, созданных для осуществления цветового анализа и с той или иной степенью эффективности применяемых на практике. Однако в большинстве из указанных областей цветовой контроль до сих пор не автоматизирован. Одной из ключевых причин сложившейся ситуации является недостаточная изученность не только процессов анализа цветовой картины посредством оптико-электронных систем технического зрения, но и особенностей ее формирования.

Таким образом, актуальной задачей является создание источника излучения с «управляемой цветностью», посредством которого возможно обеспечивать специальные условия освещения и возможность их изменения в привязке к конкретной задаче цветового анализа.

На основании результатов проведенного математического моделирования авторами разработан экспериментальный образец многоэлементного светодиодного RGB-источника излучения, реализующего аддитивное смешение трех хроматических унитарных цветов: красного, зеленого и синего. Особенностью предложенной математической модели является возможность учета взаимного расположения излучающих элементов разных цветов в светодиодной матрице, а также величин интенсивности свечения каждого отдельного элемента для создания равномерного освещения зоны анализа определенных размеров на заданном расстоянии от источника [1]. Кроме того, модель также учитывает особенности формы индикатрисы излучения реального светодиодного элемента, заключающиеся в наличии характерного энергетического провала по центру излучающей площадки.

Экспериментальный образец многоэлементного RGB-источника излучения представляет собой светодиодную матрицу размером $70 \times 36 \text{ мм}^2$, содержащую 28 красных (марки L-934SRC-J), 28 синих (марки HB3b-448ABD) и 25 зеленых (марки BL-L324PGC) светодиодов. Для обеспечения равномерного освещения зоны анализа $160 \times 120 \text{ мм}^2$ на расстоянии 100 мм от источника разработана специальная электронная схема управления на базе платформы Arduino UNO R3, реализующая возможность адресного управления элементами светодиодной матрицы.

Проведенные экспериментальные исследования образца многоэлементного RGB-источника излучения подтвердили работоспособность разработанной математической модели.

Дальнейшим направлением исследований является создание моделей многоэлементных источников излучения кольцевой и сферической формы, а также источников, содержащих излучающие элементы шести цветов для оптимизации их спектральных характеристик [2].

Работа проводилась при финансовой поддержке Министерства образования и науки Российской Федерации в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы».

1. Chertov Aleksandr N., Gorbunova Elena V., Korotaev Valery V., Peretyagin Vladimir S. and Maria G. Serikova. Simulation of the multicomponent radiation source with the required irradiance and color distribution on the flat illuminated surface // Proc. SPIE 8429, 84290D. – 2012 [Электронный ресурс]. – Режим доступа: <http://dx.doi.org/10.1117/12.922104>, свободный. Яз. англ. (дата обращения 25.06.2012).
2. Горбунова Е.В., Перетягин В.С., Чертов А.Н. Организация освещения рабочей зоны оптико-электронных систем цветового анализа промышленного назначения // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 3 (73). – С. 140.

Горбунова Елена Васильевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, ст. научный сотрудник, vredina_ia@mail.ru

Лашманов Олег Юрьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, o.lashmanov@gmail.com

Перетягин Владимир Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, peretyagin@mail.ru

УДК 535.6, 622.7, 681.5

ОСОБЕННОСТИ АЛГОРИТМА ЦВЕТОВОГО АНАЛИЗА МИНЕРАЛОВ

Е.В. Горбунова, Д.Б. Петухова

Рассмотрена структура алгоритма цветового анализа минеральных объектов для систем регистрации оптических сепараторов и оптико-электронных комплексов оценки обогатимости руд твердых полезных ископаемых.

Ключевые слова: обогащение руд твердых полезных ископаемых, фотометрический метод, анализ цвета, алгоритм анализа.

В связи с широким применением систем технического зрения и контроля в различных отраслях промышленности актуальны разработки, связанные с распознаванием объектов наблюдения на изображении и формированием соответствующих адаптивных алгоритмов. Селективными признаками, характеризующими объект, могут быть его цвет, форма, наличие поверхностных и (или) внутренних дефектов, пространственное положение и ориентация относительно системы контроля и др. Одним из примеров подобных автоматизированных систем являются оптические (фотометрические) сепараторы, используемые в горнодобывающей промышленности для обогащения руд твердых полезных ископаемых и основанные на идентификации минеральных объектов по цвету. В существующих сепараторах, таких как, например, Optosort (производства компании ALIUD GmbH), MikroSort (компания Mogensen GmbH) или CommoDas GmbH – UltraSort, решение об отсечке минерального объекта в так называемый «концентрат» (т.е. признание его содержащим искомый полезный компонент) принимается по результатам оценки процентного содержания одной или нескольких характеристических цветовых составляющих на его поверхности. При этом, как правило, в качестве базовой цветовой модели используется система RGB (или производные от нее), не являющаяся, в силу своей неравномерности, объективной с точки зрения анализа цвета. Следствием подобного подхода являются сложность установки точных цветовых порогов и допусков при настройке данного типа обогатительного оборудования, а также оценки малых цветовых различий между компонентами в исследуемом материале. В рамках работ по исследованию возможностей создания оптических сепараторов отечественного производства, а также специальных оптико-электронных комплексов для оценки обогатимости минерального сырья фотометрическим методом разработан алгоритм цветового анализа, учитывающий, в том числе, указанные недостатки существующих сортировщиков.

Предлагаемый алгоритм включает в себя следующие основные этапы. На первом этапе реализуется предобработка изображения, полученного от видеоинформационной системы регистрации сепаратора, направленная на повышение качества кадра. В зависимости от индивидуальных особенностей конкретного вида руды выбирается и применяется метод цветовой интерполяции, оптимальный с точки зрения эффективности анализа. Для контрастного материала выбор следует делать в пользу таких алгоритмов, как Nearest, Adaptive, Freeman; при необходимости различения тонких цветовых оттенков, выделения зон цветности малых размеров более эффективными будут AHD, Wemmiao Lu, Ron Kimmel и др. На втором этапе происходит поиск и выделение объекта на полученном изображении, включая вычитание фоновой составляющей, а также удаление теней и полутеней от объекта в случае их наличия. Проведенные исследования показали, что для решения поставленной задачи анализа минеральных объектов, как статичных, так и динамических, наиболее эффективен алгоритм cvCanny, отличающийся меньшим числом ошибок и являющийся более гибким и универсальным. Для анализа движущихся минеральных объектов также применим метод межкадровой разности. На последнем этапе осуществляется собственно цветовой анализ. При этом сначала производится выбор цветовой модели, которая по совокупности критериев [Л] является наиболее подходящей для анализа материала данного типа. Далее реализуется выделение зон и участков заданной цветности с последующим вычислением их процентного отношения к общей площади изображения. При превышении предустановленного порога определяется координата энергетического центра изображения объекта и формируется команда на отсечку.

В дальнейшем предполагается усовершенствовать предложенный алгоритм за счет выделения на изображении минералов дополнительных селективных признаков, таких как, например, прожилки (линейные объекты различной степени протяженности) или «блестки» (точечные объекты). Их можно использовать в качестве альтернативы цветовому анализу или в дополнение к нему с целью увеличения эффективности процесса распознавания.

Работа проводилась при финансовой поддержке Министерства образования и науки Российской Федерации в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы».

[Л]. Вакуленко А.Д., Горбунова Е.В., Чертов А.Н. Критерии применимости цветовых пространств в оптико-электронных системах цветового анализа минералов // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 3 (79). – С. 156–157.

Горбунова Елена Васильевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, ст. научный сотрудник, vredina_ia@mail.ru

Петухова Дарья Борисовна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, petuxovadarja@yandex.ru

OPTICAL AND OPTICAL ELECTRONIC SYSTEMS. OPTICAL TECHNOLOGIES SILVER NANOCLUSTERS INFLUENCE ON FORMATION OF QUANTUM DOTS IN FLUORINE PHOSPHATE GLASSES

E. Kolobkova, N. Nikonorov, V. Aseev

Silver nanoclusters influence on formation of PbSe QDs in fluorine phosphate glasses $\text{Na}_2\text{O}-\text{P}_2\text{O}_5-\text{Ga}_2\text{O}_3-\text{AlF}_3-\text{ZnO}(\text{Se})-\text{PbF}_2$ has been studied. Absorption spectra of the glasses containing silver nanoclusters and PbSe QDs (at diameter from 3,5 nm up to 13 nm) demonstrated a shift of the first excitonic peak position to the longer wavelengths than Ag-free glasses under given heat-treatment conditions. Moreover the absorption coefficients of the glasses containing silver nanoclusters and PbSe QDs are higher than ones for Ag-free glasses.

Keywords: quantum dots, surface plasmonic resonance, silver nanoclusters, absorption spectra.

Elena Kolobkova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Associate professor, kolobok106@rambler.ru

Nikolay Nikonorov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, Department head, D.Sc., Nikonorov@oi.ifmo.ru

Vladimir Aseev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, assistant, Aseev@oi.ifmo.ru

ANALYSIS OF RAY TRACING THROUGH OPTICAL SYSTEMS WITH METAMATERIAL ELEMENTS

A.Voznesenskaya, D.Kabanova

A common expression of the refraction law is proposed. The expression allows performing ray tracing through optical media with both positive and negative refraction index. Results of computer simulation are presented. It is shown that optical systems including elements from metamaterials provide low aberrations.

Keywords: optical systems computation, nanostructured optical metamaterials.

Anna Voznesenskaya - Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate Professor, PhD, voznesenskaya@mail.ifmo.ru

Darya Kabanova – «Expert-System», Inc, engineer, kabanova@devexperts.com

MATHEMATICAL DESCRIPTION OF OPTICAL SYSTEM FOR RECTANGULAR LIGHT SPOT CREATING IN OBLIQUE BEAM

A. Trofimuk

Mathematical description is suggested and optical system is designed creating rectangular light spot in wide oblique beam without tilt of light source and optical system. This description is applicable for using in automated design process and makes it possible to get satisfied illuminance uniformity in light spot.

Keywords: LED, Bezier curve, rectangular light spot, non-imaging optics, street lights.

Anatoly Trofimuk – Ellis Amalgamated LLC (Belarus), development engineer, anatoli@opticsforhire.com

LAMELLAR LATTICE SPACING AND IRIDESCENCE INTERFERENTIAL MODEL

A. Simakov, A. Smirnov, B. Fedorov, M. Ivanov

The numerical interferential model of the iridescence phenomenon in plagioclases has been developed. Iridescence spectra for group of reference labradorite samples with different iridescence colors have been calculated on the basis of interferential model. Modeling spectra of reference samples correspond to experimental ones within the error of 10-15 %. The possibility of iridescence plagioclases out of a visible range (in ultra-violet and near infrared areas) is proved as a result of modeling. The existence of interferential maxima in near infrared area is experimentally confirmed.

Keywords: iridescence, plagioclase, lamellar lattice, interferential model.

Anton Simakov – National Mineral Resources University, research fellow, PhD, simakov@spmi.ru

Alexander Smirnov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate Professor, PhD, Associate Professor, smirnav@phd.ifmo.ru

Boris Fedorov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, full professor, D.Sc., borfedorov@rambler.ru

Mikhail Ivanov – National Mineral Resources University, Vice-chancellor for education, D.Sc., Associate professor, ivan@spmi.ru

DATA PROCESSING SCHEME MODIFICATION FOR PHASE INTERFEROMETRICAL ACOUSTIC SENSOR

M. Plotnikov, I. Deineka, I. Sharkov

Results of digital scheme design for high speed recording to micro SD memory card from electronic signals processing unit of interferometrical type fiber-optic acoustic sensor with FPGA and Xilinx ISE software

development environment are presented. Experimental results are given, which show working abilities of proposed scheme with data record speed equal to 141000 b/s.

Keywords: phase interferometrical sensor, digital processing, data recording, memory card.

Michael Plotnikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, engineer-researcher, plotnikov-michael@yanex.ru

Ivan Deyneka – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, junior researcher, vanoDNK@yandex.ru

Ilya Sharkov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, ilya.sharkov@gmail.com

SINGLE-PULSE WRITING AND VISUALIZATION OF TYPE II FIBER BRAGG GRATINGS

S. Varzhel, A. Kulikov, V. Zakharov, V. Aseev

The article presents the results of single-pulse writing of type II fiber Bragg gratings in birefringent optical fiber with elliptical stress cladding by a single KrF excimer laser pulse. Gratings diffraction efficiency induced by a single pulse exceeds 99.9% with a full-width at half maximum about 1 nm. Images of type II fiber Bragg gratings using differential interference contrast method and bright-field microscopy are obtained.

Keywords: fiber Bragg grating, birefringence, phase mask, excimer laser.

Sergey Varzhel – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, vsv187@gmail.com

Andrey Kulikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, a.kulikov86@gmail.com

Viktor Zakharov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, Viktor-zah@yandex.ru

Vladimir Aseev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, scientific researcher, aseev@oi.ifmo.ru

PHOTONICS AND OPTICAL INFORMATICS

PARTICULARITIES OF FEMTOSECOND SPECTRAL SUPERCONTINUUM GENERATION IN CRYSTAL MEDIA WITH $\chi^{(2)}$ -NONLINEARITY

S. Nalegaev, V. Bespalov, S. Putilin

Results of supercontinuum generation experiments in case of excitation by optical pulses with central wavelength 830 nm, duration 40-50 fs, energy up to 1,0-1,5 mJ are given. Supercontinuum radiation is detected in crystal DKDP and other crystals with $\chi^{(2)}$ -nonlinearity at range from 400 nm to 1450 nm, as well as up to 2200 nm, where radiation within blue range 400–2200 nm is mainly caused by parametric processes. Obtained radiation is inferior to supercontinuum generators based on micro structured fibers by the width of output spectrum.

Keywords: femtosecond pulses, $\chi^{(2)}$ nonlinearity, $\chi^{(3)}$ nonlinearity, supercontinuum generation, parametric generation.

Sergei Nalegaev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, fioi@inbox.ru

Victor Bespalov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, victorbespaloff@gmail.com

Sergei Putilin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, senior scientific researcher, PhD, SEPutilin@yandex.ru

INVESTIGATION OF LUMINESCENCE KINETICS IN PbS QUANTUM DOTS

A. Litvin, P. Parfenov, E. Ushakova, A. Baranov

The dimensional dependence of the spectral and kinetic parameters of the PbS quantum dots luminescence has been investigated in a wide spectral range (0.8 - 2 microns). Registered lifetimes lie in the range of 0.25 - 2.7 ms, and decrease with the nanocrystal sizes growth. We consider a three-level model with the energy level in the forbidden zone to explain such long luminescence lifetimes. It is shown that the anomalous size dependence of the luminescence lifetimes can be related to the existence of transitions with both decreasing and increasing of energy at room temperature.

Keywords: quantum dots, luminescence, near-infrared region, luminescence lifetimes.

Alexander Litvin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, litvin88@gmail.com

Peter Parfenov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, qrspeter@pochta.ru

Elena Ushakova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, el.ushakova@gmail.com

Alexander Baranov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., senior scientific researcher, a_v_baranov@yahoo.com

COMPENSATION METHOD FOR THE SPECTRAL HETEROGENEITY OF THE RADIATION SOURCE BY ADJUSTING ITS COLOR TEMPERATURE FOR HYPERSPECTRAL APPLICATIONS IN MICROSCOPY

I. Gurov, A. Lopatin, A. Melnikov

The paper deals with the compensation method for the spectral heterogeneity of the radiation source of a hyperspectral microscope for higher effective dynamic range of detecting video array.

Keywords: microscopy, hyperspectral image, spectrophotometry.

Igor Gurov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, gurov@mail.ifmo.ru

Alexander Lopatin – LOMO PLC, Chief optician, PhD, ailopatin@gmail.com

Alexey Melnikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, melnikov.alexey@gmail.com

PLANT TISSUE STUDY BY OPTICAL COHERENCE MICROSCOPY

I. Gurov, M. Volynsky, E. Zhukova, N. Margaryants

The studying possibilities of plant tissues microstructure by optical coherence microscopy with a tunable wavelength (1305 ± 75 nm) are discussed. The images of tomograms (*B-scans*) and 3D-images of research areas for arboreous aloe (*Aloe arborescens*), ripsalidopsis (*Hatiora gaertneri*), zonal pelargonium (*Pelargonium zonale*) are given. Research results of suberization and the surface layers degradation are analyzed. The spectra of diffuse reflection from healthy, sick and dry tissues of leaves in the range of 900–1700 nm are shown.

Keywords: optical coherence microscopy, tomogram, 3D-image, structure of plant tissue.

Igor Gurov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, gurov@mail.ifmo.ru

Maxim Volynsky – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, scientific researcher, PhD, maxim.volynsky@gmail.com

Ekaterina Zhukova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, zhukova-ifmo@mail.ru

Nikita Margaryants – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, senior lecturer, fosp@grv.ifmo.ru

ANALYSIS AND SYNTHESIS OF COMPLEX SYSTEMS ADAPTIVE STABILIZING CONTROLLER OF A SIMPLE STRUCTURE

V. Nikiforov, D. Gerasimov

An adaptive stabilizing controller for a linear time invariant multi-input multi-output plant presented in a lower triangular form is proposed. The controller has a simple structure, because it has only one adjustable parameter and one simple adaptation algorithm of integral type. Robust modifications of the controller are also presented, and simulation results are given.

Keywords: linear plant, lower triangular form, adaptive and robust control.

Vladimir Nikiforov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Vice-chancellor, D.Sc., nikiforov@mail.ifmo.ru

Dmitry Gerasimov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, gerasimovdn@mail.ru

EIGENVECTORS STRUCTURE OF STATE MATRICES IN MULTICHANNEL SYSTEMS AS A DEGENERATION FACTOR

N. Dudarenko, A. Ushakov

The problem of systems degeneration of dynamic multichannel systems due to the eigenvectors structure of their state matrices is considered. The problem is solved in relation to the identification, controllability and quality estimation with the state vector norm.

Keywords: multichannel systems, controllability, degeneration, identification.

Natalia Dudarenko – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, dudarenko@yandex.ru

Anatoly Ushakov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, Ushakov-AVG@yandex.ru

OPTIMAL SYNTHESIS FOR DYNAMIC SYSTEMS WITH DELAY IN CONTROL

D. Muzyka, R. Peshcherov, V. Tertychny-Dauri

The problem of the optimal control law formation for nonlinear dynamic systems with time delay in control channel is considered. Necessary optimality condition (Bellman equation) is proved for systems with delay in control channel in accordance with the principle of optimality. Conclusions are supported by numerical simulation in the optimal stabilization task of solid body rotation.

Keywords: delay in the control, optimal control, principle of optimality, Bellman optimization.

Dmitry Muzyka – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, 146038@niuitmo.ru

Ruslan Peschero – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, rpeshchero@mail.ru

Vladimir Tertychny-Dauri – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, tertychny-dauri@mail.ru

CAPACITIVE POSITION SENSOR FOR PLANAR ELECTRIC DRIVE

A. Balkovoy, M. Tyapkin

Planar electric drive based on planar linear motor found its application in the fast accurate planar (XY) motion tasks. Operation principle of the capacitor position sensor for the planar electric drive as well as methods of the signal processing are presented, developed prototype of the capacitive position sensor integrated into the armature of the planar linear motor is given. Results demonstrate the accuracy and resolution of the sensor.

Keywords: planar electric drive, capacitive position sensor.

Alexander Balkovoy – National Research University «Moscow Power Engineering Institute», PhD, senior scientific researcher, balk1954@yahoo.com

Mikhail Tyapkin – National Research University «Moscow Power Engineering Institute», postgraduate, tiapkinmg@mail.ru

FAST BOUNDARY ELEMENT METHOD FOR MAGNETOSTATIC PROBLEMS SOLUTION

I. Stupakov, M. Royak

The article deals with solution of magnetostatic field problems by boundary element method and its accelerating based on fast multipole method. This approach can significantly reduce the computational cost and solve problems with higher degree of freedom. Due to symmetric Galerkin formulation, this method can be used together with the finite element method for solving nonlinear problems of magnetostatics. Numerical experiments results demonstrating the effectiveness of this approach are presented.

Keywords: boundary element method, finite element method, magnetostatics, incomplete scalar magnetic potential, spherical harmonics.

Ilya Stupakov – Novosibirsk State Technical University, postgraduate, istupakov@gmail.com

Mikhail Royak – Novosibirsk State Technical University, Professor, D.Sc., Associate professor, royak@fpm.ami.nstu.ru

TRANSFORMATION PROPERTIES OF n -DIMENSIONAL DIGITAL SIGNALS ON THE BASELINE OF RECTANGULAR HITS

A. Grishentsev

The article deals with decomposition properties of n -dimensional digital signals on the baseline of rectangular hits. Linear transformation property realization, the condition of conservation of signal energy during the transition from the spatial to the frequency-spatial representation is shown. The transition to an orthogonal transformation configuration is formulated.

Keywords: decomposition of n -dimensional signals, spectral analysis, digital signal processing.

Alexei Grishentsev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, tigerpost@ya.ru

QUERY DISTRIBUTION OPTIMIZATION IN CLUSTER SYSTEM AT A COMBINATION OF ANALYTICAL MODELING AND SIMULATION

I. Golubev, V. Bogatyrev

Multi-stage procedure of the query distribution optimization in cluster system using a combination of analytical modeling and simulation is proposed. This procedure makes it possible to find the optimal part of the redistributed query flow for different distributions of intervals between incoming requests and time of their service.

Keywords: load distribution, simulation, cluster, optimization.

Ivan Golubev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, www.golubev@mail.ru

Vladimir Bogatyrev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, Vladimir.bogatyrev@gmail.com

RELIABILITY EVALUATION MATHEMATICAL MODEL OF AVIONICS ITEM WITH AN UNLOADED RESERVE

O. Kuznetsova

Faultness evaluation mathematical model for avionics item with loaded duplication and an additional unloaded reserve is designed. The model takes into account incomplete operation control of devices, and time delay in the engaging of unloaded device. Comparative analysis of structural reservation variants is done.

Keywords: reliability, faultness, unloaded reserve, operation control.

Olga Kuznetsova – Federal State Unitary Enterprise «Saint Petersburg Experimental Design Bureau «Electroavtomatika» named after P.A. Efimov», Head of reliability section, kuzola@ya.ru

BIOMEDICAL TECHNOLOGIES

DRUG INTOXICATION INFLUENCE ON SPEAKERS' VOICE FEATURES

A. Raev, Yu. Matveev, T. Goloshchapova

The paper deals with analysis of speech recordings before and after speakers' drug-abuse treatment. It is shown that there are no statistically significant dependences in distortions of speaker's voice features on different groups of drugs and on the drug intoxication degree. The fundamental frequency changes are not regular and do not have general nature. The main reason for the fundamental frequency change is the speaker's emotional state change, rather than drug addiction treatment. Studying the duration effect of narcotic drugs use on the speakers' voice showed that for speakers with prolonged use of the heroin group drugs a tendency to decrease the fundamental frequency by about 3% per year was found.

Keywords: speaker, pitch frequency, drug intoxication.

Andrei Raev – «Speech Technology Center», LLC, Director of Research Department, raev@speechpro.com

Yuri Matveev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., matveev@speechpro.com

Tatiana Goloshchapova – Expert and Forensic Bureau of Drug Control Service of the Russian Federation, Section head, PhD.

COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES

FUZZY MARKOVIAN MODEL BASED TECHNOLOGY FOR ASSESSMENT OF MULTI-AGENT DISTRIBUTED INFORMATION SYSTEMS RELIABILITY STATUS

A. Masloboev, V. Bogatikov

Assessment technology for multi-agent distributed information systems reliability status based on fuzzy Markovian model, which nodes are represented as integral reliability indexes, has been developed. The index is an assessment of fuzzy equality degree between a current situation, which is described by reliability indexes system, and the situations, which corresponding to network safe functioning domain. The technology provides information and hardware system resources distribution identification for reliability goal achievement and support.

Keywords: technology, modeling, reliability status assessment, distributed information systems, fuzzy Markovian model, multi-agent systems.

Andrey Masloboev – Federal State Budgetary Establishment of Science Institute for Informatics and Mathematical Modeling of Technological Processes Kola Science Center of the Russian Academy of Sciences, senior research fellow, PhD, Associate professor, masloboev@iimm.kolasc.net.ru

Valery Bogatikov – Federal State Budgetary Establishment of Science Institute for Informatics and Mathematical Modeling of Technological Processes Kola Science Center of the Russian Academy of Sciences, leading research fellow, D. Sc., Professor, vnbgtk@iimm.kolasc.net.ru

METHOD OF SEMI-AUTOMATIC MORPHOLOGICAL DICTIONARY GENERATION

S. Lapshin, I. Lebedev

This article provides a method for semi-automatic morphological dictionary generation. Provided method makes it possible to simplify significantly the procedure of new dictionary generation and update and, thus, to increase the precision and recall values of DLP and IPC morphological parsers.

Keywords: semi-automatic dictionary generation, morphological description, DLP and IPC systems.

Sergey Lapshin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, sv.lapshin@gmail.com

Ilya Lebedev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, lebedev@cit.ifmo.ru

APPLICATION OF NONLINEAR DYNAMICS METHODS FOR JOY EMOTION RECOGNITION IN SPEECH

C. Sidorov, N. Filatova

The article deals with the task of recognizing speech samples, recorded when speakers under test display the emotion of joy, from samples of the same speakers' speech in a neutral state. Methods of nonlinear dynamics were used to solve the problem. Studies were performed on recordings from the base of Emo-DB (Berlin) and fragments of the Russian bases (Tver). A model housing of emotional speech was formed, consisting of two-level database (sentences and phonemes), served as the basis for the initial evaluation of developed algorithms performance. Stable signs of nonlinear dynamics were formed - the attractor reconstruction and the recurrence plot. New quantitative signs were proposed for speech patterns classification of the person experiencing the joy emotion, based on estimates of the maximum vectors of the attractor reconstruction for the four quadrants.

Keywords: emotion, emotional state, speech, speech signal, nonlinear dynamics, attractor reconstruction, recurrence plot.

Constantine Sidorov – Tver State Technical University, postgraduate, bmisidorov@rambler.ru, bmisidorov@mail.ru

Natalia Filatova – Tver State Technical University, Professor, D.Sc., Professor_nfilatova99@mail.ru

EFFICIENCY INCREASING METHOD OF THE EVOLUTIONARY ALGORITHMS BY REINFORCEMENT LEARNING

A. Buzdalova, M. Buzdalov

A scalar optimization method based on evolutionary algorithms controlling by reinforcement learning is proposed. Reinforcement learning is used to choose the most effective fitness function at each generation of the evolutionary algorithm. Experimental results for a model problem H-IFF are given. Comparison of the developed method with other evolutionary optimization methods is performed. According to experimental results, the proposed method increases the effectiveness of evolutionary algorithms.

Keywords: scalar optimization, multi criteria optimization, reinforcement learning, evolutionary algorithms, H-IFF.

Arina Buzdalova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, afanasyevarina@gmail.com

Maxim Buzdalov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, mbuzdalov@gmail.com

MATERIALS TECHNOLOGY AND NANOTECHNOLOGIES INVESTIGATION OF NANOPARTICLES MOVEMENT EFFECT UNDER THE INFLUENCE OF FIELDS AND CHARGES CREATED BY A FOCUSED ELECTRON BEAM

Ph. Komissarenko, A. Denisjuk

The article deals with investigation of dielectric nanoparticles movement effect observed by electron microscope. In the experiments, 120 nm size polystyrene nanoparticles were moved over a substrate and fixed on the end of a closely placed metallic needle being charged under the electron beam. Interaction mechanisms in the system «particle–needle–substrate» were analyzed. Results of calculation and modeling are in agreement with experimentally observed effects. The observed effect can serve as a basis of a new method for precise movement of nanoparticles.

Keywords: electron microscopy, methods of nanoparticles precise movement.

Philipp Komissarenko – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, malkav86@mail.ru

Andrei Denisjuk – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, aidenisjuk@gmail.com

INHOMOGENEOUS ELECTRONIC SPECTRA BROADENING FOR THE DYE MOLECULES WITHIN NANOPOROUS GLASS

Yu. Kolesnikov, A. Novikov

An impact of the exciting radiation frequency on the fluorescence spectra parameters of adsorbed molecules has been investigated. Observed effects are interpreted in the belief that electronic spectra of the dye molecules within nanoporous glass undergo an inhomogeneous broadening. In some cases, the found dependences demonstrate a non-monotonic character, that's why inhomogeneous broadening not always might be described with simple Gaussian function. It has been proved that in present case the inhomogeneous broadening appears due to potential variations of intermolecular interactions between adsorbate and adsorbent. The mentioned variations take place due to variable local curvature of the pore surface and coexistence of certain stable molecule orientations on the surface.

Keywords: molecular fluorescence and absorbance spectroscopy, organic dye molecules, molecular interactions, nanoporous silicate glasses.

Yury Kolesnikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Vice-chancellor, D.Sc., Professor, kolesnikov@mail.ifmo.ru

Alexander Novikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, alexfednov@gmail.com

PULSED FIBER LASER APPLICATION FOR SiO₂/Si SYSTEM MICROSTRUCTURING

A. Skvortsov, V. Veiko, Huynh Cong Tu

The effect of SiO₂/Si system irradiation by pulsed ytterbium fiber laser PLI-1-50 on the morphology and system structural properties has been investigated. It is shown that during surface irradiation of the monocrystalline silicon wafer covered by a thin layer of thermally grown silicon dioxide (SiO₂), there are significant structural changes associated with localization of the silicon strips of slip-lines and grid of slip-lines on the surface formed by the intersection of these strips. Phonon theory of the resulting microstructural topography has been considered.

Keywords: laser irradiation, dislocations, slip-lines, grid of slip-lines, mechanical stresses, plastic deformation, recrystallization, nanocomposite.

Albert Skvortsov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, a-skvortsov@yandex.ru

Vadim Veiko – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, Professor, D.Sc., Professor, veiko@lastech.ifmo.ru

Huynh Cong Tu – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, Picochip912@yahoo.com

THERMAL PHYSICS AND THEORETICAL HEAT ENGINEERING **POSSIBILITY INVESTIGATION OF BATTERY HEAT SENSORS APPLICATION IN NON-STATIONARY HEAT MEASUREMENT**

I. Sivakov

The article deals with experimental results showing the application possibility of the battery heat sensor (Gerashchenko sensor) in non-stationary heat measurement with the method of parametric identification of the heat transfer model in the sensor.

Keywords: non-stationary heat measurement, heat flux sensors, Gerashchenko sensor, inverse heat conduction problem, parametric identification, Kalman filter, differential-difference model.

Ivan Sivakov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, i.a.sivakov@gmail.com

INFORMATION SECURITY METHODS AND SYSTEMS **MULTIDIMENSIONAL MODEL OF DISCRETIONARY ACCESS TO THE OBJECTS IN VERSION CONTROL SYSTEM**

A. Spivak, A. Razumovskiy, I. Zikratov

A new model of discretionary access control for file version control systems is suggested. Three-dimensional access control model based on Harrison-Ruzzo-Ullman discretionary model is considered. For management and control tasks solution the base operations of Harrison-Ruzzo-Ullman model are supplemented with new operations taking into account the persons access to object versions.

Keywords: model, security, discretionary access control, versions control.

Anton Spivak – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, anton.spivak@gmail.com

Andrey Razumovskiy – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, assistant, PhD, xrew@yandex.ru

Igor Zikratov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department Head, D.Sc., Associate professor, zikratov@cit.ifmo.ru

ECONOMICS AND FINANCES. MANAGEMENT **METHODS FOR ASIAN OPTION PRICING WITH PAYMENTS OF DISCRETE DIVIDENDS OUT OF AVERAGING PERIOD**

M. Kosyakov, M. Ponomarev, D. Ivanov, Y. Shpolyanskiy

The article deals with two methods for Asian option pricing with arithmetic averaging and multiple payments of absolute discrete dividends out of averaging period: direct Monte-Carlo simulation and our suggested original method based on Crank-Nicolson finite-difference scheme with analytical Curran approximation as the terminal condition. The comparative analysis shows that the precision provided by the original approach is acceptable for practical problems with the grid settings recommended in the paper for different input parameters. The calculation takes a few milliseconds (or even fractions of ms) on a modern PC against seconds for the Monte-Carlo simulation.

Keywords: Asian option, discrete dividends, Monte-Carlo simulation, finite-difference scheme, Crank-Nicolson scheme, Curran approximation.

Mikhail Kosyakov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, mkosyakov@gmail.com

Maxim Ponomarev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, maxim.v.ponomarev@gmail.com

Dmitry Ivanov – Tbricks AB, mathematician; Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, dm.vl.ivanov@gmail.com

Yuri Shpolyanskiy – Tbricks AB, leading mathematician; Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, D.Sc., Associate professor, shpolyan@mail.ru

BRIEF REPORTS

ON THE TERMINOLOGY ISSUES IN THE FIELD OF MECHATRONICS

E. Shalobaev, R. Tolochka

The publication covers topical issues of terminology in the field of mechatronics in connection with the 25th anniversary of the term «mechatronics» appearance in Russia, as well as dissemination of the terms recommended by the Committee for standardization and terminology of the International Organization for theory of machines and mechanisms.

Keywords: mechatronics, micro- and nanomehatronika, terminology, international translator, micro-system engineering, level approach, triad «sensors–controller–actuator».

Evgeny Shalobaev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, PhD, Associate professor, shalobaev47@mail.ru

Rimantas Tolochka – Kaunas University of Technology, Professor, D.Sc., Professor, tadas.tolocka@ktu.lt

ON THE APPLICATION OF RECTIFIED LOGARITHMIC CHARACTERISTICS METHOD IN SMOOTHING ALGORITHMS

A. Loparev, O. Stepanov, O. Yashnikova

The article deals with summarizing of approximate rectified logarithmic characteristics method for stationary smoothing algorithms.

Keywords: filtering, smoothing, logarithmic characteristics method, stationary processes.

Alexei Loparev – Concern CSRI Elektropribor, JSC, Head of section, PhD, Associate professor, loparev@mail15.com

Oleg Stepanov – Concern CSRI Elektropribor, JSC, Head of the Scientific and Educational center, D.Sc., soalax@mail.ru

Olga Yashnikova – Concern CSRI Elektropribor, JSC, postgraduate, olga_evstifeeva@mail.ru

MOVING TRAIN AS A SOURCE OF ACOUSTIC WAVES PROPAGATING IN THE RAILWAY TRACK

S. Bibikov, A. Shapar

Based on experimental data, the trundle noise of the «wheel – rail» pair was selected as the dominant source of sound waves for approaching train detector.

Keywords: noise level, trundle noise, train approach.

Sergei Bibikov – Speech Technologies Centre, Ltd, Deputy technical director, bibikov@speechpro.com

Alexander Shapar – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, shapar.av@gmail.com

COLOR RENDITION SETTING FOR OPTICAL-ELECTRONIC VISION SYSTEMS BY «REFERENCE COLORS» METHOD

E. Gorbunova, D. Shitov

The method of color rendition setting for optical-electronic vision systems based on the use of test table is proposed. This method makes it possible to analyze the observed objects color with a given degree of accuracy.

Keywords: color analysis, color rendition, test table, color space conversion.

Elena Gorbunova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, senior staff scientist, PhD, vredina_ia@mail.ru

Denis Shitov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, qdenisq@gmail.com

MULTIELEMENT COLOR LIGHT SOURCE

E. Gorbunova, O. Lashmanov, V. Peretyagin

The article deals with development results of an experimental prototype of multielement color light source in the LEDs matrix form for use in optical-electronic systems implementing a color analysis of observed objects.

Keywords: color analysis, LED, RGB light source, uniform illumination.

Elena Gorbunova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, senior staff scientist, PhD, vredina_ia@mail.ru

Oleg Lashmanov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, o.lashmanov@gmail.com

Vladimir Peretyagin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, peretyagin@mail.ru

FEATURES OF MINERALS COLOR ANALYSIS ALGORITHM

E. Gorbunova, D. Petukhova

The article deals with color analysis algorithm structure of mineral objects for detection systems of optical sorters and for optical-electronic complexes for evaluation of solid minerals ore dressability.

Keywords: solid minerals ore dressing, photometric method, color analysis, analysis algorithm.

Elena Gorbunova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, senior staff scientist, PhD, vredina_ia@mail.ru

Dar'ya Petukhova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, petuxovadarja@yandex.ru