

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

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

---

**ВЕСТНИК  
конференции молодых ученых  
СПбГУ ИТМО**

**Сборник научных трудов**

**Том 1**



**САНКТ-ПЕТЕРБУРГ  
2004**

Выпуск содержит материалы I конференции молодых учёных Санкт-Петербургского государственного университета информационных технологий, механики и оптики.

Конференция была проведена 16–19 февраля 2004 г. Санкт-Петербургским государственным университетом информационных технологий, механики и оптики.

## **ПРОГРАММНЫЙ КОМИТЕТ КОНФЕРЕНЦИИ**

### **Председатель**

Ректор СПбГУ ИТМО, д.т.н., профессор **В.Н. Васильев**

### **Сопредседатели**

Проректор по развитию, д.т.н., проф. **В.О. Никифоров**  
Проректор по УО и АР, д.ф.-м.н., проф. **Ю.Л. Колесников**  
Декан факультета ППО, д.т.н., проф. **В.Л. Ткалич**

### **Члены программного комитета**

Д.ф.-м.н., проф. **В.Г. Беспалов**, д.т.н., проф. **Д.Д. Куликов**, д.т.н., проф. **В.М. Мусалимов**, д.т.н., проф. **А.А. Ожиганов**, д.т.н., проф. **Э.С. Путилин**, д.ф.-м.н., проф. **В.С. Сизиков**, д.т.н., проф. **А.М. Скворцов**, д.т.н., проф. **С.К. Стафеев**, д.т.н., проф. **В.А. Тарлыков**, д.т.н., проф. **Е.Б. Яковлев**, к.ф.-м.н., доц. **А.А. Андреева**, к.т.н., доц. **Н.Н. Горлушкина**, к.т.н., доц. **Л.А. Губанова**, к.т.н., доц. **Н.Ф. Гусарова**, к.ф.-м.н., доц. **А.А. Королёв**, к.т.н., доц. **Б.А. Крылов**, к.т.н., доц. **М.Я. Марусина**, к.т.н., проф. **Б.С. Падун**, к.ф.-м.н., доц. **О.А. Приходько**, к.ф.н., доц. **В.Н. Садовников**, к.т.н., доц. **Б.Д. Тимченко**

## **ОРГАНИЗАЦИОННЫЙ КОМИТЕТ КОНФЕРЕНЦИИ**

### **Председатель**

Зам. проректора по НР **Л.М. Студеникин**

### **Зам. председателя**

К.т.н., доц. **В.И. Соловьёв**

### **Члены организационного комитета**

**П.А. Белов**, **К.В. Богданов**, **Н.Ф. Гусарова**, **А.А. Зинчик**, **Л.Н. Казар**, **А.В. Козаченко**, **А.Г. Метляков**, **М.В. Никитина**, **Д.В. Осипов**, **А.Ю. Прокопенко**, **Т.А. Прудентова**, **К.А. Сергеев**, **Т.В. Точилина**, **В.Н. Фролков**, **Р.А. Халецкий**, **Д.Г. Штенников**

ISBN 5-7577-0150-1

© Санкт-Петербургский государственный университет  
информационных технологий, механики и оптики,  
2004

# ПРЕДИСЛОВИЕ

---

## **Уважаемый читатель!**

Вы держите в руках первый выпуск сборника научных трудов конференции молодых учёных Санкт-Петербургского государственного университета информационных технологий, механики и оптики, прошедшей в феврале 2004 г.

Это издание задумывалось как площадка, с которой мог бы стартовать молодой специалист, способный к научно-технической работе. И одновременно это попытка совместить в одном издании всё многообразие различных научных направлений, представленных в СПбГУ ИТМО. Возможно, именно эти страницы станут для кого-то отправной точкой в междисциплинарном прорыве или удачной попыткой создания единой методологии для различных отраслей науки и техники. Материалы сборника могут представлять интерес для студентов, аспирантов и молодых специалистов, чьи научные интересы лежат в сфере оптики, лазерной физики, физики кристаллов и полупроводников, механики и мехатроники, измерительной техники, технологии приборостроения, информационных технологий, САПР, системного анализа и теории управления, медицинской техники и технологии, а также целого ряда гуманитарных областей.

Сборник содержит два тома. В первом томе размещены работы секций «Оптические технологии» (посвящённые вопросам физической оптики и оптического приборостроения) и «Информационные технологии» (связанные с вопросами проектирования вычислительных систем, программного и алгоритмического обеспечения). Во втором томе представлены труды секций «Механика, приборостроение и системный анализ» (отражающие проблемы

приборостроения, точной механики, узлов и деталей машин, проектирования технологических процессов, измерительной техники, математического моделирования и теории управления), «Физика твёрдого тела» (связанные с вопросами полупроводниковой технологии и лазерной модификации материалов), «Биомедицинские технологии и томография» (затрагивающие вопросы биомедицинского приборостроения и разработки методов рентгеновской и магнитно-резонансной томографии), «Гуманитарные науки» (посвящённые задачам гуманитарных наук, представленных в научных школах СПбГУ ИТМО от экономической теории до педагогики и философии).

Работы, вошедшие в сборник, отражают широкий спектр научно-технических исследований, которые ведутся не только в СПбГУ ИТМО, но и в целом ряде сотрудничающих с ним научных и производственных организаций: ВНИЦ ГОИ им С.И. Вавилова, ФТИ им. А.Ф. Иоффе, ИАнП РАН, ОАО «ЛОМО» и т.п. В сборнике подробно представлено новое направление физических основ оптоинформатики, которое развивается в тесном взаимодействии со специалистами ГОИ им С.И. Вавилова. Традиционно для СПбГУ ИТМО представлено большое количество трудов по конструированию оптических систем и лазерной физике.

*Доктор технических наук,  
профессор*

*В.Л. Ткалич*

## ГЕНЕРАЦИЯ АНТИСТОКСОВОГО ИЗЛУЧЕНИЯ ПРИ ВЫНУЖДЕННОМ КОМБИНАЦИОННОМ РАССЕЯНИИ В УСЛОВИЯХ ФАЗОВОГО КВАЗИСИНХРОНИЗМА

Н.С. Макаров

Научный руководитель – д.ф.-м.н, проф В.Г. Беспалов

На основе выведенной системы связанных дифференциальных уравнений проведен численный анализ генерации антистоксового излучения в условиях фазового квазисинхронизма. Моделирование показало возможность реализации слоистой структуры, обеспечивающей эффективную генерацию антистоксового излучения. Изучен ряд параметров, влияющих на эффективность преобразования, а также степень их влияния. Результаты численного моделирования могут быть использованы для построения эффективных нелинейно-оптических перестраиваемых источников когерентного излучения.

### Введение

Вынужденное комбинационное рассеяние (ВКР) широко используется для дискретной перестройки излучения импульсных и непрерывных лазеров [1-5]. Как правило, в процессе ВКР генерируется ряд линий, сдвинутых как в стоксовую, так и в антистоксовую области спектра относительно линии возбуждающего излучения. Наиболее эффективная ВКР-генерация происходит на частоте первой стоксовой компоненты, для которой коэффициент преобразования может достигать предельных величин, соответствующих соотношению Мэнли-Роу [4]. Эффективность преобразования в первую антистоксовую компоненту рассеяния ( $\eta_a$ ) не превышает 3–6% [6, 7], и для ее увеличения используют сложные схемы четырехволнового смешения с выполнением условия фазового синхронизма [8], в которых экспериментальная эффективность преобразования в антистоксовую компоненту не превышает 15%.

Н. Бломбергенем в 1962 г. [9] было предложено использовать среды с периодически изменяемыми параметрами нелинейности второго порядка ( $\chi^{(2)}$ ) для повышения эффективности генерации второй гармоники лазерного излучения путем реализации фазового квазисинхронизма. Впоследствии эта идея была экспериментально реализована [10] и в настоящее время широко используется для преобразования частоты лазерного излучения [11].

Таким образом, изучение возможности генерации антистоксового излучения ВКР в среде с изменяемыми параметрами нелинейности третьего порядка ( $\chi^{(3)}$ ) вдоль продольной координаты является актуальным с точки зрения как расширения представлений о физических процессах комбинационного рассеяния, так и создания дискретно перестраиваемых по частоте источников высококогерентных импульсов.

### Теория

Многоволновое попутное и обратное ВКР можно описать в приближении медленно меняющихся огибающих системой связанных дифференциальных уравнений

$$\begin{cases} \left( \frac{\partial}{\partial z} + \frac{n(\omega_j)}{c} \frac{\partial}{\partial t} \right) E_j^+ = \frac{g_j^+ \omega_j}{2\omega_{-1} i} \left( E_{j+1}^+ q^* e^{iz\Delta_{j+1}^1} + E_{j-1}^+ q e^{-iz\Delta_j^1} + E_{j+1}^- q^* e^{iz\Delta_{j+1}^2} + E_{j-1}^- q e^{-iz\Delta_j^3} \right) \\ \left( -\frac{\partial}{\partial z} + \frac{n(\omega_j)}{c} \frac{\partial}{\partial t} \right) E_j^- = \frac{g_j^- \omega_j}{2\omega_{-1} i} \left( E_{j+1}^- q^* e^{iz\Delta_{j+1}^4} + E_{j-1}^- q e^{-iz\Delta_j^4} + E_{j+1}^+ q^* e^{iz\Delta_{j+1}^3} + E_{j-1}^+ q e^{-iz\Delta_j^2} \right), \\ \frac{\partial q}{\partial t} = \frac{1}{iT_2} \left( -iq + \sum_j E_j^+ E_{j-1}^{+*} e^{iz\Delta_j^1} + \sum_j E_j^- E_{j-1}^{-*} e^{iz\Delta_j^4} + \sum_j E_j^- E_{j-1}^{+*} e^{iz\Delta_j^2} + \sum_j E_j^+ E_{j-1}^{-*} e^{iz\Delta_j^3} \right) \end{cases}$$

где  $E_j^\pm$  – медленно меняющиеся амплитуды попутных и обратных компонент ВКР, отрицательным значениям  $j$  соответствуют стоксовы компоненты ВКР, нулевому значению – волна накачки, а положительным значениям – антистоксовы компоненты ВКР;  $\omega_j = \omega_0 + j\omega_v$ , взаимодействующие волны  $E_j^+$  распространяются в положительном направлении оси  $z$ , а волны  $E_j^-$  – в отрицательном направлении оси  $z$ ,  $T_2$  – время дефазировки молекулярных колебаний,  $\omega_v$  – частота молекулярного перехода,  $g_j^\pm$  – коэффициенты стационарного ВКР-усиления.

### Результаты численного моделирования

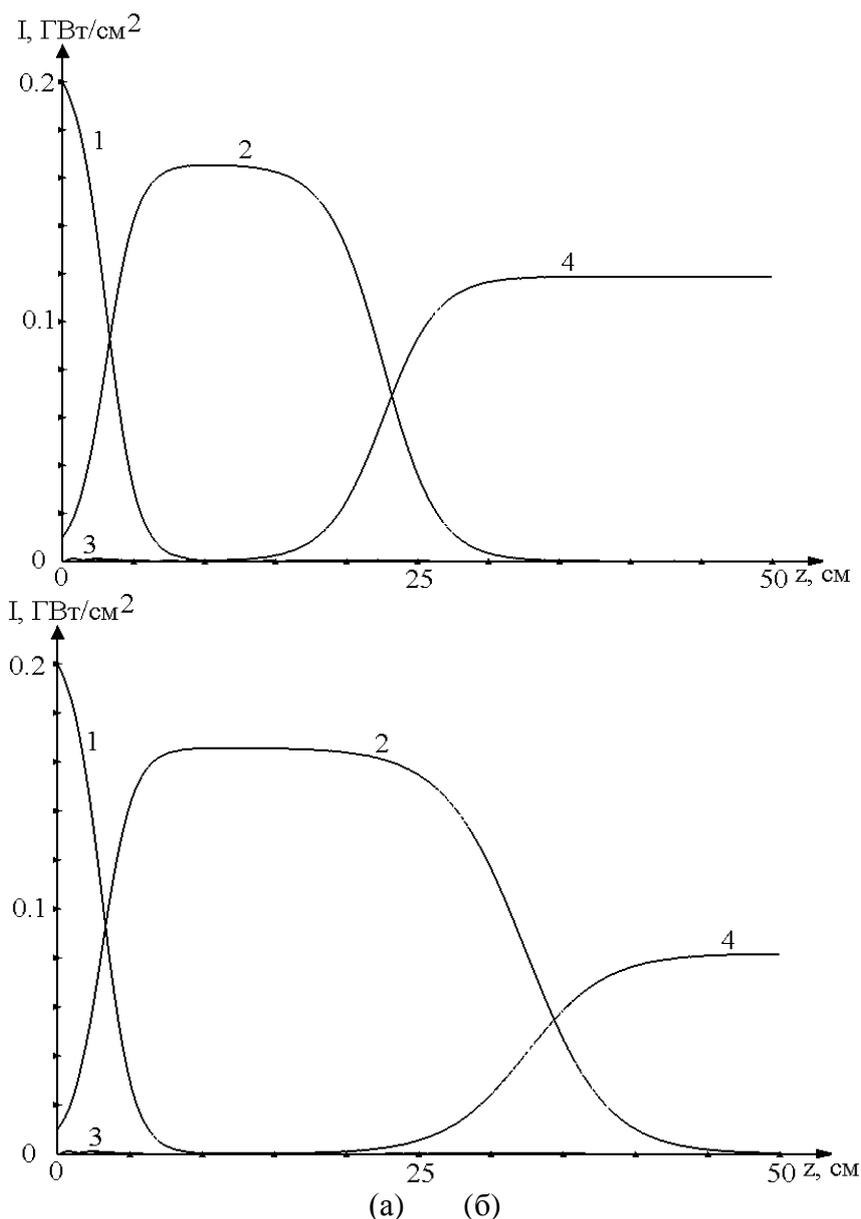


Рис. 1. Зависимость интенсивностей взаимодействующих волн от продольной координаты в водороде без учета дисперсии коэффициента стационарного ВКР-усиления (а) и с ее учетом (б): 1 – волна накачки, 2 – первая стоксовая компонента ВКР, 3 – первая антистоксовая компонента ВКР, 4 – вторая стоксовая компонента ВКР (влияние учета дисперсии коэффициента стационарного ВКР-усиления на результаты численного моделирования многоволнового ВКР в водороде)

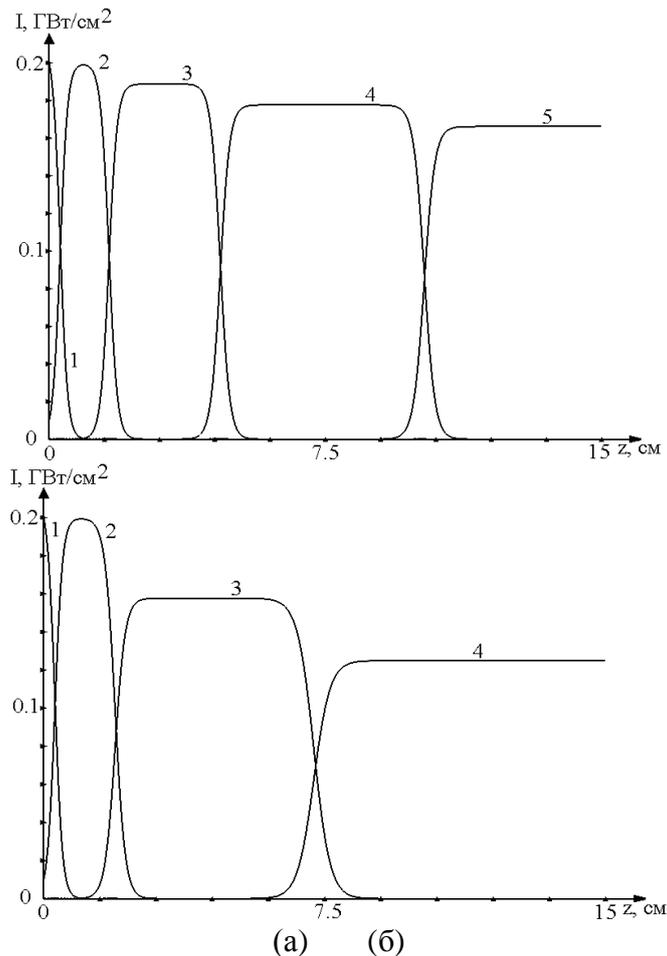


Рис. 2. Зависимость интенсивностей взаимодействующих волн от продольной координаты в нитрате бария без учета дисперсии коэффициента стационарного ВКР-усиления (а) и с ее учетом (б): 1 – волна накачки, 2 – первая стоксовая компонента ВКР, 3 – вторая стоксовая компонента ВКР, 4 – третья стоксовая компонента ВКР, 5 – четвертая стоксовая компонента ВКР (влияние учета дисперсии коэффициента стационарного ВКР-усиления на результаты численного моделирования многоволнового ВКР в нитрате бария)

Для сжатого водорода ( $g=4.42$  см/ГВт,  $\Delta=3.84$  рад/см,  $p=30$  атм) и нитрата бария ( $g=47.42$  см/ГВт,  $\Delta=77.84$  рад/см) на основании теоретической модели ВКР было проведено предварительное моделирование многоволнового ВКР с учетом дисперсии коэффициента стационарного ВКР-усиления и без ее учета (рис. 1, 2).

Из расчетов и приведенных графиков видно, что при учете дисперсии коэффициента стационарного ВКР-усиления эффективность преобразования энергии во вторую стоксовую компоненту в отсутствие слоистой структуры снижается на 30% для водорода и на 15% для нитрата бария. Поскольку в реальных комбинационно-активных средах имеет место дисперсия коэффициента стационарного ВКР-усиления, а разница в результатах, полученных с учетом и без учета дисперсии  $g$ , существенна, то для повышения точности расчетов все дальнейшие вычисления проводились с ее учетом. Кроме того, численное моделирование показало, что для повышения точности расчетов необходимо и достаточно учитывать генерацию 4–5 стоксовых и антистоксовых компонент ВКР.

Численное моделирование генерации антистоксового излучения в условиях фазового квазисинхронизма показало, что оптимальным выбором длин активных и пассивных слоев можно создать структуру для достижения эффективного антистоксового ВКР-преобразования (с эффективностью  $>20\%$ ). Зависимости длин слоев такой структуры в сжатом водороде от их номера представлены на рис. 3.

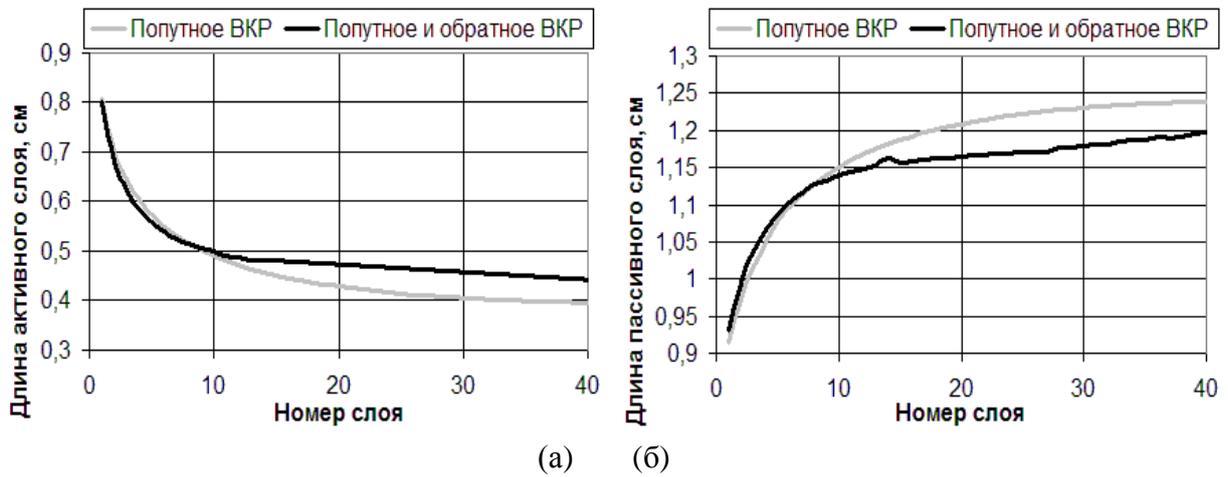


Рис. 3. Зависимости длин активных (а) и пассивных (б) слоев оптимальной слоистой структуры, обеспечивающей эффективную генерацию антистоксового излучения в сжатом водороде в условиях фазового квазисинхронизма ( $I_0^+ = 0.2$  Гвт/см<sup>2</sup>,  $I_{-1}^+ = 0.01$  Гвт/см<sup>2</sup>), от номера слоя (сравнение приближения однонаправленного взаимодействия с двунаправленным)

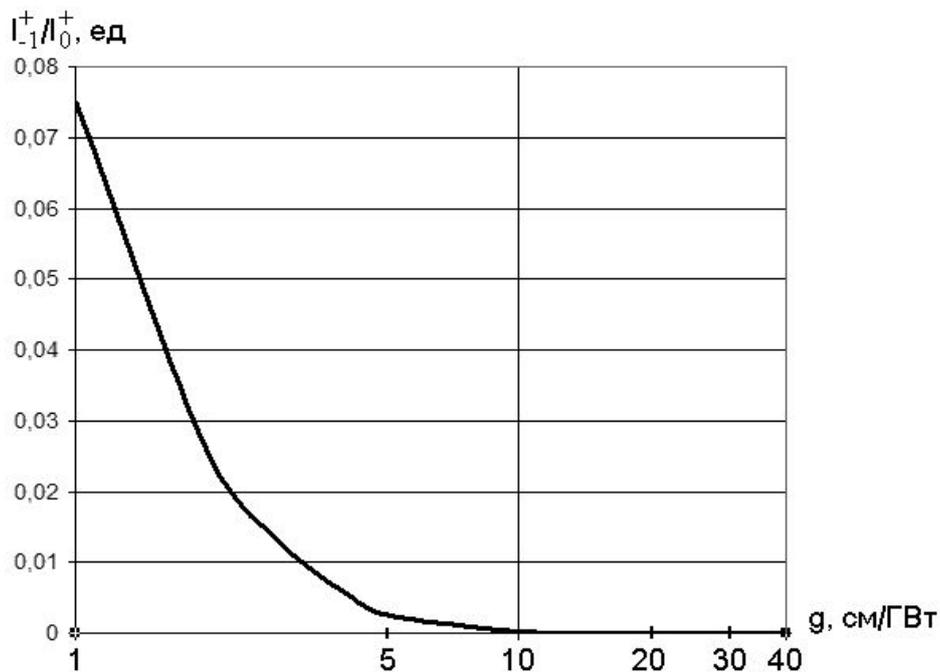


Рис. 4. Зависимость оптимального соотношения входных интенсивностей волн Стокса и накачки от коэффициента стационарного ВКР-усиления

Расчеты показали, что в сравнении с однонаправленным приближением взаимодействия учет генерации обратного ВКР приводит к незначительному уменьшению длин активных слоев и увеличению длин пассивных слоев в начале слоистой структуры и увеличению длин активных слоев и уменьшению длин пассивных слоев в ее конце. Эффективность антистоксового ВКР-преобразования снижается с 30% до 25%, эффективность обратного ВКР в слоистой структуре не превышает 0.01%. Это связано с тем, что часть энергии волны накачки тратится на генерацию обратного ВКР, что слегка нарушает условия фазового квазисинхронизма. В начале слоистой среды из-за генерации обратного ВКР снижаются длины активных слоев, что приводит к сокращению преобразования энергии из волны накачки в ВКР-компоненты. В результате к концу среды интенсивность волны накачки оказывается немного больше, чем в приближении однонаправленного взаимодействия, что позволяет немного увеличить длины активных слоев в конце среды. Исследование зависимости эффективности антистоксового ВКР-

преобразования от точности выбора длин активных и пассивных слоев показало, что при изменении длин слоев оптимальной слоистой структуры, обеспечивающей фазовый квазисинхронизм, на 5%, эффективность антистоксового ВКР-преобразования изменяется не более, чем на 0.02%. Максимально допустимая случайная центрированная погрешность в выборе длин слоев, практически не влияющая на эффективность антистоксового ВКР-преобразования, составляет 15%.

Для определения оптимальных условий антистоксовой ВКР-генерации была исследована зависимость оптимального соотношения входных интенсивностей стоксовой волны и волны накачки от волновой расстройки и коэффициента стационарного ВКР-усиления. Было установлено, что при фиксированной волновой расстройке оптимальное соотношение входных интенсивностей убывает при росте коэффициента стационарного ВКР-усиления, а при фиксированном коэффициенте стационарного ВКР-усиления оно не зависит от волновой расстройки (рис. 4). Общая зависимость оптимального соотношения входных интенсивностей от волновой расстройки и коэффициента стационарного ВКР-усиления может быть аппроксимирована формулой  $\frac{I_{-1}^+}{I_{0\text{opt}}^+} = 0.1359g^{-2.6146}$ , где коэффициент стационарного ВКР-усиления измеряется в см/ГВт.

Была изучена зависимость максимальной эффективности преобразования от входной интенсивности волны накачки. Установлено, что с ростом интенсивности волны накачки максимальная эффективность преобразования возрастает, однако для каждой среды существует критическое значение интенсивности волны накачки, при котором эффективность антистоксового преобразования существенно снижается. Наличие критического значения связано с существенным ростом интенсивности первой стоксовой компоненты с последующей генерацией высших стоксовых компонент уже в первом активном слое, что полностью истощает энергию волны накачки, и рост интенсивности антистоксовой компоненты в следующем слое не происходит.

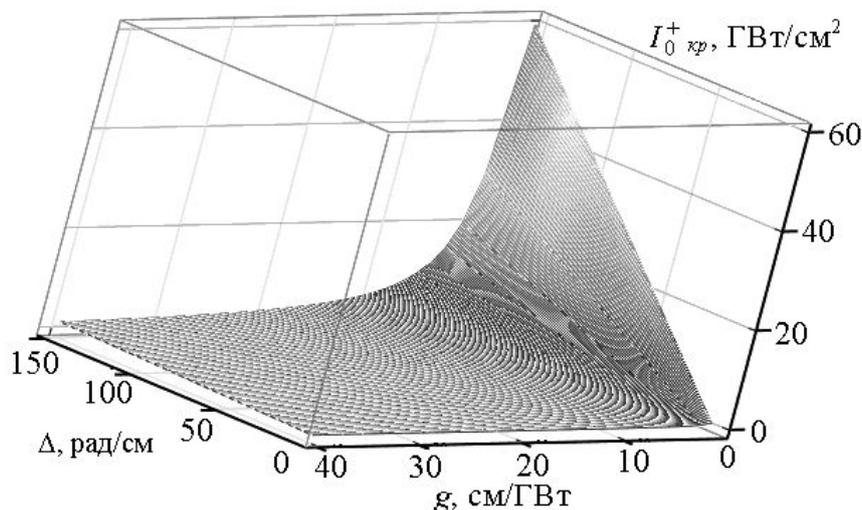


Рис. 5. Зависимость критической интенсивности волны накачки от волновой расстройки и коэффициента стационарного ВКР-усиления

Установлено, что при фиксированной волновой расстройке критическая интенсивность волны накачки обратно пропорциональна коэффициенту стационарного ВКР-усиления, а при фиксированном коэффициенте стационарного ВКР-усиления она зависит от волновой расстройки линейно (рис. 5). Общая зависимость критической интенсивности волны накачки от коэффициента стационарного ВКР-усиления и волновой расстройки может быть аппроксимирована формулой  $I_{0,кр} = 0.4022\Delta^{0.9977}g^{-1}$ , где ин-

тенсивность измеряется в ГВт/см<sup>2</sup>, волновая расстройка – в рад/см, а коэффициент стационарного ВКР-усиления – в см/ГВт.

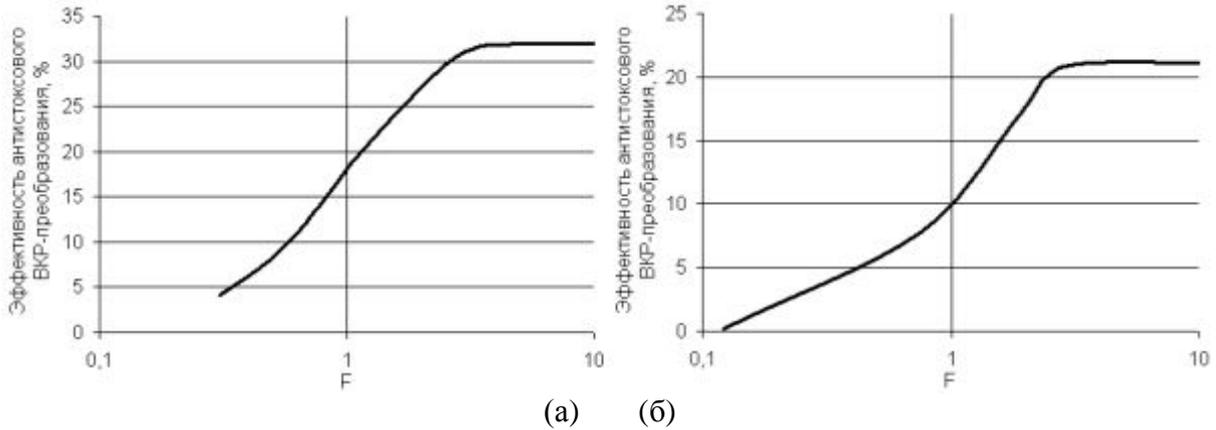


Рис. 6. Влияние дифракционных эффектов на эффективность антистоксового ВКР-преобразования в водороде (а) и нитрате бария (б)

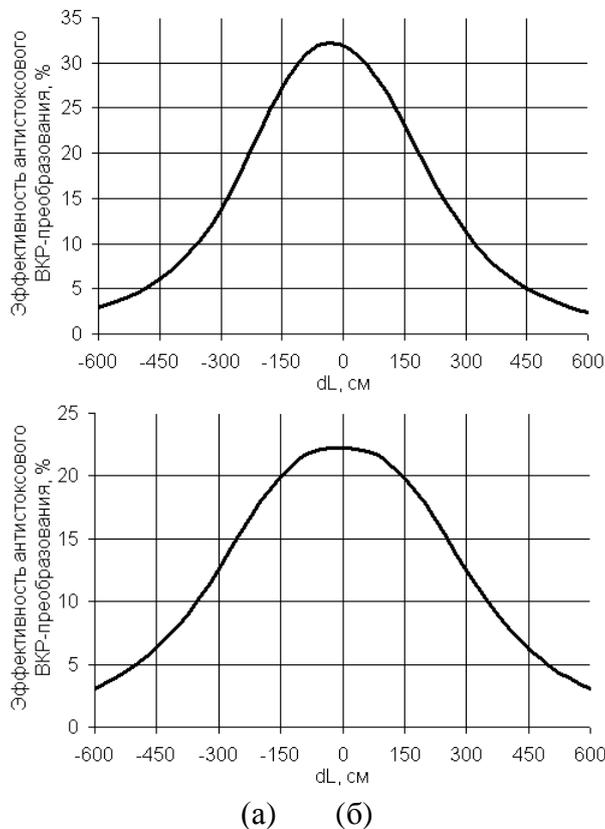


Рис. 7. Влияние положения перетяжки гауссова пучка на эффективность антистоксового ВКР-преобразования в водороде (а) и нитрате бария (б) (положительным значениям  $dL$  соответствует расположение перетяжки пучка перед ВКР-активной средой, а отрицательным – внутри нее и за ней)

Численное моделирование многоволнового стационарного ВКР в сжатом водороде и нитрате бария с учетом дифракционных эффектов показало, что для каждого значения диаметра пучков существует своя оптимальная ВКР-активная слоистая среда, в которой достигается высокая эффективность антистоксового ВКР-преобразования. Для определения оптимальных условий генерации антистоксового излучения была изучена максимальная эффективность антистоксового ВКР-преобразования в зависимости от числа Френеля активного волновода ( $F = \frac{ka^2}{L}$ , где  $k$  – волновое число,  $a$  – диаметр пучка, а  $L$  – длина ВКР-активной среды; см. рис. 6) и от положения перетяжки гауссова

пучка (см. рис. 7) в фиксированной слоистой среде. Из расчетов и приведенных графиков следует, что с увеличением числа Френеля эффективность антистоксового ВКР-преобразования в слоистой структуре возрастает и достигает своего максимума при числе Френеля, равном трем. При дальнейшем росте числа Френеля эффективность антистоксового ВКР-преобразования практически не изменяется. Максимальная эффективность антистоксового ВКР-преобразования составляет 32% для водорода и 21% для нитрата бария. Это несколько ниже, чем при стационарном взаимодействии без учета дифракции, однако разницу можно объяснить тем, что вместо плоских волн при дифракции рассматриваются гауссовы пучки. Было установлено, что максимальная эффективность антистоксового ВКР-преобразования при диаметре входных пучков 1 мм достигается в случае, когда перетяжка пучка находится внутри ВКР-активной среды (на расстоянии  $\sim 80\%$  от  $L_m$  для водорода и  $\sim 100\%$  от  $L_m$  для нитрата бария, где  $L_m$  – длина ВКР-активной среды). При этом при изменении положения перетяжки пучка от  $+30$  до  $-30$  см (до среды и в среде или за ней) эффективность антистоксового ВКР-преобразования практически не изменяется. Положительным значениям  $dL$  расположения перетяжки пучка радиуса  $r_0$  соответствует случай, при котором лазерный источник находится на расстоянии  $dL$  от ВКР-активной среды; отрицательным – случай, когда лазерный источник находится непосредственно перед ВКР-активной средой, а пучок радиуса  $R$  проходит через линзу с фокусным расстоянием  $F$ , где

$$dL = \frac{F\pi^2 R^4}{\pi^2 R^4 + F^2 \lambda^2}; r_0 = \frac{F\lambda R}{\sqrt{\pi^2 R^4 + F^2 \lambda^2}}.$$

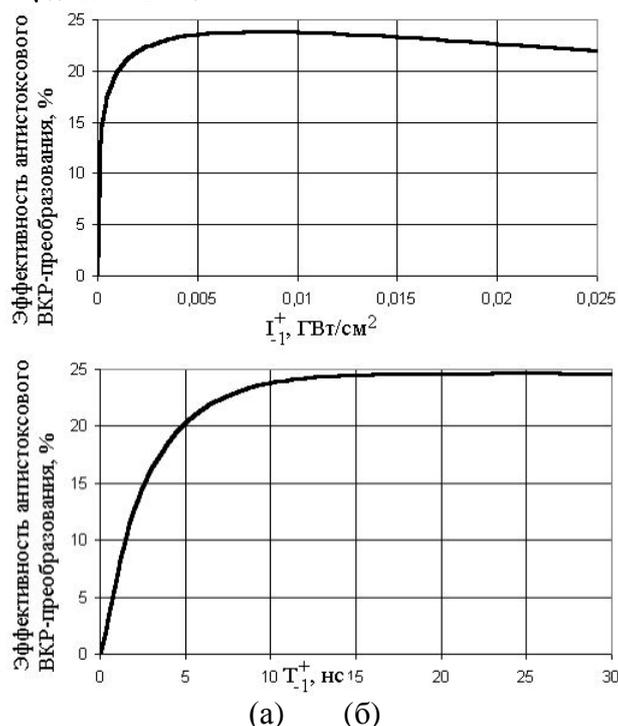


Рис. 8. Влияние интенсивности стоксового затравочного импульса (а) и его длительности (б) на эффективность антистоксового ВКР-преобразования

Для слоистой структуры в сжатом водороде, состоящей из 40 слоев, было изучено влияние интенсивности и длительности стоксового затравочного импульса (рис. 8) и импульса накачки (рис. 9) на эффективность антистоксового ВКР-преобразования. Из расчетов и графиков видно, что в достаточно большом диапазоне параметров стоксового импульса (при  $0.001 \text{ ГВт/см}^2 \leq I_{-1}^+ \leq 0.025 \text{ ГВт/см}^2$  и  $5 \text{ нс} \leq T_{-1}^+ \leq 20 \text{ нс}$ ) эффективность антистоксового ВКР-преобразования остается выше 20%. Это снижает требования, предъявляемые к системе предварительной генерации стоксового импульса, необходимого для эффективной антистоксовой ВКР-генерации. Кроме того, эффективная

антистоксовая ВКР-генерация в достаточно широком диапазоне параметров импульса накачки (при  $T_0^+ \geq 1$  нс и  $0.15 \text{ ГВт/см}^2 \leq I_0^+ \leq 0.6 \text{ ГВт/см}^2$ ) позволяет реализовать перестраиваемый источник антистоксового ВКР-излучения.

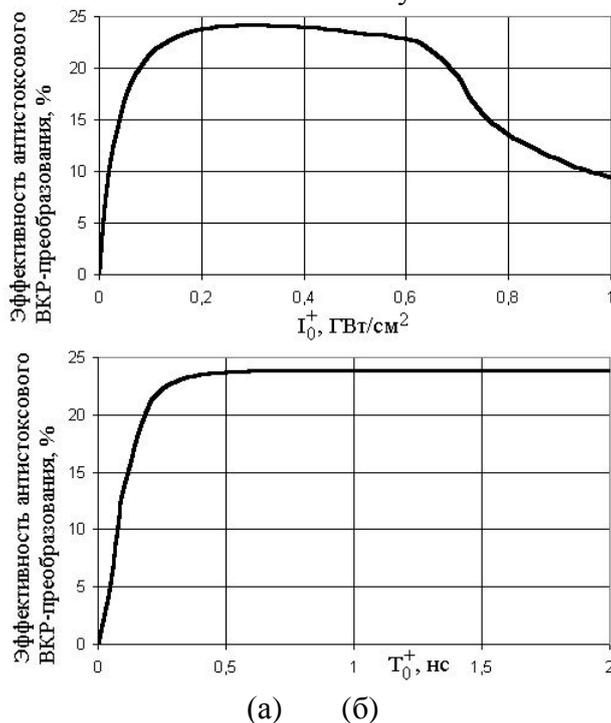


Рис. 9. Влияние интенсивности импульса накачки (а) и его длительности (б) на эффективность антистоксового ВКР-преобразования

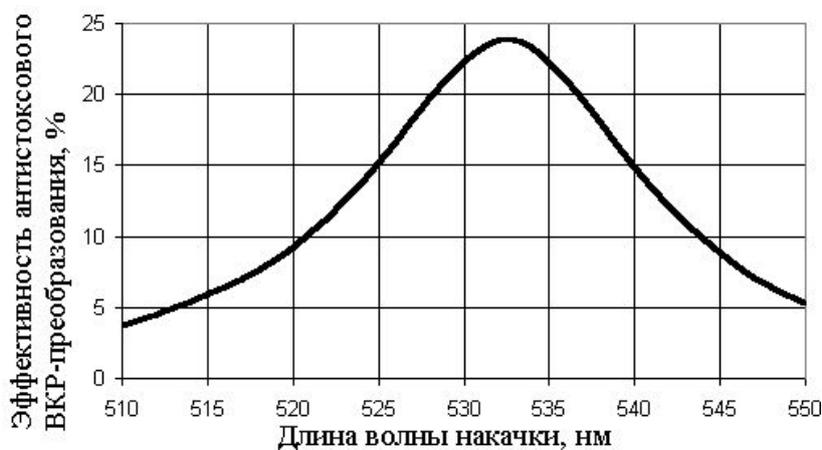


Рис. 10. Влияние длины волны накачки на эффективность антистоксового ВКР-преобразования  
 Была изучена зависимость эффективности антистоксового ВКР-преобразования в фиксированной слоистой структуре от длины волны накачки (рис. 10). Расчеты показали, что слоистая структура, обеспечивающая фазовый квазисинхронизм, обладает высокой чувствительностью к изменению длины волны накачки. Это приводит к существенному уменьшению эффективности антистоксового ВКР-преобразования даже при изменении длины волны накачки всего на 0.75%.

## Заключение

Численное моделирование генерации антистоксового излучения в условиях фазового квазисинхронизма показало, что для каждой среды существует оптимальное значение начальной интенсивности стоксового излучения, при котором достигается максимальная эффективность антистоксового ВКР-преобразования. Установлено, что с

ростом интенсивности волны накачки максимальная эффективность антистоксового ВКР-преобразования возрастает, однако для каждой среды существует критическое значение интенсивности волны накачки, при котором эффективность антистоксового преобразования существенно снижается. Установлено, что при изменении длин слоев слоистой структуры, обеспечивающей фазовый квазисинхронизм, на 5%, эффективность антистоксового ВКР-преобразования изменяется не более, чем на 0,02%. Максимально допустимая случайная центрированная погрешность в выборе длин слоев, практически не влияющая на эффективность антистоксового ВКР-преобразования, составляет 15%. Показано, что при числе Френеля активного волновода больше трех влиянием дифракции на реализацию условий фазового квазисинхронизма можно пренебречь. Моделирование показало, что в некотором пределе изменение входных параметров излучения (интенсивности, длительности) практически не влияет на эффективность антистоксового ВКР-преобразования.

## Литература

1. Minck R.W., Terhune R.W., Rado W.G. Laser-stimulated Raman effect and resonant four-photon interactions in gaseous H<sub>2</sub>, D<sub>2</sub> and CH<sub>4</sub> // *Appl. Phys. Letts.*, Vol. 3 (1963) №3, p. 181-184.
2. Бломберген Н. Вынужденное комбинационное рассеяние // *УФН*, Т. 96 (1969) №2, С. 307-352.
3. Бутылкин В.С., Каплан А.Е., Хронополо И.Г., Якубович Е.М. Резонансные взаимодействия света с веществом. - М.: Наука, 1977, 351 с.
4. Шен И.Р. Принципы нелинейной оптики: Пер. с англ. - М.: Наука, 1989, 557 с.
5. Беспалов В.Г., Крылов В.Н., Михайлов В.Н., Парфенов В.А., Стаселько Д.И. Генерация перестраиваемого излучения с высокой спектральной яркостью на основе колебательного и вращательного ВКР в газах // *Опт. и спектр.*, Т. 70 (1991) в.2, С. 332-336.
6. Беспалов В.Г., Духовный А.М., Стаселько Д.И. Исследование когерентности излучения при ВКР в сжатом водороде // *Письма в ЖТФ*, Т. 5 (1979) №20, С. 1236.
7. Андреев Р.Б., Горбунов В.А., Гулидов С.С., Паперный С.Б., Серебряков В.А. О роли параметрических эффектов при генерации высших компонент ВКР в газах // *Квант. электрон.*, Т. 9 (1982) №1. С. 56-60.
8. Ottusch J.J., Mangir M.S., Rockwell D.A. Efficient anti-Stokes Raman conversion by four-wave mixing in gases // *J. Opt. Soc. Am. B*, 8 (1991) p. 68-77.
9. Armstrong J.A., Bloembergen N., Ducuing J., Pershan P.S. Interaction between light waves in a nonlinear dielectric // *Phys. Rev.*, 127 (1962) p. 1918-1939.
10. Urschel R., Bäder U., Borsutzky A., Wallenstein R. Spectral properties and conversion efficiency of 355-nm-pumped pulsed optical parametric oscillators of  $\beta$ -barium borate with noncollinear phase matching // *J. Opt. Soc. Am.*, 16 (1999) p. 565-579.
11. Chou M.H., Parameswaran K.R., Fejer M.M., Brener I. Multiple-channel wavelength conversion by use of engineered quasi-phase-matching structures in LiNbO<sub>3</sub> waveguides // *Optics Letters*, 24 (1999) №16, p. 1157-1159.

# УШИРЕНИЕ "ФИОЛЕТОВОГО" КРЫЛА ФЕМТОСЕКУНДНОГО СПЕКТРАЛЬНОГО СУПЕРКОНТИНУУМА ИЗ-ЗА ДИСПЕРСИИ НЕЛИНЕЙНОГО ПОКАЗАТЕЛЯ ПРЕЛОМЛЕНИЯ СРЕДЫ

С.А. Штумпф

Научные руководители – д. ф.-м. н., проф. С.А. Козлов,  
к. ф.-м. н., доц. А.А. Королев,

Для описания динамики поля сверхкоротких импульсов с континуумным спектром в прозрачных нелинейных средах выведено волновое уравнение, которое учитывает частотную зависимость не только линейной, но и нелинейной части поляризационного отклика оптической среды. С его помощью проведено численное моделирование процесса самоуширения спектра предельно короткого лазерного импульса.

## Введение

Развитие фемтосекундной лазерной техники [1–3] значительно изменило облик нелинейной оптики. Прежде всего это связано с тем, что распространение фемтосекундного излучения в оптических средах без разрушения вещества (по крайней мере, за длительность импульса) оказалось возможным наблюдать при гораздо больших интенсивностях, чем для более длинных импульсов [4, 5]. Поэтому нелинейные явления фемтосекундной оптики часто приобретают качественно новые черты. Например, генерация спектрального суперконтинуума [6] – явление, редкое даже для пикосекундных импульсов, – для фемтосекундных становится фундаментальным, сопровождающим практически все остальные нелинейные эффекты [7–9]. Сфазировав фемтосекундный спектральный суперконтинуум, можно получать импульсы, состоящие лишь из нескольких колебаний светового поля [2, 3], которые обычно называют предельно короткими импульсами (под предельно малым при этом понимают число колебаний в импульсе, но не сам его временной размер) [10].

Плодотворность метода медленно меняющихся огибающих, который эффективен в нелинейной оптике квазимонохроматических импульсов, при описании динамики сверхкоротких импульсов со сверхуширенными временными спектрами становится дискуссионной [10, 11]. Поэтому в последние годы много внимания было уделено как модификации этого метода [2, 12, 13], так и разработке новых теоретических подходов. Среди последних наибольшее распространение получил полевой подход (см., например, работы обзорного характера [11, 14]), в рамках которых анализируется динамика непосредственно поля импульса, а не его огибающей). Уравнения эволюции поля обычно получают либо в приближении медленно меняющегося профиля поля излучения, либо в эквивалентном приближении слабой дисперсии и нелинейности. Не обладая существенными преимуществами перед уравнениями эволюции огибающих при описании дисперсии линейного отклика вещества, полевые уравнения проще и адекватнее описывают нелинейные процессы генерации и кросс-модуляции компонент сверхуширяющегося спектра излучения.

При теоретическом исследовании самовоздействия фемтосекундных импульсов в прозрачных изотропных средах обычно адекватно учитывается частотная зависимость линейного поляризационного отклика, а его нелинейная, кубичная по полю часть нерезонансной электронной природы предполагается не дисперсионной (безынерционной) (см., например, обзор [11] и работы [15, 16] по теоретическому описанию самого широкого фемтосекундного спектрального суперконтинуума из наблюдавшихся экспериментально, который был получен в фотоннокристаллических волокнах [13]). Однако коэффициент нелинейного показателя преломления даже нерезонансной электронной природы в диапазоне спектра реализуемых в эксперименте континуумов может изменяться заметно, что диктует необходимость анализа влияния на генерацию спектрального суперконтинуума "немгновенности" не только линейного, но и нелинейного поля-

ризационного отклика среды. В дальнейшем будем обозначать зависимость показателя преломления среды и его составляющих от частоты излучения понятием *дисперсия* показателя преломления (линейной или нелинейной его части), под *нелинейностью* среды будем понимать составляющие дисперсии, нелинейные по полю импульса. В настоящей работе проанализирован характер изменения нерезонансной нелинейности прозрачных оптических сред в зависимости от частоты излучения. Выведено волновое уравнение, которое учитывает дисперсию как линейной, так и кубичной по полю нелинейности диэлектрической среды электронной природы. Показано, что дисперсия нерезонансной нелинейности среды в процессе самоуширения спектра фемтосекундного импульса может приводить к увеличению его "фиолетовой" области, в том числе к повышению эффективности генерации третьей гармоники. При этом ослабляется "синяя" и сохраняется "красная" области спектра.

### Вывод уравнения динамики поля излучения в среде с дисперсией нелинейного показателя преломления

Для плоской однородной линейно поляризованной электромагнитной волны в изотропной диэлектрической среде волновое уравнение может быть записано в виде [10]

$$\frac{\partial^2 E}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2} = \frac{4\pi}{c^2} \frac{\partial^2 P_l}{\partial t^2} + \frac{4\pi}{c^2} \frac{\partial^2 P_{nl}}{\partial t^2}, \quad (1)$$

где  $E$  – электрическое поле волны,  $P_l$  – линейная, а  $P_{nl}$  – нелинейная часть поляризации отклика вещества,  $z$  – направление распространения,  $t$  – время,  $c$  – скорость света в вакууме.

Дисперсия линейного показателя преломления диэлектрической среды в диапазоне ее прозрачности хорошо аппроксимируется рядом [17]

$$n^2(\omega) = n_0^2 + 2cn_0a\omega^2 + 2cn_0a_1\omega^4 + \dots - 2cn_0b\omega^{-2} - 2cn_0b_1\omega^{-4} \dots, \quad (2)$$

где  $n_0$ ,  $a$ ,  $a_1$ ,  $b$ ,  $b_1$  – эмпирические параметры дисперсии среды.

Для оптической среды, характеризуемой дисперсионным соотношением (2), полное волновое уравнение (1) в приближении однонаправленного распространения в сопровождающей световой импульс системе координат излучения можно записать в виде [11]

$$\frac{\partial E}{\partial z} - a \frac{\partial^3 E}{\partial \tau^3} + a_1 \frac{\partial^5 E}{\partial \tau^5} - \dots + b \int_{-\infty}^{\tau} E d\tau' - b_1 \int_{-\infty}^{\tau} d\tau' \int_{-\infty}^{\tau'} d\tau'' \int_{-\infty}^{\tau''} E d\tau''' + \frac{2\pi}{cn_0} \frac{\partial P_{nl}}{\partial \tau} = 0, \quad (3)$$

где  $\tau = t - \frac{n_0}{c} z$ .

При изучении распространения в диэлектрической среде импульса с очень широким спектром важнейшим требованием к описанию нелинейной части поляризационного отклика  $P_{nl}$  является, как и в приведенном описании линейной части отклика  $P_l$ , правильный учет в ней дисперсии. В работах [18–20] было показано, что хорошим приближением для диэлектриков при описании дисперсии нелинейности электронной природы является трехуровневая энергетическая модель молекулы вещества (трехзонная модель кристаллического твердого тела). Первый уровень соответствует основному состоянию валентных электронов молекулы (в твердом теле – валентной зоне), второй и третий – эффективные уровни, каждый из которых отвечает группам возбужденных состояний электронов с одинаковой четностью (в твердом теле – подзонам зоны проводимости). При этом второму уровню соответствуют состояния с четностью, противоположной четности основного состояния (в твердом теле – подзоны, сильно связанные в электродипольном приближении с основным состоянием), а третьему уровню – состоя-

ния с той же четностью, что и у основного (в твердом теле – подзоны, в электродипольном приближении слабо связанные с основным состоянием, но сильно связанные с состояниями, соответствующими второму уровню). В рамках такой трехуровневой модели вещества уравнения динамики поляризованности среды  $P = P_1 + P_2$ , выводимые из уравнений для матрицы плотности с точностью до кубической нелинейности, могут быть записаны в виде [21]

$$\left\{ \begin{array}{l} \ddot{P}_1 + 2T_{21}^{-1}\dot{P}_1 + (\omega_{21}^2 + T_{21}^{-2})P_1 = 2\omega_{21}\hbar^{-1}|\mathbf{p}_{12}|^2 N_{12}E + \omega_{21}\hbar^{-1}RE - f_{21} \\ \ddot{P}_2 + 2T_{32}^{-1}\dot{P}_2 + (\omega_{32}^2 + T_{32}^{-2})P_2 = 2\omega_{32}\hbar^{-1}|\mathbf{p}_{23}|^2 N_{23}E - \omega_{32}\hbar^{-1}RE + f_{32} \\ \ddot{R} + 2T_{31}^{-1}\dot{R} + (\omega_{31}^2 + T_{31}^{-2})R = \omega_{31}\hbar^{-1}(|\mathbf{p}_{23}|^2 P_1 - |\mathbf{p}_{12}|^2 P_2)E - \\ - (\hbar\omega_{21})^{-1}|\mathbf{p}_{23}|^2 [(\ddot{P}_1 + T_{21}^{-1}\dot{P}_1)E + (\dot{P}_1 + T_{21}^{-1}P_1)(\dot{E} + T_{31}^{-1}E)] + \\ + (\hbar\omega_{32})^{-1}|\mathbf{p}_{12}|^2 [(\ddot{P}_2 + T_{32}^{-1}\dot{P}_2)E + (\dot{P}_2 + T_{32}^{-1}P_2)(\dot{E} + T_{31}^{-1}E)] \\ N_{12} + \tau_{12}^{-1}(N_{12} - N_{12}^0) - \tau_{32}^{-1}(N_{23} - N_{23}^0) = -2(\hbar\omega_{21})^{-1}(\dot{P}_1 + T_{21}^{-1}P_1)E + \\ + (\hbar\omega_{32})^{-1}(\dot{P}_2 + T_{32}^{-1}P_2)E \\ N_{23} - \tau_{21}^{-1}(N_{12} - N_{12}^0) + \tau_{23}^{-1}(N_{23} - N_{23}^0) = (\hbar\omega_{21})^{-1}(\dot{P}_1 + T_{21}^{-1}P_1)E + \\ - 2(\hbar\omega_{32})^{-1}(\dot{P}_2 + T_{32}^{-1}P_2)E \end{array} \right. , \quad (4)$$

где

$$\begin{aligned} f_{mn} = & (\hbar\omega_{31})^{-1} \{ |\mathbf{p}_{23}|^2 (\hbar\omega_{21})^{-1} [(\ddot{P}_1 + T_{21}^{-1}\dot{P}_1)E^2 + (\dot{P}_1 + T_{21}^{-1}P_1)(2\dot{E} + T_{mn}^{-1}E)E] - \\ & - |\mathbf{p}_{12}|^2 (\hbar\omega_{32})^{-1} [(\ddot{P}_2 + T_{32}^{-1}\dot{P}_2)E^2 + (\dot{P}_2 + T_{32}^{-1}P_2)(2\dot{E} + T_{mn}^{-1}E)E] + \\ & + (\ddot{R} + T_{31}^{-1}\dot{R})E + (\dot{R} + T_{31}^{-1}R)(\dot{E} + T_{mn}^{-1}E) \}, \end{aligned} \quad (5)$$

$N = N_1 + N_2 + N_3 = N_1^0 + N_2^0 + N_3^0$  – концентрация структурных элементов среды,  $N_i^0$  – населенности уровней в отсутствие поля (термодинамически равновесные),  $N_{ij} = N_i - N_j$ ,  $N_{ij}^0 = N_i^0 - N_j^0$ ,  $R$  осуществляет нелинейную параметрическую связь между электрическим полем импульса и поляризованностью среды.

При рассмотрении нерезонансного взаимодействия излучения с веществом естественны предположения  $N_1 \gg N_2, N_3$  и  $\omega_{21}, \omega_{31} \gg T_{21}^{-1}, T_{31}^{-1}, \tau_{21}^{-1}, \tau_{31}^{-1}$ . При этих условиях систему уравнений (4) можно свести к виду

$$\left\{ \begin{array}{l} \ddot{P}_1 + \omega_{21}^2 P_1 = \alpha_1 N_{12} E + \beta_1 R E - \delta P_1 E^2 - \xi \dot{P}_1 \dot{E} E - \zeta \dot{R} \dot{E} \\ \ddot{P}_2 + \omega_{32}^2 P_2 = \alpha_2 N_{23} E + \beta_2 R E + \delta P_1 E^2 + \xi \dot{P}_1 \dot{E} E + \zeta \dot{R} \dot{E} \\ \ddot{R} + \omega_{31}^2 R = \gamma P_1 E - \eta \dot{P}_1 \dot{E} - \vartheta N_{12} E^2 \\ \dot{N}_1 = -\mu_1 \dot{P}_1 E \\ \dot{N}_2 = \mu_1 \dot{P}_1 E \end{array} \right. \quad (6)$$

где  $\alpha_1 = 2\omega_{21}|\mathbf{p}_{12}|^2 \hbar^{-1}$ ,  $\alpha_2 = 2\omega_{32}|\mathbf{p}_{23}|^2 \hbar^{-1}$ ,  $\delta = |\mathbf{p}_{23}|^2 \hbar^{-2}$ ,  $\beta_1 = (\omega_{21} + \omega_{31})\hbar^{-1}$ ,  $\beta_2 = (\omega_{31} + \omega_{32})\hbar^{-1}$ ,  $\gamma = (\omega_{21} + \omega_{31})|\mathbf{p}_{23}|^2 \hbar^{-1}$ ,  $\xi = \omega_{21}^{-1}\omega_{32}^{-1}|\mathbf{p}_{23}|^2 \hbar^{-2}$ ,  $\zeta = (\hbar\omega_{31})^{-1}$ ,  $\mu_1 = (\hbar\omega_{21})^{-1}$ ,  $\vartheta = 2|\mathbf{p}_{12}|^2 |\mathbf{p}_{23}|^2 \hbar^{-2}$ ,  $\eta = \omega_{21}^{-1}|\mathbf{p}_{23}|^2 \hbar^{-1}$ .

Полагая в рамках нерезонансного приближения  $\ddot{P}_{1,2} \ll \omega_{21}^2 P_{1,2}$  и применяя итерационную процедуру по такому параметру малости (изложенную, например, в работе

[22]), из материальных уравнений (6) для нелинейной части поляризации несложно получить выражение

$$P_{nl} = \chi^{(3)} E^3 + \chi_1^{(3)} E \left( \frac{\partial E}{\partial t} \right)^2 + \chi_2^{(3)} E^2 \frac{\partial^2 E}{\partial t^2}, \quad (7)$$

где  $\chi^{(3)} = 4|\mathbf{p}_{12}|^2 \hbar^{-3} (|\mathbf{p}_{23}|^2 \omega_{21}^{-2} \omega_{31}^{-1} - |\mathbf{p}_{12}|^2 \omega_{21}^{-3}) N_1^0$ ;

$$\chi_1^{(3)} = \left( 20 \frac{|\mathbf{p}_{12}|^4}{(\hbar \omega_{21})^3} \cdot \frac{1}{\omega_{21}^2} + 2 \frac{|\mathbf{p}_{12}|^2 |\mathbf{p}_{23}|^2}{\hbar^3} \cdot \frac{1}{\omega_{21}^2} \cdot \left( \frac{8}{\omega_{21}^2 \omega_{32}^2} - \frac{2}{\omega_{31}^2 \omega_{32}^2} + \frac{1}{\omega_{21} \omega_{31} \omega_{32}} \right) (2\omega_{21} - \omega_{31}) + \frac{2}{\omega_{31}^2 \omega_{32}^2} \right) N_1^0$$

$$\chi_2^{(3)} = \left( 20 \frac{|\mathbf{p}_{12}|^4}{(\hbar \omega_{21})^3} \cdot \frac{1}{\omega_{21}^2} + 2 \frac{|\mathbf{p}_{12}|^2 |\mathbf{p}_{23}|^2}{\hbar^3} \cdot \frac{1}{\omega_{21}^2} \cdot \left( -\frac{6}{\omega_{21}^2 \omega_{31}} - \frac{1}{\omega_{21} \omega_{31}^2} + \frac{\omega_{21}^2 - 3\omega_{31}^2}{\omega_{21}^2 \omega_{31} \omega_{32}^2} \right) \right) N_1^0.$$

В выражении (7) второе и третье слагаемые характеризуют инерционную часть нерезонансной нелинейности поляризационного отклика среды.

Волновое уравнение (3) с учетом (7) принимает вид

$$\frac{\partial E}{\partial z} - a \frac{\partial^3 E}{\partial \tau^3} - a_1 \frac{\partial^5 E}{\partial \tau^5} + \frac{\partial}{\partial \tau} \left( g^{(3)} E^3 + g_1^{(3)} E \left( \frac{\partial E}{\partial \tau} \right)^2 + g_2^{(3)} E^2 \frac{\partial^2 E}{\partial \tau^2} \right) = 0, \quad (8)$$

где  $g^{(3)} = \frac{2\pi}{cN_0} \chi^{(3)}$ ,  $g_1^{(3)} = \frac{2\pi}{cN_0} \chi_1^{(3)}$ ,  $g_2^{(3)} = \frac{2\pi}{cN_0} \chi_2^{(3)}$ .

В выведенном уравнении (8) второе и третье слагаемые описывают дисперсию линейного показателя преломления среды (при необходимости к ним можно добавить сколь угодно много слагаемых, приведенных в (3)), четвертое – недисперсионную часть нелинейности; пятое и шестое характеризуют дисперсию кубичной нелинейности.

Определяя решения уравнения (8) вида плоской монохроматической волны  $E = \frac{1}{2} E_0 e^{i(kz - \omega t)} + \frac{1}{2} E_0^* e^{-i(kz - \omega t)}$ , для показателя преломления среды получаем выражение

$$n = n_0(\omega) + \frac{1}{2} n_2(\omega) E_0^2, \quad (9)$$

где

$$n_0(\omega) = N_0 + ca\omega^2 + ca_1\omega^4, \quad (10)$$

$$n_2(\omega) = \frac{1}{2} (3g^{(3)} + (g_1^{(3)} - 3g_2^{(3)})\omega^2) cE_0^2. \quad (11)$$

Выражения (9)–(11) демонстрируют дисперсию нелинейного показателя преломления (НПП) среды, описываемую уравнением динамики поля излучения (8) в этой среде.

Вводя нормированные поле, координаты и время,

$$\tilde{E} = E / E_0; \quad \tilde{z} = a \langle \omega \rangle^3 z; \quad \tilde{\tau} = \langle \omega \rangle \tau, \quad (12)$$

где  $E_0$  – максимальная амплитуда электрического поля импульса на входе в среду,  $\langle \omega \rangle$  – центральная частота входного спектра излучения,

$$\langle \omega \rangle = \frac{\int_0^\infty \omega G(\omega) d\omega}{\int_0^\infty G(\omega) d\omega}, \quad (13)$$

$G(\omega)$  – начальная спектральная плотность излучения, волновое уравнение (8) можно привести к виду

$$\frac{\partial \tilde{E}}{\partial \tilde{z}} - \frac{\partial^3 \tilde{E}}{\partial \tilde{\tau}^3} + \tilde{a}_1 \frac{\partial^5 \tilde{E}}{\partial \tilde{\tau}^5} + \tilde{g}^{(3)} \frac{\partial \tilde{E}^3}{\partial \tilde{\tau}} + \tilde{g}_1^{(3)} \frac{\partial}{\partial \tilde{\tau}} \left( \tilde{E} \left( \frac{\partial \tilde{E}}{\partial \tilde{\tau}} \right)^2 \right) + \tilde{g}_2^{(3)} \frac{\partial}{\partial \tilde{\tau}} \left( \tilde{E}^2 \frac{\partial^2 \tilde{E}}{\partial \tilde{\tau}^2} \right) = 0 \quad (14)$$

где  $\tilde{a}_1 = a_1 a^{-1} \langle \omega \rangle^2$ ;  $\tilde{g}^{(3)} = g^{(3)} a^{-1} E_0^2 \langle \omega \rangle^{-2}$ ;  $\tilde{g}_1^{(3)} = g_1^{(3)} a^{-1} E_0^2$ ;  $\tilde{g}_2^{(3)} = g_2^{(3)} a^{-1} E_0^2$ .

Соотношения коэффициентов нормированного уравнения позволяют, с одной стороны, без расчетов оценивать степень влияния различных физических процессов на самовоздействие лазерного импульса и, с другой стороны, за один численный расчет моделировать различные физические ситуации.

### Оценка параметров уравнения самовоздействия излучения с широким спектром в кварцевом стекле

Определим значения величин  $g^{(3)}$  и  $G^{(3)} = g_1^{(3)} - 3g_2^{(3)}$ , характеризующих коэффициенты при нелинейных слагаемых уравнения (14), для широко применяемого оптического материала – кварцевого стекла – по известным экспериментальным данным для  $n_2$  на различных длинах волн [23–29]. Как видно из рис. 1, при значениях коэффициентов

$$g^{(3)} = 2.0 \cdot 10^{-24} \text{ ед. СГСЭ}, \quad G^{(3)} = 9.4 \cdot 10^{-56} \text{ ед. СГСЭ}. \quad (15)$$

достигается хорошее согласие теоретической зависимости (11) с экспериментальными данными. При указанных значениях дисперсия  $n_2$  кварцевого стекла описывается выражением

$$n_2(\omega) = n_2^0 + A\omega^2, \quad (16)$$

где  $n_2^0 = 0.9 \times 10^{-13} \text{ ед. СГСЭ}$ ,  $A = 1.4 \times 10^{-45} \text{ ед. СГСЭ}$ .

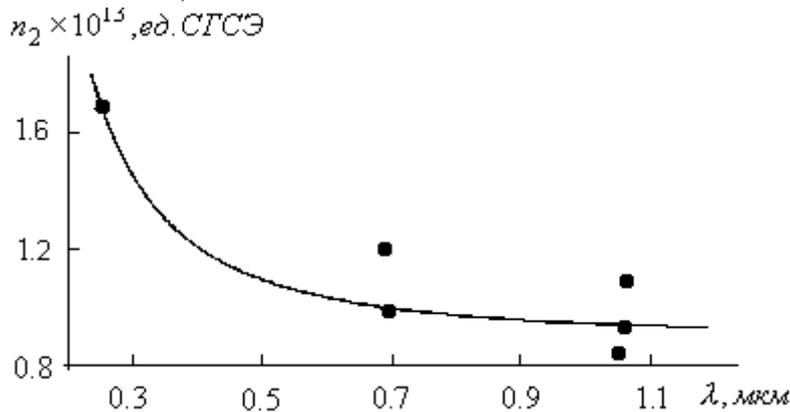


Рис. 1. Дисперсия коэффициента нелинейного показателя преломления кварцевого стекла  $n_2$ . Точками указаны экспериментальные значения  $n_2$  кварцевого стекла [23-29], сплошная линия – график зависимости (11) при  $g^{(3)} = 2.0 \cdot 10^{-24} \text{ ед. СГСЭ}$ ,  $G^{(3)} = 9.4 \cdot 10^{-56} \text{ ед. СГСЭ}$ .

Для определения соотношения между  $g_1^{(3)}$  и  $g_2^{(3)}$  используем линейризованную модель материальных уравнений (6) вида:

$$\ddot{P} + \omega_{21}^2 P = 2|p_{12}|^2 \omega_{21} \hbar^{-1} N E. \quad (17)$$

Из (17) следует дисперсионное соотношение для линейного показателя преломления:

$$n_0^2(\omega) = 1 + \frac{8\pi}{\hbar} N \frac{\omega_{21} |p_{12}|^2}{\omega_{21}^2 - \omega^2}. \quad (18)$$

Воспользовавшись экспериментальными данными по дисперсии линейного показателя преломления кварцевого стекла [30] и учитывая, что концентрация структурных элементов среды составляет  $N = 2.7 \times 10^{23} \text{ см}^{-3}$ , несложно получить оценки:  $\omega_{21} = 1.9 \times 10^{16} \text{ с}^{-1}$ ,  $|p_{12}| = 1.8 \times 10^{-18} \text{ ед. СГСЭ}$ . С учетом (15) далее можно получить:

$\omega_{31} = 3.1 \times 10^{16} \text{ c}^{-1}$ ,  $|\mathbf{p}_{23}| = 7.6 \times 10^{-18} \text{ ед. СГСЭ}$ . В итоге получаем значения нелинейных коэффициентов  $g_1^{(3)} = 3.6 \cdot 10^{-56} \text{ ед. СГСЭ}$ ,  $g_2^{(3)} = -1.9 \cdot 10^{-56} \text{ ед. СГСЭ}$

Вычислим коэффициенты нормированного волнового уравнения (14), задавшись определенными характеристиками входного импульса. Рассматривая излучение с центральной длиной волны, соответствующей второй гармонике титан-сапфирового лазера ( $\lambda = 390 \text{ нм}$ ,  $\langle \omega \rangle = 2.45 \cdot 10^{14} \text{ c}^{-1}$ ) при максимальной напряженности поля излучения  $E_0 = 2 \cdot 10^6 \text{ В/м}$  (что соответствует пиковой интенсивности  $I_0 = 1 \cdot 10^{13} \text{ Вт/см}^2$ ) получаем:  $\tilde{g}^{(3)} = 3.8$ ,  $\tilde{g}_1^{(3)} = 1.6$ ,  $\tilde{g}_2^{(3)} = -0.8$ . Отметим, что при данных условиях  $\tilde{g}_1^{(3)}/\tilde{g}^{(3)} = 0.44$ ,  $\tilde{g}_2^{(3)}/\tilde{g}^{(3)} = -0.22$ , т.е. эффекты дисперсии нелинейности при наличии в спектре импульса компонент, частота которых находится в видимом диапазоне и выше, могут иметь тот же порядок, что и эффекты, обусловленные безынерционной частью нелинейности третьего порядка.

### Численное решение волнового уравнения, учитывающего дисперсию ННП в кварцевом стекле

Изучение изменения формы и спектра импульса в процессе его самовоздействия в диэлектрической среде с дисперсией и линейного, и нелинейного показателя преломления в настоящей работе проводилось с помощью численного моделирования решений волнового уравнения (14) с помощью алгоритмов, изложенных в [31].

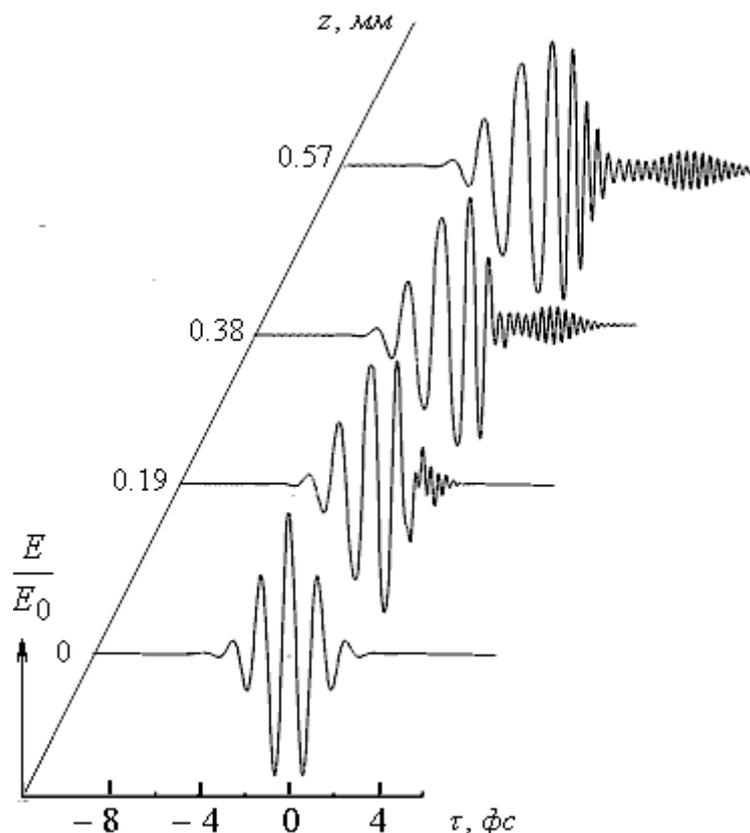


Рис. 2. Эволюция временного профиля электрического поля светового импульса входной пиковой интенсивности  $I_0 = 1 \cdot 10^{13} \text{ Вт/см}^2$  с центральной длиной волны  $\lambda = 390 \text{ нм}$  в кварцевом стекле.

На рис. 2 и 3 приведена эволюция в кварцевом стекле формы и спектра гауссова на входе в среду импульса с начальной длительностью 8 фс, пиковой напряженностью поля  $E_0 = 2 \cdot 10^6 \text{ В/м}$  и центральной длиной волны  $\lambda = 390 \text{ нм}$ . Из рис. 2 видно, что в нелинейной среде происходит уменьшение фазовой скорости импульса, расплывание его переднего фронта и укручение заднего, а также выделение сопровождающего импульса на утроенных частотах. Из рис. 3 видно, что уширение спектра импульса происходит как в низкочастотную, так и в высокочастотную область, наблюдается смещение максимума спектральной плотности излучения в область более низких частот, а также генерация спектральных компонент на утроенной частоте.

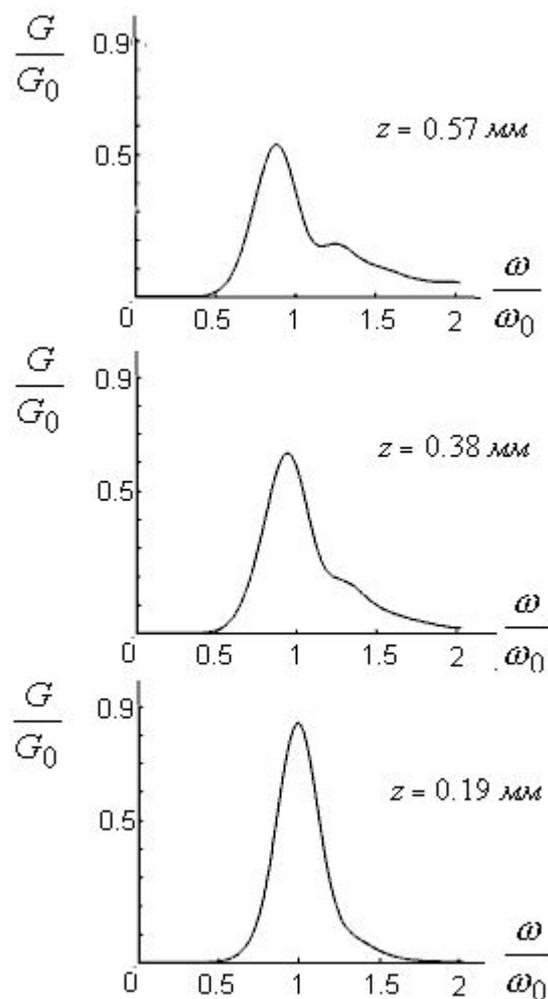


Рис. 3. Эволюция спектра импульса входной интенсивности  $I_0 = 1 \cdot 10^{13} \text{ Вт/см}^2$  с центральной длиной волны  $\lambda = 390 \text{ нм}$  в кварцевом стекле

Рис. 4 показывает, что при учете дисперсии нелинейности расчетная генерация компонент спектра в области третьей гармоники ("фиолетовой" части спектра) усиливается, а в высокочастотной ("синей") части основного спектра, напротив, ослабляется. Для относительно высокочастотного излучения это может, в частности, приводить к тому, что увеличивается доля поглощенной энергии импульса (так как высшие гармоники спектра оказываются в этом случае в зоне поглощения вещества; граница этой зоны для кварцевого стекла разных видов составляет  $\lambda = 180 \div 220 \text{ нм}$  [30]).

Из рис. 5 видно, что для излучения с исходно большей центральной длиной волны вклад дисперсионных компонент уравнения (14) существенно слабее по сравнению с

предыдущим случаем и заметен в основном в области утроенных компонент уширенного спектра.

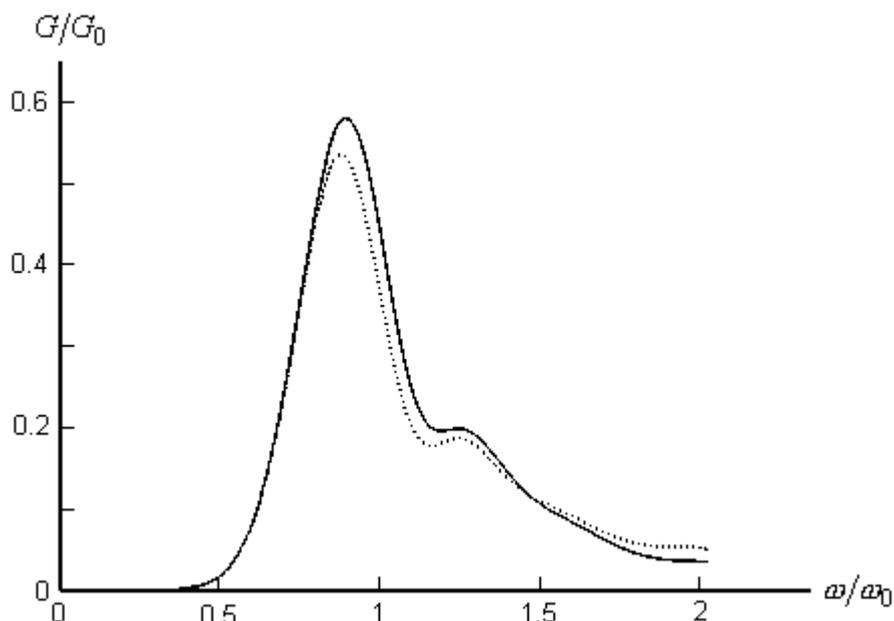


Рис. 4. Фемтосекундный спектральный суперконтинуум в кварцевом стекле: сплошная кривая – расчет без учета дисперсии, пунктирная – с учетом дисперсии (пиковая интенсивность входящего импульса  $I_0 = 1 \cdot 10^{13} \text{ Bm/cm}^2$ , центральная длина волны  $\lambda = 390 \text{ нм}$ )

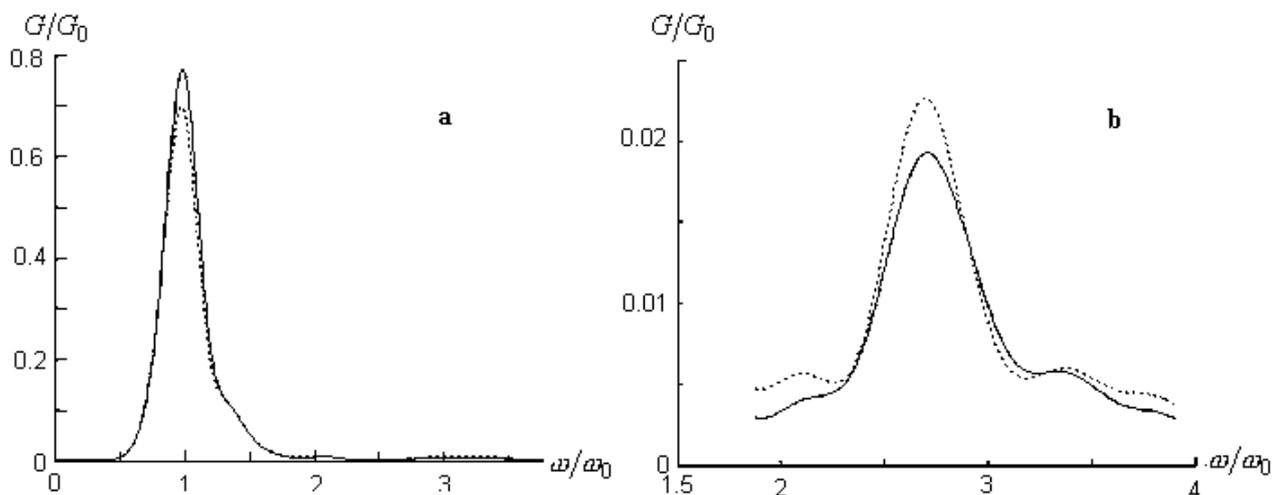


Рис. 5. Фемтосекундный спектральный суперконтинуум в кварцевом стекле: сплошная кривая – расчет без учета дисперсии, пунктирная – с учетом дисперсии (пиковая интенсивность  $I_0 = 1 \cdot 10^{13} \text{ Bm/cm}^2$ , центральная длина волны импульса  $\lambda = 780 \text{ нм}$ ). Рисунок а – общий вид спектра, б – детализация в области высоких частот.

## Заключение

В работе проанализирована дисперсия нелинейной поляризованности диэлектриков электронной природы. В рамках нерезонансного приближения получено выражение для нелинейной поляризованности, определен вид ее инерционности. Приведено уравнение динамики поля, учитывающее дисперсию как линейного, так и нелинейного показателя преломления диэлектрической среды. Промоделировано самовоздействие импульса с малым числом колебаний светового поля в кварцевом стекле. Показано, что

дисперсия нелинейного отклика приводит к уширению высокочастотной части генерируемого спектрального суперконтинуума, в том числе к увеличению эффективности генерации утроенных частот.

Работа поддержана грантами РФФИ № 01-02-17841 (2001 г.) и Федеральной программы "Университеты России" № 01.01.030.

## Литература

1. Steinmeyer G., Sutter D.H., Gallman L., Matuschek N., Keller U. Frontiers in ultrashort pulse generation: pushing limits in linear and nonlinear optics. // *Science*, 1999. v.286. p.1507-1512.
2. Brabec Th., Krausz F. Intense few-cycle laser fields: Frontiers of nonlinear optics // *Rev. Mod. Phys.* 2000., v.72, N2, p.545-591.
3. Gerullo G., De Silvestri S., Nisoli M., Sartania S., Stagira S., Svelto O. Few-optical-cycle laser pulses: From high peak power to frequency tunability. // *IEEE J. of Selected Topics in Quantum Electronics*. 2000, v.6, N6, p.948-958.
4. Corkum P.B., Burnett N.H., Bruehl F. Above-threshold ionization in the long-wavelength limit. // *Phys. Rev. Lett.* 1989. v. 62. N11. p. 1259-1262.
5. Burnett N.H., Corkum P.B. Cold-plasma production for recombination extreme-ultraviolet lasers by optical field induced ionization. // *J. Opt. Soc. Am. B*. 1989. v.6. N6. p. 1195-1199.
6. Alfano R.R. The supercontinuum laser source. Springer, New-York, 1989.
7. Chin S.L., Brodeur A., Petit S., Kosareva O.G., Kandidov V.P. Filamentation and supercontinuum generation during the propagation of powerful ultrashort laser pulses in optical media (white light laser). // *J. Nonl. Opt. Phys. And Mater.*, 1999, v. 8, N1, p. 121-146.
8. Brodeur A., Chin S.L. Ultrafast white-light continuum generation and self-focusing in transparent condensed media. // *J. Opt. Soc. Am. B*, 1999, v. 16, N4, p. 637-650.
9. Ranka J.K., Windeler R.S., Stentz A.J. Visible continuum generation in air-silica microstructure optical fiber with anomalous dispersion at 800 nm // *Opt. Lett.* 2000. v.25, N1, p.25-27.
10. Ахманов С.А., Выслоух В.А., Чиркин А.С. Оптика фемтосекундных лазерных импульсов. М., Наука, 1988. 312 с.
11. Bepalov V.G., Kozlov S.A., Shpolyasnsky Yu.A., Walmsley I.A. Simplified field wave equations for nonlinear propagation of extremely short light pulses // *Phys. Rev.*, 2002, v.66, 013811
12. Brabec Th., Krausz F. Nonlinear optical pulse propagation in the single-cycle regime. // *Phys. Rev. Lett.* 1997, v.78, N17, p.3282-3285.
13. Ranka J.K., Gaeta A.L. Breakdown of the slowly-varying envelope approximation in the self-focusing of ultrashort pulses. // *Opt. Lett.* 1998, v.23, N7, p.534-536.
14. Маймистов А.И. Некоторые модели распространения предельно коротких электромагнитных импульсов в нелинейной среде. // *Квантовая электроника*, 2000, т. 30, 287-304.
15. Husakou A.V., Herrmann J. Supercontinuum generation, four-wave mixing, and fission of higher-order solitons in photonic-crystal fibers. // *J. Opt. Soc. Am. B* 2002, v. 19, p. 2171-2175.
16. Gaeta A.L. Nonlinear propagation and continuum generation in microstructured optical fibers // *Opt. Lett.* 2002, v. 27, p. 924-928.
17. Азаренков А.А., Альтшулер Г.Б., Козлов С.А. Нерезонансный нелинейный поляризационный отклик среды в поле предельно коротких световых импульсов. // *Оптика и спектроскопия*, 1991, т. 71, №2, с. 334-339.
18. Braunstein R. Nonlinear optical effects // *Phys. Rev.* 1962, v. 125, p. 475-477.
19. Fournier J.T., Snitzer E. The nonlinear refractive index of glass. // *IEEE J. Quant. Electron.* v. QE-10, 1974, №5, p.473-475.
20. Альтшулер Г.Б. Нелинейность показателя преломления диэлектриков при электронном и электронно-ядерном механизмах. // *Оптика и спектроскопия*, 1983, т. 55, №1, с. 83-89.
21. Козлов С.А., Сазонов С.В. Нелинейное распространение импульсов длительностью в несколько колебаний светового поля в диэлектрических средах. // *ЖЭТФ*, 1997, с. 404-418.
22. M. Born, A. Wolf. Principles of Optics. Pergamon Press, N.Y., 1968, 4 ed.
23. Ross I.N., Toner W.T., Hooker C.J., Barr J.R.M., Coffey I. Nonlinear properties of silica and air for picosecond ultraviolet pulses. // *J. Modern. Opt.* 1990, v. 37, p. 555-573.
24. Owyong A., Hellwarth R.W., George N. Intensity-Induced Changes in Optical Polarizations in Glasses // *Phys. Rev. B.*, 1972, v. 5. p.628-633.
25. Owyong A. Ellipse rotation studies in laser host materials. // *IEEE J. Quant. Electron.*, 1973, v. 9, p. 1064-1069.
26. Альтшулер Г.Б., Карасев В.Б., Шарлай С.Ф. Управление характеристиками излучения с помощью элементов с нелинейным преломлением. Л.: ЛДНТП, 1978.

27. Milam D., Weber M.J. Measurement of nonlinear refractive-index coefficients using time-resolved interferometry: Application to optical materials for high-power neodymium lasers // J. Appl. Phys., 1976, v. 47, p. 2497-2501.
28. Milam D., Weber M.J. Time-resolved interferometric measurements of the nonlinear refractive index in laser materials // Optics Comms., 1976, v. 18, p. 172-173.
29. Adair L., Chase L.L., Payne S.A. Nonlinear refractive-index measurements of glasses using three-wave frequency mixing // J. Opt. Soc. Am. B, 1987, v. 4, p. 875-881.
30. Оптическое стекло. Каталог. М., 1979.
31. Штумпф. С.А. Влияние дисперсии нерезонансной нелинейности диэлектрика на характер самовоздействия в нем импульса из нескольких колебаний светового поля. // в кн.: Проблемы нелинейной и когерентной оптики, СПб, 2002, с.236-244.

# МЕТОДИКА ОБРАБОТКИ ОТРАЖАТЕЛЬНЫХ ГОЛОГРАММ, ЗАПИСАННЫХ НА ФОТОПЛАСТИНАХ ВРП ПРИ ИМПУЛЬСНОМ ЭКСПОНИРОВАНИИ

М.К. Серебрякова

Научный руководитель – к.ф.-м.н., ст.н.с. В.Н. Михайлов

Исследована возможность применения для обработки голограмм, записанных при помощи импульсного лазера с наносекундной длительностью экспозиции, метода, основанного на формировании в эмульсионном слое микропустот.

## Введение.

Использование импульсных лазеров позволяет записывать голограммы не только стационарных (неподвижных), но и нестационарных (подвижных) объектов. Однако известно, что поведение галогидосеребряных материалов сильно зависит от длительности экспозиции.[1] В частности, при наносекундной длительности засветки чувствительность таких материалов значительно снижается. Повысить дифракционную эффективность можно с помощью химической обработки экспонированных пластин. Одним из эффективных методов является метод, основанный на получении так называемых галогидосеребряных желатиновых голограмм. Он базируется на достижении большой разницы показателей преломления в эмульсии за счет образования в эмульсионном слое микропустот. Ранее эта методика использовалась лишь для обработки голограмм, записанных в непрерывном режиме. В данной работе исследовалась возможность применения такого метода для обработки голограмм, записанных при помощи импульсного лазера с наносекундной длительностью экспозиции.

## Экспериментальная часть

В качестве источника излучения для записи отражательных голограмм зеркальных объектов использовался лазер на основе тетрафторида иттрия-лития, активированного неодимом, генерирующий импульсы длительностью 20 нс, с преобразованием излучения во вторую гармонику (длина волны 527 нм). Схема экспериментальной установки записи голограмм изображена на рис.1.

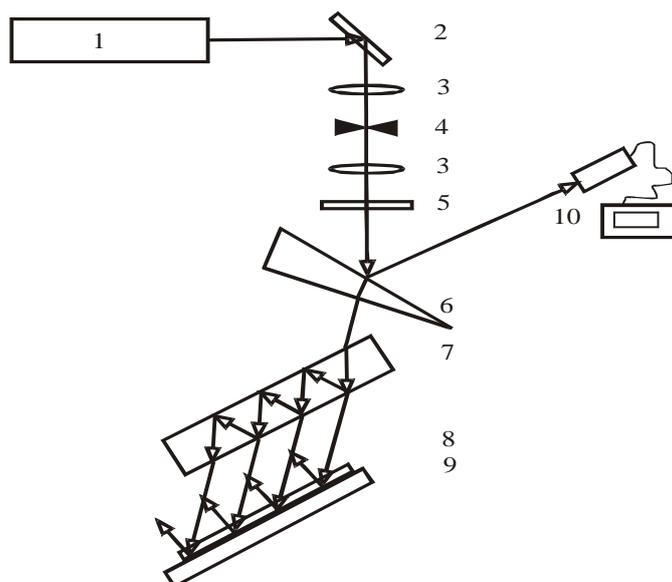


Рис. 1. Схема записи отражательных голограмм зеркальных объектов. 1 – лазер YLF:Nd, 2 – поворотное зеркало, 3 – телескоп, 4 – точечная диафрагма, 5 – фильтр, 6 – оптический клин, 7 – делитель, 8 – фотопластина, 9 – зеркало, 10 – ИЛД.

Угол схождения опорного и объектного пучков составлял 125°. Соотношение пучков при записи отражательных голограмм было M=1:1. Энергия опорного пучка составляла от  $6 \times 10^{-4}$  до  $1 \times 10^{-5}$  Дж/см<sup>2</sup>. Голограммы записывались на пластинках ВРП производства завода "Славич". Эти пластины являются универсальными и могут использоваться как для непрерывной, так и для импульсной засветки. Согласно [2], пластины ВРП являются одними из наиболее эффективных отечественных галогидосеребряных фотоматериалов для импульсной голографической записи. Максимум спектральной чувствительности этих пластин лежит в зеленой области спектра. Средний размер микрокристаллов галогенида серебра фотопластинок ВРП составляет около 40 нм.

Основными стадиями обработки галогидосеребряных желатиновых голограмм являются проявка, отбеливание, задубливание, фиксирование и дегидратация [3]. Роль стадии задубливания при получении таких голограмм очень важна. Процесс получения галогидосеребряных желатиновых голограмм базируется на факте, что микропустоты, сформированные во время обработки, остаются после обработки и создают модуляцию показателя преломления. Структура микропустот может легко разрушиться во время или после обработки, во время стадии дегидратации. Существенной проблемой рассматриваемого процесса является вопрос, как заменить молекулы воды и использующихся при дегидратации спиртов на воздух без усадки эмульсии или разрушения структуры микропустот в обработанной эмульсии. Добиться этого можно при достаточной степени задубливания желатинового слоя, которая, в свою очередь, может быть получена несколькими способами:

- предварительным задубливанием эмульсии перед проявлением,
- промежуточным задубливанием,
- задубливанием в процессе обработки путем использования проявляющих и отбеливающих растворов, обладающих дубящими свойствами.

Для некоторых фотоматериалов может быть использовано сочетание нескольких этих способов.

За основу обработки экспонированных голограмм была выбрана схема, состоящая из стадии проявления, отбеливания и дегидратации.

Известно, что наиболее эффективными для обработки импульсных голограмм являются проявители HD и SM6 [4, 5]. Их состав приведен в табл. 1. Важным различием этих растворов являются их дубящие свойства. Проявитель HD является сильно дубящим, в то время как SM6 содержит большое количество сульфита натрия, предотвращающего задубливание. Результаты экспериментов показали, что проявление в недубящем проявителе SM6 предпочтительнее с точки зрения светорассеяния голограмм, прошедших все стадии обработки.

	SM-6	HD
Сульфит натрия	-	5 г
Метол	-	5 г
Гидрохинон	-	10 г
Аскорбиновая кислота	18 г	10 г
Двухосновный фосфат натрия	28.4 г	-
Фенидон	6 г	1 г
Гидроксид натрия	12 г	10 г
Вода	До 1 литра	До 1 л

Таблица 1. Состав проявляющих растворов

Как известно, стадия отбеливания проводится для преобразования голограммы из амплитудной в фазовую и повышения разности показателей преломления экспониро-

ванной и неэкспонированной областей эмульсии [6]. В зависимости от типа используемого отбеливателя отбелка может быть реверсивной, с полным удалением металлического проявленного серебра из эмульсионного слоя, или регалогенирующей, переводящей проявленное непрозрачное серебро в прозрачный галогенид. В данной работе использовались три различных отбеливателя:

- недубящий реверсивный отбеливатель на основе перманганата калия (ПМО-1),
- недубящий регалогенирующий отбеливатель также на основе перманганата калия (ПМО-2),
- дубящий реверсивный отбеливатель на основе бихромата калия.

	ПМО-1	ПМО-2	Бихромат
Перманганат калия	0.4 г	0.4 г	-
Бихромат калия	-	-	0.8 г
Серная кислота	0.4 мл	0.4 мл	1 мл
Хлорид натрия	-	10 г	-
Вода	До 1 л	До 1 л	До 1 л

Таблица 2. Состав отбеливающих растворов

Составы отбеливающих растворов указаны в таблице 2.

В результате проведенных экспериментов нами был выбран недубящий регалогенирующий отбеливатель ПМО-2, дающий наибольшие значения дифракционной эффективности при относительно небольшом светорассеянии голограмм.

Таким образом, после описанной выше обработки и градиентной сушки в этаноле были получены фазовые голограммы, содержащие галогенид серебра. Для исследования возможности дальнейшего увеличения яркости и контраста отражательных изображений голограмм была предложена схема химико-фотографической обработки, использующая стадию второго экспонирования и последующей реверсивной отбелки. Стадия второго некогерентного экспонирования и проявления в разбавленном проявляющем растворе (см. табл. 3) позволяет снизить рассеяние в фотослое за счет получения мелкодисперсного коллоидного серебра. Последующая реверсивная отбелка является дубящей, что позволяет дополнительно увеличить разность показателей преломления в фотослое. Таким образом, в окончательном варианте последовательность стадий химико-фотографической обработки заключалась в следующем:

1. проявление в недубящем проявителе SM6,
2. отбеливание в перманганатном отбеливателе ПМО-2 или ПМО-1,
3. градиентная сушка в этаноле,
4. второе проявление в разбавленном дубящем проявителе HD,.
5. отбеливание в бихроматном отбеливателе,
6. сушка в изопропанолe.

Во время заключительной стадии использовали градиентную последовательную дегидратацию в 50%-ом, 90%-ом и 100%-ом изопропиловом спирте. При обработке части пластинок в проявитель (при втором проявлении) добавляли роданид аммония, позволяющий подрастворить микрокристаллы галогенида серебра при проявлении. Последовательность стадий химико-фотографической обработки фотопластин приведена в табл. 3.

В дальнейшем проводились измерения дифракционной эффективности и ширины спектра записанных голограмм, обработанных согласно стадиям, приведенным в табл. 3. На спектрофотометре СФ18 измерялся спектр пропускания полученных отражательных голограмм. Графики приведены на рис. 2.

	1 группа	2 группа	3 группа
Проявление SM6 не перемешивая	4 мин	4 мин	4 мин
Отбелка	ПМО-2 До осветления + 30 с	ПМО-2 До осветления + 30 с	ПМО-1 До осветления + 30 с
Сушка Ethanol	50%-3 мин 100%-3 мин	50%-3 мин 100%-3 мин	50%-3 мин 100%-3 мин
2-е экспонирование (60 Вт, L = 30см)	5 мин	5 мин	5 мин
Проявление HD 1:15 не перемешивая	15 мин	+ 0.5 г/л (NH <sub>4</sub> )SCN (роданид аммония) 15 мин	15 мин
Отбелка Бихромат	До осветления + 30 с	До осветления + 30 с	До осветления + 30 с
Сушка ISO	50% -3 мин 90% -3 мин 100%-3мин	50% -3 мин 90% -3 мин 100%-3мин	50% -3 мин 90% -3 мин 100%-3мин

Таблица 3. Последовательность химической обработки голограмм в эксперименте.

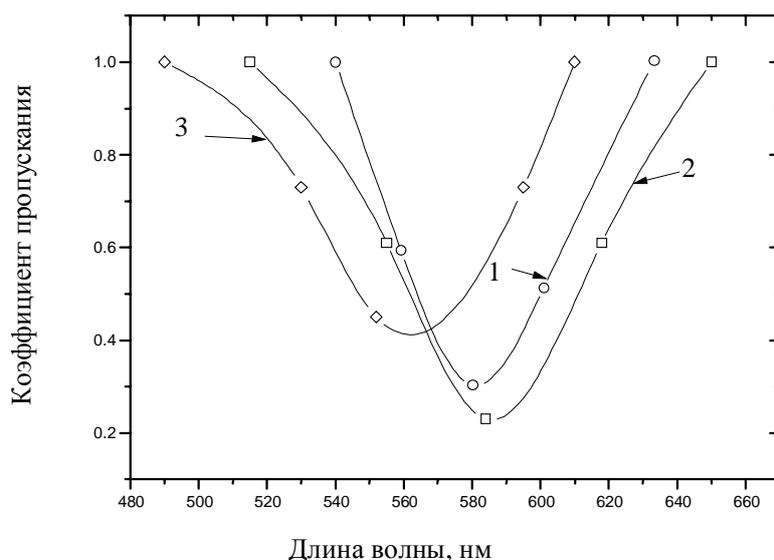


Рис. 2. Спектр пропускания отражательных голограмм зеркальных объектов.

	Дифракционная эффективность	Спектральная селективность
1 группа	40-45%	40-45 нм
2 группа	40-45%	65 нм
3 группа	40-45%	63 нм

Таблица 4. Дифракционная эффективность и спектральная селективность отражательных импульсных голограмм.

Исходя из построенных графиков, определялась величина спектральной селективности голограмм. Дифракционная эффективность измерялась с помощью аргонового лазера (длина волны 514.5 нм). Результаты измерений приведены в табл. 4. Полученная дифракционная эффективность всех голограмм составляла 40–45%, спектраль-

ная селективность голограмм первой группы составляла 40–45 нм. Меньшая спектральная селективность (ширина спектра около 65 нм) была получена на голограммах второй и третьей группы.

### Обсуждение результатов

Дифракционная эффективность отражательных голограмм зависит от разности показателей преломления экспонированной и неэкспонированной области эмульсионного слоя. В данной работе разность показателей преломления в зоне, экспонированной лазерным излучением, и в зоне, засвеченной лишь некогерентным излучением, достигалась одновременно несколькими способами. Обработка в регалогенирующем отбеливателе после первого проявления приводит к концентрированию галогенида серебра в зоне последующей некогерентной засветки. Задубленность желатины в этой области, являющаяся результатом обработки пластин в отбеливателе на основе бихромата, также приводит к увеличению разности показателей преломления. Еще большая разница достигается после дегидратации в изопропиловом спирте, приводящая к образованию микропустот в незадубленной области желатины, расположенной в зоне когерентной засветки пластины. Следует заметить, что отбелка в бихромате приводит к неравномерности дубления эмульсионного слоя по глубине, в результате чего уменьшается (относительно теоретической) спектральная селективность отражательных голограмм, что в свою очередь приводит к повышению яркости формируемого голограммой изображения при его восстановлении белым (немонохроматическим) светом. Еще большие расхождения практической и теоретической спектральной селективности достигаются в результате незначительных изменений химической обработки. Как видно из табл. 4, использование реверсивной отбелки после первого проявления позволяет получить спектральную селективность 65 нм (при теоретической 22 нм). Введение роданида аммония в проявитель во время второго проявления также уменьшает спектральную селективность отражательных голограмм, что, по-видимому, является результатом неравномерного по глубине второго проявления. Во всех приведенных случаях поверхность фотослоя повреждается, и толщина рабочего слоя эмульсии уменьшается, что и приводит к уменьшению спектральной селективности голограмм.

### Заключение

В результате проделанной работы для импульсных отражательных голограмм была подобрана схема химической обработки, позволяющая получить дифракционную эффективность более 40 % и варьировать спектральную селективность таких голограмм от 40 до 65 нм. Использование разработанной методики химико-фотографической обработки для обработки импульсных отражательных голограмм диффузных объектов показали перспективность ее использования, особенно при импульсном копировании изобразительных голограмм.

### Литература

1. Джеймс Т. Теория фотографического процесса. 4 изд. Л.: Химия, 1980., 672 с.
2. Шварцвальд А.И., Стаселько Д.И. и др. // Тез. докл. XIII Всес. конф. "Высокоскоростная фотография, фотоника и метрология быстропотекающих процессов". М., 1987. С. 165.
3. Усанов Ю.Е., Шевцов М.К., Кособокова Н.Л., Кириенко Е.А. Механизм образования микропустотной структуры и методы получения галогенидосеребряных желатиновых голограмм. // Оптика и спектроскопия. 1991. 71(4)б 651-658б
4. V.N. Mikhailov, J.Y.Son, O.V.Grinevitskaya, H.S.Lee, Y.J.Choi Improved processing for silver halide pulse holography. // Proc. SPIE. 1996. Vol. 2688, pp. 155-161.
5. H.I. Bjelkhagen Silver halide recording materials for holography and their processing. Springer, Berlin, 1993.
6. P. Hariharan Basic processes involved in the production of bleached holograms. // The Journal of Photographic Science. 1990. Vol. 38.

# **КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ И АНАЛИЗ ЭФФЕКТИВНОСТИ ОПТИЧЕСКОГО СОЕДИНЕНИЯ ТИПА КАНАЛЬНЫЙ ВОЛНОВОД – ОПТИЧЕСКОЕ ВОЛОКНО**

**Н.Н. Вознесенская**

**Научный руководитель – д.т.н., проф. В.П. Вейко**

Рассмотрено новое направление в оптоэлектронике и фотонике - оптические интегральные схемы, являющиеся частью гибридных оптических устройств, а также проблема световой эффективности оптических связей или оптических межсоединений (ОМ). Для вычисления световой эффективности ОМ предложен новый метод моделирования. Метод основан на послойном вычислении световых волн, проходящих через набор неоднородных сред. Разработана программа, которая позволяет промоделировать прохождение света через достаточно сложные структуры с размером порядка длины волны. Метод применяется к новым оптическим элементам для ОМ – микролинзам, сформированным на торце оптического волокна, и оптическим канальным волноводам. Показано, что применение основного приближения геометрической оптики – предположение о бесконечной малости длины волны – неприемлемо для решения поставленной в работе задачи. Расчеты показывают неоднозначность влияния формы и взаимного расположения микрооптических элементов на картину распределения интенсивности проходящего света и световую эффективность ОМ.

## **Введение**

Проектирование оптических межсоединений (Optical Interconnects) (далее – ОМ) представляет собой набор важнейших задач в создании оптических интегральных схем, которые применяются для усиления и фильтрации оптической связи и при создании сверхъёмкой памяти для компьютера. Данная работа посвящена одному из аспектов разработки оптических микросхем, а именно математическому моделированию и исследованию микрооптических узлов различной конфигурации, используемых для передачи излучения между современными устройствами фотоники и оптической связи – оптическими межсоединениями и оптическими микросхемами. Специфика задачи заключается в том, что, в отличие от чисто электронных соединений, переход излучения из одной части соединения в другую определяется достаточно сложным механизмом дифракции и интерференции. Как правило, различные компоненты оптических соединений имеют совершенно разную геометрическую форму и различный состав оптических материалов. Из-за этого при переходе излучения из одной части в другую неизбежны значительные потери энергии оптического излучения. Поэтому в работе предполагается разработать математическую модель для расчета прохождения излучения через оптические соединения различной конфигурации с целью их оптимизации на предмет увеличения их эффективности, что является весьма актуальным.

## **Описание и анализ существующих подходов к решению поставленной в работе задачи**

Принцип действия ОМ может быть описан двумя способами – на основе расчета лучей или волноводной теории. К сожалению, ни один из этих подходов не достаточен для полного объяснения действия оптических межсоединений. Принцип трассировки лучей имеет место только в дальнепольной зоне, что соответствует условию, при котором длина волны стремится к нулю, и, напротив, волноводная теория хорошо реализована только для правильных волноводов и не предлагает никакого решения для соединений оптических элементов различной конфигурации [1–3].

Метод расчета лучей основан на решении общих уравнений для преломления и отражения лучей, используемых в различных системах программного обеспечения для проектирования линз (например, ZEMAX, SOLTIS, Light Tools и т.д.).

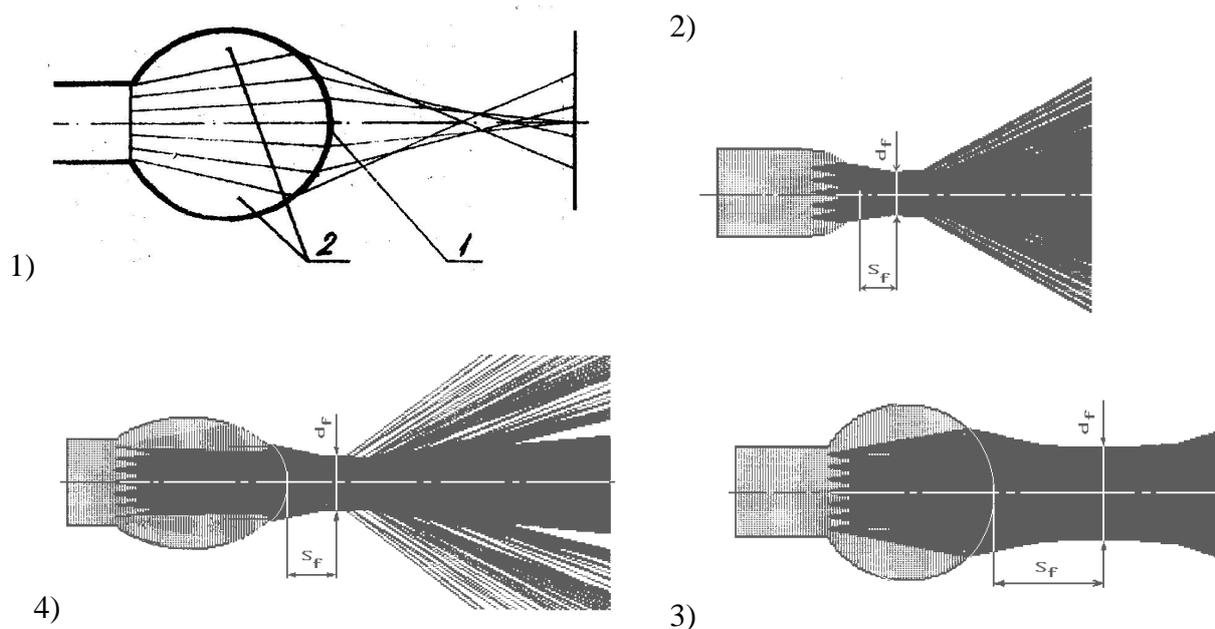


Рис. 1. Теоретическая модель анализа концевой оптоволоконной линзы (1): 1 – область преломления, 2 – область полного внутреннего отражения. Теоретические каустики 2) арковой, 3) сферической и 4) эллиптической концевых оптоволоконных линз

На рис. 1 представлена картина расчета лучей, которая показывает сферически-абберационный луч света, появляющийся от микролинзы на конце волокна, и пучок света, подобный форме гауссова пучка лучей. Однако такой подход едва ли можно применить для сужения лазерных пучков из-за значительных интерференционных эффектов, описываемых только в эрмит-гауссовой концепции [3,4]. Главное различие между расчетом лучей и эрмит-гауссовыми решениями заключается в том, что в тех областях пучка лучей, где лучи проходят очень плотно друг к другу, имеют место специфические распределения света, которые абсолютно отличаются от данных расчета лучей. Эти лучевые области очень важны для анализа ОМ. Заметим, что, как вытекает из параксиальной оптики, инвариант Лагранжа-Гельмгольца является точной величиной только в рамках метода расчета лучей [4, 5, 7].

Для улучшения адекватности модели важно разумно использовать волновую теорию света и попробовать расширить рамки волноводной теории на наиболее актуальные случаи ОМ. Как следует из основных принципов дифракции света, инвариант Лагранжа-Гельмгольца является асимптотическим дальнепольным приближением поперечных измерений дифрагированного луча, удовлетворяющего условию Фраунгофера. Иначе, формализм гауссовых пучков соответствует либо дальне-, либо ближнепольной области, которая является необходимой для анализа ОМ. К сожалению, нет никаких аналитических решений для общих форм конца волновода и тем более для различных оптических связей. Поэтому в настоящей работе используется простая методика для числовой оценки светового распространения через сложные формы микроструктурных оптических элементов [5].

### Разработка новой теории и компьютерное моделирование

Известно, что существующие математические модели, которые используются в зарубежных компьютерных программах, не всегда доступны и не являются универсальными. Использование хорошо известных волноводных моделей наталкивается на сложность конфигурации оптических трактов и поэтому не может дать адекватные результаты для анализа эффективности прохождения света через оптические межсоеди-

нения и ввода излучения в оптические микросхемы. В данной работе рассмотрен общий волновой подход к анализу распространения света через микроструктуры различных конфигураций. Так как использование преобразования Фурье при расчете прохождения излучения через неоднородную среду затруднено, в работе используется послойное моделирование прохождения волнового поля через участки (домены) с различными оптическими свойствами. Принцип метода плоских доменов основан на дискретизации структуры в продольном (осевом) направлении тонкими слоями, в пределах которых световая волна описывается на основе теории дифракции Френеля. Помимо этого, каждый слой рассматривается как состоящий из однородных сред. В этом случае метод Фурье-преобразование может применяться для того, чтобы вычислить прохождение плоских волн комплексной амплитуды поля в пределах каждой однородной области. Под плоской волной понимается набор решений уравнения Гельмгольца в общем случае, что дает легкую возможность для вычисления их распространения через среду. Рис. 2 иллюстрирует алгоритм последовательного вычисления распространения света через ту или иную структуру.

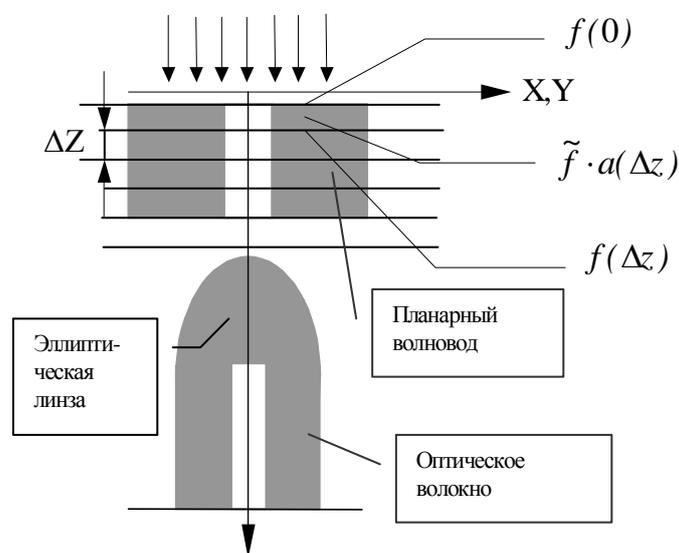


Рис. 2. Принцип дискретного и последовательного вычисления распространения светового потока через неоднородную структуру:  $dZ$  – толщина уровня,  $f(0)$  – световая амплитуда на верхней плоскости уровня,  $\tilde{f}$  – преобразование Фурье  $f(0)$ ,  $a(\Delta z)$  – коэффициент распространения,  $f(\Delta z)$  – обратное преобразование Фурье  $\tilde{f} \cdot a(\Delta z)$ , дающее световую амплитуду на нижней уровня

Преобразование Фурье остается оптимальной процедурой для математического синтеза распределений поля в доменах в единое целое. При этом возможна визуализация трехмерного распределения не только скалярной амплитуды, но и векторной комплексной амплитуды светового поля с последующим вычислением более адекватной картины интенсивности. Основным преимуществом выбранного метода решения неоднородного волнового уравнения, которое для данной работы весьма важно, является органичное сочетание в нем геометрического и волнового описания светового поля в пространстве. Несмотря на математическую сложность этого метода, он в данном случае намного проще, чем расчет лучей, так как не приводит к неопределенностям при расчете прохождения света через границы между различными средами и дает целостную картину распространения поля в пространстве [5–8].

Математический аппарат метода сводится к следующему: во-первых, решается неоднородное волновое уравнение относительно скалярного поля, во-вторых, скалярное поле раскладывается по плоским волнам при помощи дискретного преобразования

Фурье и, в-третьих, учитываются оптические параметры сред для вычисления распространения плоских волн. Этими параметрами являются диэлектрические проницаемости, которые входят в дифракционные множители распространения поля внутри различных сред.

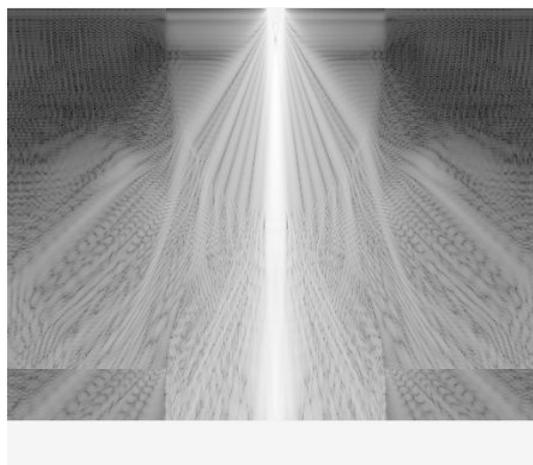
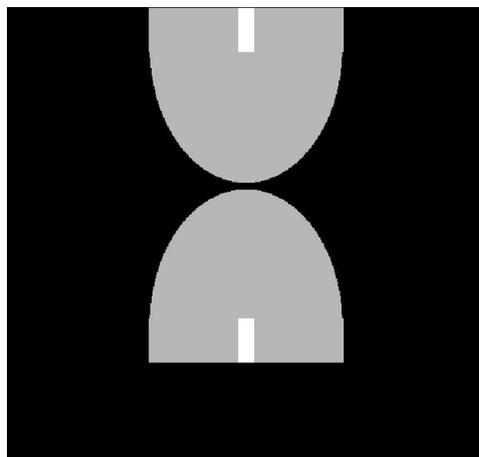


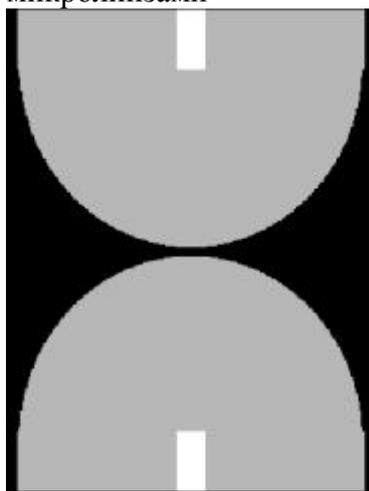
Рис. 3. Свет появляется из верхнего конца оптоволоконной коллимирующей эллиптической линзы диаметром  $124 \mu$  и фокальной длиной  $88 \mu$  ( $\lambda = 1 \mu$ ) и проходит вниз до подобной фокусирующей линзы, установленной на конце противоположного оптического волокна. Промежуток между линзами –  $4 \mu$

2 оптических волокна без микролинз.



Зазор между волокнами =  $4 \mu$ .

2 оптическое волоконных волокна с микролинзами



L1:  $D_{11}=125 \mu$ ,  $H_{11}=88 \mu$

L2:  $D_{12}=125 \mu$ ,  $H_{12}=88 \mu$

Зазор =  $4 \mu$ .

Распределение интенсивности света.



$\lambda = 1 \mu$ . Эффективность = 79%.



Эффективность = 59%

$\lambda = 1 \mu$ .

Рис. 4. Результаты расчета прохождения света через различные варианты оптических межсоединений

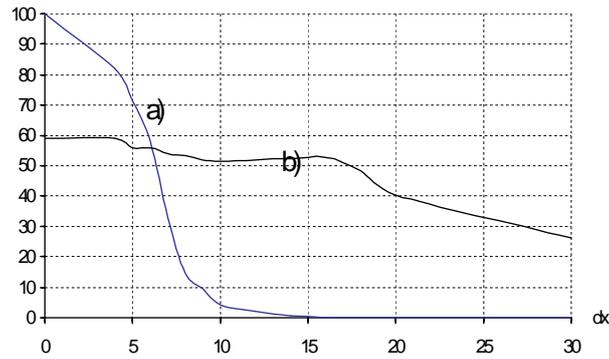
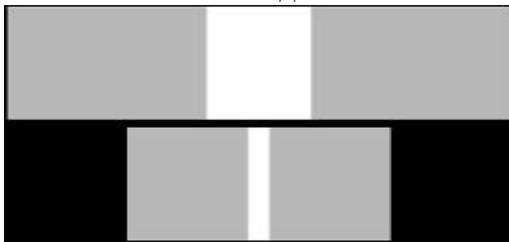
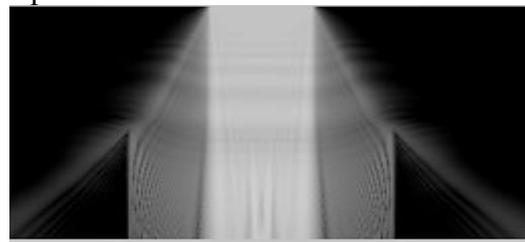


Рис. 5. Компьютерное моделирование эффективности оптического контакта двух одномодовых оптических волокон: а) без коллимирующей и фокусирующей линзы, б) с коллимирующей и фокусирующей линзы. Примечание: метод плоских доменов показывает более низкую границу величин эффективности, которая в действительности может быть выше

Канальный волновод и оптическое волокно без микролинзы

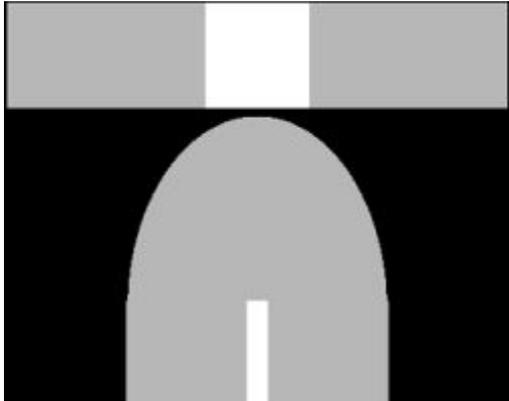


Внутренний диаметр волновода 50x25  $\mu$ , зазор 4  $\mu$ . Внутренний диаметр оптоволокна 9,5  $\mu$

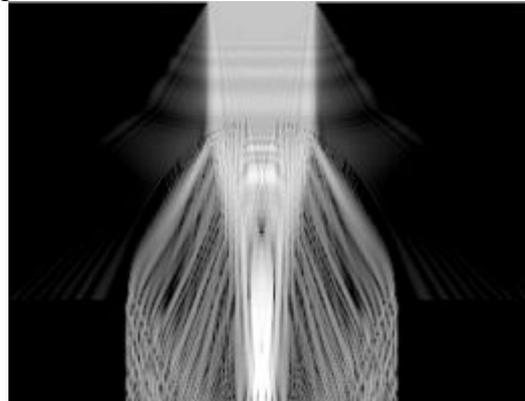


$\lambda=1 \mu$   
Эффективность = 12%

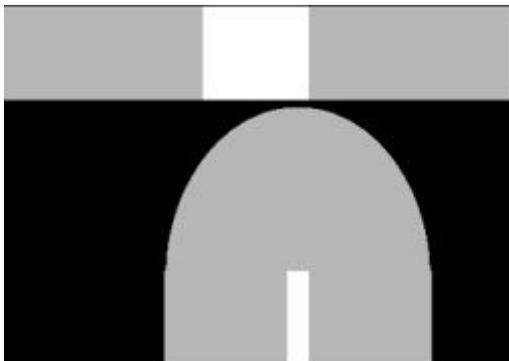
Канальный волновод и оптическое волокно с микролинзой



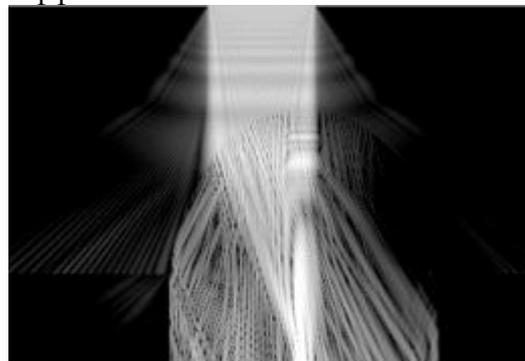
Размер линзы:  $D_{12}=125 \mu$ ,  $H_{12}=88 \mu$



$\lambda=1 \mu$   
Эффективность = 55.2%



Размер линзы:  $D_{12}=125 \mu$ ,  $H_{12}=88 \mu$  Поперечный сдвиг по оси = 20  $\mu$



$\lambda=1 \mu$   
Эффективность 54.5%

Рис. 6. Результаты расчета прохождения света через различные варианты оптических межсоединений

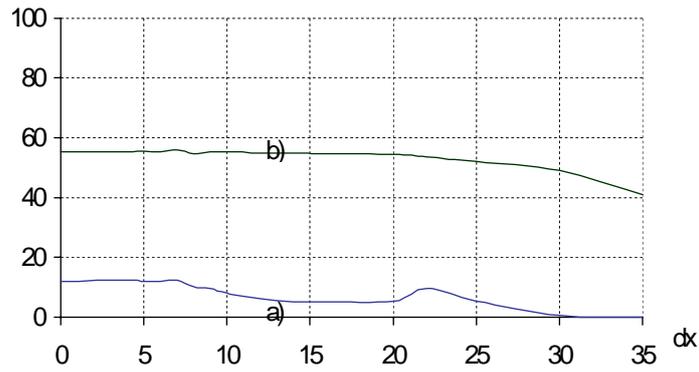
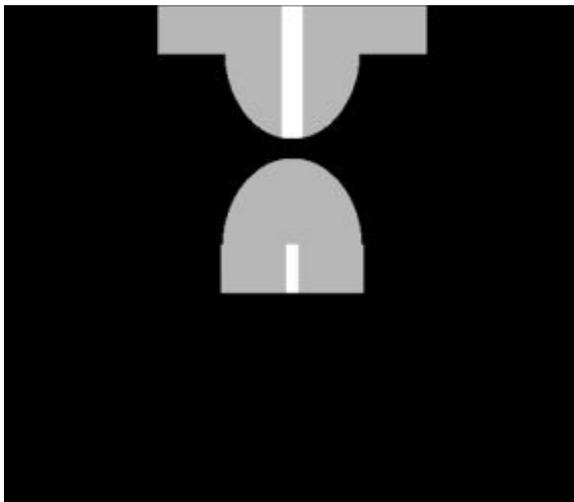
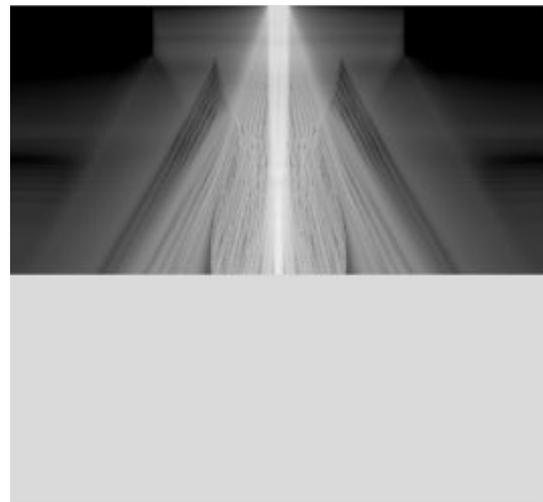


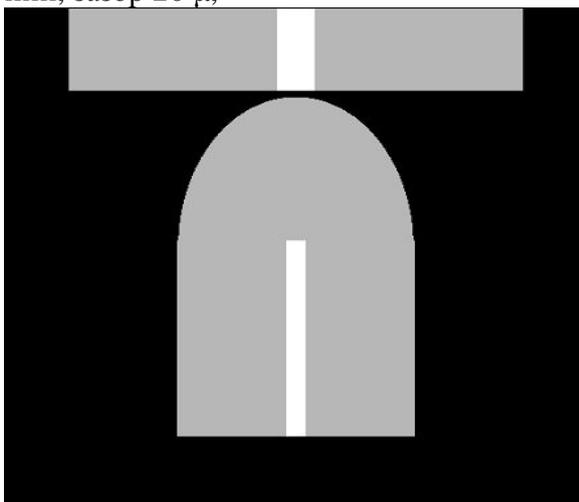
Рис.7. Компьютерное моделирование эффективности оптического контакта канального волновода  $50 \times 25 \mu$  с одномодовым оптическим волокном в различных поперечных сечениях: а) плоский конец, б) с линзой на конце волокна, диаметр линзы  $124 \mu$



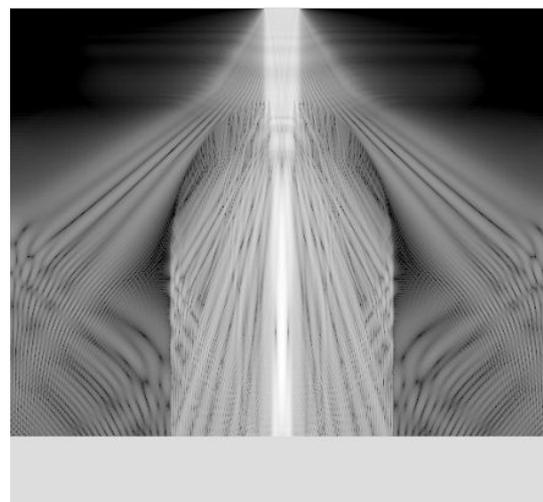
Размер сердцевинки канального волновода  $30 \times 30 \mu$ . Диаметр микролинзы =  $120 \mu$ . Параметры волокна (см. выше) без изменений, зазор  $20 \mu$ ,



$\lambda = 1,5 \mu$   
Эффективность – 68%



Канальный волновод без микролинзы, оптическое волокно вытянуто в продольном направлении, зазор =  $4 \mu$ ,



$\lambda = 1.5 \mu$   
Эффективность = 57,77 %

Рис.8. Результаты расчета прохождения света через различные варианты оптических межсоединений

Рис. 3 иллюстрирует пример вычисления структуры интенсивности пучка, идущего от концевой оптоволоконной коллимирующей эллиптической линзы диаметром 124 мкм и фокальной длиной 88 мкм и проходящего через подобную линзу, установленную на конце другого волокна. Зазор между линзами равен 4 мкм. В отличие от примера расчета лучей, показанного на рисунке, эти линзы безабберационные [13].

Все результаты вычислений представлены в логарифмическом масштабе, чтобы показать детально все интерференционные и дифракционные эффекты при прохождении света через оптические соединения. Разумеется, невозможно получить такую картину методом расчета лучей.

На рис. 4–8 представлены результаты расчета прохождения света через различные варианты оптических межсоединений.

На рис. 8 показан пример расчета прохождения света через структуру канальный волновод - оптическое волокно для двух случаев: в первом линза сделана как на конце одномодового оптического волокна, так и на волноводе; во втором линза присутствует только на волокне, а длина волокна больше, но зазор меньше, чем в первом случае. Из рисунка видно, что эти два соединения мало уступают друг другу по эффективности.

## Заключение

Для вычисления световой эффективности ОМ предложен новый метод моделирования. Метод основан на послойном вычислении световых волн, проходящих через набор неоднородных сред. Разработана программа, которая позволяет промоделировать прохождение света через достаточно сложные структуры с размером порядка длины волны. Метод применен к новым оптическим элементам для ОМ – микролинзам, сформированным на торце оптического волокна и оптическим канальным волноводам. Показано, что применение основного приближения геометрической оптики – предположение о бесконечной малости длины волны – неприемлемо для решения поставленной в работе задачи. Расчеты показывают неоднозначность влияния формы и взаимного расположения микрооптических элементов на картину распределения интенсивности проходящего света и световую эффективность ОМ.

Представленные результаты имеют не только важное практическое значение, но и могут быть использованы для дальнейшего развития теории дифракции света в оптических системах, что позволит укрепить международное сотрудничество в области создания оптических интегральных схем. Создание адекватных компьютерных моделей микрооптических узлов может также способствовать более быстрому развитию фотонных технологий, что в настоящее время очень важно.

## Литература

1. Kogelnik H. On the propagation of Gaussian beams of light through lenslike media including those with a loss and gain variation. // *Appl. Opt.*, Vol. 4, 1965, pp. 1562.
2. Arnaud J.A. Hamiltonian theory of beam mode propagation. In: *Progress in Optics* (ed. E.Wolf), North Holland, Amsterdam, 1973.
3. SPIE's OE Magazine, Sept., 2002.
4. Joseph W. Goodman Introduction to Fourier Optics, McGraw-Hill, 1996, p. 441.
5. Voznesensky N.B. Simulation model for light propagation through nanometer-sized structures. // *Optical Memory and Neural Networks*, Vol. 9, No.3, 2000, pp.175-183.
6. S. Todoroki I, A. Nukui and S. Inoue. Formation of optical coupling structure between silica glass waveguide and molten tellurite glass proplet. // *Proceedings of SPIE*, 2002.
7. Thomas Ammer, Michael T. Gale and Marcus Rossl. Chip-level integrated diffractive optical microlenses for multimode vertical-cavity surface-emitting laser to fiber coupling. // *Journal of Optical Engineering*, Vol. 14, No. 12, December 2002, pp. 3141-3150 .
8. V.N. Voznesenskaya, V.P. Veiko, A. Petrov, N.B. Voznesensky, V.F. Pashin, S.M. Metev, C. Wochnovski. Optical interconnections optimization based on a classical approach. // *Proceedings of SPIE*, Vol. 4977, March 2003.

# ЭФФЕКТ ФАРАДЕЯ В ВОЛОКОННЫХ СВЕТОВОДАХ

М. С. Ерофеева

Научный руководитель – д.т.н., проф. А.Л. Дмитриев

Описана экспериментальная установка для исследования фарадеевского вращения плоскости поляризации оптического излучения в двулучепреломляющем волоконном световоде. Угол вращения плоскости поляризации измеряется посредством поляризационного балансного фотодетектора. Установлена нелинейная зависимость азимута плоскости поляризации света на выходе оптического волокна от величины и направления магнитного поля. Предложено объяснение наблюдаемой нелинейности.

## Введение

В настоящее время известно сравнительно мало работ, посвящённых анализу магнитооптического эффекта Фарадея в одно- и многомодовых волоконных световодах. Краткий обзор ранних исследований приведен в [1]. Недавно в [2] исследовался поворот спекл-картины в маломодовом оптическом волокне на длине волны 0,633 мкм в продольном магнитном поле. В [1, 3, 4] рассмотрены возможности использования эффекта Фарадея в волоконных световодах при разработках оптических датчиков электрического тока и измерителей сильных магнитных полей. Экспериментальные и теоретические работы по исследованиям эффекта Фарадея в протяжённых (длиной сотни метров) маломодовых волоконных световодах при относительно небольших напряженностях магнитного поля практически отсутствуют. Между тем эффект Фарадея в оптическом волокне представляет большой научный и практический интерес, например, для разработок оптических датчиков магнитных полей, а также волоконно-оптических гироскопов, при определении условий достижения их предельной чувствительности.

## Эффект Фарадея

Эффект Фарадея состоит во вращении плоскости поляризации линейно-поляризованного света, распространяющегося в веществе, вдоль вектора напряженности магнитного поля (продольный эффект Фарадея). Согласно феноменологической модели, линейно-поляризованный свет представляется когерентной суперпозицией пучков света с циркулярной поляризацией разного знака вращения. Под действием магнитного поля показатели преломления ( $n_+$  и  $n_-$ ) для циркулярно право- и лево-поляризованного света становятся различными. Вследствие этого при прохождении через среду (вдоль магнитного поля) линейно поляризованного излучения его циркулярно лево- и право-поляризованные составляющие распространяются с разными фазовыми скоростями, приобретая разность хода, линейно зависящую от оптической длины пути. В результате плоскость поляризации линейно поляризованного монохроматического света с длиной волны  $\lambda$ , прошедшего в среде путь  $L$ , поворачивается на некоторый угол  $\varphi$ . Величина этого угла равна [5]

$$\varphi = VHL, \quad (1)$$

где  $H$  – напряженность магнитного поля,  $L$  – длина образца,  $V$  – постоянная Верде.

В волоконных световодах наблюдение эффекта Фарадея упрощается благодаря значительной (до нескольких сотен – тысяч метров) оптической длине пути света в волокне и, вместе с тем, сопряжено с трудностями вследствие необходимости учета модового характера оптического излучения.

Так как число мод  $N$  волоконного световода связано с длиной волны света  $\lambda$  соотношением [6]

$$N = \frac{1}{2}V^2 = \frac{2\pi a}{\lambda} \sin \theta_{\max}, \quad (2)$$

где  $V$  – нормированная частота,  $a$  – радиус сердцевины,  $\theta_{max}$  – числовая апертура волокна, то наблюдение эффекта Фарадея предпочтительно вести в ИК-диапазоне длин волн, что и было реализовано в настоящей работе. Для сравнения влияния модового состава излучения на результаты измерения было проведено исследование эффекта Фарадея в видимом диапазоне.

## Эксперимент

В настоящей работе экспериментально исследовался эффект Фарадея в многомодовом двулучепреломляющем оптическом волокне длиной 800 м на двух длинах волн – 0,63 мкм и 0,85 мкм. Схема экспериментальной установки при исследованиях на длине волны 0,85 мкм показана на рис. 1.

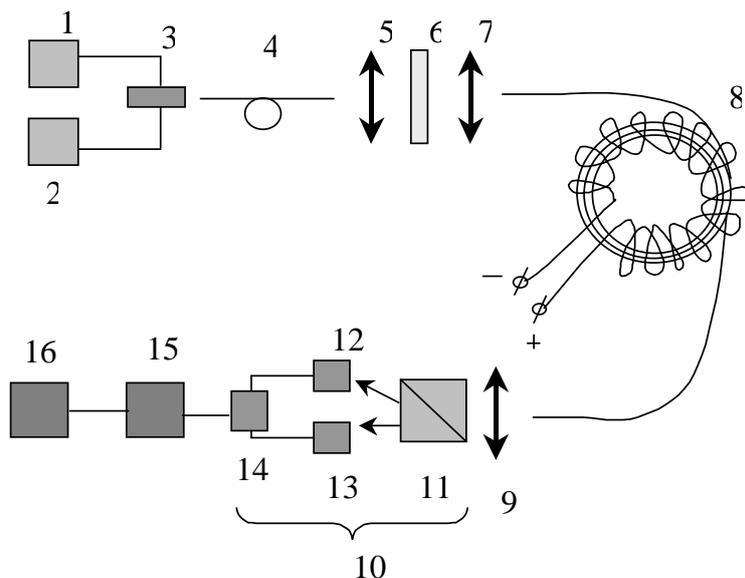


Рис. 1. Схема эксперимента для исследования эффекта Фарадея на длине волны 0,85 мкм. 1 – генератор импульсов Г5-54, 2 – блок питания лазера, 3 – полупроводниковый лазер, 4 – волоконный кабель, 5, 7, 9 – микрообъективы, 6 – линейный поляризатор (призма Глана), 8 – катушка оптического волокна с тороидальной электрической обмоткой, 10 – балансный фотосмеситель (включает расщепитель поляризации 11 и полупроводниковые фотоприемники 12, 13 марки ФД9-К), 14 – сопротивление нагрузки, 15 – селективный усилитель У2-8, 16 – осциллограф С1-49.

Амплитудно-модулированное излучение инжекционного лазера (полупроводниковый модуль типа МПД-1-1А, длина волны 0,85 мкм, мощность излучения 1 мВт, частота модуляции – 20 кГц) коллимируется, проходит через линейный поляризатор 6 и посредством микрообъектива 7 вводится в исследуемое оптическое волокно 8. Входной и выходной торцы волоконных световодов (ВС) закреплены на трехкоординатных микропозиционерах с вращающимися вокруг оси волокна держателями, что позволяет устанавливать желаемый азимут поляризации вводимого оптического излучения относительно осей большой и малой скорости двулучепреломляющего световода. Внешний и внутренний диаметр катушки оптического волокна равны соответственно 180 и 90 мм, высота катушки 45 мм. Диаметр сердцевины волокна 9 мкм, диаметр оболочки ВС 125 мкм. Тороидальная электрическая катушка для создания магнитного поля, ориентированного вдоль оси световодов, содержит 2000 витков медного провода диаметром 0,35 мм, сопротивление обмотки 85 Ом. Выходное оптическое излучение коллимируется и направляется на входную грань расщепителя поляризации 11 (призмы Рошона), разделяющего падающий луч света на компоненты с ортогональными ( $0^\circ$  и  $90^\circ$ ) плоскостями линейной поляризации. В балансном фотоприемнике оптические сигналы детектиру-

ются и регистрируемые электрические сигналы вычитаются на сопротивлении нагрузки 14. При постоянной мощности линейно-поляризованного света, падающего на призму Рошона, выходной сигнал поляризационного балансного фотодетектора (ПБФ) прямо пропорционален величине угла ориентации плоскости поляризации падающего света относительно поперечной ( $45^\circ$ ) оси пропускания призмы Рошона 11. (Принцип действия ПБФ с предельной чувствительностью описан в [7]). Выходной сигнал фотодетектора на частоте следования импульсов 20 кГц усиливается избирательным усилителем 15 типа У2-8 и контролируется на экране осциллографа 16. Уровень выходного сигнала ПБФ определяется непосредственно по шкале индикатора прибора У2-8.

Калибровка выходного сигнала (вольт/градус) производилась при отсчетах углов вращения призмы 11 по лимбу угломерного устройства, связанного с призмой. При этом на балансный фотодетектор направлялся линейно-поляризованный свет. Результирующая погрешность отсчетов изменений углов азимута поляризации света по шкале индикатора У2-8 обычно не превышала  $0,3^\circ$ .

Типичная экспериментальная зависимость изменения азимута поляризации излучения на выходе ВС от величины и направления магнитного поля в волокне приведена на рис. 2.

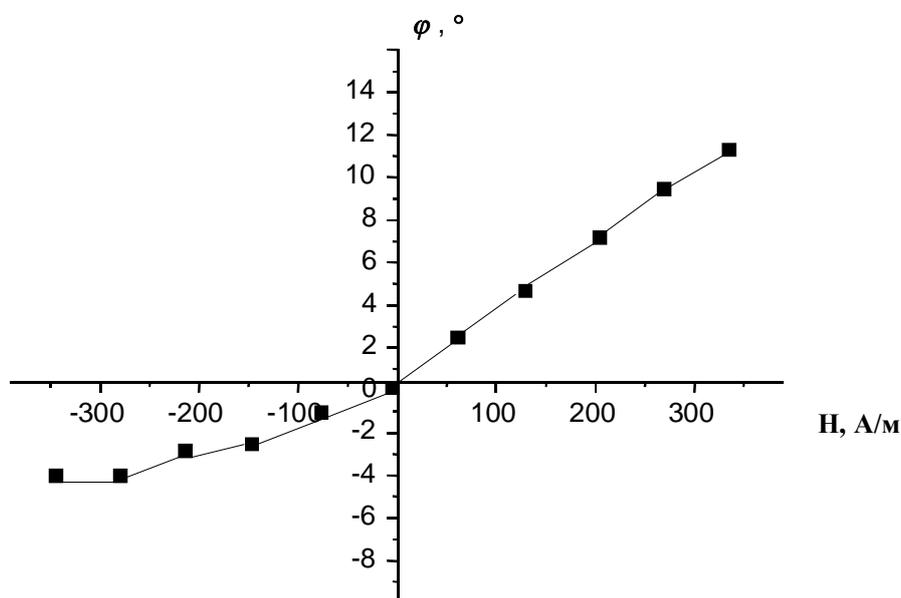


Рис. 2. Экспериментальная зависимость величины угла  $\varphi$  поворота плоскости поляризации от напряженности  $H$  магнитного поля в волокне ( $\lambda = 0,85 \text{ мкм}$ ).

### Обсуждение результатов

Из рис. 2 видно, что в области положительных значений магнитного поля  $H$  выполняется почти линейная зависимость  $\varphi \sim H$ , однако при отрицательных значениях  $H$  величина фарадеевского угла вращения существенно уменьшается. В настоящее время причина замеченной нелинейности не вполне ясна. Например, нелинейную зависимость  $\varphi$  от  $H$  можно объяснить особенностями действия балансного фотодетектора при регистрации эллиптически-поляризованного света. Выходной сигнал ПБФ прямо пропорционален мощности полностью линейно-поляризованной составляющей падающего пучка света. Если же на ПБФ падает эллиптически-поляризованное излучение, его можно представить суперпозицией линейно-поляризованного и циркулярно-поляризованного световых пучков. Отношение  $\gamma$  мощности линейно поляризованной компоненты к полной мощности света равно

$$\gamma = (1 - e^2)/(1 + e^2), \quad (3)$$

где  $e$  – эллиптичность (отношение малой и большой полуосей эллипса поляризации). При увеличении эллиптичности величина  $U$  уменьшается, что и приводит к уменьшению амплитуды регистрируемого ПБФ электрического сигнала. Циркулярно-поляризованная компонента излучения влияет лишь на уровень дробовых шумов фотодетектора. Кажущееся уменьшение угла фарадеевского вращения в области отрицательных значений магнитного поля на рис. 2 может быть следствием указанного увеличения эллиптичности регистрируемого света при отрицательных значениях магнитного поля.

Необходимо принимать во внимание, вообще говоря, сложный характер векторной структуры электромагнитного поля в сечении двулучепреломляющего маломодового волоконного световода, которое лишь в грубом приближении может быть представлено суперпозицией неполяризованной и полностью эллиптически-поляризованной компонент. Существенное влияние на преобразование поляризации света в волоконном световоде оказывают его изгибы и скрутки. Изгиб световода вследствие явления фотоупругости приводит к двулучепреломлению в волокне, при этом у кварцевых световодов величина двулучепреломления  $\Delta n = n_x - n_y$  равна  $\Delta n \approx -0,133(r/R)^2$  [1], где  $r$  – радиус световода,  $R$  – радиус изгиба волокна. В условиях описываемого эксперимента (радиус сердцевинки световода около 5 мкм, радиус катушки 50 мм) величина двулучепреломления, вызванная изгибом волокна, равна  $1,33 \cdot 10^{-9}$ . При длине световода 800 м разность фаз ортогонально поляризованных световых волн, обусловленная наведённым двулучепреломлением, на длине волны 0,85 мкм равна около 7,4 рад. Полная картина преобразования поляризации излучения в изогнутом двулучепреломляющем световоде под влиянием эффекта Фарадея носит сложный характер даже в одномодовом режиме распространения света. Эта картина существенно усложняется при многомодовом режиме распространения излучения в волокне.

При измерениях эффекта Фарадея в многомодовом волокне на длине волны 0,63 мкм в установке на рис. 1 вместо элементов 1-4 использован гелий-неоновый лазер. Амплитудная модуляция излучения лазера осуществлялась с помощью электрооптического модулятора типа МЛ-102.

Результаты измерений показаны на рис. 3.

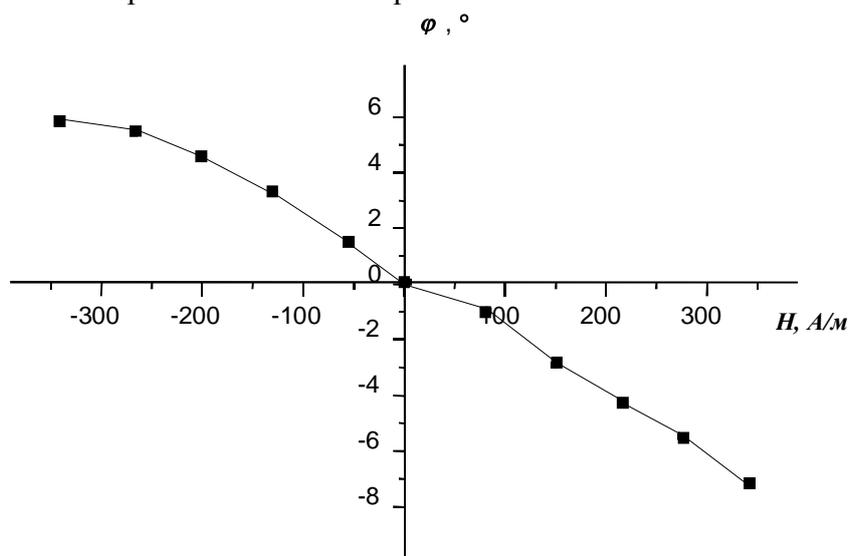


Рис. 3. Экспериментальная зависимость величины угла  $\phi$  поворота плоскости поляризации от напряженности  $H$  магнитного поля в волокне ( $\lambda = 0,63$  мкм).

При проведении измерений на длине волны 0,63 мкм оптическое излучение вводилось в волокно со стороны выходного конца катушки волокна, показанной на рис. 1, чем и объясняется изменение знака угла вращения. Сравнение рис. 2 и рис. 3 указыва-

ет, что вопреки ожиданиям величина угла фарадеевского вращения на длине волны 0,63 мкм меньше, чем на длине волны 0,85 мкм, а нелинейность зависимости угол вращения - напряженность магнитного поля в коротковолновой области менее выражена.

Рассчитаем угол поворота плоскости поляризации для длины волны 0,63 мкм по формуле (1), принимая, что постоянная Верде для плавленого кварца на длине волны  $\lambda = 0,63\text{ мкм}$  равна  $V \approx 2,7 \cdot 10^{-4}$  град/А [2]. С учетом спектральной зависимости ее значения от длины волны, получим, что при  $\lambda = 0,85\text{ мкм}$  постоянная Верде примерно в 1,35 раза меньше, чем при  $\lambda = 0,63\text{ мкм}$ . Согласно (1), при напряженности магнитного поля 300 А/м величина угла поворота плоскости поляризации излучения, прошедшего через оптическое волокно длиной 800 м, равна 48°. Экспериментальное значение угла вращения на длине волны 0,85 мкм равно 14°, а на длине волны 0,63 мкм - 10°. Сравнительно большое расхождение расчетных и экспериментальных данных, по-видимому, объясняется неточностями оценки величины постоянной Верде, а также приближенной оценкой величины магнитного поля, создаваемого электрической тороидальной катушкой. При расчете угла вращения в формулу (1) подставлялось максимальное значение напряженности магнитного поля  $H$  (значение напряженности в центре контура), в то время как точное рассмотрение этой задачи подразумевает неравномерное распределение напряженности магнитного поля в разных точках. На разные участки волоконного световода, в зависимости от их расположения относительно электрической обмотки, действуют разные величины магнитного поля и, следовательно, в этих участках катушки волокна различны и величины углов вращения плоскости поляризации.

Следует также отметить, что на результаты измерений угла вращения плоскости поляризации излучения на выходе оптоволоконного кабеля могут влиять и механические напряжения в катушке волокна, величина которых изменяется при деформациях соленоида, вызванных магнитным взаимодействием токов.

Более полное объяснение наблюдаемых зависимостей составит предмет последующих экспериментальных и теоретических исследований.

## Заключение

В настоящей статье описана экспериментальная установка для исследования фарадеевского вращения плоскости поляризации оптического излучения в диапазонах длин волн 0,63 мкм и 0,85 мкм в двулучепреломляющем маломодовом волоконном световоде. Угол вращения плоскости поляризации был измерен посредством поляризационного балансного фотодетектора. При оценке угла вращения плоскости поляризации была установлена нелинейная зависимость азимута плоскости поляризации света на выходе оптического волокна от величины и направления магнитного поля. В настоящее время причина замеченной нелинейности не вполне ясна и является предметом более тщательного анализа исследуемой проблемы. На сегодняшний день предложено, по крайней мере, два объяснения наблюдаемой нелинейной зависимости  $\varphi$  от  $H$ . Во-первых, причиной может быть особенность действия балансного фотодетектора при регистрации эллиптически-поляризованного света; во-вторых, влияние механических напряжений в катушке волокна, вызванных магнитным взаимодействием токов в электрической обмотке. Возможен также совокупный эффект.

Исследование эффекта Фарадея в оптическом волокне представляет большой научный и практический интерес для современных разработок оптических датчиков, волоконно-оптических гироскопов, при определении условий достижения их предельной чувствительности.

Автор благодарит А.Л. Дмитриева за руководство работой и полезные рекомендации.

## Литература

1. Гуляев Ю. В. и др. Модуляционные эффекты в волоконных световодах и их применение. М.: Радио и связь, 1991. 152 с.
2. Ардашева Л. И. и др. Поворот спекл-картины в маломодовом оптическом волокне в продольном магнитном поле // Оптический журнал, 2002, т. 69, № 7.
3. Papp A. and Harms H. Magnitooptical current transformer 1: Principles// Appl. Opt. 1980. Vol. 19, N 22. – pp. 3729-3734.
4. Smith A. M. Polarization and magneto optic properties of single-mode optical fiber // Appl. Opt. 1978. Vol. 17. Pp. 2301-2307
5. Калитеевский Н. И. Волновая оптика. М.: Наука, 1971. 376 с.
6. Снайдер А., Лав Дж. Теория оптических волноводов. М.: Радио и связь, 1987. 656 с.
7. Александров Е. Б., Запаский В. С. Миллисекундная чувствительность поляриметрических измерений // Оптика и спектроскопия. 1976. Т. 41. Вып. 5. С. 855–858.

# РАСЧЕТ ФУНКЦИИ ПРОПУСКАНИЯ И ЧУВСТВИТЕЛЬНОСТИ ВОЛОКОННООПТИЧЕСКОГО ДАТЧИКА С МОДУЛИРУЮЩИМ ОТРАЖАТЕЛЬНЫМ ЭЛЕМЕНТОМ

А.О. Вознесенская

Научный руководитель – д.т.н., проф. И.К. Мешковский

Представлены расчетные зависимости функции пропускания амплитудного волоконнооптического датчика (ВОД) отражательного типа от расстояния  $z$  между выходным торцом передающего оптического волокна (ОВ) и модулирующим отражательным элементом (МОЭ). Показано, что чувствительность функции пропускания ВОД к изменению расстояния  $z$  повышается при уменьшении диаметра сердцевины ОВ и уменьшении радиуса кривизны профиля МОЭ.

## Введение

Развитие волоконнооптических систем обусловило появление целого класса функциональных устройств – волоконнооптических датчиков (ВОД) физических величин [1–3]. Такие датчики обладают рядом преимуществ – электрической пассивностью, термической стабильностью, коррозионной стойкостью, устойчивостью к электромагнитным помехам и радиационным воздействиям, компактностью и многофункциональностью.

Функция преобразования в ВОД представляет собой сложную многоступенчатую зависимость тока фотоприемника  $I_{\Phi\Pi}$  от внешнего воздействия  $F_{вх}$  [2] (рис. 1):

$$I_{\Phi\Pi} = P_{ИИ} (I_{ИИ}) f_{\Phi\Pi} \{y_i [x_i (F_{изм} (F_{вх}))]\} S_{\Phi\Pi} k_{\Pi}, \quad (1)$$

где  $P_{ИИ} (I_{ИИ})$  – мощность оптического излучения, создаваемого источником излучения (ИИ) при протекании по нему тока  $I_{ИИ}$ , Вт;  $f_{\Phi\Pi}$  – функция фотоприемного преобразования;  $S_{\Phi\Pi}$  – интегральная чувствительность фотоприемника, мкА/Вт;  $k_{\Pi}$  – коэффициент потерь мощности оптического излучения при  $F_{вх} = 0$ .

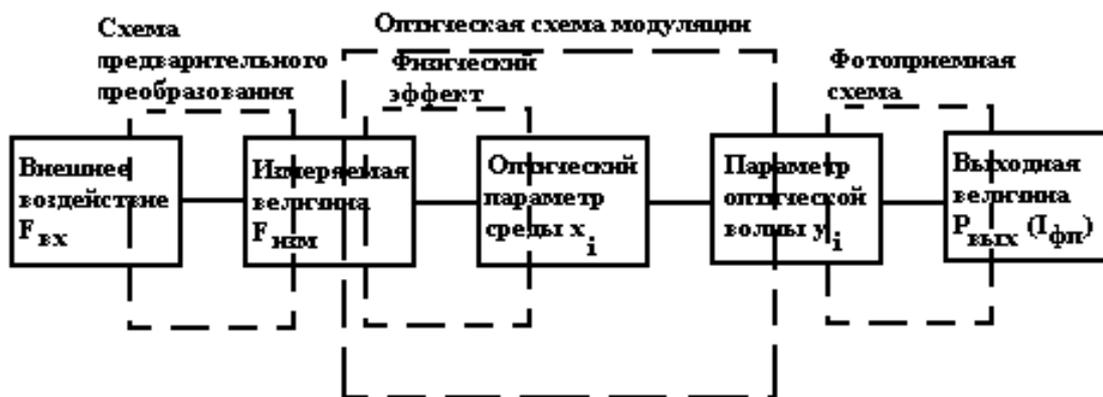


Рис. 1. Схема преобразования ВОД

Наиболее распространенными являются амплитудные ВОД, принцип действия которых основан на модуляции интенсивности света в зоне измерений [4]. Преимуществами амплитудных ВОД являются относительная простота принципа действия и изготовления, а также широкий диапазон регистрируемых физических величин – акустическое давление, температура, смещение, ускорение и др.

В амплитудных ВОД внешнее воздействие приводит к изменению функции пропускания  $T = I_i / I_0$ , где  $I_0$ ,  $I_i$  – суммарные интенсивности света на входе и на выходе ВОД, соответственно.

Основной характеристикой амплитудного ВОД является чувствительность функции пропускания к единичному изменению внешнего воздействия  $\partial T_j / \partial F_{вх}$ , определяемая через промежуточные коэффициенты  $\partial x_i / \partial F_{вх}$ ,  $\partial y_i / \partial x_i$ ,  $\partial T_j / \partial y_i$ :

$$\frac{\partial T_j}{\partial F_{вх}} = \frac{\partial x_i}{\partial F_{вх}} \frac{\partial y_i}{\partial x_i} \frac{\partial T_j}{\partial y_i}.$$

### Расчет функции пропускания ВОД

В настоящей работе рассмотрен амплитудный ВОД отражательного типа, в котором модуляция интенсивности света происходит благодаря изменению положения модулирующего отражательного элемента (МОЭ)  $z$ , находящегося на пути расходящегося светового пучка в продольном направлении, под влиянием изменения входной физической величины  $F_{вх}$  (см. рис. 2).

В этом случае функция пропускания  $T$  определяется отношением площадей излучаемого и отраженного световых пятен в плоскости выходного торца передающего оптического волокна (ОВ) [3]:  $S_{изл} = \pi a^2$ , где  $a$  – радиус сердцевины волокна;  $S_{отраж} = \pi r^2$ , где  $r$  – радиус отраженного светового пятна в плоскости выходного торца волокна:  $T = \frac{S_{изл}}{S_{отраж}} = \left(\frac{a}{r}\right)^2$ .

Посредством геометрического анализа схемы ВОД было получено выражение для расчета радиуса отраженного светового пятна  $r$ , и функция пропускания имеет вид:

$$T = \frac{a^2}{\left\{ a + z \left[ Na + \sin(\arcsin(Na)) + 2 \arcsin\left(\frac{zNa + a}{R}\right) \right] \right\}^2}, \quad (2)$$

где  $z$  – расстояние между выходным торцом волокна и отражающей поверхностью;  $Na$  – числовая апертура ОВ,  $Na = \sqrt{n_1^2 - n_2^2}$ , где  $n_1$  и  $n_2$  – показатели преломления сердцевины и оболочки ОВ соответственно;  $R$  – радиус кривизны отражателя.

Однако действительная интенсивность света на выходе ВОД будет меньше, чем расчетное значение, получаемое по приведенной формуле (2). Это обусловлено особенностями распространения оптического излучения в ОВ, которое поддерживает существование мод только в пределах своей числовой апертуры  $Na$ , а остальные моды являются вытекающими и не достигают фотоприемника. Поэтому суммарная интенсивность вернувшегося на фотоприемник света  $I_{верн}$  оказывается меньше суммарной интенсивности отраженного света в плоскости выходного торца ОВ  $I_{отраж}$ .

Для расчета действительной функции пропускания ВОД была создана математическая модель. При моделировании область выходного торца ОВ была разбита на  $M$  равных элементарных площадок, излучающих световую энергию внутри конуса, определяемого числовой апертурой ОВ. Внутри каждого светового конуса было сформировано  $N$  лучей, отличающихся направлением. Таким образом, при расчете была организована выборка из  $M \times N$  лучей [5], несущих равную световую энергию  $j_{oi}$  – поскольку в схеме ВОД рассматривалось многомодовое ОВ и предполагалось, что распределение

интенсивности света на выходном торце ОВ – равномерное. Суммарная интенсивность света на входе ВОД определялась выражением  $I_0 = \sum_{i=0}^{M \times N} j_{oi}$ .

Функции пропускания ВОД  $T_{отраж}$  и  $T_{верн}$ , соответствующие суммарной интенсивности отраженного света в плоскости выходного торца ОВ  $I_{отраж}$  и суммарной интенсивности вернувшегося на фотоприемник света  $I_{верн}$ , при моделировании имели вид

$$T_{отраж} = \frac{I_{отраж}}{I_0} = \frac{\sum_{i=0}^{M \times N} j_{i(1)}}{\sum_{i=0}^{M \times N} j_{0i}}, \quad (3.a)$$

$$T_{верн} = \frac{I_{верн}}{I_0} = \frac{\sum_{i=0}^{M \times N} j_{i(1,2)}}{\sum_{i=0}^{M \times N} j_{0i}}, \quad (3.б)$$

где  $j_{i(1)}$  – световая энергия, переносимая единичным отраженным лучом при соблюдении условия попадания отраженных лучей в область сердцевины ОВ,

$$(x^2 + y^2) \leq (2a)^2. \quad (4)$$

Здесь  $(x, y)$  – координаты проекции луча в плоскости выходного торца ОВ;  $j_{i(1,2)}$  – световая энергия, переносимая единичным отраженным лучом при соблюдении двух условий: 1) условия (4), 2) условия волноводного распространения оптического излучения:

$$0 \leq \sin \Theta \leq Na, \quad (5)$$

$\Theta$  – угол падения отраженного луча на выходной торец ОВ.

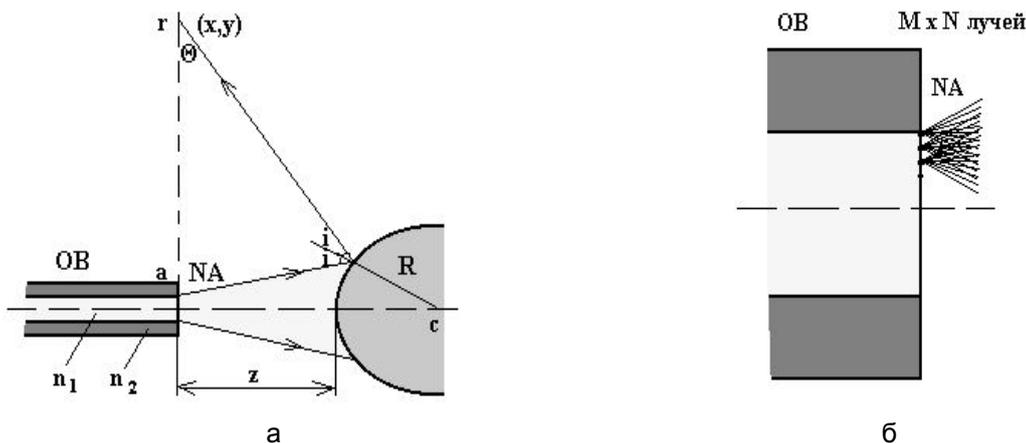


Рис. 2. Пояснение к расчету функции пропускания ВОД: а – схема амплитудного ВОД отражательного типа; б – распределение оптического излучения в плоскости выходного торца ОВ

### Результаты расчета функции пропускания ВОД

Созданная математическая модель была реализована в среде программирования Borland C++. Расчет был проведен как для случая, когда МОЭ имеет сферический про-

филь (радиус 750 мкм), так и для случая, когда он плоский. Также были рассмотрены ОВ с различными диаметрами сердцевины – 50 и 200 мкм.

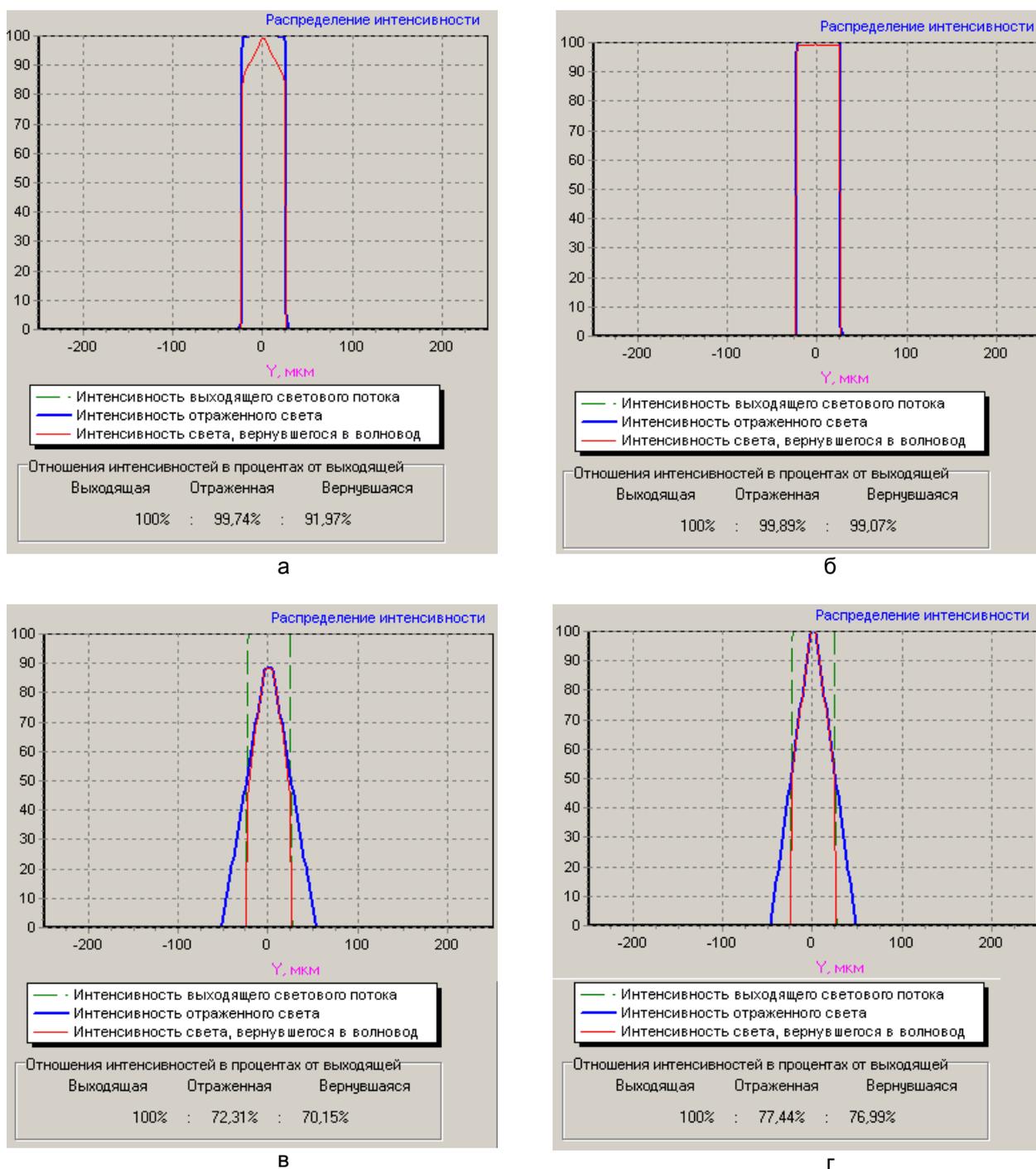


Рис. 3. Результаты компьютерного моделирования: а –  $R=750$  мкм,  $z=0$  мкм,  $2a=50$  мкм; б –  $R=\infty$ ,  $z=0$  мкм,  $2a=50$  мкм; в –  $R=750$  мкм,  $z=50$  мкм,  $2a=50$  мкм; г –  $R=\infty$ ,  $z=50$  мкм,  $2a=50$  мкм

На рис. 3 приведены результаты компьютерного моделирования для случая, когда диаметр сердцевины ОВ равен 50 мкм. Очевидно, что даже когда МОЭ находится вплотную к выходному торцу ОВ ( $z=0$ ), суммарная интенсивность света на выходе ВОД  $I_{\text{верн}}$  уменьшается при уменьшении радиуса кривизны МОЭ (рис. 3.а, б) и всегда меньше суммарной интенсивности на входе ВОД  $I_0$ . Кроме того, при увеличении рас-

стояния  $z$  наибольшее изменение функции пропускания наблюдается, когда МОЭ имеет сферический профиль (рис. 3.в, г).

На рис. 4 приведены результаты расчета функции пропускания ВОД для случаев, когда ОВ имеет диаметры сердцевины 50 и 200 мкм и числовую апертуру  $Na=0,22$ .

Очевидно, что при расстояниях между выходным торцом ОВ  $z < 2,5 \cdot (2a)$  суммарная интенсивность вернувшегося на фотоприемник света  $I_{\text{верн}}$  меньше суммарной интенсивности отраженного света в плоскости выходного торца ОВ  $I_{\text{отраж}}$ ; при увеличении расстояния  $z$  наблюдается выравнивание их значений  $I_{\text{верн}} = I_{\text{отраж}}$ .

Сопоставляя полученные расчетные зависимости функции пропускания ВОД для ОВ с различными диаметрами сердцевины, можно сделать вывод, что при увеличении диаметра сердцевины ОВ интервал линейности функции пропускания увеличивается. Однако при этом динамический диапазон ВОД значительно уменьшается. Например, для ОВ с диаметром сердцевины  $2a=50$  мкм на интервале  $z = 0-600$  мкм изменение выходной мощности  $dA \approx 13$  дБ, тогда как для ОВ с диаметром сердцевины  $2a = 200$  мкм  $dA \approx 5$  дБ. Кроме того, для ОВ с диаметром сердцевины  $2a = 200$  мкм отмечено, что как для случая сферического ( $R = 750$  мкм), так и для случая плоского ( $R = \infty$ ) профиля поверхности МОЭ динамический диапазон ВОД остается постоянным. Однако при малом диаметре сердцевины ОВ  $2a=50$  мкм сферичность профиля поверхности МОЭ вносит увеличение изменения выходной мощности ВОД.

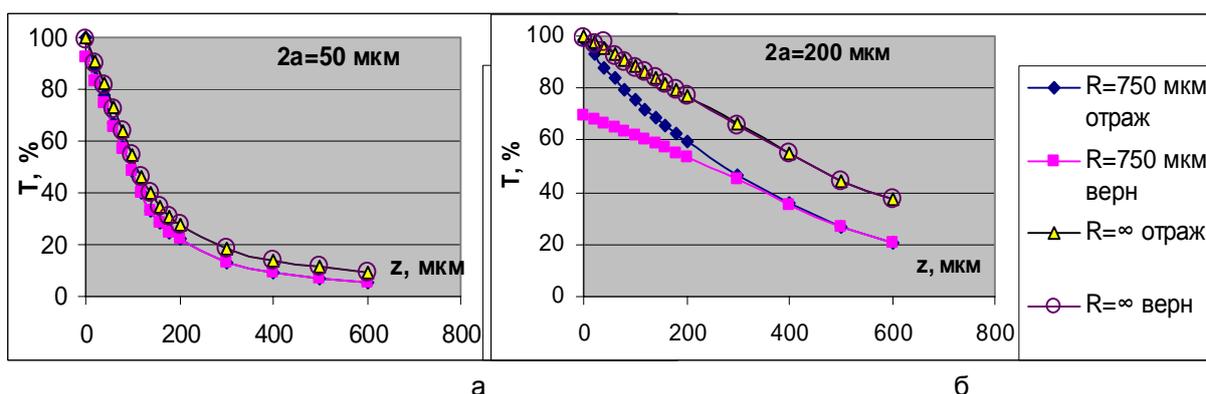


Рис. 3. Функция пропускания ВОД: а –  $2a = 50$  мкм; б –  $2a = 200$  мкм

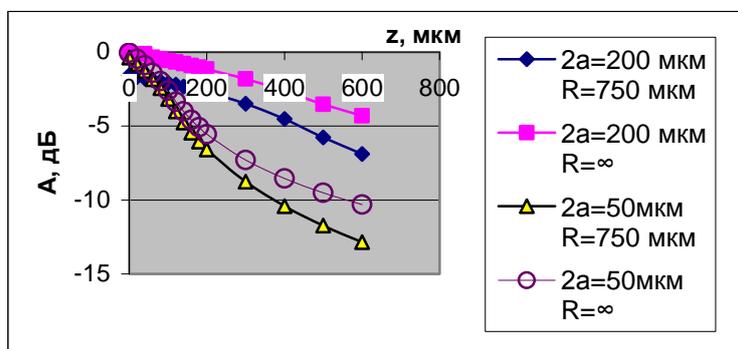


Рис. 4. Динамический диапазон ВОД

## Заключение

В результате компьютерного моделирования преобразования оптического излучения в амплитудном ВОД отражательного типа были получены зависимости функции

пропускания ВОД от расстояния  $z$  между выходным торцом ОБ и МОЭ для двух значений диаметра сердцевины ОБ 50 и 200 мкм и различных профилей поверхности МОЭ. Установлено, что наибольшая чувствительность функции пропускания ВОД к изменению расстояния  $z$  соответствует случаю, когда диаметр сердцевины ОБ минимален и МОЭ имеет сферический профиль.

## Литература

1. Волоконно-оптические датчики/ Т. Окоси, К. Окамото, М. Оцу, Х. Нисихара, К. Кюма, К. Хататэ; Под ред. Т. Окоси: Пер. с япон. Л.: Энергоатомиздат. Ленинград. отд-ние, 1990.
2. Бусурин В.И., Носов Ю.Р. Волоконно-оптические датчики: Физические основы, вопросы расчета и применения. М.: Энергоатомиздат, 1990.
3. Волоконная оптика и приборостроение/ М.М. Бугусов, С.Л. Галкин, С.П. Оробинский, Б.П. Пал; Под общ. Ред. М.М. Бугусова. Л. : Машиностроение. Ленинград. отд-ние, 1987.
4. Е.А. Бадеева, В.А. Мещеряков, Т.И. Мурашкина. Классификация амплитудных волоконно-оптических преобразователей // Датчики и системы, №2, 2003. – с.20 – 25.
5. Родионов С.А. Автоматизация проектирования оптических систем: Учеб. пособие для приборостроительных вузов. Л.: Машиностроение, Ленинград. отд-ние, 1982.

# ОСНОВНЫЕ МЕТОДЫ ИЗГОТОВЛЕНИЯ ПРЕФОРМ ДЛЯ ВЫТЯЖКИ ОПТИЧЕСКОГО ВОЛОКНА

В.Н. Фролков, Ю.И. Рядчин

Научный руководитель – д.т.н., проф. Ю.А. Гатчин

Рассмотрены основные методы изготовления заготовок для вытягивания оптического волокна. Проведен сравнительный анализ MCVD, PCVD, OVD методов.

## Введение

Применение оптических волокон в последнее время привело к революционным изменениям в отрасли связи. Сегодня оптическое волокно, обладая огромной пропускной способностью, является фактически единственной средой для передачи больших объемов информации. Эволюция методов и технологии производства оптических волокон не столь заметна, но современные заводы, выпускающие оптические волокна, имеют огромные преимущества по производительности перед своими предшественниками, созданными в 80-х годах. Такое совершенствование процессов изготовления оптических волокон позволило существенно снизить цены на него. Но за эти два десятилетия с начала промышленного производства оптических волокон преимущественная технология их изготовления окончательно еще не определилась. На рынке предлагается волокно, изготовленное четырьмя отличными друг от друга процессами – OVD, VAD, MCVD и PCVD.

Современное производство оптических волокон обычно разбивают на несколько этапов. Первый этап называется процессом производства заготовки (преформы), которая представляет собой стеклянный стержень длиной порядка метра и диаметром около 10-40 мм. Из этой преформы в дальнейшем методом вытяжки производится непосредственно само волокно. В одном процессе, относящемся к многокомпонентным стеклам с низкой температурой плавления, стекло готовят в большой плавке, а волоконный световод формируют на втором, отдельном, этапе. Во втором процессе, обычно относящемся к кварцевому стеклу, используется заготовка, содержащая материал как сердцевины, так и оболочки, образованных непосредственно из газовой фазы. В дальнейшем заготовка перетягивается в волоконный световод. В третьем процессе вытягивается волоконный световод из заготовки однородного состава и покрывается пластиком, образующим отражающую оболочку [1].

Для получения типичных характеристик оптических волокон необходимо обеспечить чистоту материала не более 1 частицы на миллиард для ионов металла и десяти ионов гидроксильной группы  $OH^-$ . Практика показывает, что именно недостаточная чистота производства была основным препятствием для широкомасштабного внедрения ОВ как среды передачи информации в 70-е годы.

Существующие сегодня технологии позволяют изготавливать оптические волокна с минимальным затуханием вплоть до 0,18 дБ/км в диапазоне 1550 нм, что близко к теоретическим показателям для кварца. Поэтому современные исследования по снижению потерь в волокне сместились в область разработок в среднюю (2–5 мкм) и дальнюю (8–12 мкм) инфракрасную область (волокно, оптические приемники и излучатели). Характерные значения диаметра сердцевины одномодовых волокон лежат в пределах 4–10 мкм, многомодовых – 50–100 мкм. Диаметр оболочки обычно превышает диаметр сердцевины в 5–10 раз.

## Модифицированный метод химического парофазного осаждения (MCVD)

Наиболее распространенным методом производства заготовок является описанный в 1974 г. американским физиком Джоном Макчеснеем, работавшим в лаборатори-

ях Белла (AT&T Bell Labs), модифицированный метод химического парофазного осаждения (MCVD).

Этот метод стал одним из наиболее испытанных и проверенных способов получения заготовок [2,3]. Сам процесс довольно прост, гибок и поэтому нетруден для освоения. Простота его выполнения наиболее подходит для научных исследований в области оптических волноводов. Крупномасштабное использование этого метода осуществлено фирмой Lucent Technologies в США и многими другими изготовителями США, Японии и Европы. При использовании этого метода возможно изменение целого ряда параметров, например, вида волокна, окончательного диаметра волокна, числовой апертуры и профиля показателя преломления. Эти параметры изменяются посредством регулирования расхода паров двуокиси кремния и различных примесей [4]. Благодаря этому можно изготавливать оптическое волокно, соответствующее определенным техническим условиям.

Метод MCVD – Modified Chemical Vapour Deposition, модифицированное химическое парофазное осаждение – заключается в следующем (рис. 1).

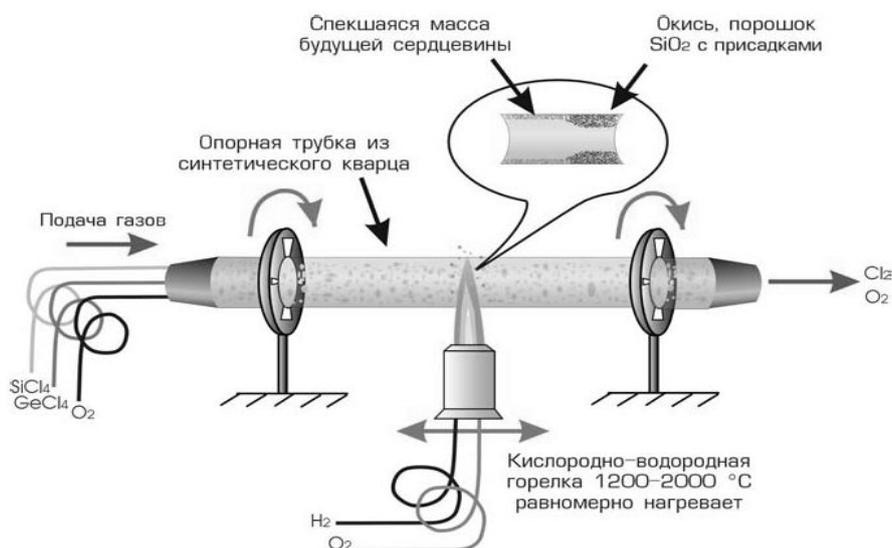


Рис. 1. Изготовление волокна. Метод MCVD

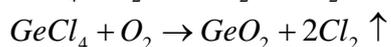
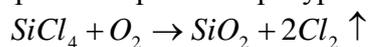
На внутренней стороне стеклянной трубки происходит осаждение сверхчистой двуокиси кремния (при изготовлении сердцевины всегда добавляются определенные примеси), а затем – усаживание трубки, вызываемое повышением температуры (с помощью пламени газа или токов сверхвысокой частоты). В результате этого трубка сжимается и превращается в сплошной стеклянный стержень диаметром 30–40 мм и длиной около 1000 мм. Таким образом, этот стеклянный стержень получает окончательный профиль, который будет иметь и готовое волокно [5].

Процесс осаждения основан на высокотемпературном окислении  $\text{SiCl}_4$  и веществ, используемых в качестве примеси. Он занимает, в зависимости от размера заготовки, от четырех до восьми часов, в течение которых происходит осаждение того стекла, которое превратится в светопроводящую сердцевину. Процесс осаждения одинаков и для одномодовых, и для многомодовых волокон.

В качестве примесей к сердцевине волокна чаще всего используется германий, способствующий увеличению показателя преломления. Окись фосфора применяется для снижения требующейся температуры процесса, а фтор – прежде всего для уменьшения показателя преломления оболочки волокна. Как правило, волокно образуется из наружной кремнеземной трубки.

Процесс начинается с того, что трубка из сверхчистого высококачественного кремнезема (двуокись кремния, синтетический кварц  $SiO_2$ ) промывается в кислотной ванне, а затем зажимается в устройстве, напоминающем токарный станок. В нем трубка может вращаться вокруг своей центральной оси. Кислородно-водородная газовая горелка движется вдоль трубки взад и вперед, сильно и равномерно нагревая трубку. Входной конец трубки через вращающуюся газонепроницаемую соединительную муфту присоединен к комплексному устройству подачи газов. В это устройство входят смеситель газов и автоматизированные регуляторы массового расхода. Крайне важно, чтобы это устройство было абсолютно герметичным. Это требуется и для того, чтобы предотвратить попадание внутрь загрязняющих веществ, и для того, чтобы обеспечить точное дозирование газов. Из резервуаров поступают химические реактивы  $SiCl_4$  и  $GeCl_4$ , которые путем нагрева переводятся из жидкой фазы в газообразную и подаются в трубку. Туда же подается кислород  $O_2$ . Из другого (выходного) конца трубки удаляется избыточный материал.

В ходе процесса точно контролируется количество химических веществ, подаваемых в трубку либо с помощью газов-носителей ( $Ar$ ,  $He$  и  $O_2$ ), либо в виде отдельного потока. Окисление  $SiCl_4$  и примесей происходит в зоне нагревания, непосредственно над горелкой при температуре 1000 – 1400°C



Окись, образующаяся в виде мелкого порошка, – это оксиды кремния и германия  $SiO_2$  и  $GeO_2$  – уносится потоком газа вдоль трубки, потом происходит ее осаждение, а пары хлора улетучиваются. Этот процесс называется термофорезом. Когда тепло от горелки достигает осажденного порошка, то он плавится и превращается в сплошной, без включения пузырьков, прозрачный слой двуокиси кремния (с примесями или без них). Когда горелка доходит до конца трубки, то она меняет направление и быстро движется вдоль трубки в обратную сторону к начальной точке, чтобы обеспечить осаждение нового слоя порошка. Во время осаждения (и сердцевины, и оболочки) осуществляется введение неодинакового количества различных примесей. Путем последовательного осаждения одного слоя двуокиси кремния за другим и тщательного дозирования примесей можно получить оптическое волокно с различными профилями показателя преломления. При этом, насколько это технически возможно, предотвращается попадание в систему загрязняющих веществ. Особенно это касается загрязняющих веществ, содержащих водород, который образует в волокне соединения  $OH^-$ , приводящие к увеличению затухания.

Основным преимуществом процесса MCVD является то, что и структуру, и свойства световода можно обеспечить еще в заготовке, а потом сохранить в готовом волокне. Относительные размеры и профиль показателя преломления заготовки передаются готовому волокну в процессе вытяжки [4].

После того как осаждение завершается, начинается следующий важный этап процесса изготовления – усаживание трубки. Этот этап состоит из нескольких шагов. Чтобы обеспечить усаживание, температуру поднимают (с помощью кислородно-водородной горелки – MCVD) до 1500–2000°C.

При такой температуре трубка постепенно размягчается и усаживается, образуя сплошной стержень – заготовку. Этот шаг является решающим для обеспечения окончательных геометрических параметров заготовки. Усаживание происходит тогда, когда кислородно-водородная горелка или токи сверхвысокой частоты воздействуют на трубку по всей ее длине. Усаживание обеспечивается вязким внутренним течением, вызываемым поверхностным натяжением, которое возрастает по мере того, как стекло становится более горячим и менее вязким.

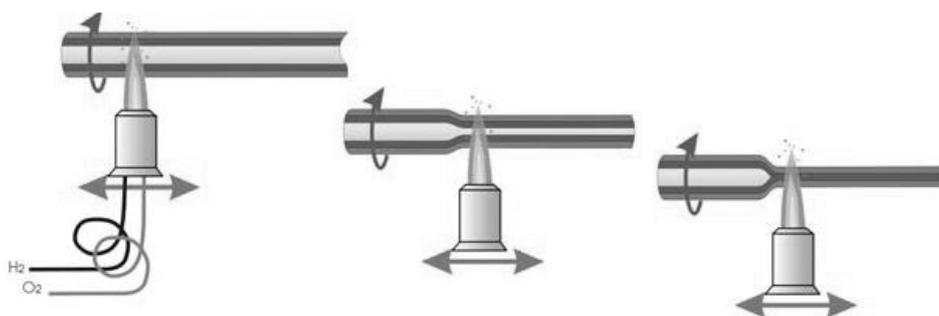


Рис. 2. Усаживание заготовки

Как уже было сказано, метод самый простой, и в отечественной промышленности только он реально и работал. Однако для него нужна труба-заготовка очень хорошего качества, без включений, так как включения – это центры напряжений, из которых может начать расти трещина. С этим довольно успешно борются путем химической или огневой полировки поверхности трубок. Недостатком метода MCVD следует считать наличие большого температурного градиента между внутренней поверхностью опорной кварцевой трубы в месте реакции и наружной поверхностью [4,6].

### Плазменный метод химического парофазного осаждения (PCVD)

Метод PCVD, т.е. Plasma-activated Chemical Vapor Deposition (плазменный метод химического парофазного осаждения), также относится к группе методов внутреннего осаждения.

Изготовление заготовок плазменным методом химического парофазного осаждения осуществляется, в основном, аналогично модифицированному методу химического парофазного осаждения. Различие заключается в технике реакции. Исходными материалами являются  $\text{SiCl}_4$ ,  $\text{O}_2$ ,  $\text{C}_2\text{F}_6$  и  $\text{GeCl}_4$ . Опорную кварцевую трубу помещают в электрическую печь, которая разогревает ее до температуры около  $1000^\circ\text{C}$ . Резонатор перемещается вдоль опорной кварцевой трубы и токами сверхвысокой частоты дополнительно производит разогрев газовой смеси до  $1500^\circ\text{C}$ . Плазму получают путем возбуждения газа с помощью, например, микроволн (сверхвысокие частоты). При этом газ ионизируется, т.е. разлагается на свои носители электрических зарядов. При воссоединении этих носителей выделяется тепловая энергия, которая может быть использована для плавления материалов с высокой температурой точки плавления. Так, при плазменном методе галогениды вступают в реакцию с помощью плазмы низкого давления (давление газа примерно 10 мбар), и в результате соединения с кислородом образуется  $\text{SiO}_2$ . С помощью системы насосов внутри кварцевой трубы создается поток с давлением 100–400 Па для удаления продуктов реакции ( $\text{Cl}_2$ , избыток кислорода). Продукты реакции в виде стеклообразного слоя без образования порошкообразных оксидов осаждаются на опорной кварцевой трубе. Поскольку плазменная плазма движется быстро вдоль трубки назад и вперед, то за счет перемещения резонатора можно изготовить более 1000 тонких слоев, благодаря чему можно повысить точность профиля показателя преломления.

Основное преимущество метода заключается в проведении всего процесса осаждения при более низкой температуре, чем при методе MCVD. Точность изготовления профиля показателя преломления повышается [4].

### Метод внешнего парофазного осаждения (OVD)

Метод внешнего парофазного осаждения (Outside Vapour Deposition – OVD) более сложен, чем метод MCVD [7]. Изобретателями данного метода являются группа ученых компании Corning Glass Work – Роберт Маурер, Дональд Кек и Петер Шульц, которым

принадлежит честь открытия оптического волокна. Поскольку химические аспекты этого метода, в основном, те же самые, что и у предыдущего, описание будет более кратким. Из-за сложности, а также ввиду всесторонней патентной защиты метод применяется исключительно разработчиком - компанией Corning Glass Work и ее лицензиатами. Однако общее количество волокна, изготовленного этим методом, на сегодняшний день не меньше, а может быть, и больше, чем волокна, изготовленного методом MCVD. Это объясняется тем, что при массовом производстве метод OVD намного эффективнее, чем MCVD.

Процесс изготовления заготовки методом OVD делится на два этапа.

На первом этапе происходит осаждение порошкообразной двуокиси кремния (часто называемой белой сажей) – с примесями или без них – на тонком стержне. Состав газа и температура пламени регулируются по времени в целях получения ступенчатого или градиентного профиля показателя преломления. Горячий поток частиц порошкообразной двуокиси кремния движется над поверхностью стержня, при этом некоторые частицы осаждаются на стержне. Сам стержень в это время вращается и одновременно движется в осевом направлении, проходя через горелку. Некоторые из частиц окажутся при этом в спекшемся состоянии. Когда произойдет осаждение такого количества стекла, которого будет достаточно для образования и сердцевинки, и оболочки, процесс остановится. Исходный стержень (обычно из  $Al_2O_3$ ), имеющий более высокий ТКЛР, чем осаждаемый на него материал, будет аккуратно извлечен из заготовки [4,6].

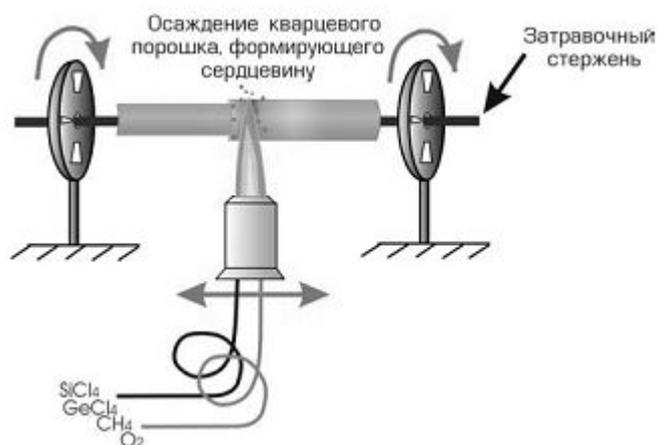


Рис. 3. Изготовление заготовки. Метод OVD.

На втором этапе – спекании – пористую стеклянную заготовку высокой чистоты, но с достаточно высоким содержанием воды и гидроксильных групп, подвергают операции дегидратирования при температуре 1200 °С в среде газообразного гелия или хлора. Затем пористую заготовку нагревают до 1400-1600 °С для осуществления процесса стеклования и схлопывания, в результате чего получают заготовку из стекла, свободного от пузырей. Весь этот процесс происходит в печи, в контролируемых условиях.

Таким образом, процесс OVD состоит из пяти основных шагов:

- очистка сырья,
- подача реакционных веществ в зону нагревания,
- химические реакции и образование частиц,
- осаждение частиц на подложке (исходный стержень или уже осажденная заготовка),
- дегидратация и спекание [4]

Существенным недостатком метода внешнего парофазного осаждения является повышенное содержание ионов OH в получаемой заготовке, что приводит к возрастанию потерь в оптическом волокне.

## Метод осевого парофазного осаждения VAD

Метод осевого парофазного осаждения (VAD) [8] был разработан в Японии для того, чтобы не нарушать патент компании Corning Glass Work на метод OVD и наладить непрерывное получение заготовок для изготовления оптического волокна. Изобретателем метода считается группа японских ученых корпораций NTT и "Сумитомо" во главе с Татсуо Изавой, который в 2001 г. за это открытие был удостоен знаменитой премии имени Тиндалла.

Метод VAD [9–12] аналогичен методу OVD в той степени, в которой осаждение является не внутренним, а наружным процессом, при котором образуется пористая заготовка, которая затем обезвоживается и спекается. Однако при использовании метода VAD кварцевый порошок осаждается в осевом направлении, а не в радиальном. Это обстоятельство затрудняет модификацию профиля показателя преломления, но облегчает получение более длинных заготовок. В принципе этот процесс может быть непрерывен - по мере изготовления заготовки из нее же можно вытягивать волокно [13].

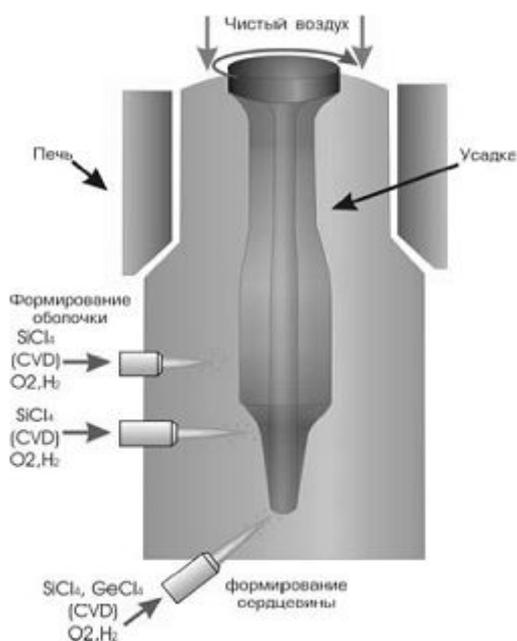


Рис. 4. Изготовление заготовки. Метод VAD

При изготовлении используются механизм осевого перемещения заготовки, реакционная камера, горелка, испаритель для сырья и блок управления. Заготовка медленно протягивается вверх, проходя при этом через используемое в данном процессе оборудование. Сырье ( $SiCl_4$ ,  $GeCl_4$  и  $POCl_3$ ) вводится так же, как и при использовании метода OVD: применяется кислородно-водородная газовая горелка, а очень мелкие частицы стекломассы, образовавшиеся в результате реакции гидролиза в пламени, осаждаются на торцевой поверхности уже осажденной заготовки, которая используется в качестве подложки для дальнейшего наращивания заготовки. Пористая заготовка движется в осевом направлении, наращивая свою массу в ходе этого движения. В кольцевой электрической печи сопротивления с графитовым нагревательным стержнем в атмосфере сульфурилхлорида  $SO_2Cl_2$  заготовка обезвоживается, а затем в кольцевой графитовой печи для зонного остекловывания превращается в прозрачный твердый стержень [4, 6]. Важными отличительными чертами процесса является следующее:

- постоянное наращивание заготовки,
- точное регулирование расхода сырья,
- поддержание постоянной температуры пламени,

- поддержание постоянной температуры в зоне нагревания заготовки,
- поддержание постоянной скорости поворота заготовки,
- сохранение зоны наращивания заготовки в определенном месте.

Перемещение зоны наращивания заготовки в другое место приводит к изменениям профиля показателя преломления готовой заготовки. Скорость перемещения в осевом направлении (т.е. скорость наращивания заготовки) составляет примерно 40–60 мм/час.

Чтобы обеспечить осаждение, используются от одной до трех горелок. Изменяя число горелок и состав и пропорции сырья, можно обеспечить получение различных профилей показателя преломления. В ходе отверждения заготовка обезвоживается в среде газообразного хлора. Более толстую оболочку волокна можно получить путем дополнительного осаждения на поверхности спеченной стержневой заготовки. Это означает, что можно получать очень большие заготовки [4].

Главная сложность этого метода заключается в контроле градиентного профиля показателя преломления из-за одновременного наращивания сердцевины и оболочки, который оказывает значительное влияние на модовый состав и импульсный отклик оптического волокна [6].

При вытяжке не происходит смешивания отдельных слоев - при этом происходит, выражаясь математическим языком, преобразование подобия. Вытяжка световодов проводится в столь же чистых помещениях, как и при производстве микросхем, чтобы на поверхность волокна не попадали пылинки – те же самые включения. После того, как волокно остынет, на него наносится защитная пленка полимера. Полимер призван защитить волокно от механических воздействий и от окружающей среды, прежде всего водорода и воды. Дело в том, что вода ускоряет рост трещин, а водород, диффундируя вглубь волокна, повышает коэффициент затухания. Однако по-настоящему могут защитить стекло только два покрытия – металлическое и углеродное. И японские фирмы сейчас начинают покрывать стекловидным углеродом каждое волокно в отдельности. Пока же, для достижения полной защиты, лучше всего помещать уже готовый кабель в герметичную металлическую оболочку, но это практически реализуемо только в наиболее дорогих кабелях для трансокеанских линий.

### **Наращивание заготовок. Внутреннее осаждение и плазменное жакетирование (APVD метод)**

Просто сказать, что волокно было изготовлено с помощью процесса MCVD или PCVD, было бы недостаточно. При производстве волокон методом MCVD или PCVD, особенно одномодовых, крайне нерациональным является использование полноразмерной трубки, т.е. трубки, из которой производится полная заготовка для оптического волокна. Все процессы получения заготовок путем осаждения из газовой фазы являются сегодня двухэтапными.

На первом этапе производятся сердечник и часть оболочки, нужная для обеспечения оптических свойств изготавливаемого волокна. Назовем ее для простоты заготовкой сердечника. Из заготовки сердечника, в зависимости от ее размера, может быть получено несколько стержней, на которые потом наращивается оболочка. Затем заготовка с наложенной оболочкой вытягивается в волокно. Различают несколько различных способов изготовления такой оболочки: жакетирование, золь-гель, плазма, кварцевый порошок.

Способ жакетирования представляет собой обжимание кремнеземной трубки вокруг заготовки сердечника; это самый популярный процесс для изготовления оболочки сердечников, полученных способами MCVD и PCVD. Однако надеваемые наружные трубки весьма дороги. Две компании – Alcatel и Lucent, изготавливающие заготовки сердечников в основном способом MCVD, недавно организовали – в качестве альтер-

нативы жакетированию – серийное производство с получением оболочки волокна путем плазменного распыления и с помощью золь-геля. Все производители сердечников волокна, пользующиеся процессами OVD и VAD (кроме одного, применяющего процесс VAD), получают оболочку волокна с помощью кварцевого порошка.

Процесс изготовления заготовок начал становиться двухэтапным (с жакетированием заготовки) после того, как в серийном производстве стал использоваться VAD. Компания Corning стала использовать в серийном производстве альтернативу жакетированию заготовки – получение оболочки с помощью кварцевого порошка. Поскольку все изготовители, прибегавшие к процессу VAD, перешли на кварцевый порошок, то происходило уменьшение доли волокна, произведенного с помощью жакетирования. Ожидается, что вследствие введения компанией Alcatel плазменного распыления и компанией Lucent – золь-гелевого способа использования способа жакетирования сердечника еще более сократится. Одной из причин этого является стоимость надеваемой наружной трубки. Экономические анализы показывают, что, для того чтобы способ жакетирования стал конкурентоспособным по отношению к другим способам получения оболочки, цена трубки должна упасть до уровня около 100 долларов за килограмм. Другой серьезной причиной являются трудности с изготовлением трубок для больших заготовок. Несколько изготовителей, использующих способ жакетирования, заявили, что они готовы выпускать заготовки большего размера, если только будут доступны соответствующие им по размеру трубки. По мере снижения цен на волокно вопрос о заготовках большого размера будет становиться все более важным и существенным для решения проблемы выживания. Это означает, что, подобно компаниям Alcatel и Lucent, и другие изготовители одномодового волокна откажутся от способа жакетирования и выберут какую-нибудь экономную технологию, которая позволит производить крупные заготовки в большом объеме. Большинство из них, вероятно, выберет способ с использованием кварцевого порошка, для которого уже имеется оборудование и разработанная технология.

## **Заключение**

Наибольшее распространение на сегодняшний день получил метод MCVD. Одними из главных его достоинств являются простота и гибкость метода, а также возможность сохранения структуры и свойств преформы в готовом волокне. Они обусловили применение этого метода не только для производства, но и для проведения научных исследований в большинстве лабораторий мира. Недостаток метода – наличие большого температурного градиента между внутренней поверхностью опорной кварцевой трубы в месте реакции и наружной поверхностью. При методе PCVD процесс осаждения происходит при более низкой температуре, чем при методе MCVD, что уменьшает величину названного выше температурного градиента. Кроме того, повышается точность изготовления профиля показателя преломления. Несмотря на большую сложность по сравнению с предыдущими методами, при массовом производстве метод OVD является более производительным. Однако патентная защита пока не дала этому методу широко распространиться. Другим его недостатком можно назвать повышение потерь в оптическом волокне из-за высокого содержания ионов OH в преформе. Существенным преимуществом метода VAD является получение преформы любой длины, так как процесс осевого осаждения может быть бесконечным. Среди недостатков можно выделить трудность получения заданного профиля показателя преломления.

Авторы статьи выражают огромную благодарность Константину Владимировичу Дукельскому и Юрию Николаевичу Кондратьеву за предоставленные материалы, возможность ознакомления с оборудованием для получения оптического волокна, а также большую помощь при написании данной работы.

## Литература

1. Мидвинтер Дж. Э. Волоконные световоды для передачи информации: пер. с англ. М.: Радио и связь, 1983. 336 с.
2. Nagel S. R., MacChesney J. B., Walker K. L., in: Optical Fiber Communications, vol. 1, ed. by T. Li, Academic, Orlando, 1985, Ch. 1.
3. Ларин Ю.Т., Рязанов И.Б. Расчет параметров оптических кабелей. М.: МЭИ, 1992. 122 с.
4. Иоргачев Д. В., Бондаренко О. В. Волоконно-оптические кабели и линии связи. М.: Эко-Трендз, 2002. 282 с.
5. Агравал Г. Нелинейная волоконная оптика: пер. с англ. М.: Мир, 1996. 323 с.
6. Гроднев И. И., Ларин Ю.Т., Теумин И. И. Оптические кабели. М., 1991.
7. Morrow A. J., Sarkar A., Schultz P. C., in: Optical Fiber Communications, vol. 1, ed. by T. Li, Academic, Orlando, 1985, Ch. 2.
8. Niizeki N., Inagaki N., Edahiro T., in: Optical Fiber Communications, vol. 1, ed. by T. Li, Academic, Orlando, 1985, Ch. 3.
9. Кацуяма, Тосё, Мацумура, Хироси, Инфракрасные волоконные световоды: пер. с англ. М., 1992.
10. Takahashi H., Sugimoto I. J. Lightwave Technol., LT-2, 613 (1984).
11. Takahashi H., Sugimoto I., Sato T. Electron. Lett., 18, 398 (1982).
12. Izawa T. Technical Digest of the Conference of Integrated Optics and Optical Fiber Communications (То-кью), CI-1, 373 (1977).
13. Вейнберг В. Б., Саттаров Д. К. Оптика световодов. М.: Машиностроение, 1977, 320 с.

# ВЫТЯГИВАНИЕ ВОЛОКОННЫХ СВЕТОВОДОВ

В.Н. Фролков, Ю.И. Рядчин

Научный руководитель – д.т.н., проф. Ю.А. Гатчин

Рассмотрены основные проблемы вытягивания оптического волокна. Проведен анализ установки вытягивания волоконных световодов для передачи информации. Предложены пути решения ряда проблем.

## Введение

За два последних десятилетия применение оптических волокон привело к революционным изменениям в отрасли связи. Обладая поистине неограниченной широкополосностью, оптическое волокно на сегодняшний день является, по сути, единственной средой для передачи больших объемов информации. Для различных применений разрабатываются и производятся различные типы волокон, отличающиеся как по конструктивным параметрам (профиль показателя преломления, геометрия), так и по материалам, из которых они изготовлены. Постоянное совершенствование систем передачи обеспечивает увеличение пропускной способности сетей. Вместе с совершенствованием систем передачи улучшаются и оптические компоненты и волокно. Эта тенденция продолжается и сегодня. Таким образом, конструкции волокон продолжают совершенствоваться почти на всех ступенях сетевой иерархии. В 1980-х гг. преобладающую роль играли многомодовые волокна, а в 1990-х гг. основным видом волокна стало одномодовое волокно, работающее на длине волны 1,3 микрона.

Получение волокна состоит из двух этапов:

Первый этап – получение заготовки, которая представляет собой стеклянный стержень (преформу) длиной порядка метра и диаметром около 10–60 мм. Для этого существует несколько способов, каждый из них имеет свои преимущества и недостатки.

Второй этап – процесс превращения толстой преформы в тонкое волокно заданного диаметра, который и называется вытяжкой.

## Процесс вытягивания волокна из преформы

Физическая сущность процесса заключается в следующем. Стекло (конец преформы) нагревается в печи до пластичного состояния, при котором вязкость составляет  $10^4$ – $10^6$  пуаз, при этом тянущее устройство вытягивает высоковязкий расплав в тонкую нить. Для осуществления непрерывности процесса объемный расход вытянутого волокна, равный  $V_{\text{выт.}} \cdot d_{\text{в.}}^2$ , где  $V_{\text{выт.}}$  – скорость вытягивания,  $d_{\text{в.}}$  – диаметр волокна, компенсируется подачей заготовки в зону высоких температур со скоростью подачи  $V_{\text{заг.}}$ , при этом строго соблюдается равенство объемных расходов  $V_{\text{выт.}} \cdot d_{\text{в.}}^2 = V_{\text{заг.}} \cdot d_{\text{заг.}}^2$ . Устройство нагрева построено таким образом, чтобы по его оси формировался четко выраженный температурный максимум, в зоне которого формируется "луковица" перетягиваемой преформы. Силы поверхностного натяжения стремятся превратить сечение волокна в круг, независимо от формы заготовки, однако, меняя температуру максимума в осевом профиле печи, скорость вытяжки и подачи, можно получать и несферическое волокно, повторяющее форму многогранной заготовки.

Сохранение формы исходной заготовки принципиально зависит от соотношения вязкости и поверхностного натяжения и соотношений скорости подачи заготовки и вытягивания волокна, определяющего время пребывания луковицы в зоне формирования.

Температурные зависимости  $\sigma = f(T)$  и  $\eta = f(T)$  для кварцевого стекла принципиально различны. Если поверхностное натяжение кварцевого стекла слабо зависит от температуры и составляет около 300 дН/см в области 1400–2000 °С, то в этом интервале

вязкость меняется на 4 порядка, так как зависимость  $\eta = \eta_0 e^{-\frac{E_a}{RT}}$  экспоненциальна [1]; только при температуре  $\sim 2000$  °С силы поверхностного натяжения начинают преобладать и форма сечения начинает меняться.

Луковица перетягиваемой преформы имеет сложную форму тела вращения. По мере прогрева толстой преформы при ее движении вниз стекло нагревается, достигая (рис. 1) на участке "а-б" максимальной температуры. В точке "в" кривизна луковицы меняет свое значение, и на участке "в-б" происходит образование волокна под действием тянущей силы. Температура в точках "а" и "б" близки к температурам, соответствующим верхним значениям области отжига. При температуре ниже "а" и "б" стекло находится в упругом состоянии и не деформируется.

Нужно сказать, что процесс вытягивания есть процесс топологического (конформного) преобразования, при котором свойства преформы по радиусу не меняются, а только масштабируются в размере волокна. В то же время нагрев преформы до высоких температур не может не вызвать диффузионных потоков, которые сглаживают профиль показателя преломления волокна [2].

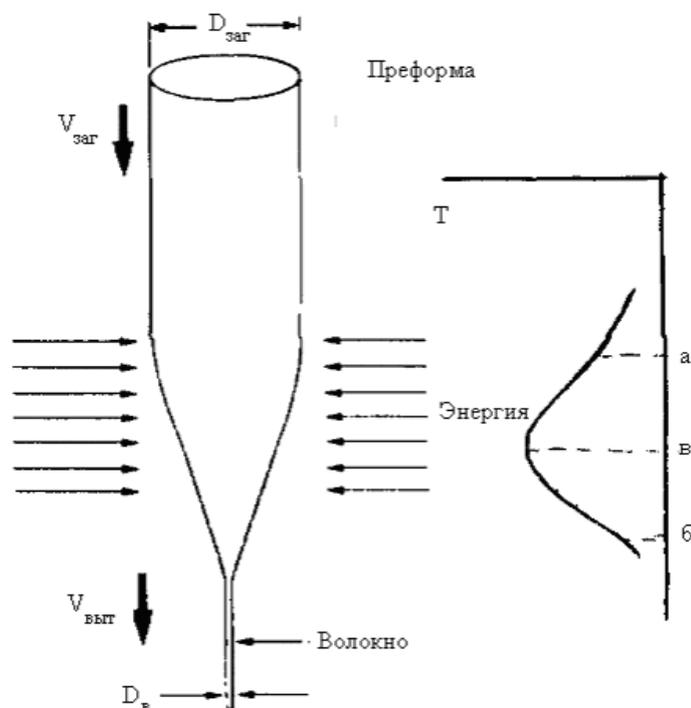


Рис. 1. Схема процесса вытягивания волокон

Существующая система обратной связи между скоростью подачи заготовки и скоростью вытягивания позволяет с частотой 2,5 кГц проверять постоянство соотношений скоростей и отслеживать непостоянство диаметра преформы, корректируя скорость вытягивания волокна при прохождении утолщенных диаметров преформы. Более того, системы вытягивания могут работать в ручном и автоматическом режимах, причем последний достигается за счет программируемой скорости вытягивания, которая по обратной связи выбирает единственную соответствующую данному режиму скорость подачи.

Техника вытягивания стеклянных световодов постепенно улучшается от лабораторных образцов до высокопроизводительных образцов промышленного оборудования заводов. Если первые образцы башен вытягивания имели по высоте 3–4 м, то современная башня вытягивания составляет 12–18 м и представляет собой систему единого комплекса оборудования, включающего следующие основные части:

- система подачи и центрирования преформ;
- система вытягивания и намотки волокна;
- печь нагрева;
- датчик диаметра волокна;
- системы нанесения покрытия и датчики концентричности покрытия;
- датчик диаметра волокна в покрытии;
- датчик величины натяжения волокна;
- система согласования скорости намотки катушек и скорости - вытягивания;
- система подачи обеспыленного воздуха;
- башня, на которой смонтированы системы.

Схематически система вытягивания представлена на рис. 2.



Рис. 2 Схема башни для вытягивания оптического волокна

Подающий механизм представляет собой уравновешенную трехкулачковую головку, закрепленную на двух винтовых направляющих, снабженных безлюфтовыми шариковыми парами. Движение головки вверх-вниз осуществляется шаговым электродвигателем, причем движение вниз – подача преформы – может осуществляться с любой заданной скоростью. Маршевое движение вверх – подъем заготовки – осуществляется с постоянной скоростью.

Головка представляет собою трехкулачковый патрон. Кроме степени свободы вверх-вниз, головка может вращаться со скоростью от 0 до 60 об/мин. Вращение осуществляется с помощью самостоятельного шагового двигателя с реверсом вращения. Кроме этих движений головка может перемещаться по направлениям  $x$ – $y$  к оси башни и вправо-влево от нее. Этот механизм позволяет совмещать ось преформы с осью башни, а правильнее сказать, с осью высокотемпературной печи. Эта процедура может проводиться вручную и в автоматическом режиме, в последнем случае команду на перемещение по направлениям  $x$ – $y$  каретка получает от датчика диаметра волокна.

Заданная скорость подачи преформы обеспечивается с точностью  $\pm 0,2\%$ . Система противовеса полностью разгружает винтовой механизм от веса перемещаемой каретки, механизма вращения и центрирования, так что, кроме потерь на трение шариковых пар, электродвигателю приходится преодолевать только вес преформы, причем только при движении каретки вверх. Система двух направляющих предохраняет механизм подачи от перекоса по координате  $z$  [11].

## Печь

Метод получения горячей зоны с требуемыми характеристиками целиком зависит от материалов, используемых в конкретном процессе. Для вытягивания заготовок из кварцевого стекла необходима температура около 2100–2400 °С. Требуемая температура достигается путем применения нагревательных графитовых элементов. Графитовый нагреватель запитывается от двух медных шин, позволяющих подать ток величиной до 2500 А от специального печного трансформатора. Верхняя и нижняя крышка печи – водоохлаждаемые. Внутренняя боковая поверхность печи предохраняется от перегрева специальным набором из молибденово-вольфрамовых экранов. Для защиты графита, экранов и других частей печи, находящихся в зоне высоких температур, в печь подается поток особо чистого аргона, подвергаемого в системе автоматического управления печью дополнительной очистке от примесей кислорода. Очистка аргона от примесей кислорода осуществляется пропусканием его через геттерную печь, нагретую до 600 °С и заполненную титановой стружкой. Разогретая стружка количественно поглощает примеси кислорода в аргоне, доводя концентрацию кислорода до  $10^{-9}$  ат. %.

Регулировка температуры в печи производится по заданной мощности. Для определения температуры в центре печи последняя оснащена пирометром, позволяющим определить температуру с точностью  $\pm 5$  °С.

Подача аргона в печь разделена на два потока, один из которых защищает нагреватель, а второй осуществляет газовый затвор, предотвращающий попадание внешней атмосферы в печь на выходе волокна из печи.

Система питания печи обеспечивает заданное значение мощности печи при работе на температуре вытягивания и возможность быстрого достижения этой температуры [10].

В идеале зона нагрева должна быть хорошо контролируема и минимальна по размерам. Таким требованиям удовлетворяет лазерный нагрев (рис. 3).

Мощность лазерного луча достаточно легко контролировать, следовательно, легко контролировать скорость и область плавления волокна. Лазерный луч падает на вращающийся зеркальный перископ, так что выходящий пучок сканирует по образующей цилиндра диаметром несколько сантиметров, попадает на наклонное зеркало с отвер-

стием в центре, через которое пропускается световодное волокно. Затем сканирующий пучок с помощью вогнутого зеркала сводится во вращающееся фокальное пятно в фиксированной точке у конца заготовки, где образуется луковица. При условии, что заготовка точно центрирована относительно фокальной точки оптической системы, луковица однородно нагревается со всех сторон и сильно уменьшается благодаря тому, что сфокусированный пучок очень узок. Подобная вытяжная установка позволяет справиться с переходными флуктуациями диаметра волоконного световода, характерными для вытяжных установок с протяженной горячей зоной и возникающими из-за механического дрейфа положения луковицы [13].

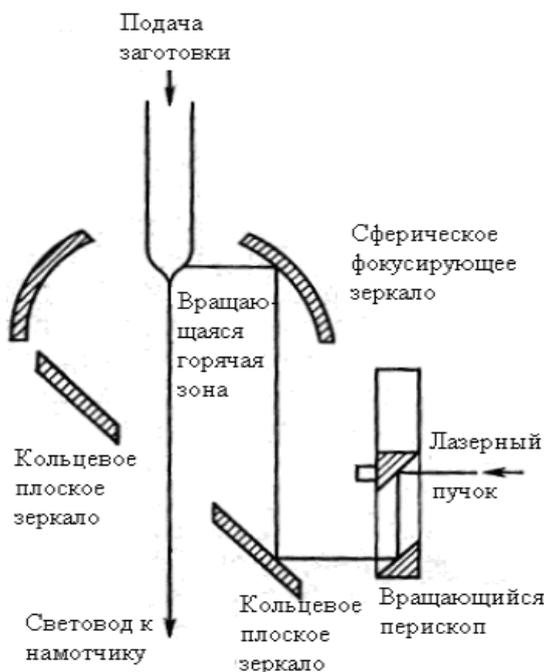


Рис 3. Схема расположения узлов вытяжной установки с лазерным нагревателем

### Датчик диаметра волокна

Для успешного проведения процесса вытягивания и обеспечения постоянства диаметра волокна башня снабжена датчиком диаметра волокна, который основан на измерении ширины теневого изображения волокна на многоэлементном приемнике. Освещение движущегося волокна осуществляется стабилизированным гелий-неоновым лазером, модулированный пучок которого позволяет определить текущее значение диаметра волокна с точностью 0,5 мкм на уровне 125 мкм с частотой 2,5 кГц. Оптоэлектронная часть датчика снабжена специализированным процессором, который запоминает текущее значение фактических диаметров волокна, производя дифференцирование этой величины по времени, и при пересечении заданных номинальных значений волокна дает команду на тахогенератор системы вытягивания. Величина и знак этой команды определяется величиной  $\delta d_v / \delta t$ , получаемой при пересечении фактических значений диаметра волокна, и величиной заданного номинала.

Вторая важная функция датчика диаметра заключается в определении положения волокна по координате  $x$ - $y$ . Только при совмещении оси волокна с осью печи можно рассчитывать на образование осесимметричной луковицы волокна. В противном случае волокно будет эллиптическим и механически непрочным. Выход от датчика диаметра по координатам  $x$ - $y$  подается на тахогенераторы системы центрирования заготовки механизма подачи.

Данные по работе этой системы выводятся на панель управления диаметром волокна, где отображаются фактические значения диаметра волокна и его положения по отношению к центру печи.

### **Нанесение первичных покрытий**

Кварцевое волокно является одним из самых прочных материалов, позволяющим иметь относительную деформацию около 10% при напряжении около 10 ГПа [5]. Однако это потенциально высокое значение реализуется только в специальных условиях. Даже в стерильных условиях лаборатории, где производится вытягивание волокна, прочность кварцевого волокна значительно меньше. Поэтому в процессе вытягивания кварцевое волокно покрывается специальными первичными защитно-упрочняющими покрытиями. Никакого дополнительного упрочнения при этом не происходит, покрытие позволяет консервировать потенциально высокие значения прочности кварцевого волокна.

Для нанесения покрытий башня вытягивания оснащается системами для аппликации покрытий на волокно. В качестве покрытий используют два вида органических полимеров:

- термореактивные полисилоксановые каучуки;
- фотоотверждаемые эпоксиакрилаты.

Разная химическая природа этих покрытий, естественно, требует различных технологий отверждения. Полисилоксановые каучуки наносятся из растворов, содержащих каталитическую добавку – платинохлористоводородную кислоту. После нанесения покрытия удаление растворителя и сшивка молекул мономера происходит под действием температуры в специальной термопечи при температуре 360-400 °С. Большая скорость вытягивания приводит к необходимости увеличения длины печи. Это обстоятельство следует учитывать при проектировании башен вытягивания большой производительности.

Силоксановые каучуки имеют сравнительно низкий показатель преломления, меньший, чем у кварцевого стекла, поэтому они позволяют получать особый класс оптических кварцевых волокон – волокон "кварц-полимер". Эти волокна характеризуются ступенчатым профилем показателя преломления, низкопреломляющая часть волокна – его оболочка – выполняется из полисилоксанового каучука. В данном случае, кроме функции защитного упрочнения, оболочка выполняет оптические функции образования светопередающей структуры. Поскольку при распространении света по волокну часть электромагнитного излучения проникает в оболочку [6], небезразличен факт поглощения света в полимерной оболочке. Это заставляет получать композиции полисилоксановых каучуков в достаточно чистом виде с малым коэффициентом светопоглощения.

Другим видом защитно-упрочняющих покрытий являются эпоксиакрилаты. Принципиальным отличием их является почти полное отсутствие растворителя в их составе и способ сшивки мономера в сплошную полимерную массу. Таким фактором полимеризации является достаточно мощное ультрафиолетовое излучение. УФ квант способствует образованию свободных радикалов в структуре мономера. Высокая реакционная способность свободных радикалов приводит к их полимеризации и отверждению полимера при комнатной температуре. При этом длина зоны отверждения существенно меньше, чем для термореактивных силоксановых каучуков.

Общим узлом для всех видов покрытий является аппликатор, в котором осуществляется контакт холодного (успешного охладиться) волокна и жидкого полимерного компаунда.

В работах [7, 8] описаны различные виды аппликаторов. Первичное защитно-упрочняющее покрытие должно образовывать концентричную оболочку вокруг волокна, причем в этой оболочке должны отсутствовать пузыри, так как механические свой-

ства покрытия определяются разнотолщинностью и пузырностью покрытия. Так как полимерные компаунды при комнатной температуре представляют собой сравнительно вязкие жидкости, процесс подачи их в фильеру и обеспыливание требуют нагрева и термостатирования. В некоторых случаях для этого используют повышенное давление [9].

Толщина покрытия должна быть такой, чтобы нейтрализовать абразивное действие пылевых частиц, находящихся в атмосфере помещения башни вытягивания.

Конструктивно фильерный узел представляет собой коническую воронку с калиброванным отверстием (рис 4.). Диаметр отверстия составляет 240 мкм. В систему нанесения покрытия входит емкость для заливки полимера, термостатируемая при определенной температуре, система подачи давления, собственно фильерный узел размещается на салазках, снабженных  $x$ - $y$  позиционером и датчиком толщины покрытия, работающим в двух взаимно перпендикулярных направлениях. Датчик концентричности в своей основе представляет оптическую систему, в которой луч гелий-неонового лазера освещает волокно в двух взаимно перпендикулярных направлениях. Изображение волокна и покрытия формируется на многоэлементном приемнике, выход которого дает информацию на осциллограф, фиксирующий отражение от поверхности волокна и от поверхности покрытия. Добываясь симметричной картины, получают концентричное покрытие, для чего каретку в случае необходимости перемещают по координатам  $x$ - $y$ .

Толщину покрытия определяют специальным оптическим датчиком, установленным после системы отверждения. В связи с тем, что толщина покрытия значительно слабее определяет свойства оптического волокна, диаметр его в покрытии определяется со средней точностью  $\pm 10$  мкм.

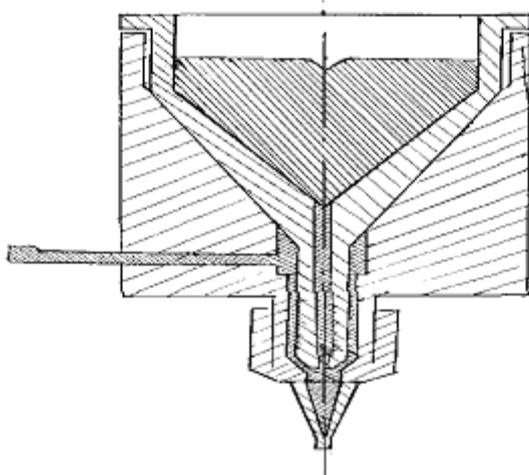


Рис. 4. Схема фильеры с подачей полимера под давлением

### Усилие вытягивания

Оптические и особенно механические характеристики волокна сильно зависят от величины натяжения, передаваемого от тянущего устройства волокну. Принципиально в пределах вязкости  $10^{5,5}-10^3$  пуаз образование волокна возможно, однако при вытягивании в области высоких вязкостей и, соответственно, больших времен релаксации, образование волокна происходит в условиях больших величин сдвиговых деформаций, способствующих образованию дефектов структуры. В этих условиях легко выдерживать геометрию волокна, но структура стекла в нем далека от равновесной. При высоких температурах, по существу, стекло течет, и дефекты из-за наличия сдвиговых напряжений не образуются.

Требования повышения производительности установок заставляют использовать высокотемпературную область графитовой печи в интервале 2100–2300 °С. Для определения величины усилия вытягивания в процессе предназначен датчик натяжения,

представляющий конструкцию из трех легких дисков, укрепленных на прецизионных подшипниках, позволяющих вращаться практически без трения. Волокно охватывает по полуокружности эти три диска, так что точки входа волокна к первому диску и выхода с третьего диска находятся на одной оси. Второй диск установлен на оси, соединенной с тензодатчиком, ось которого перпендикулярна оси диска.

Важной частью башни вытягивания является система вытяжного устройства (трактора) и накопления волокна на катушку. Усилие вытягивания передается волокну в тракторе, который представляет собой привод (редуктор с электродвигателем), на низкооборотной стороне которого установлен вытяжной барабан, прижатый через пассивные ролики к бесконечному ремню. Волокно прижимается к тянущему ремню на дуге  $90^\circ$  таким образом, что усилие вытягивания передается не в одной точке волокна, а на дуге окружности длиной около 20 см. Это уменьшает усилие прижима до минимума, что важно для минимального травмирования волокна за счет воздействия абразивных частиц наружной атмосферы. Электродвигатель привода обеспечивает возможность варьирования скорости вытягивания в области 0–200 м/мин. Тахогенератор двигателя соединен с датчиком диаметра, который передает команду на изменение скорости вытягивания, указанным выше образом.

### **Приемное устройство**

Одновременно с вытягиванием волокна осуществляется его накопление на катушку, установленную на намоточном устройстве. Намоточное устройство позволяет закрепить на нем приемную катушку, которая вращается с тангенциальной скоростью, равной скорости вытягивания. Кроме того, катушка может перемещаться возвратно-поступательно для осуществления раскладки волокна в заданных системой конечных выключателей пределах так, чтобы волокно не ложилось на предыдущий виток. Движение раскладчика осуществляется от самостоятельного привода, согласованного электрически с приводом вращения катушки. Система предусматривает возможность изменения шага намотки.

Так как задача уравнивания скорости вытягивания и подмотки с высокой точностью не является тривиальной, между трактором и намоточным устройством стоит компенсатор, который и осуществляет уравнивание окружных скоростей этих двух частей системы. Допустимый разбаланс скорости вытягивания и подмотки составляет  $\pm 3\%$ . Естественно, что этот разбаланс недопустим в течение долгого периода времени, так как волокно может быть порвано. Компенсатор выбирает эту разницу, уравнивая скорости до их полного равенства. От привода трактора работает информационное устройство определения длины вытянутого волокна, которое отображает информацию на пульте управления башней.

### **Заключение**

В современных башнях вытягивания предусмотрена система очистки воздуха и подача его ламинарного потока сверху вниз для уменьшения влияния внешних климатических условий помещения на качество волокна. При этом в зоне действия клин-бокса осуществляется класс "100" по степени запыленности. Класс "100" соответствует техническим требованиям 100 частиц в кубическом футе подаваемого воздуха. Определяются все частицы с размером более 0,5 мкм.

В автоматическом режиме, при заданных скорости вытягивания и диаметре волокна с заданным допуском, система выбирает необходимую скорость подачи заготовки. Начало процесса, как правило, осуществляют в ручном режиме, а при достижении номинальных значений диаметра волокна переходят к автоматизированному процессу. Как было показано выше, увеличение температуры при вытяжке и снижение усилия

вытягивания приводят к получению более качественного волокна. Также желательно использование лазерного нагрева. Увеличение скорости вытягивания приведет к снижению времени пребывания оптического волокна в термопечи или УФ оптической системе, что, в свою очередь, должно привести к неизбежному увеличению общей высоты вытяжной башни.

Авторы статьи выражают огромную благодарность Константину Владимировичу Дукельскому и Юрию Николаевичу Кондратьеву за предоставленные материалы, возможность ознакомления с оборудованием для получения оптического волокна, а также большую помощь при написании данной работы.

## Литература

1. Лeko B.K., Мазурин O.B. Свойства кварцевого стекла, Л., 1985
2. Унгер X. Г. Планарные и волоконные световоды, М., 391 с., 1980
3. Дукельский К.В. разработка специальных типов оптических волокон для нетрадиционных областей использования, канд. дисс., 2003
4. Данилов Е.Б. Вытягивание волокон, канд. дисс., 1989
5. Foord S.G., Lees J., Proc. IEE., 1976, 597
6. Wood D.L., Walcer K.L., Sipson J.R., Conf.ProcEur.Conf. Opt.Fiber, 1979, 121
7. Muto R., Akijama N., Sakata H., J Non-Cryst.Solids, 1975, 19, 269
8. Ansel R.E., Stanton J.J., Physics Fiber Optics, 1981, 2, 36
9. Ерофеева Е.В., Шевандин В.С. V Всесоюзная конференция "Волоконно-оптические системы передачи" Сборник тезисов, М., 1988, с.91-92
10. Отчет ГОИ по теме "Бирюза", 1987 Разработка многокомпонентных оптических стекол для ВОЛС.
11. Гроднев И. И., Ларин Ю.Т., Теумин И. И. Оптические кабели. – М., 1991.
12. Иоргачёв Д. В., Бондаренко О. В. Волоконно-оптические кабели и линии связи. – М.: Эко-Трендз, 2002. – 282 с.
13. Мидвинтер Дж. Э. Волоконные световоды для передачи информации: пер. с англ. - М.: Радио и связь, 1983. - 336 с.

# КОНТРОЛЬ КАЧЕСТВА ОПТИЧЕСКОГО ВОЛОКНА МЕТОДОМ ГАЗОРАЗРЯДНОЙ ВИЗУАЛИЗАЦИИ

В.Н. Фролков

Научный руководитель – д.т.н., проф. Ю.А. Гатчин

Приводятся результаты исследования оптического волокна методом газоразрядной визуализации. Получены изображения дефектов оптического волокна. Сравниваются усредненные ГРВ-граммы для различного уровня шума.

## Введение

Волокна характеризуются достаточно большим количеством параметров. Наиболее важные из них – диаметр волокна, состояние поверхности, толщина оболочки, механическая прочность, тепловые характеристики, спектральное пропускание, рассеяние света, неоднородность, двойное лучепреломление. Данные параметры влияют в первую очередь на разрешающую способность волоконного жгута, его светопропускание и, в конечном счете, на качество передаваемого жгутом изображения [1].

В настоящее время под термином "эффект Кирлиан" понимается визуальное наблюдение или регистрация на фотоматериале свечения газового разряда, возникающего вблизи поверхности исследуемого объекта при помещении последнего в электрическое поле высокой напряженности.

## Принцип газоразрядной визуализации

При внесении какого-либо объекта в электрическое поле в разрядном промежутке, образованном одним из электродов и поверхностью объекта исследования, возникает разряд, несущий информацию об объекте. Этот разряд наблюдают через прозрачный электрод или фиксируют на фоточувствительном материале, помещенном в разрядный промежуток [3]. Наличие разрядного промежутка предложенной системы является обязательным. Если исследуется токопроводящий объект, то один из электродов покрывается тонким слоем диэлектрика. Было выделено два варианта газоразрядного фотографирования – контактный и вакуумный [4].

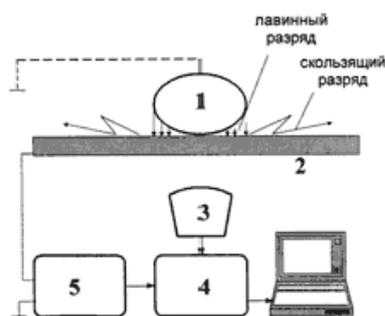


Рис 1. Схематическое изображение устройства для исследования ГРВ характеристик:  
1 – объект исследования, 2 – прозрачный электрод, 3 – оптическая система,  
4 – видеопреобразователь, 5 – электронные блоки

Принцип газоразрядной визуализации (ГРВ) заключается в следующем (рис. 1). Между исследуемым объектом 1 и диэлектрической пластиной 2, на которой размещается объект, подаются импульсы напряжения от генератора электромагнитного поля (ЭМП) 5, для чего на обратную сторону пластины 2 нанесено прозрачное токопроводящее покрытие. При высокой напряженности поля в газовой среде пространства контакта объекта 1 и пластины 2 развивается лавинный и/или скользящий газовый разряд, параметры которого определяются свойствами объекта. Свечение разряда с помощью оптической системы и ПЗС-камеры 3 (ПЗС – прибор с зарядовой связью) преобразуется в видеосигналы, которые записываются в виде одиночных кадров (ГРВ-грамм) или

AVI-файлов в блок памяти, связанный с компьютерным процессором обработки. Процессор обработки представляет собой специализированный программный комплекс, который позволяет вычислять набор параметров и на их основе делать определенные диагностические заключения [3].

При всем многообразии конкретных технических решений сущность процесса визуализации может быть сведена к некоторой теоретической схеме. Первичным процессом является процесс взаимодействия ЭМП с объектом исследования, в результате которого при определенной напряженности ЭМП возникает эмиссия с поверхностью объекта заряженных частиц, участвующих в иницировании начальных фаз газового разряда. Газовый разряд, в свою очередь, может влиять на состояние объекта, вызывая вторичные эмиссионные, деструктивные и тепловые процессы. Таким образом, в процессе газоразрядной визуализации формируется некоторая последовательность информационных преобразований. Состояние объекта характеризуется процессами, среди которых определяющую роль с точки зрения процесса ГРВ играют физико-химические и эмиссионные процессы, а также процессы газовой выделенности, которые зависят от изменений импеданса объекта как единого целого, импеданса участков его поверхности, их структурных и эмиссионных свойств. Неоднородность поверхности и объема, процессы эмиссии заряженных частиц или выделения газов оказывают влияние на параметры электромагнитного поля, за счет чего изменяются параметры газового разряда [5]. Такими параметрами являются характеристики тока разряда и оптического излучения. При этом основная информация извлекается из характеристик свечения, которое представляет собой пространственно распределенную группу участков различной яркости. Приемник излучения преобразует пространственное распределение яркости в изображение, а анализ амплитудных характеристик видеосигналов приводит к формированию набора параметров.

Для получения результатов использовались данные от динамической ГРВ видеосъемки. Полученные AVI-файлы, длительностью 10 секунд, обрабатывались с помощью программы Video Analyzer. Эта программа обрабатывает видеофайл и переводит результаты обработки в графическую форму, разбивая данные по ГРВ параметрам. Также с помощью данной программы были получены усредненные ГРВ-граммы для кремниевых пластин, преформ и волоконных жгутов при разном уровне шума.

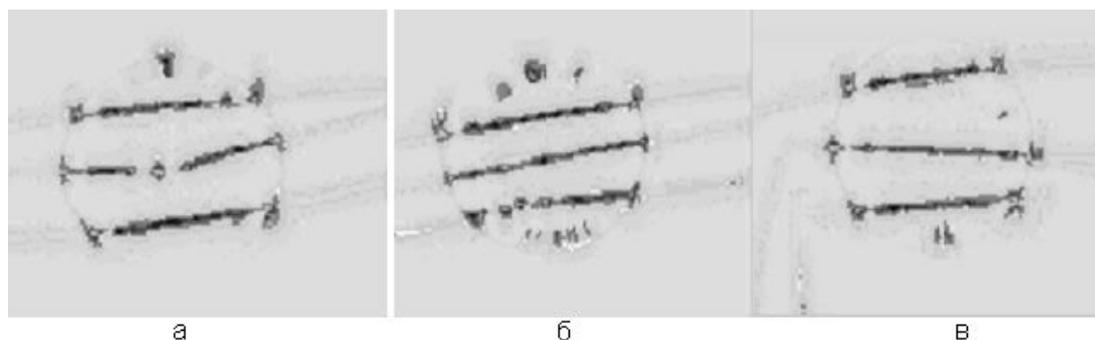


Рис 2. Усредненные ГРВ-граммы жгутов при уровне шума 50 а) сломан средний жгут; б) сломан нижний жгут; в) сломаны средний и нижний

На рис. 2. приведены усредненные картины свечения жгутов при уровне шума 50. В жгуты был искусственно привнесен дефект (жгуты были надломлены). Месту надлома соответствуют яркие точки голубого цвета на картинах свечения (точки наибольшей интенсивности). Также наблюдается чередование менее ярких точек по жгуту. Объяснение этого факта пока не находит обоснованной физической теории. Можно предложить способ дефектоскопии, заключающийся в определении точки с наибольшей интенсивностью – предполагаемым местом дефекта. Современная конструкция прибора недостаточно эффективна для использования данного метода, поэтому рекомендуется дальнейшее усовершенствование конструкции.

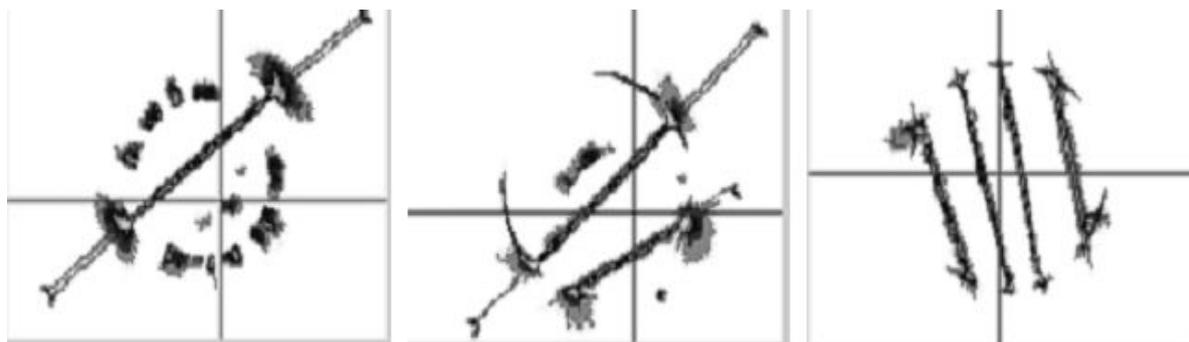


Рис 3. Усредненные ГРВ-граммы жгутов при уровне шума 90

На рис. 3 приведены усредненные картины свечения жгутов при уровне шума 90. На рисунках совместно расположены, соответственно, один, два и четыре жгута. Можно отметить, что на картине, где изображены четыре жгута, отсутствует паразитное свечение от электрода и краев отверстия диэлектрической пластины с зазором. Возможно, увеличение числа контролируемых волокон или жгутов повышает качество и информативность картины. ГРВ-граммы жгутов при уровне шума 90 отличаются более ярким и наглядным изображением по сравнению с шумом 50. Неоднородное распределение свечения вдоль волоконно-оптического жгута позволяет сделать выводы о возможности визуального контроля качества металлического покрытия, нанесенного на волокно или волоконный жгут.

Для исследования полупроводниковых структур, преформ и волоконного жгута методом газоразрядной визуализации была проведена серия экспериментов. Визуальный и параметрический анализ показал, что с помощью данной методики возможен контроль металлического покрытия волокна.

Для оценки разрешающей способности ГРВ-методики при параметрическом анализе требуются дополнительные исследования.

При анализе дефектной структуры объекта на основе какого-либо ГРВ-параметра необходимо учитывать временные и энергетические режимы получения данного параметра. Возможно, следует увеличить время экспозиции и амплитуду подающего напряжения.

ГРВ метод представляет интерес как быстрый, недорогой метод диагностики полупроводниковых структур и оптического волокна и требует дальнейшего, более детального рассмотрения физических особенностей формирования газового разряда, учета информационного шума, который может возникать от посторонних объектов и газовой среды.

## Заключение

Таким образом, показана принципиальная возможность контроля дефектов оптического волокна методом газоразрядной визуализации. Рекомендуется увеличивать число контролируемых оптических волокон, повышение уровня шума. Возможно дальнейшее усовершенствование конструкции прибора.

## Литература

1. Мидвинтер Дж. Э. Волоконные световоды для передачи информации: пер. с англ. М.: Радио и связь, 1983. - 336 с.
2. Гроднев И. И., Ларин Ю.Т., Теумин И. И. Оптические кабели. М., 1991.
3. Коротков К.Г. Основы ГРВ биоэлектрографии СПб. СПбГИТМО (ТУ), 2001
4. Адаменко В.Г. Исследование механизма формирования изображений, получаемых с помощью высокочастотного электрического разряда, Дисс. Минск, 1975, 140 с.
5. Романий С.Ф., Черный З.Д. Неразрушающий контроль материалов по методу Кирлиан. Днепропетровск 1991, 144 с.

# РАСШИФРОВКА ИНТЕРФЕРОГРАММ ФАЗОВОГО СДВИГА

А.В. Ильин

Научный руководитель – к.т.н. Д.А. Гаврилин

В работе рассматривается математическая модель формирования серии измерений, проводимых с применением интерферометров фазового сдвига. Предлагается алгоритм анализа измерений и аппроксимации функции деформации контролируемых поверхностей.

## Введение

В современном оптическом производстве для высокоточного контроля качества оптических систем широкое применение получили интерференционные методы. Метод получения интерферограмм при последовательном изменении оптической длины плеча в одном из каналов интерферометра и последующей совместной их обработке позволяет существенно повысить точность интерференционного метода контроля. Полученную таким образом серию интерферограмм будем называть интерферограммами фазового сдвига. Возможность применения этого метода контроля определяется наличием соответствующей компьютерной технологии. Важным этапом формирования компьютерной технологии для применения в рассматриваемом случае является разработка алгоритмов расшифровки интерферограмм фазового сдвига и аппроксимации функции деформации волнового фронта. Вполне очевидно, что задачи этого этапа должны быть решены в первую очередь.

## Математическая модель

Рассмотрим математическую модель формирования интерференционной картины. Выражение, описывающее распределение интенсивности по интерферограмме, выглядит следующим образом [6]:

$$I(x, y) = 1 + \cos[2\pi W(x, y)], \quad (1)$$

где  $I(x, y)$  – интенсивность,  $W(x, y)$  – функция деформации волнового фронта. Здесь  $x$  и  $y$  – координаты точки на интерференционной картине. В дальнейшем, для простоты, функции  $I(x, y)$  и  $W(x, y)$  будем записывать как  $I$  и  $W$ . Формула (1) описывает идеальный случай, поскольку в ней не учитывается множество факторов, присутствующих при образовании реальных интерференционных картин [2], но для описания математической модели формирования серии интерферограмм фазового сдвига она является достаточной.

Распределение относительной интенсивности для каждой интерферограммы, получаемой на интерферометре фазового сдвига, может быть описано следующим образом [7]:

$$I_i = \frac{1}{2} \left\{ 1 + \cos \left[ 2\pi \left( W \pm i \frac{\lambda}{N} \right) \right] \right\} \quad (2)$$

Здесь  $i=0,1,\dots,k-1$ , где  $k$  – число интерферограмм в серии,  $\lambda/N$  – величина единичного сдвига в длинах волн. Значение  $i=0$  соответствует первой интерферограмме. Знак "+" в формуле (2) отражает случай, когда расстояние между контролируемой и эталонной поверхностями увеличивается, знак "-" – когда расстояние уменьшается.

## Обработка интерферограмм

Задача обработки интерферограмм сводится к определению функции  $W$ . Однако, как видно из выражения (1), взаимосвязь между видом интерференционной картины

(распределением интенсивности  $I$ ) и функцией деформации волнового фронта  $W$  является нелинейной. Для устранения нелинейности проводят расшифровку интерферограммы, заключающуюся в определении точек с известным значением функции деформации волнового фронта. Такими точками являются максимумы и минимумы интерференционных полос, так как известно, что от полосы к полосе функция деформации волнового фронта увеличивается на одну длину волны.

Обработка серии интерферограмм фазового сдвига должна выполняться совместно. В процессе обработки на каждой интерферограмме определяются экстремумы полос (так называемая идентификация полос) – точки с известными значениями функции деформации волнового фронта. При этом номера полос проставляются относительно полос первой интерференционной картины с учетом знака функции деформации волнового фронта, так как известны величина и направление фазового сдвига. Таким образом, после обработки серии интерферограмм получаем для аппроксимации набор точек, в  $k$  раз больший набора точек, получаемых от одной интерферограммы (причем эти точки равномерно распределены по всей области локализации интерференционной картины). Рассмотрим процесс перемещения интерференционной полосы при сдвиге контролируемой поверхности относительно эталонной (рис.1).

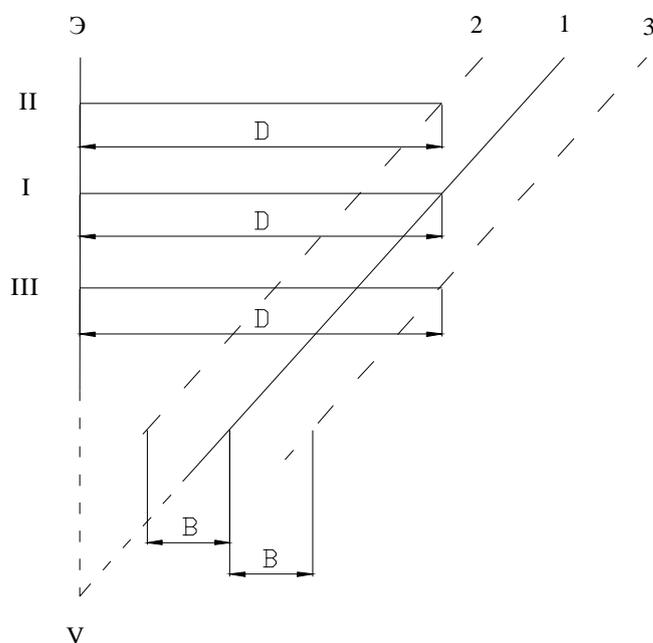


Рис. 1. Смещение контролируемой поверхности относительно эталонной

На рис.1 показана эталонная поверхность  $\mathcal{E}$  и три различных положения (1, 2 и 3) контролируемой поверхности. Положение 1 соответствует первой интерферограмме, относительно которой нумеруется всё последующие. Контролируемая и эталонная поверхности образуют между собой клин, на рисунке вершина клина обозначена как  $V$ . Пусть интерференционная полоса соответствует расстоянию  $D$  между эталонной и контролируемой поверхностями (полоса I). Изменим расстояние между контролируемой и эталонной поверхностями на величину  $b$ . Если расстояние между поверхностями уменьшается на величину  $B$  (положение 2), то интерференционная полоса I сдвигается от вершины клина  $V$  (полоса II). Если расстояние между поверхностями увеличивается на величину  $B$  (положение 3), то интерференционная полоса сдвигается к вершине клина  $V$  (полоса III). Полосы на всех интерферограммах нумеруются, начиная от вершины клина. На первой интерферограмме ближайшей к вершине клина полосе будет присвоен номер 1, и так далее. На следующих интерферограммах этой серии полосы будут нумероваться в пересчёте относительно первой интерферограммы (рис.2).

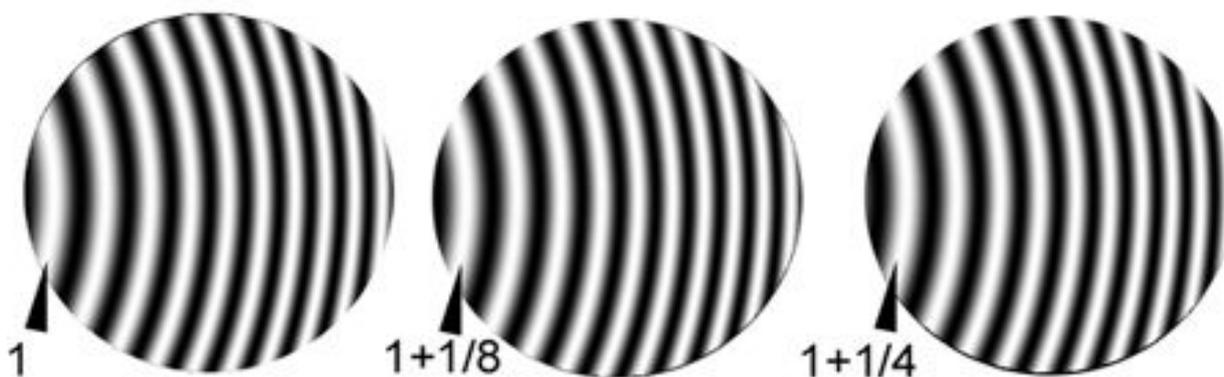


Рис. 2. Нумерация полос

Номер полосы соответствует значению функции деформации волнового фронта, выраженному в волновой мере, на этой полосе. Если  $a$  - значение функции деформации волнового фронта для одной из полос на  $i$ -той интерферограмме, то относительно первой интерферограммы значение функции деформации волнового фронта для этой же полосы будет составлять  $W_i - \frac{i}{N}$  (расстояние между контролируемой и эталонной поверхностями увеличивается), либо  $W_i + \frac{i}{N}$  (расстояние между контролируемой и эталонной поверхностями уменьшается).

Таким образом, для определения функции деформации волнового фронта необходимо на каждой из зарегистрированных интерферограмм определить положение экстремумов и пересчитать значения функции деформации волнового фронта относительно первой интерферограммы ( $i = 0$ ).

Функцию деформации волнового фронта удобно представлять в виде разложения по некоторому базису. Наиболее предпочтительным в оптике является использование ортогонального базиса в виде так называемых полиномов Цернике [4], преимущества которых заключаются также в том, что каждый член разложения представляет собой оптимальную форму аберрации в смысле минимума среднего квадрата функции деформации волнового фронта [1]. Функцию деформации волнового фронта  $W$  можно представить в полярных координатах  $(\rho, \varphi)$  в виде:

$$W(\rho, \varphi) = \sum_{n=0}^p \sum_{m=0}^n c_{nm} R_n^m(\rho) \cos(m\varphi) + \sum_{n=1}^p \sum_{m=1}^n s_{nm} R_n^m(\rho) \sin(m\varphi), \quad (3)$$

где  $R_n^m(\rho)$  – радиальные полиномы Цернике, зависящие только от  $\rho$ ,  $m+n$  – четное число,  $p$  определяет порядок полиномов,  $c_{nm}$  и  $s_{nm}$  – косинусные и синусные коэффициенты Цернике соответственно. Переход от декартовых координат к полярным осу-

ществляется по формулам 
$$\begin{cases} \rho = \sqrt{x^2 + y^2} \\ \varphi = \arctan\left(\frac{x}{y}\right). \end{cases}$$

С помощью вышеприведённых выражений можно получить значение функции деформации волнового фронта в любой точке зрачка. Так как эта функция изменяется плавно по зрачку, то для полного суждения о ней обычно достаточно иметь её значения в небольшом количестве точек. Как было сказано выше, функция деформации волнового фронта может быть полностью описана при помощи коэффициентов разложения по выбранному базису, причём требуется сравнительно небольшое количество коэффици-

ентов. Имея значения этих коэффициентов, можно с помощью формулы (3) найти требуемые значения функции деформации волнового фронта в любой точке зрачка. Таким образом, задача заключается в нахождении значений этих коэффициентов по известным значениям функции деформации волнового фронта в отдельных точках – экстремумах интерференционных полос. Так как полиномы Цернике ортогональны на круге единичного радиуса [5], то координаты найденных экстремумов пересчитываются в относительные для контура интерферограммы.

### Аппроксимация функции деформации волнового фронта

Задача определения коэффициентов Цернике может быть выполнена аппроксимацией типа интерполяции. Однако функция деформации волнового фронта имеет в общем случае сложный характер, и для повышения точности следует использовать избыточный набор данных, а аппроксимацию производить методом наименьших квадратов [3]. При этом для каждой интерферограммы формируется система уравнений, которая в матричном виде для  $i$ -той интерферограммы может быть записана следующим образом:  $\mathbf{A}_i \cdot \mathbf{x} = \mathbf{b}_i$ . Такая система уравнений является при этом переопределённой и содержит больше уравнений, чем неизвестных коэффициентов.

Представим выражение (3) в следующем виде:

$$W(\rho, \varphi) = \sum_{n=0}^p \sum_{m=0}^n c_{nm} P_{c_n}^m(\rho, \varphi) + \sum_{n=1}^p \sum_{m=1}^n s_{nm} P_{s_n}^m(\rho, \varphi),$$

где  $P_{c_n}^m(\rho) = R_n^m(\rho) \cos(m\varphi)$  и  $P_{s_n}^m(\rho) = R_n^m(\rho) \sin(m\varphi)$ . Тогда структурная матрица  $\mathbf{A}_i$  и вектор правых  $\mathbf{b}_i$  частей будет заполняться следующим образом:

$$\mathbf{A}_i = \begin{pmatrix} P_{c_0}^0(x_1, y_1) & \dots & P_{c_p}^p(x_1, y_1) & P_{s_0}^0(x_1, y_1) & \dots & P_{s_p}^p(x_1, y_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{c_0}^0(x_j, y_j) & \dots & P_{c_p}^p(x_j, y_j) & P_{s_0}^0(x_j, y_j) & \dots & P_{s_p}^p(x_j, y_j) \end{pmatrix},$$

$$\mathbf{b}_i = \begin{pmatrix} N_1 + \frac{i}{N} \\ \vdots \\ N_j + \frac{i}{N} \end{pmatrix}$$

при уменьшении расстояния между контролируемой и эталонной поверхностями,

$$\mathbf{b}_i = \begin{pmatrix} N_1 - \frac{i}{N} \\ \vdots \\ N_j - \frac{i}{N} \end{pmatrix}$$

при увеличении расстояния между контролируемой и эталонной поверхностями:

При этом количество столбцов структурной матрицы зависит от порядка полиномов  $p$ , а количество строк, также как и для вектора правых частей, от количества точек, используемых для аппроксимации функции деформации волнового фронта. Значения элементов вектора  $\mathbf{b}$  определяются номером полосы  $N_j$ , на которой расположена  $j$ -ая точка, пересчитанным относительно первой интерферограммы. Вектор неизвестных коэффициентов  $\mathbf{x}$ , соответственно, состоит из определяемых коэффициентов:

$$\mathbf{x}^T = \{c_{00} \quad \dots \quad c_{pp} \quad s_{11} \quad \dots \quad s_{pp}\}$$

Результирующая система уравнений, составленная для всего набора интерферограмм, будет состоять из блочных матриц  $\mathbf{A}$  и  $\mathbf{b}$  и иметь вид:  $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ , где

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_i \end{pmatrix}, \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_i \end{pmatrix} \quad (4)$$

## Заключение

Интерферометры фазового сдвига позволяют повысить точность оценки качества исследуемых оптических элементов, в том числе определять знак деформаций, что необходимо для сложных оптических поверхностей. Однако их использование возможно только совместно с компьютерной обработкой интерферограмм, получаемых при аттестации оптики. Разработанный алгоритм расшифровки интерферограмм и полученные математические соотношения аппроксимации функции деформации волнового фронта определяют основу формирования требуемой компьютерной технологии.

## Литература

1. Родионов С.А. Автоматизация проектирования оптических систем. Ленинград, Машиностроение, 1982 270 стр.
2. Гаврилин Д.А. Дипломная работа “Исследование и реализация методов компьютерного моделирования и обработки интерферограмм”, 2000.
3. Дж. Голуб, Ч. Ван Лоун. Матричные вычисления. Москва, Мир, 1999 548 стр.
4. Домненко В.М. Математическое моделирование формирования фотолитографического изображения. Диссертация на соискание ученой степени кандидата технических наук. СПбГИТМО(ТУ), Санкт-Петербург, 1999. 132 стр.
5. Агурок И.П. Автоматизация обработки интерферограмм при контроле оптических систем. Диссертация на соискание ученой степени кандидата технических наук. ЛОМО, Ленинград, 1986. 159 стр.
6. Калигиевский Н.И. Волновая оптика: учебное пособие для вузов. Москва, "Высшая школа", 1995. 463 стр.
7. Timothy L. Pennington. Miniaturized 3-D Mapping System Using a Fiber Optic Coupler as a Young's Double Pinhole Interferometer. Virginia Polytechnic Institute, June 2000

# РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ КОМПЬЮТЕРНОЙ ИЗОФОТОМЕТРИИ ФРЛ И ЧКХ ОПТИЧЕСКОЙ СИСТЕМЫ

Ю.В.Беляев

Научные руководители – к.т.н., доцент В.М.Домненко,  
д.т.н., профессор В.К. Кирилловский

## Введение

Первичными характеристиками качества изображения, сформированного оптической системой, считаются:

- функция рассеяния точки (ФРТ); обозначим ее как  $D(y',z')$ ;
- функция рассеяния линии (ФРЛ); обозначим ее как  $D_L(y')$ .

ФРЛ является одномерной функцией, которая описывает распределение относительной освещенности в двумерном изображении тест - объекта типа "светящаяся линия" (по нормали к оси направления линии).

Кроме того, в практике оптического приборостроения традиционными являются такие характеристики качества изображения, как оптическая передаточная функция (ОПФ) и связанная с ней функция передачи модуляции (ФПМ), называемая также частотно-контрастной характеристикой (ЧКХ) изображающей системы. Как правило, такие характеристики могут быть определены путем математической обработки функций рассеяния. Так, ОПФ можно рассчитать как Фурье - преобразование ФРЛ.

$$d_L(\mu) = \int_{-\infty}^{\infty} D_L(y) \cdot e^{-2\pi i \mu y} dy,$$

где  $D_L$  - ФРЛ,  $\tilde{\psi}$  - пространственная частота в оптическом изображении,  $y$  - пространственная координата в ФРЛ.

Модуль ОПФ есть ФПМ; она связывает пространственную частоту и коэффициент передачи контраста в сформированном изображении.

Метод изофотометрии ФРЛ нашел применение при исследованиях оптических систем различных классов, включая фото- и кино-объективы, прецизионные объективы микроскопов, оптические системы для цветного телевидения [1, 2].

Сущность способа изофотометрии ФРЛ состоит в том, что в качестве тест - объекта исследуемому объективу предъявляется объект типа "светящаяся линия переменной яркости". Для реализации такого тест-объекта в коллиматоре установлена щель, совмещенная с линейным фотометрическим клином, ось которого совпадает с расположением щели [1, 2].

Построенное исследуемым объективом изображение этого тест - объекта переносится с увеличением при помощи вспомогательного микрообъектива на чувствительную площадку передающей телевизионной камеры, включенной в замкнутую телевизионную систему. Видеосистема работает в режиме выделения контура при постоянных световых характеристиках приемника изображения. Для этого в ее составе подключен блок выделения контура (БВК), выделяющий изофоту изображения тест-объекта типа "линия переменной яркости", сформированного исследуемой оптической системой.

Эта изофота имеет форму графика функции рассеяния линии (ФРЛ) исследуемого объектива, где по оси ординат располагаются значения логарифма относительной освещенности ( $\lg E_{\text{отн}}$ ), а по оси абсцисс - пространственная протяженность в плоскости изображения.

Так как график ФРЛ [ $\lg(E_{\text{отн}}) = f(y')$ ] формируется на экране телевизионного монитора в реальном масштабе времени, данная установка позволяет оперативно определять плоскость наилучшей установки (фокусировки) объектива, осуществлять его юстировку до достижения наилучшего качества изображения и оценивать качество изображения по ширине ФРЛ и соответствующему значению фотографической разрешающей способности (ФРС).

Компьютерной обработкой данных получают весь комплекс требуемых характеристик качества изображения объектива: оптическую передаточную функцию ОПФ, модуляционную передаточную функцию ФПМ (называемую также частотно-контрастной характеристикой ЧКХ). Определяют требуемые критерии качества изображения исследуемой системы, такие как значение коэффициента передачи контраста на заданной критической частоте.

### **Метод компьютерной изофотометрии ФРЛ**

В данной работе выполнено дальнейшее развитие указанного метода средствами компьютерных технологий. Таким образом, полученные результаты можно считать шагом в развитии новой ветви компьютерной оптики в рамках направления, получившего название “Современные оптические измерения”.

Автоматизация расчета и проектирования систем получила свое развитие с появлением электронно-вычислительных машин (ЭВМ) в 40-х годах нашего столетия. В наши дни применение компьютеров получает все более широкое распространение, в том числе и в условиях действующего оптического производства, несмотря на ряд нерешенных проблем и несбывшихся надежд, связанных с попытками полностью автоматизировать труд создателя оптической системы. Наряду с вопросами автоматизации проектирования, выполняются и задачи расчетной оценки качества изображения оптических систем.

Согласно рассмотренным концепциям, авторами предложены алгоритмы и разработано программное обеспечение, позволяющее реализовать компьютерную интерпретацию метода изофотометрии ФРЛ на созданной установке.

Сформированное телевизионной камерой изофотометра изображение тест-объекта “светящаяся линия переменной яркости” вводится в компьютер при помощи телевизионного тюнера с опцией захвата кадра. Изображение, записанное в BMP-файл, далее вызывается программой “LSF\_MTF”, которая подвергает его оконтуриванию, причем формируется график ФРЛ как функции  $\lg(E_{\text{отн}}) = f(y')$ , после чего программа подвергает график оцифровке, определяя координаты его точек. При этом формируется таблица массивов значений функции. Далее программой выполняется расчет спектра пространственных частот ФРЛ как функция передачи модуляции (МПФ или ЧКХ) исследуемой оптической системы (рис.4).

### **Разработка программного обеспечения для реализации метода компьютерной изофотометрии ФРЛ и определения МПФ оптической системы**

В процессе данного исследования было разработано и реализовано программное обеспечение для обработки результатов изофотометрии функции рассеяния линии и вычисления частотно-контрастной характеристики. Программа позволяет быстро и просто загружать изображение, зафиксированное с помощью телекамеры, производить его обработку, выводить результаты на экран монитора и на печать, а также сохранять данные в файлах различных форматов.

Для получения графика ФРЛ в логарифмическом виде был разработан и реализован алгоритм, заключающийся в постолбцовом сканировании двумерного массива значений пиксельной яркости как двумерной выборки, полученной в результате

декодирования изображения, сохранённого в стандартном графическом формате (BMP). При этом осуществляется поиск в каждом столбце элемента, характеризующего перепад яркости на уровне 50% от максимальной для изображения. По существу, происходит оконтуривание изображения, транслированного с установки. В ходе такой фильтрации формируется график функции рассеяния линии, имеющий логарифмические значения относительной освещённости  $IgE$  по оси ординат, а также одномерный массив данных, содержащий в себе элементы со значениями логарифма относительной освещённости. Номер элемента в этом массиве определяет пространственную координату  $y'$  для данной точки графика (первоначально - в пикселах; перевод в миллиметры производится через масштабный коэффициент, являющийся паспортной характеристикой установки).

Массив значений модуляционной передаточной функции определяется путём одномерного дискретного сдвигового Фурье-преобразования, алгоритм которого заложен в библиотеку FFTW, предназначенную для сходных преобразований в среде разработки Microsoft Visual C++.

Для получения графика ФРЛ в линейном виде программа вычисляет значения антилогарифмов относительной освещённости.

При отображении графиков ФРЛ и МПФ происходит нормирование осей в соответствии с введёнными величинами плотности клина и коэффициента масштабирования, предназначенного для перехода к реальным координатам.

### **Методика работы с программой**

Ввод изображения в программу для обработки осуществляется нажатием кнопки "Открыть" (или Меню-> Файл-> Открыть... , или цифры 1). Открывается диалоговое окно, в котором нужно выбрать файл, содержащий снимок изображения с установки и имеющий графический bmp-формат. При этом в правой части окна существует возможность предпросмотра изображения, находящегося в том или ином файле.

После окончательного выбора файла для обработки, происходит его загрузка в программу. Картинка из него отображается в левой части экрана.

Далее нужно ввести так называемый коэффициент масштабирования, определяющий соответствие размера пиксела представленной на экране картинке значению в миллиметрах. Это нужно для правильной градуировки пространственных осей в реальных единицах.

Можно производить обработку как изображения целиком, так и его части (в таком случае нужно выделить эту область).

Нажатием на кнопку "Контур" запускается процесс оконтуривания. В левой части экрана выводится график функции рассеяния линии в логарифмическом масштабе (по оси абсцисс - единицы логарифма освещённости, по оси ординат - единицы пространственных координат). В правой части - табличные значения координат этого графика.

Нажатие на кнопку "Step1. ФРЛ" позволяет увидеть графическую и табличную информацию о функции рассеяния линии в линейном масштабе. Это достигается путём взятия антилогарифма для каждого значения кривой ФРЛ в логарифмическом масштабе. Градуировка по оси ординат осуществляется в единицах освещённости, по оси абсцисс - единицы прежние (пространственная координата).

Вид модуляционной передаточной функции можно получить, если нажать " Step2. МПФ". Здесь в основу математической модели положен механизм одномерного сдвигового быстрого преобразования Фурье. График по горизонтали - в единицах частоты, по вертикали - в единицах контраста.

В любой момент можно увидеть исходную картинку путём нажатия соответствующей кнопки (Image или Меню-> Вид-> Отображение-> Исходная

картинка, либо 2). Также существует возможность вернуться к любому из ранее увиденных графиков (Контур, ФРЛ или МПФ) для данного обрабатываемого изображения.

Для удобства считывания полученных в результате обработки значений, в программе реализованы такие возможности, как:

- вывод вспомогательной сетки;
- информация о положении курсора на графике.

Открытие из файла изображения для обработки: Меню-> Файл-> Открыть...

Вывод на экран исходной картинке (изображения, для обработки, загруженного из файла): Меню-> Вид-> Отображение-> Исходная картинка

Выделение контура исходного изображения (получение ФРЛ в логарифмическом масштабе): Меню-> Вид-> Отображение-> Контур картинка

Получение ФРЛ в линейном масштабе: Меню-> Вид-> Отображение-> ФРЛ

Вывод на экран МПФ: Меню-> Вид-> Отображение-> МПФ

Результаты обработки сохраняются автоматически в формате HTML в виде табличных значений соответственно для логарифмической и линейной ФРЛ, а также для МПФ. Существует возможность сохранения пользователем этих же значений в txt-формате. Также можно зафиксировать результаты в файле \*.html, который будет содержать изображение графика, выведенного в данный момент на экран, и значения координат в табличном виде. При этом график сохраняется в графическом JPEG-формате.

Экспортировать результаты обработки изображения можно также в формат PDF. Для этого необходимо, чтобы на производящем обработку компьютере были установлены программные продукты компании Adobe Systems Inc. (например, Adobe Acrobat). Чтобы получить pdf-файл нужно выбрать Печать или Меню-> Печать..., далее в качестве имени принтера указать название Adobe Acrobat или Acrobat Distiller, затем нажать ОК.

Результаты обработки можно вывести на печать (для размера бумаги А4 и книжной ориентации) непосредственно из программы. Будет напечатан график и табличные значения.

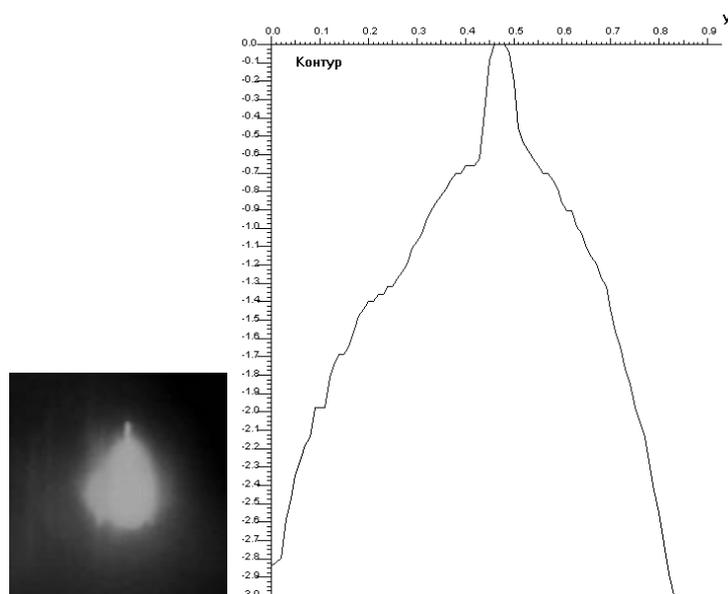


Рис. 2. Кривая функции рассеяния линии, сформированная в результате обработки изображения в телевизионно-компьютерном изофотометре ФРЛ; таблица значений ФРЛ получена путем компьютерного анализа изображения кривой.

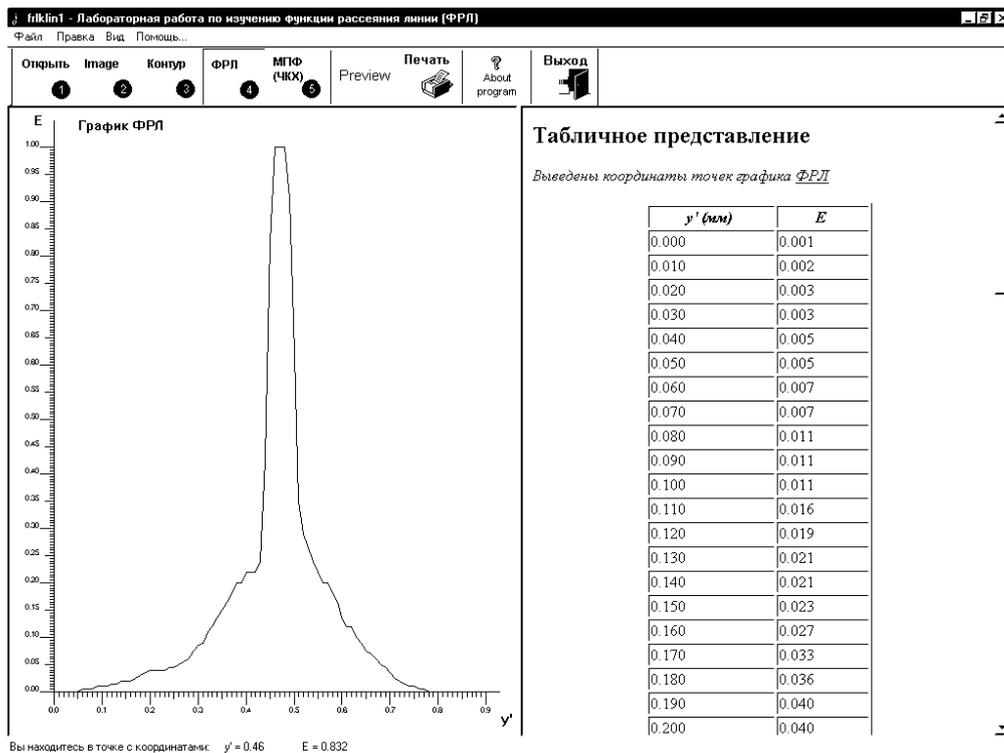


Рис. 3. Функция рассеяния линии в линейном масштабе по оси значений относительной освещённости.

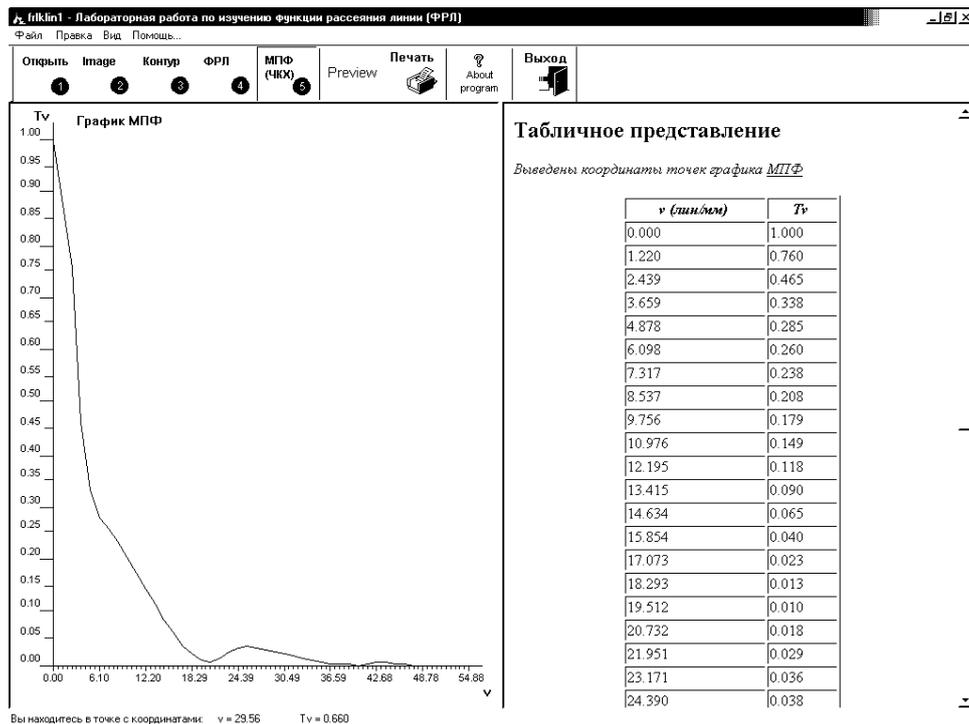


Рис. 4. Модуляционная передаточная функция (МПФ), вычисленная в результате Фурье-преобразования экспериментально полученной ФРЛ

В случае выбора варианта установки страницы, не предусмотренного в данной программе, на экран будет выведено об этом предупреждающее сообщение. Например, при выборе альбомной ориентации листа:

Для получения ФРЛ в логарифмическом масштабе в программу заложен алгоритм оконтуривания по столбцам, заключающийся в сканировании массива двумерной

выборки и нахождении элемента с минимальным номером в каждом столбце, которому присущ перепад яркости RGB- пиксела в 50%. Эта выборка получается в процессе загрузки исследуемого изображения и приведения его к полутоновому виду. Такое преобразование (к полутоновому виду) осуществляется в первую очередь для того, чтобы сократить информационную избыточность изображения, оставить в нем только ту информацию, которая нужна для решения конкретной задачи.

### **Заключение**

Проведенные испытания показали эффективность применения программы для реализации метода компьютерной изофотометрии ФРЛ. На кафедре Прикладной и компьютерной оптики поставлена лабораторная работа по измерению ФРЛ телевизионной передающей камеры и определению ФПМ системы "фотообъектив - телевизионный приемник изображения". В настоящее время ведется работа по созданию установки и программного обеспечения для реализации метода изофотометрии пограничной (переходной) функции и определения на ее базе ФПМ оптической системы.

### **Литература**

1. Кирилловский В.К. Контроль качества объективов. Л. ЛИТМО. 1984 г.
2. Иванова Т.А., Кирилловский В.К. Проектирование и контроль оптики микроскопов. Л. Машиностроение. 1984 г.

# РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ КОМПЬЮТЕРНОЙ ИЗОФОТОМЕТРИИ ФРЛ И ЧКХ ОПТИЧЕСКОЙ СИСТЕМЫ

Ю.В. Беляев

Научные руководители – к.т.н., доц. В.М. Домненко,  
д.т.н., проф. В.К. Кирилловский

## Введение

Первичными характеристиками качества изображения, сформированного оптической системой, считаются:

- функция рассеяния точки (ФРТ); обозначим ее как  $D(y',z')$ ;
- функция рассеяния линии (ФРЛ); обозначим ее как  $D_L(y')$ .

ФРЛ является одномерной функцией, которая описывает распределение относительной освещенности в двумерном изображении тест-объекта типа "светящаяся линия" (по нормали к оси направления линии).

Кроме того, в практике оптического приборостроения традиционными являются такие характеристики качества изображения, как оптическая передаточная функция (ОПФ) и связанная с ней функция передачи модуляции (ФПМ), называемая также частотно-контрастной характеристикой (ЧКХ) изображающей системы. Как правило, такие характеристики могут быть определены путем математической обработки функций рассеяния. Так, ОПФ можно рассчитать как Фурье-преобразование ФРЛ.

$$d_L(\mu) = \int_{-\infty}^{\infty} D_L(y) \cdot e^{-2\pi i \mu y} dy,$$

где  $D_L$  – ФРЛ,  $\mu$  – пространственная частота в оптическом изображении,  $y$  – пространственная координата в ФРЛ.

Модуль ОПФ есть ФПМ; она связывает пространственную частоту и коэффициент передачи контраста в сформированном изображении.

Метод изофотометрии ФРЛ нашел применение при исследованиях оптических систем различных классов, включая фото- и кинообъективы, прецизионные объективы микроскопов, оптические системы для цветного телевидения [1, 2].

Сущность способа изофотометрии ФРЛ состоит в том, что в качестве тест-объекта исследуемому объективу предъявляется объект типа "светящаяся линия переменной яркости". Для реализации такого тест-объекта в коллиматоре установлена щель, совмещенная с линейным фотометрическим клином, ось которого совпадает с расположением щели [1, 2].

Построенное исследуемым объективом изображение этого тест-объекта переносится с увеличением при помощи вспомогательного микрообъектива на чувствительную площадку передающей телевизионной камеры, включенной в замкнутую телевизионную систему. Видеосистема работает в режиме выделения контура при постоянных световых характеристиках приемника изображения. Для этого в ее состав введен блок выделения контура (БВК), выделяющий изофоту изображения тест-объекта типа "линия переменной яркости", сформированного исследуемой оптической системой. Эта изофота имеет форму графика функции рассеяния линии (ФРЛ) исследуемого объектива, где по оси ординат располагаются значения логарифма относительной освещенности ( $\lg E_{\text{отн}}$ ), а по оси абсцисс – пространственная протяженность в плоскости изображения.

Так как график ФРЛ [ $\lg(E_{\text{отн}}) = f(y')$ ] формируется на экране телевизионного монитора в реальном масштабе времени, данная установка позволяет оперативно опреде-

лять плоскость наилучшей установки (фокусировки) объектива, осуществлять его юстировку до достижения наилучшего качества изображения и оценивать качество изображения по ширине ФРЛ и соответствующему значению фотографической разрешающей способности (ФРС).

Компьютерной обработкой данных получают весь комплекс требуемых характеристик качества изображения объектива: оптическую передаточную функцию ОПФ, модуляционную передаточную функцию ФПМ (называемую также частотно-контрастной характеристикой ЧКХ). Определяют требуемые критерии качества изображения исследуемой системы, такие как значение коэффициента передачи контраста на заданной критической частоте.

### **Метод компьютерной изофотометрии ФРЛ**

В данной работе выполнено дальнейшее развитие указанного метода средствами компьютерных технологий. Таким образом, полученные результаты можно считать шагом в развитии новой ветви компьютерной оптики в рамках направления, получившего название "Современные оптические измерения".

Автоматизация расчета и проектирования систем начала свое развитие с появлением электронно-вычислительных машин (ЭВМ) в 40-х годах прошлого столетия. В наши дни применение компьютеров получает все более широкое распространение, в том числе и в условиях действующего оптического производства, несмотря на ряд нерешенных проблем и несбывшихся надежд, связанных с попытками полностью автоматизировать труд создателя оптической системы. Наряду с вопросами автоматизации проектирования, решаются и задачи расчетной оценки качества изображения оптических систем.

Согласно рассмотренным концепциям, авторами предложены алгоритмы и разработано программное обеспечение, позволяющее реализовать компьютерную интерпретацию метода изофотометрии ФРЛ на созданной установке.

Сформированное телевизионной камерой изофотометра изображение тест-объекта "светящаяся линия переменной яркости" вводится в компьютер при помощи телевизионного тюнера с опцией захвата кадра. Изображение, записанное в BMP-файл, далее вызывается программой "LSF\_MTF", которая подвергает его оконтуриванию, причем формируется график ФРЛ как функции  $\lg(E_{\text{отн}}) = f(y')$ , после чего программа подвергает график оцифровке, определяя координаты его точек. При этом формируется таблица массивов значений функции. Далее программой выполняется расчет спектра пространственных частот ФРЛ как функции передачи модуляции (МПФ или ЧКХ) исследуемой оптической системы (рис.1–3).

### **Разработка программного обеспечения для реализации метода компьютерной изофотометрии ФРЛ и определения МПФ оптической системы**

В процессе исследования было разработано и реализовано программное обеспечение для обработки результатов изофотометрии функции рассеяния линии и вычисления частотно-контрастной характеристики. Программа позволяет быстро и просто загружать изображение, зафиксированное с помощью телекамеры, производить его обработку, выводить результаты на экран монитора и на печать, а также сохранять данные в файлах различных форматов.

Для получения графика ФРЛ в логарифмическом виде был разработан и реализован алгоритм, заключающийся в постолбцовом сканировании двумерного массива значений пиксельной яркости как двумерной выборки, полученной в результате декодирования изображения, сохраненного в стандартном графическом формате (BMP). При этом осуществляется поиск в каждом столбце элемента, характеризующего перепад яркости

на уровне 50% от максимальной для изображения. По существу, происходит оконтуривание изображения, транслированного с установки. В ходе такой фильтрации формируется график функции рассеяния линии, имеющий логарифмические значения относительной освещенности  $IgE$  по оси ординат, а также одномерный массив данных, содержащий в себе элементы со значениями логарифма относительной освещенности. Номер элемента в этом массиве определяет пространственную координату  $y'$  для данной точки графика (первоначально – в пикселах; перевод в миллиметры производится через масштабный коэффициент, являющийся паспортной характеристикой установки).

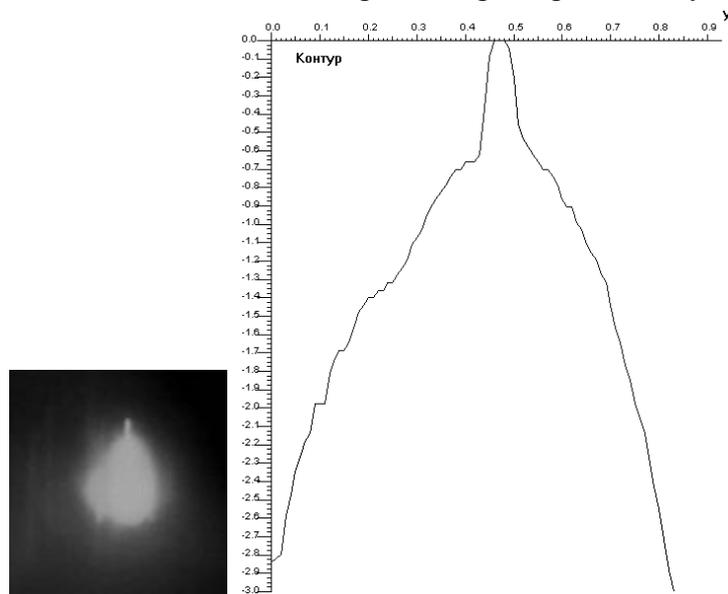


Рис. 1. Кривая функции рассеяния линии, сформированная в результате обработки изображения в телевизионно-компьютерном изофотометре ФРЛ; таблица значений ФРЛ получается путем компьютерного анализа изображения кривой

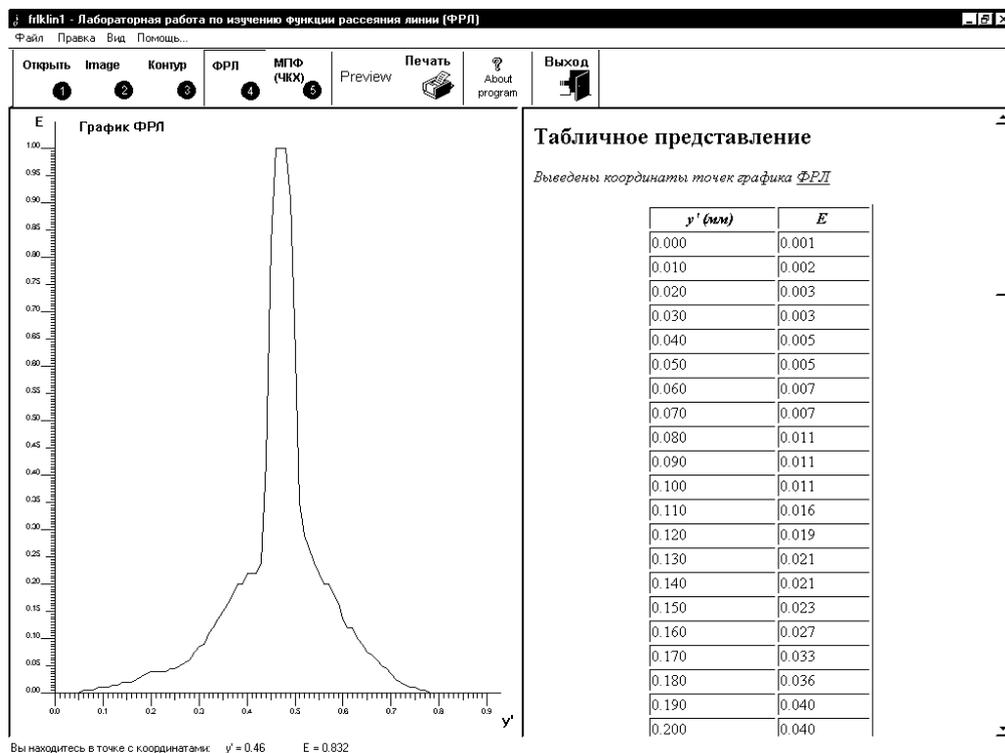


Рис. 2. Функция рассеяния линии в линейном масштабе по оси значений относительной освещенности

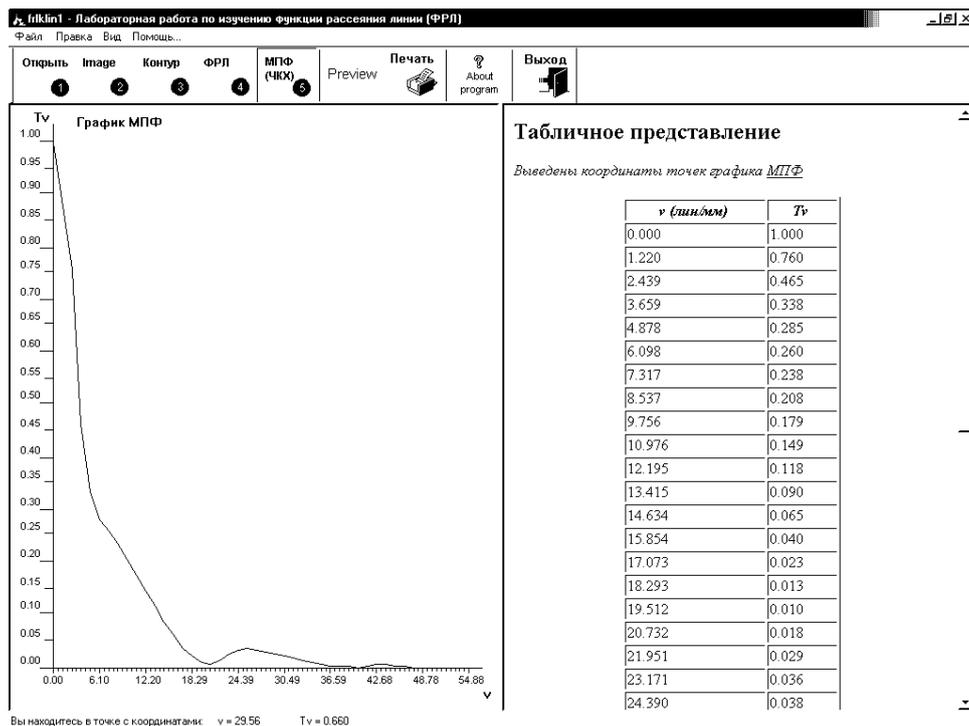


Рис. 3. Модуляционная передаточная функция (МПФ), вычисленная в результате Фурье-преобразования экспериментально полученной ФРЛ

Массив значений модуляционной передаточной функции определяется путем одномерного дискретного сдвигового Фурье-преобразования, алгоритм которого заложен в библиотеку FFTW, предназначенную для сходных преобразований в среде разработки Microsoft Visual C++.

Для получения графика ФРЛ в линейном виде программа вычисляет значения антилогарифмов относительной освещенности.

При отображении графиков ФРЛ и МПФ происходит нормирование осей в соответствии с введенными величинами плотности клина и коэффициента масштабирования, предназначенного для перехода к реальным координатам.

### Методика работы с программой

Ввод изображения в программу для обработки осуществляется нажатием кнопки "Открыть" (или Меню-> Файл-> Открыть... , или цифры 1). Открывается диалоговое окно, в котором нужно выбрать файл, содержащий снимок изображения с установки и имеющий графический bmp-формат. При этом в правой части окна существует возможность предпросмотра изображения, находящегося в том или ином файле.

После окончательного выбора файла для обработки происходит его загрузка в программу. Картинка из него отображается в левой части экрана.

Далее нужно ввести так называемый коэффициент масштабирования, определяющий соответствие размера пиксела представленной на экране картинке значению в миллиметрах. Это нужно для правильной градуировки пространственных осей в реальных единицах.

Можно производить обработку как изображения целиком, так и его части (в таком случае нужно выделить эту область).

Нажатием на кнопку "Контур" запускается процесс оконтуривания. В левой части экрана выводится график функции рассеяния линии в логарифмическом масштабе (по

оси абсцисс - единицы логарифма освещенности, по оси ординат – единицы пространственных координат). В правой части – табличные значения координат этого графика.

Нажатие на кнопку "Step1. ФРЛ" позволяет увидеть графическую и табличную информацию о функции рассеяния линии в линейном масштабе. Это достигается путем взятия антилогарифма для каждого значения кривой ФРЛ в логарифмическом масштабе. Градуировка по оси ординат осуществляется в единицах освещенности, по оси абсцисс единицы прежние (пространственная координата).

Вид модуляционной передаточной функции можно получить, если нажать "Step2. МПФ". Здесь в основу математической модели положен механизм одномерного сдвигового быстрого преобразования Фурье. График по горизонтали – в единицах частоты, по вертикали – в единицах контраста.

В любой момент можно увидеть исходную картинку путем нажатия соответствующей кнопки (Image или Меню-> Вид-> Отображение-> Исходная картинка, либо 2). Также существует возможность вернуться к любому из ранее увиденных графиков (Контур, ФРЛ или МПФ) для данного обрабатываемого изображения.

Для удобства считывания полученных в результате обработки значений в программе реализованы такие возможности, как:

- вывод вспомогательной сетки;
- информация о положении курсора на графике.

В программе предусмотрены следующие процедуры:

- открытие из файла изображения для обработки – Меню-> Файл-> Открыть...;
- вывод на экран исходной картинки (изображения, для обработки, загруженного из файла) – Меню-> Вид-> Отображение-> Исходная картинка;
- выделение контура исходного изображения (получение ФРЛ в логарифмическом масштабе) – Меню-> Вид-> Отображение-> Контур картинки;
- получение ФРЛ в линейном масштабе – Меню-> Вид-> Отображение-> ФРЛ;
- вывод на экран МПФ – Меню-> Вид-> Отображение-> МПФ;

Результаты обработки сохраняются автоматически в формате HTML в виде табличных значений соответственно для логарифмической и линейной ФРЛ, а также для МПФ. Существует возможность сохранения пользователем этих же значений в txt-формате. Также можно зафиксировать результаты в файле \*.html, который будет содержать изображение графика, выведенного в данный момент на экран, и значения координат в табличном виде. При этом график сохраняется в графическом JPEG-формате.

Экспортировать результаты обработки изображения можно также в формат PDF. Для этого необходимо, чтобы на производящем обработку компьютере были установлены программные продукты компании Adobe Systems Inc. (например, Adobe Acrobat). Чтобы получить pdf-файл, нужно выбрать Печать или Меню-> Печать... , далее в качестве имени принтера указать название Adobe Acrobat или Acrobat Distiller, затем нажать ОК.

Результаты обработки можно вывести на печать (для размера бумаги А4 и книжной ориентации) непосредственно из программы. Будет напечатан график и табличные значения.

В случае выбора варианта установки страницы, не предусмотренного в данной программе (например, при выборе альбомной ориентации листа), на экран будет выведено об этом предупреждающее сообщение.

Для получения ФРЛ в логарифмическом масштабе в программу заложен алгоритм оконтуривания по столбцам, заключающийся в сканировании массива двумерной выборки и нахождении элемента с минимальным номером в каждом столбце, которому присущ перепад яркости RGB- пиксела в 50%. Эта выборка получается в процессе загрузки исследуемого изображения и приведения его к полутоновому виду. Такое преобразование (к полутоновому виду) осуществляется в первую очередь для того, чтобы

сократить информационную избыточность изображения, оставив в нем только ту информацию, которая нужна для решения конкретной задачи.

### **Заключение**

Проведенные испытания показали эффективность применения программы для реализации метода компьютерной изофотометрии ФРЛ. На кафедре прикладной и компьютерной оптики поставлена лабораторная работа по измерению ФРЛ телевизионной передающей камеры и определению ФПМ системы "фотообъектив – телевизионный приемник изображения". В настоящее время ведется работа по созданию установки и программного обеспечения для реализации метода изофотометрии пограничной (переходной) функции и определения на ее базе ФПМ оптической системы.

### **Литература**

1. Кирилловский В.К. Контроль качества объективов. Л. ЛИТМО. 1984 г.
2. Иванова Т.А., Кирилловский В.К. Проектирование и контроль оптики микроскопов. Л. Машиностроение. 1984 г.

# КОМПОЗИЦИЯ ЗЕРКАЛЬНО-ЛИНЗОВОГО ОБЪЕКТИВА, ОСНОВАННАЯ НА ПРИМЕНЕНИИ КОРРЕКЦИОННЫХ СВОЙСТВ ПЛОСКОПАРАЛЛЕЛЬНОЙ ПЛАСТИНКИ

Г.Э. Романова

Научный руководитель – д.т.н, проф. В.А. Зверев

Рассмотрена возможность композиции зеркально-линзового объектива из двух сферических отражающих поверхностей и плоскопараллельной пластины. Исследованы варианты афокальной двухзеркальной системы с плоскопараллельной пластинкой, рассмотрена возможность сочетания афокальной насадки с дополнительным объективом. Представлен пример расчета зеркально-линзового объектива.

## Введение

Зеркальные поверхности широко применяются в системах различного назначения; их использование в составе длиннофокусных систем объективов позволяет сократить габариты системы. Система, составленная из поверхностей наиболее простой формы – сферических, обладает довольно значительной сферической aberrацией, которую обычно исправляют с помощью компенсаторов, например, с помощью афокальных двухлинзовых, расположенных в параллельном или сходящемся пучке лучей [1], или с помощью пластинки Шмидта [2]. В работе [3] было предложено использовать для компенсации сферической aberrации сферического зеркала плоскопараллельную пластинку. В данной работе предлагается использовать плоскопараллельную пластинку для коррекции сферической aberrации системы из двух сферических зеркал – как афокальной, так и фокусирующей. На основе афокальной системы из двух сферических зеркал с пластинкой можно построить компактный зеркально-линзовый объектив с апланатической коррекцией aberrаций, пример которого приведен в работе.

## Теоретическая часть

Известно [3, 4], что сферическая aberrация плоскопараллельной пластинки определяется выражением:

$$\delta g'_p = \Delta s'_p t g u' = \frac{d}{n} \left[ 1 - n \sqrt{\frac{1 - \sin^2 u'}{n^2 - \sin^2 u'}} \right] t g u'. \quad (1)$$

Это выражение можно представить в виде ряда Тейлора [3]:

$$\delta g'_p = \frac{1}{2} \frac{n^2 - 1}{n^3} d \cdot t g^3 u' - \frac{3}{8} \frac{(n^2 - 1)^2}{n^5} d \cdot t g^5 u' + \frac{5}{16} \frac{(n^2 - 1)^3}{n^7} d \cdot t g^7 u' - \frac{35}{28} \frac{(n^2 - 1)^3}{n^9} d \cdot t g^9 u' + \dots \quad (2)$$

Для зеркальной линзы, ограниченной сферической и плоской отражающими поверхностями, разделенными средой с показателем преломления  $n = -1$ , при предмете, удаленном на бесконечно большое расстояние, сферическая aberrация равна

$$\delta g'_3 = \frac{-r}{2} \left[ 1 - \sqrt{\frac{2\sqrt{1 + t g^2 u'}}{1 + \sqrt{1 + t g^2 u'}}} \right] t g u', \quad (3)$$

где  $\frac{-r}{2} = f'_3$  – фокусное расстояние зеркальной линзы. Дополнение сферического зеркала плоским до зеркальной линзы исключает неудобства со знаками. Разложив (3) в ряд Тейлора, получим:

$$\delta g'_3 = -\frac{1}{8} f'_3 t g^3 u' + \frac{9}{128} f'_3 \cdot t g^5 u' - \frac{49}{1024} f'_3 \cdot t g^7 u' + \frac{1173}{32768} f'_3 \cdot t g^9 u' + \dots \quad (4)$$

Из формул (2) и (4) видно, что возможна взаимная компенсация сферической aberrации сферического зеркала и пластинки: зеркало обладает отрицательной сферической aberrацией, а плоскопараллельная пластинка, наоборот, имеет положительную сферическую aberrацию.

Рассмотрим такую систему в области aberrаций третьего порядка.

Коэффициенты aberrаций третьего порядка определим выражениями вида [5]:

$$S_I^* = B_0 \quad (5)$$

$$S_{II}^* = K_0 + qB_0 \quad (6)$$

$$S_{III}^* = C_0 + 2qK_0 + q^2B_0 \quad (7)$$

$$S_{IV}^* = D_0 \quad (8)$$

где

$$B_0 = \sum_{i=1}^n h_i P_i, \quad K_0 = -\sum_{i=1}^n W_i + \sum_{i=1}^n h_i S_i P_i, \quad C_0 = \sum_{i=1}^n \frac{v_{i+1} \alpha_{i+1} - v_i \alpha_i}{h_i} - 2 \sum_{i=1}^n S_i W_i + \sum_{i=1}^n h_i S_i^2 P_i,$$

$$D_0 = \sum_{i=1}^n \frac{v_i \alpha_{i+1} - v_{i+1} \alpha_i}{h_i}, \quad W_i = \frac{\alpha_{i+1} - \alpha_i}{v_{i+1} - v_i} (v_{i+1} \alpha_{i+1} - v_i \alpha_i), \quad P_i = \frac{\alpha_{i+1} - \alpha_i}{v_{i+1} - v_i} W_i, \quad v_i = \frac{1}{n_i},$$

$$S_i = \sum_{k=2}^i \frac{d_{k-1}}{n_k h_k h_{k-1}},$$

а величина  $q$  связана с положением зрачка и предмета; если предмет расположен на бесконечно большом расстоянии в воздухе, то  $q = -s_{p1}$ , где  $s_{p1}$  – расстояние от первой поверхности системы до входного зрачка.

На рис. 1 показана система, состоящая из сферического зеркала и плоскопараллельной пластинки, сферическая зеркальная поверхность представлена в виде эквивалентного тонкого компонента – зеркальной линзы. Тогда такую систему можно рассматривать как систему вида:

$\alpha_1 = 0$		$n_1 = 1$
$\alpha_2 = -1$	$d_1 = 0$	$n_2 = -1$
$\alpha_3 = 1$	$d_2$	$n_3 = 1$
$\alpha_4 = v$	$d_3 = d$	$n_4 = n$
$\alpha_5 = 1$		$n_5 = 1$

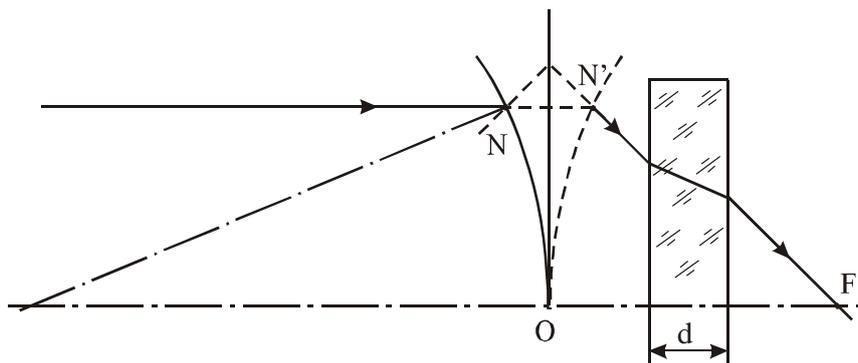


Рис.1. Зеркальная линза с плоскопараллельной пластинкой

Высоты первого параксиального луча на поверхностях:

$$h_1 = 1, \quad h_2 = h_1 - \alpha_2 d_1 = h_1 = 1, \quad h_3 = h_2 - \alpha_3 d_2 = 1 - d_2, \quad h_4 = h_3 - \alpha_4 d_3 = h_3 - vd.$$

Параметры  $W_i, P_i$ :

$$W_1 = \frac{1}{2}, \quad W_2 = 0, \quad W_3 = \frac{v-1}{v-1}(v^2-1) = v^2-1, \quad W_4 = \frac{1-v}{1-v}(1-v^2) = 1-v^2, \quad P_1 = \frac{1}{4},$$

$$P_2 = 0, P_3 = \frac{v-1}{v-1}(v^2-1) = v^2-1, \quad P_4 = \frac{1-v}{1-v}(1-v^2) = 1-v^2.$$

Вспомогательные величины  $S_i$ :

$$S_1 = 0, \quad S_2 = \frac{d_1}{-h_2} = 0, \quad S_3 = \frac{d_2}{h_2 h_3} = \frac{d_2}{h_3}, \quad S_4 = \frac{d_2}{h_3} + \frac{d_3}{nh_3 h_4} = \frac{1}{h_3} \left( d_2 + \frac{d}{nh_4} \right).$$

Полученные соотношения позволяют найти выражение для коэффициента  $B_0$ , определяющего сферическую aberrацию третьего порядка:

$$B_0 = \frac{1}{4} + (1-v^2)(-vd).$$

Положив  $B_0 = 0$ , получаем толщину пластинки  $d$ , необходимую для компенсации сферической aberrации сферического зеркала [6]:

$$d = \frac{1}{4v(1-v^2)} = \frac{n^3}{4(n^2-1)} \quad (9)$$

или

$$\tilde{d} = dv = \frac{n^2}{4(n^2-1)} = \frac{1}{4(1-v^2)}. \quad (10)$$

При исправленной сферической aberrации ( $B_0 = 0$ ) коэффициент комы  $S_{II}^* = K_0$ , при этом

$$K_0 = -\frac{1}{2} + dv(1-v^2). \quad (11)$$

Оптическая толщина пластинки определяется формулой (10), тогда:

$$K_0 = -\frac{1}{2} + \frac{n^2}{4(n^2-1)} \cdot \frac{n^2-1}{n^2} = -\frac{1}{4}. \quad (12)$$

Таким образом, вторая сумма Зейделя в два раза меньше, чем для параболического зеркала.

Выполненный анализ позволяет сделать вывод, что плоская преломляющая поверхность вполне решает задачу коррекции сферической aberrации, вносимой отражающей поверхностью сферической формы.

Рассмотрим задачу коррекции сферической aberrации при помощи плоскопараллельной пластинки в афокальной системе из двух сферических зеркал (рис. 2).

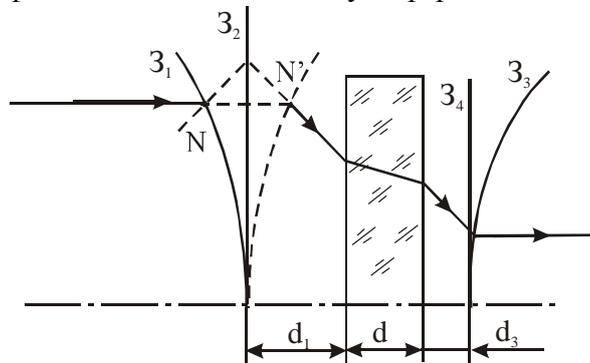


Рис. 2. Афокальная система с плоскопараллельной пластинкой

Систему удобно описать в виде:

$$\begin{array}{lll}
\alpha_1 = 0 & & n_1 = 1 \\
\alpha_2 = -1 & d_{1k} = 0 & n_2 = -1 \\
\alpha_3 = 1 & d_1 & n_3 = 1 \\
\alpha_4 = v & d & n_4 = n \\
\alpha_5 = 1 & d_3 & n_5 = 1 \\
\alpha_6 = 0 & d_{2k} = 0 & n_6 = -1 \\
\alpha_7 = 0 & & n_7 = 1
\end{array}$$

где  $d_{1k}, d_{2k}$  – толщины первой и второй зеркальных линз соответственно. Высоты:

$$h_i: h_1 = 1, h_2 = 1 - d_1, h_3 = 1 - d_1 - \frac{d}{n}, h_4 = h_3 - d_3,$$

где  $h_2, h_3$  – высоты первого парааксиального луча на первой и второй поверхностях пластинки соответственно;  $h_4 = \eta$  – экранирование поверхности первого зеркала второй поверхностью.

Пусть  $P_1 = P_{1_1} + P_{1_2}$ ,  $P_4 = P_{4_1} + P_{4_2}$ , где  $P_{i_1}, P_{i_2}$  ( $i = 1, 4$ ) – параметры для первой и второй поверхностей зеркальных линз. Тогда

$$P_1 = \frac{1}{4}, P_2 = v^2 - 1, P_3 = 1 - v^2, P_4 = -\frac{1}{4}; W_1 = \frac{1}{2}, W_2 = v^2 - 1, W_3 = 1 - v^2, W_4 = -\frac{1}{2}.$$

Вспомогательные величины  $S_i$ :

$$S_1 = 0, S_2 = \frac{d_1}{h_2}, S_3 = \frac{d_1}{h_2} + \frac{d_2}{nh_2h_3}, S_4 = \frac{d_1}{h_2} + \frac{d_2}{nh_2h_3} + \frac{d_3}{h_3h_4} = \frac{1 - h_4}{h_4}.$$

Тогда

$$B_0 = \frac{1}{4} + h_2(v^2 - 1) + \left(h_2 - \frac{d}{n}\right)(1 - v^2) + \frac{\eta}{4}.$$

При  $B_0 = 0$  получим:

$$\tilde{d} = \frac{d}{n} = \frac{1 - \eta}{4(1 - v^2)} = \frac{1 - \eta}{4} \cdot \frac{n^2}{n^2 - 1}. \quad (13)$$

Толщина пластинки в масштабе фокусного расстояния первого зеркала равна

$$d = \frac{1 - \eta}{4} \cdot \frac{n^3}{n^2 - 1}. \quad (14)$$

Коэффициент комы третьего порядка:

$$K_0 = -\sum W_i + \sum h_i S_i P_i = (1 - v^2) \cdot \frac{d}{n} - \frac{1 - \eta}{4}.$$

Подставим в последнее уравнение полученное выражение (13) для оптической толщины плоскопараллельной пластинки, тогда

$$K_0 = (1 - v^2) \cdot \frac{1 - \eta}{4(1 - v^2)} - \frac{1 - \eta}{4} = 0.$$

Таким образом, система из двух зеркал сферической формы и плоскопараллельной пластинки обладает апланатической коррекцией аббераций.

Нарушая телескопичность в такой системе (подобно нарушению телескопичности в системе Мерсенна), можно построить зеркально-линзовые объективы типа Кассегрена или Грегори [7]. На рис. 3 представлена система кассегреновского типа с плоскопараллельной пластиной; зеркальные поверхности представлены в виде зеркальных линз. Так как сходящийся пучок лучей проходит пластинку дважды – после отражений от первого и второго зеркал, то в эквивалентной схеме, показанной на рисунке, необходимо рассматривать две одинаковые пластинки.

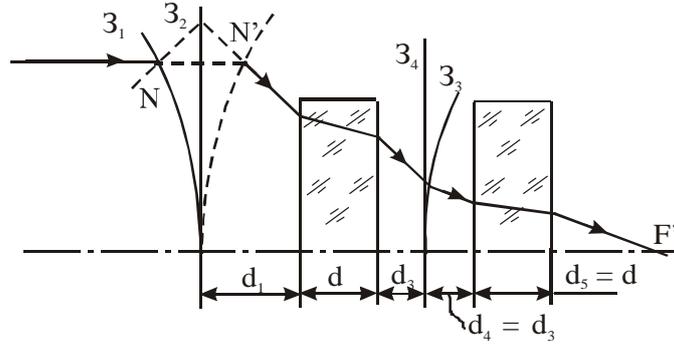


Рис. 3. Система типа Кассегрена с плоскопараллельной пластинкой

Такая схема описывается:

$\alpha_1 = 0$		$n_1 = 1$
$\alpha_2 = -\alpha$	$d_{1k} = 0$	$n_2 = -1$
$\alpha_3 = \alpha$	$d_1$	$n_3 = 1$
$\alpha_4 = \alpha v$	$d_2 = d$	$n_4 = n$
$\alpha_5 = \alpha$	$d_3$	$n_5 = 1$
$\alpha_6 = -1$	$d_{2k} = 0$	$n_6 = -1$
$\alpha_7 = 1$	$d_4 = d_3$	$n_7 = 1$
$\alpha_8 = v$	$d_5 = d$	$n_8 = n$
$\alpha_9 = 1$		$n_9 = 1$

Параметры  $P_i, W_i$ :

$$P_1 = \frac{\alpha^3}{4}, \quad P_3 = \alpha^3(v^2 - 1), \quad P_4 = -\alpha^3(v^2 - 1), \quad P_5 = \frac{(1 + \alpha)^2(1 - \alpha)}{4}, \quad P_7 = v^2 - 1,$$

$$P_8 = -(v^2 - 1), \quad P_2 = P_6 = 0; \quad W_1 = \frac{\alpha^2}{2}, \quad W_3 = \alpha^2(v^2 - 1), \quad W_4 = -\alpha^2(v^2 - 1),$$

$$W_5 = \frac{1 - \alpha^2}{2}, \quad W_7 = v^2 - 1, \quad W_8 = -(v^2 - 1), \quad W_2 = W_6 = 0.$$

Высоты первого параксиального луча на поверхностях:

$$h_1 = h_2 = 1, \quad h_3 = 1 - \alpha d_1, \quad h_4 = h_3 - \alpha d v, \quad h_5 = h_6 = \eta = h_4 - \alpha d_3, \quad h_7 = \eta - d_3,$$

$$h_8 = h_7 - d v.$$

Величина  $B_0$ :

$$B_0 = \sum h_i P_i = \frac{\alpha^3}{4} + (v^2 - 1) d v (\alpha^4 + 1) + \eta \frac{(1 + \alpha)^2(1 - \alpha)}{4}.$$

При  $B_0 = 0$  оптическая толщина пластинки равна

$$d v = \frac{d}{n} = \frac{\eta(1 + \alpha)^2(1 - \alpha) + \alpha^3}{4(\alpha^4 + 1)(1 - v^2)}. \quad (15)$$

Так как исправлена сферическая аберрация ( $S_1^* = B_0 = 0$ ), то коэффициент комы третьего порядка  $S_{II}^* = K_0$ .

$$S_{II}^* = K_0 = -\sum W_i + \sum h_i S_i P_i, \quad \text{причем} \quad \sum_{i=1}^m W_i = \frac{1}{2}.$$

Тогда

$$K_0 = -\frac{1}{2} + \sum h_i S_i P_i. \quad (16)$$

Величины  $S_i$  :

$$S_1 = 0, \quad S_2 = 0, \quad S_3 = \frac{d_1}{h_3}, \quad S_4 = \frac{d_1}{h_3} + \frac{dv}{h_3 h_4} = \frac{d_1 + dv}{h_4} \quad S_5 = \frac{d_1}{h_3} + \frac{dv}{h_4 h_3} + \frac{d_3}{h_4 \eta} = \frac{1 - \eta}{\alpha \eta},$$

$$S_6 = S_5, \quad S_7 = \frac{1 - \eta}{\alpha \eta} + \frac{d_3}{\eta h_6}, \quad S_8 = S_7 + \frac{dv}{h_7 h_8}.$$

Подставив все величины в выражение (16) и преобразовав, получим:

$$S_{II}^* = -\frac{1}{2} + \frac{\eta(1 + \alpha + \alpha^4 - \alpha^5) - \alpha^2(1 - \alpha)}{4\eta(\alpha^4 + 1)}. \quad (17)$$

Снова рассмотрим отдельную отражающую поверхность. Если вторую поверхность плоскопараллельной пластинки совместить с фокальной плоскостью зеркальной линзы, то ее значение для коррекции сферической aberrации утрачивается. Если при этом с центром в заднем фокусе оптической системы сформировать сферу произвольного радиуса, величина которого не превосходит реальной толщины пластинки, и отбросить последнюю из двух полученных таким образом линз, то характер коррекции сферической aberrации не изменится. Для сохранения масштаба изображения (фокусного расстояния) дополним оставшуюся систему апланатическим мениском из того же материала (формально говоря, из материала с тем же показателем преломления). Построенная таким образом система показана на рис. 4.

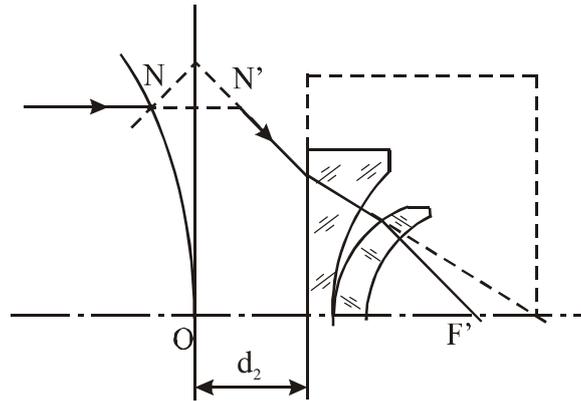


Рис. 4. Сферическое зеркало с компенсатором (зеркало представлено в виде зеркальной линзы)

Такую систему можно описать углами первого параксиального луча в виде:

$\alpha_1 = 0$		$n_1 = 1$
$\alpha_2 = -1$	$d_1 = 0$	$n_2 = -1$
$\alpha_3 = 1$	$d_2$	$n_3 = 1$
$\alpha_4 = v$	$d_3$	$n_4 = n$
$\alpha_5 = v$	$d_4$	$n_5 = 1$
$\alpha_6 = 1$	$d_5$	$n_6 = n$
$\alpha_7 = 1$		$n_7 = 1$

При этом  $P_1 = \frac{1}{4}$ ,  $P_2 = 0$ ,  $P_3 = v^2 - 1$ ,  $P_4 = 0$ ,  $P_5 = P_6 = 0$ , а величина  $B_0 = \frac{1}{4} + h_3(v^2 - 1)$ , причем  $h_3 = h_2 - \alpha_3 d_2 = 1 - d_2$ .

Тогда

$$B_0 = \frac{1}{4} + (1 - d_2)(v^2 - 1).$$

При  $B_0 = 0$  воздушный промежуток  $d_2$ :

$$d_2 = 1 + \frac{1}{4(v^2 - 1)} = 1 - \frac{n^2}{4(n^2 - 1)} = \frac{3n^2 - 4}{4(n^2 - 1)} \quad (18)$$

Коэффициент комы третьего порядка:

$$K_0 = -\frac{1}{2} - v^2 + 1 + h_3 \frac{d_2}{h_3} (v^2 - 1) = \frac{1}{2} - v^2 - (1 - v^2) \left( 1 - \frac{1}{4(1 - v^2)} \right) = -\frac{1}{4} \quad (19)$$

Аналогичным образом можно видоизменить и двухзеркальные системы с плоско-параллельной пластинкой. Рассмотрим еще один вариант афокальной системы – схему из двух отражающих поверхностей сферической формы, где вторая отражающая поверхность в сочетании с плоской преломляющей поверхностью образует зеркально-линзовый компонент, как показано на рис. 5.

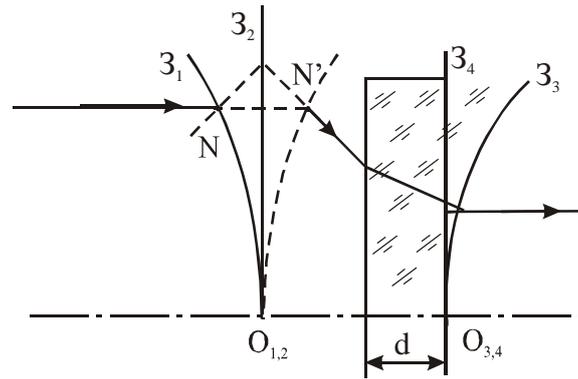


Рис. 5. Вариант афокальной системы с пластинкой

Эту систему можно описать в виде:

$\alpha_1 = 0$		$n_1 = 1$
$\alpha_2 = -1$	$d_1 = 0$	$n_2 = -1$
$\alpha_3 = 1$	$d_2$	$n_3 = 1$
$\alpha_4 = v$	$d_3 = d$	$n_4 = n$
$\alpha_5 = 0$	$d_4 = 0$	$n_5 = -n$
$\alpha_6 = 0$		$n_6 = 1$

В этом случае

$$P_1 = \frac{1}{4}, P_2 = 0, P_3 = v^2 - 1, P_4 = -\frac{1}{4}v^2; W_1 = \frac{1}{2}, W_2 = 0, W_3 = v^2 - 1, W_4 = -\frac{1}{2}v^2;$$

$$S_1 = 0, S_2 = 0, S_3 = \frac{d_2}{h_3}, S_4 = \frac{1}{h_3} \left( d_2 + \frac{d}{nh_4} \right).$$

Величина  $B_0$

$$B_0 = \frac{1}{4} + h_3(v^2 - 1) - \frac{1}{4}h_4v^2.$$

Высота  $h_4 = h_3 - vd$ , отсюда  $h_3 = h_4 + vd$ . Тогда:

$$B_0 = \frac{1}{4} + h_4(v^2 - 1) + v(v^2 - 1)d - \frac{1}{4}h_4v^2.$$

Относительная величина высоты  $h_4$  определяет центральное экранирование по диаметру. Приняв  $h_4 = \eta$ , при  $B_0 = 0$  получаем

$$d = n \left( \frac{1}{4} \cdot \frac{n^2 - \eta}{n^2 - 1} - \eta \right). \quad (20)$$

Воздушный промежуток  $d_2$  при этом равен  $d_2 = h_2 - h_3 = 1 - h_3$ , где  $h_3 = h_4 + \alpha_4 d = \eta + \nu d$ .

Тогда

$$d_2 = 1 - \eta - \frac{1}{4} \cdot \frac{n^2 - \eta}{n^2 - 1} + \eta = \frac{3n^2 - 4 + \eta}{4(n^2 - 1)} \quad (21)$$

или

$$d_2 = \frac{1}{4} \left( 3 - \frac{1 - \eta}{n^2 - 1} \right). \quad (22)$$

Оптическая толщина компенсационной плоскопараллельной пластинки зависит как от материала, из которого она изготовлена, так и от экранирования. Значение показателя преломления  $n_{\max}$ , при котором толщина пластинки обращается в ноль, определяется выражением

$$n_{\max} = \sqrt{\frac{3\eta}{4\eta - 1}}. \quad (23)$$

При исправленной сферической аберрации ( $B_0 = 0$ ) коэффициент комы третьего порядка равен

$$S_{II}^* = K_0 = -\sum W_i + \sum h_i S_i P_i,$$

где

$$K_0 = \frac{1 - \nu^2}{2} + d_2(\nu^2 - 1) - \left( d_2 + \frac{d\nu(1 - d_2)}{h_3} \right) \frac{1}{4} \nu^2 = \frac{1 - \nu^2}{2} + d_2(\nu^2 - 1) - \frac{1}{4} \nu^2 (d_2 + d\nu).$$

Подставляя в последнее уравнение формулы (20) и (22), получим:

$$\begin{aligned} K_0 &= \frac{1 - \nu^2}{2} - \frac{3n^2 - 4 + \eta}{4n^2} - \frac{1}{4n^2} \left( \frac{3n^2 - 4 + \eta}{4(n^2 - 1)} + \frac{n^2 - \eta - 4\eta(n^2 - 1)}{4(n^2 - 1)} \right) = \\ &= -\frac{1 - n^2}{2n^2} - \frac{3n^2 - 4 + \eta}{4n^2} - \frac{1 - \eta}{4n^2} = -\frac{n^2 - 1}{4n^2}. \end{aligned}$$

Окончательно получим

$$K_0 = \frac{1 - n^2}{4n^2}. \quad (24)$$

При  $n = 1.5$   $K_0 = -0.14$ , а при  $n = 2$   $K_0 = -0.1875$ . Таким образом, хотя система и не обладает строго апланатической коррекцией аберраций, коэффициент комы достаточно мал.

Достоинством и отличительной особенностью такой системы (второго типа) по сравнению с уже рассмотренной афокальной схемой (первого типа) является то, что толщина плоскопараллельной пластинки может быть любой за счет соответствующего выбора величины экранирования и значения показателя преломления. Например, можно сделать толщину пластинки минимальной для полного диаметра первого зеркала, тогда пластинка будет несущей для вторичного зеркала и, следовательно, не будет необходимости в стяжках для крепления вторичного зеркала.

Для системы касегреновского типа также возможна конфигурация системы с пластинкой, когда вторая зеркальная поверхность нанесена на вторую поверхность пластинки, как показано на рис. 6.

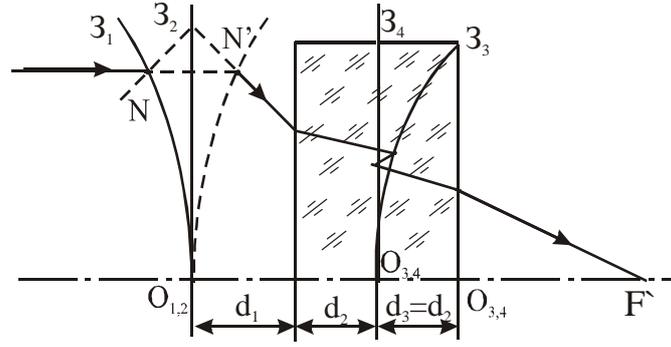


Рис. 6. Вариант объектива типа Кассегрена с плоскопараллельной пластинкой

В этом случае система описывается так:

$$\begin{array}{lll}
 \alpha_1 = 0 & & n_1 = 1 \\
 \alpha_2 = -\alpha & d_{1k} = 0 & n_2 = -1 \\
 \alpha_3 = \alpha & d_1 & n_3 = 1 \\
 \alpha_4 = \alpha v & d & n_4 = n \\
 \alpha_5 = -v & d_{2k} = 0 & n_5 = -n \\
 \alpha_6 = v & d & n_6 = n \\
 \alpha_7 = 1 & & n_7 = 1
 \end{array}$$

Тогда

$$P_1 = \frac{\alpha^3}{4}, \quad P_2 = 0, \quad P_3 = \alpha^3(v^2 - 1), \quad P_4 = \frac{(1 + \alpha)^2(1 - \alpha)}{4}v^2, \quad P_5 = 0, \quad P_6 = 1 - v^2;$$

$$W_1 = \frac{\alpha^2}{2}, \quad W_2 = 0, \quad W_3 = \alpha^2(v^2 - 1), \quad W_4 = \frac{1 - \alpha^2}{2}v^2, \quad W_5 = 0, \quad W_6 = 1 - v^2.$$

Высоты  $h_i$ :  $h_1 = h_2 = 1$ ,  $h_3 = 1 - \alpha d_1$ ,  $h_4 = h_5 = \eta = h_3 - \alpha v d$ ,  $h_6 = \eta - v d$ .

Первая сумма Зейделя:

$$\begin{aligned}
 S_I = B_0 &= \frac{\alpha^3}{4} + (\eta + \alpha v d)\alpha^3(v^2 - 1) + \eta \frac{(1 + \alpha)^2(1 - \alpha)}{4}v^2 + (\eta - v d)(1 - v^2) = \\
 &= \frac{\alpha^3}{4} + \eta \frac{(1 + \alpha)^2(1 - \alpha)}{4}v^2 + \eta(v^2 - 1)(\alpha^3 - 1) + v d(v^2 - 1)(1 + \alpha^4).
 \end{aligned}$$

Приравняв  $S_I = 0$ , получим

$$v d = -\frac{\alpha^3 + \eta(1 + \alpha)^2(1 - \alpha)v^2 + 4\eta(v^2 - 1)(\alpha^3 - 1)}{4(v^2 - 1)(1 + \alpha^4)}. \quad (25)$$

Коэффициент комы третьего порядка:

$$S_{II}^* = K_0 = -\sum W_i + \sum h_i S_i P_i,$$

где  $\sum W_i = \frac{\alpha^2}{2} + \alpha^2(v^2 - 1) + \frac{1 - \alpha^2}{2}v^2 + 1 - v^2 = \frac{(1 - \alpha^2)(1 - v^2) + 1}{2}$ .

Величины  $S_i$ :

$$S_1 = 0, \quad S_2 = 0, \quad S_3 = \frac{d_1}{h_3}, \quad S_4 = \frac{d_1}{h_3} + \frac{d v}{h_3 h_4} = \frac{d_1 + d v}{\eta}, \quad S_5 = S_4, \quad S_6 = S_4 + \frac{d v}{\eta h_6}.$$

Подставляя эти выражения в формулу, определяющую вторую сумму Зейделя, и используя уравнение (25), получим:

$$K_0 = -\frac{1}{2} + \frac{\eta(1 - \alpha^6)(2 - v^2) + \eta \alpha^2(1 - \alpha^2)(2 - 3v^2) + \eta \alpha[\alpha(1 + \alpha^4) + v^2(1 - \alpha^4)] - \alpha^2(1 - \alpha)}{4\alpha(1 + \alpha^4)}. \quad (26)$$

Афокальную систему, показанную на рис. 2 (типа I) или на рис. 5 (типа II), можно сочетать с дополнительным тонким компонентом. Полученная схема может использоваться, например, в качестве длиннофокусного компактного объектива. Для достижения апланатической коррекции aberrаций в схеме компонент в обоих случаях должен быть исправлен в отношении сферической aberrации. Но в первом случае (для афокальной насадки, показанной на рис. 2) должна быть исправлена также и кома, во втором случае (афокальная система рис. 5) компонент можно использовать, чтобы исправить кому в системе в целом. Для этого необходимо, чтобы параметр  $W_k$  компонента определялся формулой

$$W_k = -\frac{\alpha^2(n^2 - 1)}{4n^2}. \quad (27)$$

Здесь  $n$  – показатель преломления материала пластинки, а угол  $\alpha$  – угол первого параксиального луча в пространстве после первого зеркала (после зеркальной линзы). Величина этого угла определяет, во сколько раз фокусное расстояние первого зеркала короче фокусного расстояния системы, состоящей из афокальной насадки и тонкого компонента.

### Примеры расчета

В качестве иллюстрации были рассчитаны схемы афокальной и фокусирующей схем с плоскопараллельной пластинкой. В обеих системах относительное отверстие первого зеркала было выбрано 1:1.5, диаметр первого зеркала 100 мм, центральное экранирование 0.3. В афокальной системе типа I максимальная волновая сферическая aberrация составляла  $0.1\lambda$ , при этом неизопланатизм не превышал величины 0.0052%. В афокальной системе типа II волновая сферическая aberrация  $0.055\lambda$ , но значительно больший неизопланатизм – 0.813%.

В схемах типа Кассегрена с плоскопараллельной пластинкой при хорошей коррекции сферической aberrации неизопланатизм остается довольно значительным.

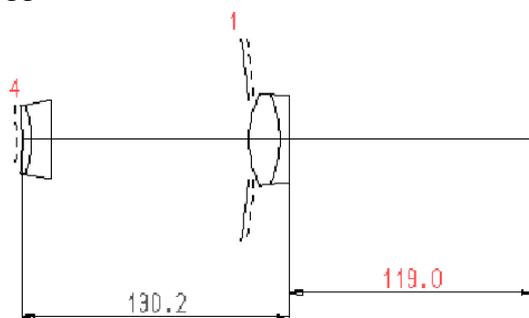


Рис. 7. Зеркально-линзовый объектив  $f' = 450$ , 1:4.5,  $2\omega = 2^\circ$

При сочетании афокальной насадки типа II с дополнительным объективом (склежкой) был рассчитан зеркально-линзовый объектив (рис. 7) с фокусным расстоянием  $f' = 450$ , относительным отверстием 1:4.5 и угловым полем в пространстве предметов  $2\omega = 2^\circ$ . Aberrации осевого пучка приведены в табл. 1.

H	$\Delta Y'$	W	$\eta, \%$
50	0.0112	-0.0678	$-0.7 \cdot 10^{-4}$
43.3	0.00102	-0.214	-0.0374
35.4	-0.00127	-0.17	-0.0481
25	-0.002	-0.0607	-0.035
0	0	0	0

Таблица 1. Aberrации осевого пучка зеркально-линзового объектива

При хорошей коррекции сферической аберрации и комы остается неисправленным астигматизм, анализ возможностей исправления которого будет являться предметом следующей работы.

### **Заключение**

Плоскопараллельная пластинка позволяет решить задачу компенсации сферической аберрации зеркальных систем, состоящих из одной или двух сферических поверхностей. Афокальную систему со стигматической или апланатической коррекцией аберраций можно сочетать с дополнительным компонентом, в этом случае полученная схема может использоваться в качестве компактного длиннофокусного объектива. Системы типа Кассегрена (или Грегори) из двух сферических зеркал с плоскопараллельной пластинкой могут применяться в качестве объективов телескопических систем.

### **Литература**

1. Чуриловский В. Н. Теория хроматизма и аберраций третьего порядка. Л.: Машиностроение, 1968.
2. Слюсарев Г. Г. Расчет оптических систем. Л.: Машиностроение, 1975
3. Зверев В. А., Хлусова Н. И. Применение плоскопараллельной пластинки для исправления сферической аберрации // ОМП, ОНТИ ГОИ, № 9, 1972, с. 24– 25
4. Русинов М. М. Техническая оптика. Л.: Машиностроение, 1979
5. Зверев В. А. Основы геометрической оптики. Санкт-Петербург: ИТМО, 2002
6. Зверев В. А. Оптическая система сферического зеркала с плоскопараллельной пластинкой // ОМП, ОНТИ ГОИ, № 11, 1986, с. 30–32
7. Русинов М. М. Композиция оптических систем. Л.: Машиностроение, 1989

# ВЗАИМОСВЯЗЬ ПАРАМЕТРОВ ПЕРЕМЕЩЕНИЯ ЭЛЕМЕНТОВ ПРИНЦИПАЛЬНОЙ СХЕМЫ ОПТИЧЕСКОЙ СИСТЕМЫ ПЕРЕМЕННОГО УВЕЛИЧЕНИЯ

Т.В. Точилина

Научный руководитель – д.т.н., проф. В.А. Зверев

В результате анализа взаимосвязи параметров перемещения элементов получена параметрическая модель принципиальной схемы оптической системы переменного увеличения. Показано, что величину расфокусировки изображения можно выразить через параметры  $n$ -компонентной схемы оптической системы переменного увеличения и, следовательно, определить значения параметров при заданной или требуемой величине расфокусировки.

Простейшим вариантом базовой схемы оптической системы переменного увеличения является однокомпонентная схема [1]. Любую оптическую систему из произвольного числа элементов (линз, отражающих поверхностей) при конечном расстоянии  $d_{HH'}$  между её главными плоскостями и отличной от нуля оптической силе  $\varphi$  будем называть однокомпонентной, если при всех возможных подвижках она перемещается как единое целое. При  $d_{HH'} \approx 0$  оптическую систему принято называть тонким компонентом. Не нарушая общности подхода к решению задачи, будем считать оптическую систему в рассматриваемом случае тонким компонентом.

Определим взаимосвязь параметров линейных перемещений в однокомпонентной схеме оптической системы переменного увеличения. Для этого обратимся к рис.1, где точками  $A_{0i}$  и  $A'_0$  обозначено положение осевых точек предмета и изображения при исходном положении оптического компонента  $\varphi_{0i}$ , точка пересечения совмещённых главных плоскостей которого с оптической осью расположена в точке  $O_{0i}$ .

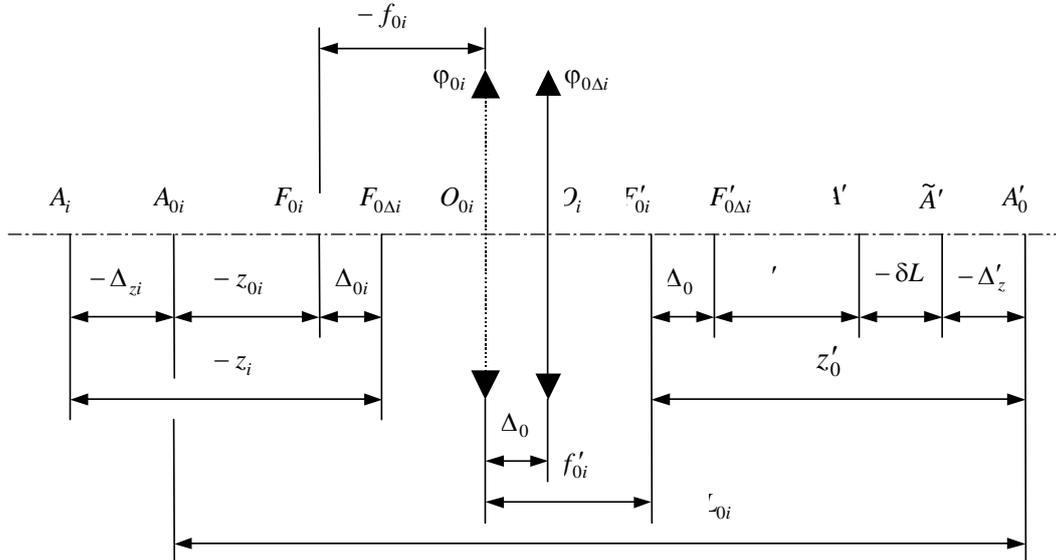


Рис. 1. Однокомпонентная схема оптической системы переменного увеличения

При смещении осевой точки предмета из точки  $A_0$  в точку  $A_i$ , а оптической системы  $\varphi_{0i}$  из положения, определяемого точкой  $O_{0i}$ , в положение, определяемое точкой  $O_i$ , параксиальное изображение точки  $A_i$  сместится из точки  $A'_0$  в точку  $A'$ . При смещении плоскости фиксирования изображения на расстояние  $A'_0\tilde{A}'$  отрезок  $A'\tilde{A}'$  определит величину остаточной расфокусировки изображения  $\delta L$ . Величины рассматриваемых

ваемых смещений элементов схемы обозначены через  $\Delta_0, \Delta_{zi}$  и  $\Delta'_z$ , смысл которых вполне очевиден из рисунка.

Показатели преломления сред в пространствах предметов и изображений будем считать равными единице. При этом взаимосвязь отрезков  $z_{0i}$  и  $z'_{0i}$  определяется формулой Ньютона:

$$z_{0i}z'_{0i} = -f_{0i}'^2, \quad (1)$$

а поперечное увеличение изображения, образованного оптической системой, фокусное расстояние которой равно  $f_0'$ , определяется соотношением

$$V_{0i} = \frac{f_{0i}'}{z_{0i}} = -\frac{z'_{0i}}{f_{0i}'}. \quad (2)$$

В соответствии с рисунком имеем:

$$z_i = z_{0i} + \Delta_{zi} - \Delta_{0i} \quad (3)$$

$$z'_{0i} + \Delta'_{zi} = \Delta_{0i} + z'_i - \delta L_i. \quad (4)$$

Из соотношения (4) следует, что

$$\delta L_i = \Delta_{0i} + z'_i - z'_{0i} - \Delta'_{zi}.$$

Используя соотношения (2) и (3), преобразуем это выражение к виду

$$\delta L_i = \Delta_{0i} - \Delta'_{zi} + \frac{f_{0i}'(\Delta_{zi} - \Delta_{0i})V_{0i}^2}{f_{0i}' + (\Delta_{zi} - \Delta_{0i})V_{0i}}. \quad (5)$$

В соответствии с рис. 1 расстояние

$$L_i = -z_i + 2f_{0i}' + z'_i = L_{0i} - \Delta_{zi} + \Delta'_{zi} + \delta L_i, \quad (6)$$

где  $L_{0i} = -z_{0i} + 2f_{0i}' + z'_{0i}$ .

Используя формулу Ньютона (1), легко показать, что

$$L_i = -\frac{(f_0' - z_i)^2}{z_i} = \frac{(f_0' + z'_i)^2}{z'_i}. \quad (7)$$

При этом

$$L_i = \frac{(\Delta_{0i} - \Delta_{zi} - z_{0i} + f_{0i}')^2}{\Delta_{0i} - \Delta_{zi} - z_{0i}} = \frac{(f_0' + z'_{0i} + \Delta'_{zi} - \Delta_{0i} + \delta L_i)^2}{z'_{0i} + \Delta'_{zi} - \Delta_{0i} + \delta L_i}$$

и, соответственно,

$$L_{0i} = -\frac{(f_0' - z_{0i})^2}{z_{0i}} = \frac{(f_0' + z'_{0i})^2}{z'_{0i}}.$$

Для оптического компонента с оптической силой  $\varphi_{0i+1}$  при смещениях элементов схемы отрезок  $\delta L_{i+1}$  определится выражением, аналогичным выражению (5), в виде

$$\delta L_{i+1} = \Delta_{0i+1} - \Delta'_{zi+1} + \frac{f_{0i+1}'(\Delta_{zi+1} - \Delta_{0i+1})V_{0i+1}^2}{f_{0i+1}' + (\Delta_{zi+1} - \Delta_{0i+1})V_{0i+1}}. \quad (8)$$

Вполне очевидно, что в двухкомпонентной схеме оптической системы переменного увеличения, образованной сочетанием компонентов  $\varphi_{0i}$  и  $\varphi_{0i+1}$ , смещение

$$\Delta_{zi+1} = \Delta'_{zi} + \delta L_i, \quad (9)$$

при этом в соответствии с выражением (5) величина

$$\Delta_{zi+1} = \Delta_{0i} + \frac{f_{0i}'(\Delta_{zi} - \Delta_{0i})V_{0i}^2}{f_{0i}' + (\Delta_{zi} - \Delta_{0i})V_{0i}}. \quad (10)$$

Применяя последовательно выражения (3), (4), (5), (8) и (9) или (10), можем определить смещение осевой точки изображения в общем случае, когда схему оптической

системы переменного увеличения образует сочетание  $n$  тонких компонентов  $\Phi_{01}, \Phi_{02}, \dots, \Phi_{0n+1}, \Phi_{0n}$ , выражением (5) в виде:

$$\delta L_n = \Delta_{0n} - \Delta'_{zn} + \frac{f'_{0n}(\Delta_{zn} - \Delta_{0n})V_{0n}^2}{f'_{0n} + (\Delta_{zn} - \Delta_{0n})V_{0n}}, \quad (11)$$

где

$$\begin{aligned} \Delta_{zn} &= \Delta'_{zn-1} + \delta L_{n-1}, \\ \Delta_{zi+1} &= \Delta'_{zi} + \delta L_i, \\ \Delta_{z3} &= \Delta'_{z2} + \delta L_2, \\ \Delta_{z2} &= \Delta'_{z1} + \delta L_1. \end{aligned}$$

При этом в соответствии с выражением (6) имеем

$$L_\Sigma = \sum_{i=1}^n L_i = \sum_{i=1}^n (L_{0i} - \Delta_{zi} + \Delta'_{zi} + \delta L_i) = \sum_{i=1}^n (L_{0i} - \Delta_{zi} + \Delta_{zi+1}). \quad (12)$$

В общем случае линейное смещение элементов схемы оптической системы переменного увеличения определяется зависимостью  $\Delta = \Delta(m)$ , где  $m$  – независимая изменяемая величина. В простейшем случае эта зависимость имеет вид

$$\Delta = km,$$

где  $k$  – коэффициент пропорциональности (передаточное отношение).

При этом выражения (10) и (11) можно переписать:

$$\delta L_n = (k_{0n} - k'_{zn})m + \frac{f'_{0n}(k_{zn} - k_{0n})mV_{0n}^2}{f'_{0n} + (k_{zn} - k_{0n})mV_{0n}} \quad (13)$$

$$L_\Sigma = \sum_{i=1}^n [L_{0i} + \delta L_i + (k'_{zi} - k_{zi})m]. \quad (14)$$

При  $n=1$ , т.е. для однокомпонентной схемы оптической системы переменного увеличения, имеем

$$\delta L_1 = (k_{01} - k'_{z1})m + \frac{f'_{01}(k_{z1} - k_{01})mV_{01}^2}{f'_{01} + (k_{z1} - k_{01})mV_{01}}.$$

Выразив линейные величины в масштабе фокусного расстояния, получаем

$$\delta L_1 = (k_{01} - k'_{z1})m + \frac{(k_{z1} - k_{01})mV_{01}^2}{1 + (k_{z1} - k_{01})mV_{01}}. \quad (15)$$

Таким образом, в работе получены соотношения, определяющие зависимость величины остаточной расфокусировки изображения в  $n$ -компонентной схеме оптической системы переменного увеличения при линейной взаимосвязи перемещения компонентов от параметров схемы. Показано, что остаточная величина расфокусировки при этом определяется выражением, подобным дроби Чебышева, что позволяет применить его для минимизации величины расфокусировки изображения, образованного  $n$ -компонентной оптической системой переменного увеличения.

## Литература

1. Журова С.А., Зверев В.А. Основы композиции принципиальных схем оптических систем переменного увеличения. // Оптический журнал. 1999. №10. С.68-85.

# ВЛИЯНИЕ ПАРАМЕТРОВ ЭЛЕМЕНТОВ ОПТИЧЕСКОЙ СИСТЕМЫ МИКРОСКОПА НА ЕГО РАЗРЕШАЮЩУЮ СПОСОБНОСТЬ

О.А. Виноградова, Д.Н. Фролов

Научный руководитель – д.т.н., проф. В.А. Зверев

Рассматривается применение результатов теории формирования частично-когерентного изображения для получения оптимального соотношения числовых апертур конденсора и микрообъектива с целью достижения максимальной разрешающей способности на микроскопе.

Известно [1], что теория формирования частично-когерентного изображения до недавнего времени применялась только в проекционной фотолитографии, где необходимым для получения качественных результатов является создание условий образования частично-когерентного освещения. Применяя подобные решения при построении осветительных систем микроскопов, можно получить дополнительные возможности повышения разрешения в микроскопии. При этом важную роль играет выбор оптимальных значений выходной числовой апертуры конденсора и входной апертуры объектива.

Ограничение разрешающей способности в микроскопе обусловлено влиянием основных элементов его оптической системы. Рассмотрим влияние основных частей этой системы, для чего приведем формулы, отражающие основные соотношения в микроскопе. Так, видимое увеличение изображения, образованного оптической системой микроскопа, определяется формулами вида:

$$\Gamma_m = \frac{tg\omega'}{tg\omega} = \frac{\ell}{f'_{ок}} \cdot \frac{L}{\ell} = V_{об} \cdot \Gamma_{ок} \quad \text{или} \quad \Gamma_m = \frac{L}{f'_m},$$

где фокусное расстояние оптической системы микроскопа  $f'_m = \frac{1}{V_{об}} \cdot f'_{ок}$ . Но

$$V_{об} = -\frac{z'}{f'_{об}} = -\frac{\Delta_0}{f'_{об}}, \quad \text{при этом} \quad \Gamma_m = -\frac{\Delta_0}{f'_{об}} \cdot \frac{L}{f'_{ок}}, \quad \text{где } \Delta_0 \text{ – оптическая длина тубуса.}$$

При формальном подходе видимое увеличение изображения на световом микроскопе можно сделать сколь угодно большим. Это было бы так, если бы оптика микроскопа в сочетании с глазом наблюдателя обладала абсолютной разрешающей способностью.

При наблюдении с помощью микроскопа одних и тех же предметов при одном и том же видимом увеличении изображения можно использовать либо объектив большого увеличения с окуляром малого увеличения, либо объектив с той же числовой апертурой, но меньшего увеличения и более сильный окуляр. Однако микроскопистам известно, что в первом случае качество наблюдаемого изображения значительно лучше. Эти соображения позволяют рекомендовать в качестве штатного применять окуляр  $\Gamma_{ок}=10^x$ .

Таким образом, из последней формулы следует, что видимое увеличение изображения при постоянной оптической длине тубуса  $\Delta_0$ , определяемое произведением фокусных расстояний объектива и окуляра, остается неизменным при  $f'_{об} \cdot f'_{ок} = \text{const}$ .

Изображение наблюдаемого предмета, образованное сочетанием микрообъектива большого увеличения с окуляром меньшего увеличения, обладает более высоким качеством [2]. Кроме того, в случае применения окуляров небольших и средних увеличений обеспечивается большее удобство при работе из-за значительного удаления выходного зрачка.

Известно, что разрешающая способность оптической системы, отображающей распределение яркости на предмете в распределение освещенности на изображении, определяется радиусом центрального максимума в дифракционном изображении точки.

Диаметр выходного зрачка микроскопа определяет диаметр входного зрачка глаза. Размер диаметра зрачка глаза непостоянен, зависит от яркости фона и размера картины, воспринимаемой глазом [3], и лежит в пределах 2–8 мм.

На рис. 1 представлены полученные А. Арнюльфом кривые, характеризующие разрешающую способность оптической системы микроскопа в сочетании с глазом для ряда значений контраста мира Фучо, равных 1, 0.5 и 0.06.

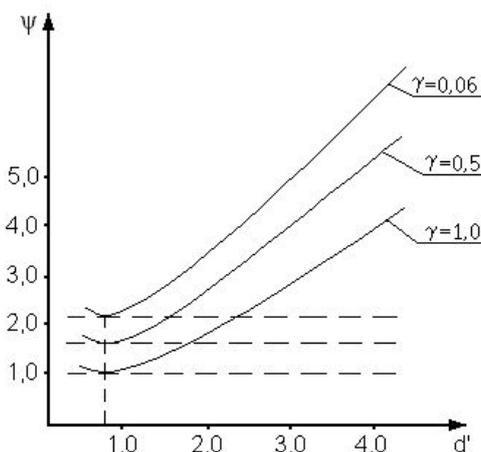


Рис. 1 Зависимость разрешающей способности микроскопа от диаметра выходного зрачка

Исследования, выполненные А. Арнюльфом, показали, что реальное разрешение структуры изображения, образованного безабберационной оптикой микроскопа, различно при различных диаметрах выходного зрачка. При этом в общем случае минимальное расстояние  $\rho$  между двумя разрешаемыми штрихами мира (тест-объекта) Фуко в плоскости предмета можно определить выражением вида:  $\rho = \psi \frac{\lambda}{2n \sin \sigma}$ .

Анализ результатов исследований, выполненных А. Арнюльфом, позволил оценить зависимость разрешающей способности микроскопа от диаметра выходного зрачка приближенным эмпирическим выражением следующего вида:

$$\psi = 1.03 + 1.4 \frac{(d' - 0.7)^2}{d'}$$

Из габаритных соотношений в оптической схеме микроскопа следует, что диаметр выходного зрачка микроскопа равен

$$d' = 2f'_m \cdot n \sin \sigma = 2f'_m \cdot V_{об} \cdot n' \cdot \sin \sigma' = 2f'_{ок} \cdot n' \cdot \sin \sigma'$$

При  $f'_{ок} = 25$  мм ( $\Gamma_{ок} = 10^x$ ) и при соблюдении условия  $0.5 \text{ мм} \leq d' \leq 1 \text{ мм}$  должно соблюдаться условие  $0.01 \leq \sin \sigma' \leq 0.02$ .

Отсюда следует, что полезное значение передней числовой апертуры микрообъектива должно удовлетворять следующему условию:

$$0.01 \cdot V_{об} \leq n \sin \sigma \leq 0.02 \cdot V_{об}$$

При этом полезная величина поперечного увеличения изображения, образованного микрообъективом, должна выбираться в пределах

$$50 \cdot n \sin \sigma \leq V_{об} \leq 100 \cdot n \sin \sigma$$

Изображение в микроскопе зависит от метода освещения объекта. Рассмотрим влияние метода освещения на разрешающую способность микроскопа. Для исследования под микроскопом прозрачного объекта необходимо осветить его проходящим светом. При этом свет, прошедший через объект, фокусируется в плоскость изображения,

образованного объективом микроскопа. Для получения требуемой освещенности на объекте определенной площади применяют осветительную систему.

Существуют различные методы освещения. При критическом освещении величина наблюдаемого предмета значительно больше эффективного размера пятна Эри в изображении точки источника, образованном конденсором, и комплексная степень когерентности для любой пары точек плоскости предмета объектива совпадает со степенью когерентности, обусловленной некогерентным источником света, проходящего через конденсор. Кроме того, степень когерентности не зависит от aberrаций конденсора [4]. Поскольку разрешающая сила микроскопа зависит только от степени когерентности света, освещающего предмет, и от свойств объекта, aberrации конденсора совершенно не влияют на разрешающую силу микроскопа.

В микроскопии используется также метод освещения по Келеру. Можно показать, что комплексная степень когерентности света, падающего на плоскость предмета, наблюдаемого с помощью микроскопа, одинакова как при критическом освещении, так и при освещении по Келеру.

Известная формула [4] распределения интенсивности света в изображении от двух точек позволяет исследовать зависимость разрешающей способности микроскопа от отношения числовых апертур конденсора и микрообъектива  $\mu$ :

$$I(P') = \left( \frac{2J_1(v_1)}{v_1} \right)^2 + \left( \frac{2J_1(v_2)}{v_2} \right)^2 + 2 \left( \frac{2J_1(\mu \cdot v_{12})}{v_{12}} \right) \left( \frac{2J_1(v_1)}{v_1} \right) \left( \frac{2J_1(v_2)}{v_2} \right).$$

На рис. 2 изображена кривая, отражающая эту зависимость. Из вида кривой следует, что наилучшее разрешение получается при  $\mu \approx 1.5$ . Величина  $\psi$  в этом случае даже несколько меньше значения 0.61, получающегося при некогерентном освещении. Однако важно заметить, что при изменении величины  $\mu$  в диапазоне от  $\mu \approx 1.0$  до  $\mu \approx 1.8$  разрешающая способность микроскопа практически постоянна, что определяет требования к параметрам осветительной системы микроскопа.

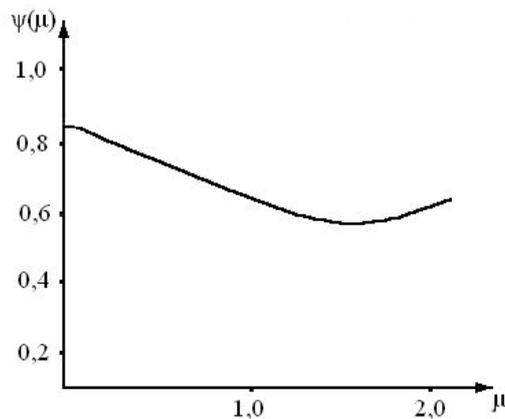


Рис. 2. Зависимость разрешающей способности микроскопа от отношения числовых апертур конденсора и микрообъектива

В соответствии с инвариантом Лагранжа-Гельмгольца параметры микрообъектива взаимосвязаны следующим соотношением:

$$n_o l_o \sin \sigma_o = n'_o l'_o \sin \sigma'_o = J_o.$$

Для осветительной системы имеем:

$$n_s l_s \sin \sigma_s = n'_s l'_s \sin \sigma'_s = n_o l_o \sin \sigma'_o = J_s.$$

Из анализа влияния параметра  $\psi(\mu)$  на разрешающую способность микроскопа следует, что инвариант  $J_s$  должен удовлетворять условию:

$$J_o \leq J_s \leq 1.8 \cdot J_o.$$

Эти соотношения определяют взаимосвязь размера источника излучения и входной числовой апертуры осветительной оптической системы микроскопа (коллектора) и требуемые их значения.

В работе с применением результатов теории формирования частично-когерентного изображения получено оптимальное соотношение числовых апертур конденсора и микрообъектива с целью достижения максимальной разрешающей способности на микроскопе.

### **Литература**

1. Вычислительная оптика: Справочник / М.М. Русинов, А.П. Грамматин, П.Д. Иванов и др. Л.: Машиностроение, 1984.
2. Андреев Л.Н., Грамматин А.П., Соколова Т.И. Влияние распределения увеличения в микроскопе между объективом и окуляром на качество изображения. // ОМП. 1979. №1. С.51–52.
3. Пинегин Н.И. Влияние яркости и размеров поля зрения на диаметр зрачка // Тр. I конф. по физиологической оптике. М.-Л. 1936, с.369
4. Борн М., Вольф Э. Основы оптики. М.: Наука, 1970.

# АНАЛИЗ ДЕЦЕНТРИРОВОК ПОВЕРХНОСТЕЙ В ОПТИЧЕСКОЙ СИСТЕМЕ НА ОСНОВЕ МЕТОДОВ ГАУССОВОЙ ОПТИКИ

К.В. Ежова

Научный руководитель – к.т.н., доц. Н.Д. Толстоба

Рассматривается возможность применения аппарата матричной оптики для анализа влияния децентрировок и наклонов поверхностей элементов оптической системы в процессе сборки и юстировки. Приводятся формулы для учета влияния децентрировок и наклонов при прохождении луча через децентрированные поверхности оптической системы.

## Введение

На протяжении долгого времени методы гауссовой оптики применяются для расчета нулевых лучей через поверхности оптической системы, что, в свою очередь, позволяет найти матрицу преобразования системы и вычислить параксиальные характеристики системы.

Существуют всего два основных вида матриц преобразования, описывающих два простых преобразования – перенос луча в свободном пространстве и преломление луча на преломляющей поверхности или в оптической системе. Эти матрицы и формируют матрицу преобразования оптической системы  $\mathbf{G}$ .

Для расчета нулевого луча формируют вектор входных координат луча  $\mathbf{b}$ , элементами которого являются высота  $h$  и угол наклона падающего луча  $\varphi$ . Затем, используя матрицу преобразования оптической системы  $\mathbf{G}$ , производят вычисление выходных координат луча:

$$\mathbf{b}' = \mathbf{G} \cdot \mathbf{b}. \quad (1)$$

Если матрицу преобразования формировать не сразу для всей оптической системы в целом, а постепенно по поверхностям, то можно вычислить координаты луча на всех поверхностях оптической системы. В этом случае аппарат матричной оптики возможно применять и при анализе децентрировок и наклонов поверхностей элементов оптической системы в процессе сборки и юстировки.

## Матрицы преломления и переноса

Любую оптическую систему можно представить совокупностью оптических поверхностей, разделенных промежутками.

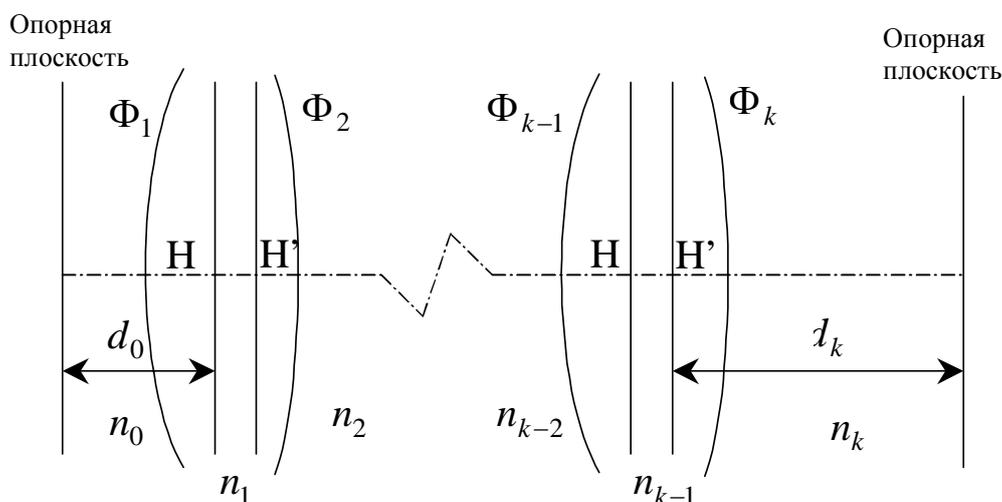


Рис. 1. Оптическая система

Матрица преобразования такой системы состоит из произведения матриц преломления на оптических поверхностях  $\mathbf{R}_n$  и матриц переноса между поверхностями  $\mathbf{T}_n$ :

$$\mathbf{G} = \mathbf{R}_n \mathbf{T}_n \dots \mathbf{R}_1 \mathbf{T}_1 \mathbf{T}_0. \quad (2)$$

Здесь  $\mathbf{R}_n$  и  $\mathbf{T}_n$  вычисляются по формулам:

$$\mathbf{R}_n = \begin{pmatrix} 1 & 0 \\ -\Phi_n & 1 \end{pmatrix}, \quad (3)$$

$$\mathbf{T}_n = \begin{pmatrix} 1 & d_n \\ 0 & 1 \end{pmatrix}; \quad (4)$$

$\Phi_n$  – оптическая сила преломляющей поверхности,  $d_n$  – расстояние между поверхностями,  $n_n$  – показатель преломления среды между поверхностями оптической системы. Зная значение элементов матрицы преобразования оптической системы, можно без труда определить параксиальные характеристики оптической системы, используя для этого известные выражения.

### Влияние децентрировок и наклонов на координаты луча

Естественно, что и при децентрировке оптической поверхности и при ее наклоне изображение осевой точки в плоскости анализа оказывается смещенным с оси. Это происходит из-за изменения хода параксиального луча в такой системе по сравнению с расчетной.

В этом случае, приняв номинальную оптическую ось исследуемой системы за параксиальный луч, можно провести анализ децентрировок и наклонов оптических поверхностей. Для этого нужно вычислять матрицу преобразования не сразу для всей оптической системы в целом, а постепенно, по поверхностям, с учетом изменения входных характеристик луча, падающего на поверхность, в результате влияния наклона или децентрировки поверхности. Результатом расчета будут являться координаты параксиального луча в плоскости анализа, положение которой задается. Фактически плоскостью анализа может являться плоскость изображения. Другими словами, в плоскости изображения параксиальный луч – оптическая ось, имеющий нулевые входные координаты, при учете наклонов и децентрировок оптических поверхностей будет отклоняться от своего номинального положения, т.е. иметь уже не нулевые координаты.

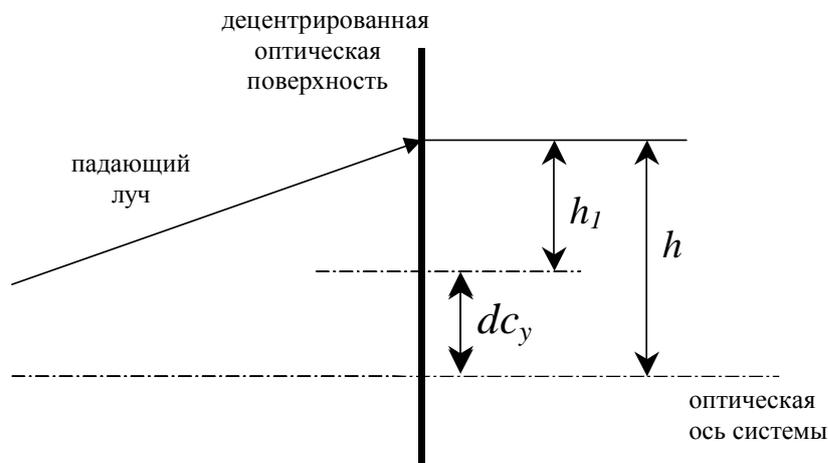


Рис. 2. Децентрировка оптической поверхности

Так как используется расчет параксиальных лучей, при влиянии децентрировок изменяется только высота падения луча на децентрированную поверхность.

Численно изменение высоты на каждой децентрированной поверхности можно определить как

$$h_1 = h - dc_y, \quad (5)$$

где  $h$  – это та высота, на которой приходит луч на поверхность после прохождения предыдущих поверхностей,  $h_1$  – высота луча с учетом децентрировки поверхности,  $dc_y$  – непосредственно сама децентрировка оптической поверхности с учетом направления смещения поверхности ("+" – выше оптической оси системы; "-" – ниже). Таким образом, для вычисления влияния децентрировок поверхностей оптической системы необходимо при вычислении хода параксиального луча вносить на каждой поверхности поправки высоты падения луча на эту поверхность.

Если поверхность наклонена к оптической оси системы (рис. 3), то изменяется не только высота падения луча на поверхность, но и угол падения луча на наклоненную поверхность.

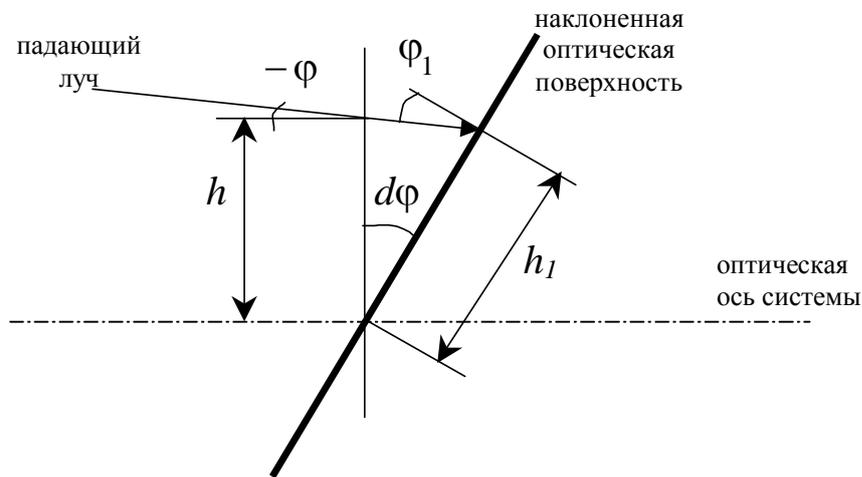


Рис. 3. Наклон оптической поверхности

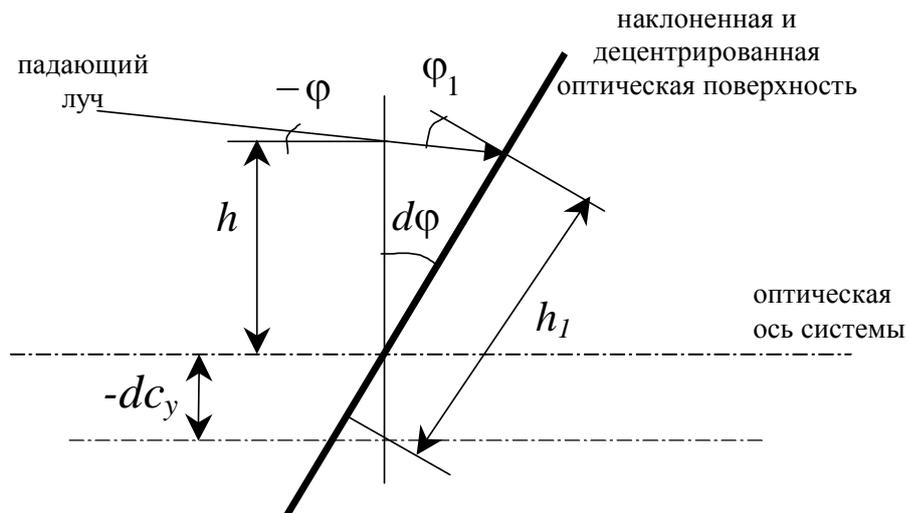


Рис. 4. Децентрировка и наклон оптической поверхности

При расчете нулевого луча также необходимо учитывать наклон не только определенной поверхности, но и наклон предыдущих поверхностей. Причем угол падения луча на поверхность изменится следующим образом:

$$\varphi_1 = \varphi + d\varphi, \quad (6)$$

где  $\varphi$  – угол падения луча на поверхность после прохождения предыдущих поверхностей без учета наклона поверхности,  $\varphi_1$  – угол падения луча на поверхность с учетом наклона поверхности. Высота луча на поверхности численно будет выражаться с учетом наклона поверхности:

$$h_1 = \frac{(h \cdot \cos \varphi)}{\cos(\varphi_1)}, \quad (7)$$

где  $h$  – высота луча на поверхности без учета наклона поверхности,  $h_1$  – высота луча с учетом наклона поверхности. Следовательно, необходимо вносить в расчет хода нулевого луча поправку высоты и угла падения луча на каждой поверхности в зависимости от ее наклона.

Обычно в оптической системе могут присутствовать и наклон поверхностей, и их децентрировка. В этом случае, если поверхность наклонена и децентрирована, высота будет определяться по формуле

$$h_1 = \frac{((h - dc_y) \cdot \cos \varphi)}{\cos(\varphi_1)}, \quad (8)$$

а угол падения луча – по формуле

$$\varphi_1 = \varphi + d\varphi. \quad (9)$$

В процессе юстировки возможно изменение положения плоскости анализа, в которой вычисляются отклонения параксиального децентрированного изображения, от идеального.

Поскольку прохождение параксиального луча через оптическую систему описывается линейной системой матриц, то по измеренным в плоскости анализа поперечным отклонениям узкого лазерного пучка лучей, проходящего вблизи номинальной оптической оси, возможно обратное вычисление децентрировок или наклонов любой из выбранных поверхностей. В этом случае при проведении юстировки оптической системы измеряются координаты параксиального луча – оптической оси в плоскости изображения, и затем с помощью обратной матрицы Гаусса вычисляется децентрировка или наклон заданной оптической поверхности, которые и необходимо исправить

## Заключение

В работе рассмотрена возможность применения аппарата матричной оптики для анализа влияния децентрировок и наклонов поверхностей оптической системы в процессе сборки и юстировки. Показано, что при появлении наклонов и децентрировок поверхностей изменяются координаты параксиальных лучей на этих поверхностях и в плоскости анализа.

## Литература

1. Джерард А., Берч Дж.М. Введение в матричную оптику. М.: Мир, 1978.
2. Родионов С.А. Автоматизация проектирования оптических систем. Л.: Машиностроение. 1982.
3. Donis Flagello et. al., "Towards a comprehensive control of full-field image quality in optical lithography", in Optical Microlithography X, Gene E. Fuller, proceedings of SPIE Vol. 3051 (1997).

# МЕТОДЫ ПРОЕКТИРОВАНИЯ И КОНТРОЛЯ ЗЕРКАЛЬНОГО ОБЪЕКТИВА ДЛЯ РЕНТГЕНОВСКОЙ ФОТОЛИТОГРАФИИ

Е.В. Гаврилов

Научные руководители – д.т.н., проф. Н.Б. Вознесенский,  
д.т.н., проф. В.К. Кирилловский

Рассматривается объектив для рентгеновской фотолиитографии. Дан анализ возможностей классических методов контроля применительно к контролю высокоточных поверхностей. Рассмотрены достоинства и недостатки интерферометра с дифракцией на точечном отверстии. Рассмотрены интерферограммы волновых фронтов от поверхностей имеющих большие отклонения от сферичности. Интерферограммы получены на интерферометре, где в качестве эталонного волнового фронта используется волна, дифрагировавшая на отверстии с диаметром, сравнимым с размерами дифракционного максимума первого порядка (дискон Эри).

## Методы проектирования

В последнее время как в России, так и за рубежом проявляется исключительно высокий интерес к фотолиитографии, в частности к технологии, которая получила название EUV (Extreme Ultra Violet). Длина волны EUV-излучения 13 нм, т.е. примерно в 25 раз меньше, чем излучение светового диапазона. Требования к оптическим системам для данной области спектра следующие:

- система реализуется с помощью зеркал;
- в системе отсутствует центральное экранирование;
- необходима строгая ортоскопичность;
- требуется телецентрический ход лучей;
- средний квадрат деформации волнового фронта  $\Delta W_{скв} \leq 0,04\lambda$ .

Естественно, что как разработка такого рода объектива, так и контроль представляют собой достаточно интересные технические задачи.

Рассмотрим объектив, расчет которого для данной технологии был выполнен в ходе настоящей работы. Объектив имеет следующие характеристики:

- фокусное расстояние  $f' = -310,5810$  мм,
- поле зрения в пространстве изображений – 56 мм,
- числовая апертура в пространстве изображений – 0,16.

Все поверхности объектива – асферические высшего порядка. Объектив имеет высокое качество изображения, в частности,  $\Delta W_{скв} = 0,04\lambda$ .

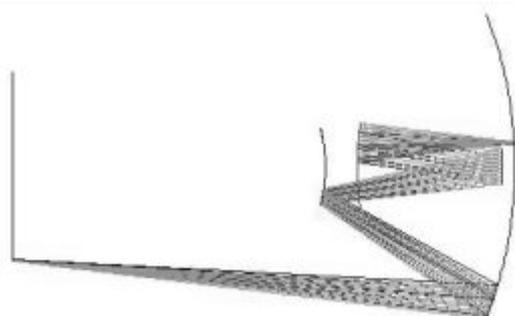


Рис. 1. Схема объектива

Согласно техническому заданию на объектив, среднеквадратическая деформация волнового фронта не должна превышать  $\Delta W_{скв} \leq 0,04\lambda$ . Отсюда порядок величины максимально допустимой деформации волнового фронта составляет  $\Delta W_{max} = 0,16\lambda$ . Так как схема 4-х зеркальная, то деформацию волнового фронта, вносимую одним из зер-

кал, определим как  $\Delta W_{зер} = 0,04\lambda$ . Имея в виду то, что деформация при отражении от зеркала равна удвоенной ошибке зеркала, максимальной допустимой ошибкой зеркала будем считать  $\Delta N = 0,02\lambda$ . При  $\lambda = 0,013$  мкм  $\Delta N \approx 0.0002$  мкм.

## Методы контроля

Хроматизм в данной системе отсутствует, и контроль можно осуществлять в световом диапазоне. Требуемая точность лазерного интерферометра для контроля поверхности зеркала в долях длины волны при рабочей  $\lambda = 0,450$  мкм. (Cd–Ar лазер) составляет по крайней мере  $\Delta W_{LS} = 0.0001\lambda$ .

Указанная точность при измерении интерферограммы может быть достигнута путем применения фазового интерферометра. Традиционные интерферометры для контроля ошибок формы оптических поверхностей обычно содержат образцовую деталь, необходимую для формирования опорного волнового фронта. При этом по умолчанию предполагается, что ее погрешностями можно пренебречь по сравнению с погрешностями контролируемого элемента. Таким образом, реально интерферограмма представляет собой дифференциальное топографирование контролируемого образца совместно с погрешностями самого интерферометра и другими факторами.

Применяемые методы обработки интерферограмм, построенные на принципе вычитания эталонного волнового фронта из результирующего, не позволяют получить погрешности интерферометра до уровня менее чем  $\lambda/20 - \lambda/30$ , так как помимо очевидных источников ошибок – нестабильности источника излучения, механических смещений, турбулентности воздуха – существуют ошибки образцовой поверхности. А именно, эталонная поверхность не может быть изготовлена точнее, чем  $\lambda/20 - \lambda/30$ . Кроме того, образцовая поверхность обычно аттестуется в процессе изготовления интерферометра, а в ходе выполнения сеанса измерения указанная образцовая поверхность может претерпевать неконтролируемые деформации, вызванные, например, действием термических, динамических и гравитационных факторов и соизмеримые с ее точностью.

В качестве альтернативы можно предложить для контроля объектива и его элементов использовать дифракционный лазерный интерферометр [1]. В данном интерферометре отсутствуют ошибки, связанные с деформацией опорного волнового фронта образцовой оптической поверхностью, так как отсутствует сама образцовая поверхность. Опорный волновой фронт здесь формируется путем дифракции лазерного пучка, сфокусированного на точечном отверстии, диаметр которого соизмерим с длиной волны излучения.

## Дифракционный интерферометр

Схемные решения традиционных интерферометров основаны на формировании опорного волнового фронта путем использования образцовых оптических поверхностей. К таким устройствам относятся ставшие классическими интерферометры Тваймана и Физо. Известные достоинства таких схемных решений послужили основой их широкого применения.

Однако в последние десятилетия повысился спрос на оптические системы и элементы наивысшего класса точности. К таким изделиям можно отнести оптические системы космических телескопов, современных исследовательских микроскопов, технологической аппаратуры микроэлектроники и оптико-электронных устройств записи информации. Требования к точности расчета и выполнения этих систем находятся на уровне  $1/20 - 1/30\lambda$ ; где  $\lambda$  – длина волны оптического излучения. Таким образом, для контроля при их изготовлении и применении необходимы приборы, обеспечивающие точность на уровне  $1/100 - 1/200\lambda$ , т.е. на порядок точнее традиционных.

Недостатком классических интерферометров при контроле оптики такого класса является необходимость присутствия в их схемах образцового оптического элемента, точность изготовления которого ограничена. При этом нет гарантии отсутствия неконтролируемого изменения точности образцового оптического элемента в процессе эксплуатации интерферометра.

Альтернативным направлением интерферометрии стало создание интерферометров, в которых опорный сферический волновой фронт формируется путем дифракции пучка излучения, сфокусированного на точечной диафрагме, соизмеримой по диаметру с длиной волны. Первая известная схема такого рода была предложена академиком Линником [5].

Высокую эффективность при контроле высокоточных оптических поверхностей и систем показал лазерный интерферометр с дифрагированным опорным волновым фронтом и разнесенными ветвями [4, 6] (рис. 2).

Разработаны разновидности схем интерферометра с дифрагированным эталонным волновым фронтом. Этот прибор имеет традиционную методику настройки и интерпретации интерференционной картины, обеспечивает высокое качество интерферограммы, гибкость при исследовании оптических систем и элементов с различными коэффициентами пропускания (отражения) от 98% (зеркальные покрытия) до 0.15 % (просветляющие покрытия).

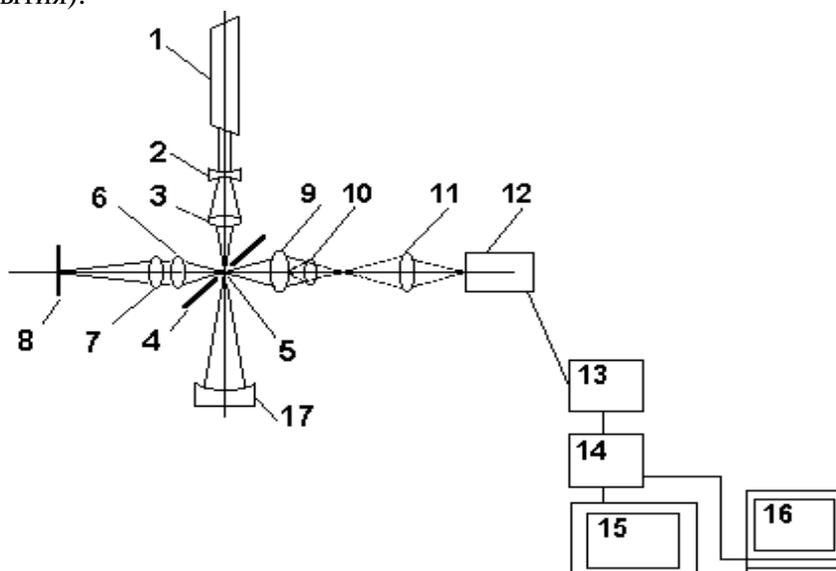


Рис. 2. Интерферометр с дифрагированным опорным фронтом : 1 – лазер, 2,3 – осветительная система, 4 – наклонное зеркало, точечная диафрагма, 6, 7 – встречные коллиматорные объективы, 8 – плоское автоколлимационное зеркало, 9–11 – наблюдательный микроскоп с линзой Бертрана, 12 – видеокамера, 13 – блок обработки изображения, 14 – блок выделения центров интерференционных полос, 15 – телевизионный монитор, 16 – компьютер, 17 – измеряемая поверхность

Прибор отличается отсутствием образцовых оптических элементов и ошибок их изготовления, конструктивная простота, малогабаритность и легкость в эксплуатации.

Недостаток, свойственный всем двухлучевым интерферометрам, состоит в ограниченной точности измерения координат интерференционных полос, что вызвано их нерезкостью вследствие синусоидального характера распределения интенсивности в интерференционном изображении.

В работе [3] дан анализ возможностей и ограничений методов и средств исследования прецизионных поверхностей. Недостатки плавных характеристик преобразования:

- ограниченный динамический диапазон, лежащий в рамках линейного участка ФП;

- низкая чувствительность визуальных измерительных наводок, связанная с нечеткостью элементов оптико-измерительного изображения (теневая картина, пятно рассеяния, пятна гартманогаммы);
- высокая избыточность информации при автоматизированной расшивке и интерпретации таких оптико-измерительных изображений.

Для устранения выявленных недостатков и получения эффективных методов и средств исследования прецизионных поверхностей, обладающих новыми свойствами и расширенными возможностями, дано развитие системы одномерных функций преобразования (ФП) [3]. Предложен аппарат развитых ФП для исследования прецизионных поверхностей.

а) ФП типа "импульс", которая может быть наглядно представлена выражением с участием  $\delta$ -функции:

$$I = \delta(W) = \lim N \exp(-N^2 \pi W^2)$$

или

$$I(W) = I_{\max} \quad \text{при } W = 0$$

$$I(W) = 0 \quad \text{при } W \neq 0;$$

б) периодическая ФП типа "гребенка", которая может быть наглядно представлена выражением

$$I_{y,z} = \text{comb}(W / \lambda_w) = \sum_{n=1}^N \delta(W - n\lambda_w)$$

где  $\lambda_w \leq W \leq n \lambda_w$ , или, на практике,

$$I(W) = I_{\max} \quad \text{при } W/\lambda_w = 0, 1, 2, 3, \dots$$

$$I(W) = 0 \quad \text{при } W/\lambda_w \neq 0, 1, 2, 3, \dots$$

Анализ показал, что разработанная система ФП дает ряд специальных положительных эффектов, таких как сокращение избыточности информации, улучшение наглядности и выявление новых деталей и свойств объекта, повышение чувствительности контроля и точности измерений, повышение оперативности и производительности, упрощение задачи автоматизации контроля и исследований.

Взаимовлияние синтезируемых методов исследования прецизионных поверхностей с достижением новых результатов и свойств иллюстрируется также на примере предложенного (и реализованного в виде серийного прибора) метода изометрии второй производной функции пространственного распределения интенсивности, позволяющего, например, выделять центры (оси) интерференционных полос. Эта разработка дала возможность создать автоматизированный интерферометр нового типа, позволяющий повысить точность расшивки интерферограмм в 10–40 раз в реальном времени.

Операция электронного выделения и визуализации центров интерференционных полос (рис.3) с умножением частоты полос осуществлена согласно ФП вида:

$$I_{ok} = \text{comb} \left[ \frac{1}{\lambda} (y \sin \theta + kW) \right],$$

где  $I_{ok}$  – относительная интенсивность в оконтуренной интерферограмме;  $\theta$  – угол заклона образцового волнового фронта интерферометра относительно рабочего волнового фронта;  $W$  – волновая абберрация;  $y$  – координата в интерферограмме, ориентированная по нормали к направлению интерференционной полосы;  $k$  – коэффициент умножения полос.

Если  $k = 2$ , то при этом цена шага интерференционной полосы составляет  $\lambda/4$ , что создает эффект интерферометрии с использованием излучения, имеющего  $\lambda = 300$  нм.

Точность определения координат интерференционных полос возрастает, как показали исследования, в 40 раз по сравнению с точностью измерения необработанной ин-

терферограммы, что соответствует возможности визуального обнаружения деформации волнового фронта на уровне  $\lambda/200$  в реальном времени.

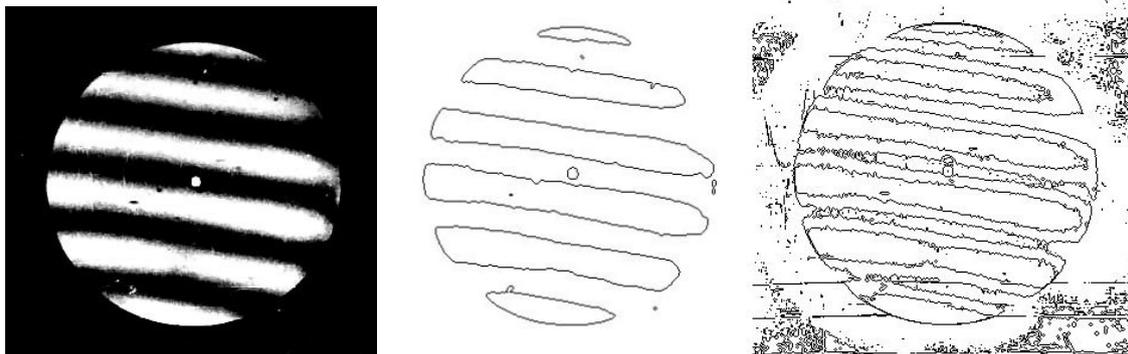


Рис. 3. Операция электронного выделения и визуализации центров интерференционных полос

Точностные характеристики прибора исследования прецизионных поверхностей зависят от принципа и конструкции измерительного преобразователя и могут быть определены при анализе функции преобразования.

Итак, в данных дифракционных интерферометрах генерация неискаженного сферического эталонного фронта волны основана на использовании хорошо воспроизводимого физического явления дифракции лазерного пучка на точечном отверстии диаметром, соизмеримым с длиной волны. В этих приборах отсутствует образцовая оптическая деталь и присущие ей неизбежные остаточные погрешности, возникающие при изготовлении.

### Контроль зеркал объектива

Для контроля асферических поверхностей предусмотрены два способа:

- использование анаберрационных схем контроля с применением технологических зеркал, изготовленных под данный интерферометр;
- применение коррекционного объектива, также изготовленного и аттестованного под данный интерферометр.

Для того чтобы использовать в полной мере преимущества PDI-интерферометра, создание коррекционных объективов для контроля асферических поверхностей не имеет никакого смысла, так как при этом мы ограничиваем сами себя точностью изготовления этого корректора, которая экспериментально не может быть оценена лучше, чем это дают известные интерферометры типа Физо или Тваймана.

Для построения схем контроля зеркал на основе анаберрационных схем поверхностей необходимо подобрать такую сферу или асферику 2-го порядка, чтобы ее отступление от асферики высшего порядка не превышало 3 мкм., что соответствует 6 интерференционным полосам. Но такое условие может быть выполнено только для поверхностей с малым отступлением от сферичности.

Задача контроля асферических зеркал данного объектива с отступлениями от сферы или другой поверхности второго порядка в несколько десятков длин волн состоит в том, что ни при каких обстоятельствах не удастся получить на интерферограмме настолько малое количество полос, чтобы процедура поиска максимумов или минимумов полос могла уверенно идентифицировать все полосы, как это происходит в случае поверхности более высокого качества. Между тем этот случай является весьма важным в интерферометрическом контроле оптики. Несмотря на большие отклонения исследуемой поверхности от сферичности, их необходимо определить с не менее высокой точностью, чем в первом случае. Поэтому приходится оценивать интерференционные кар-

тины с большим количеством искривленных полос. В качестве примера можно привести интерферограмму поверхности с большим отклонением от сферичности (рис 4).

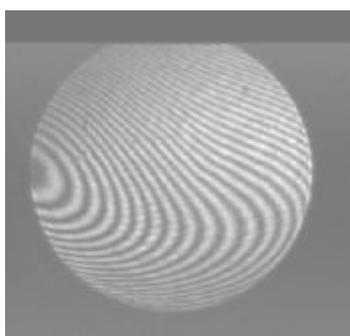


Рис. 4. Интерферограмма поверхности с большим отклонением от сферичности

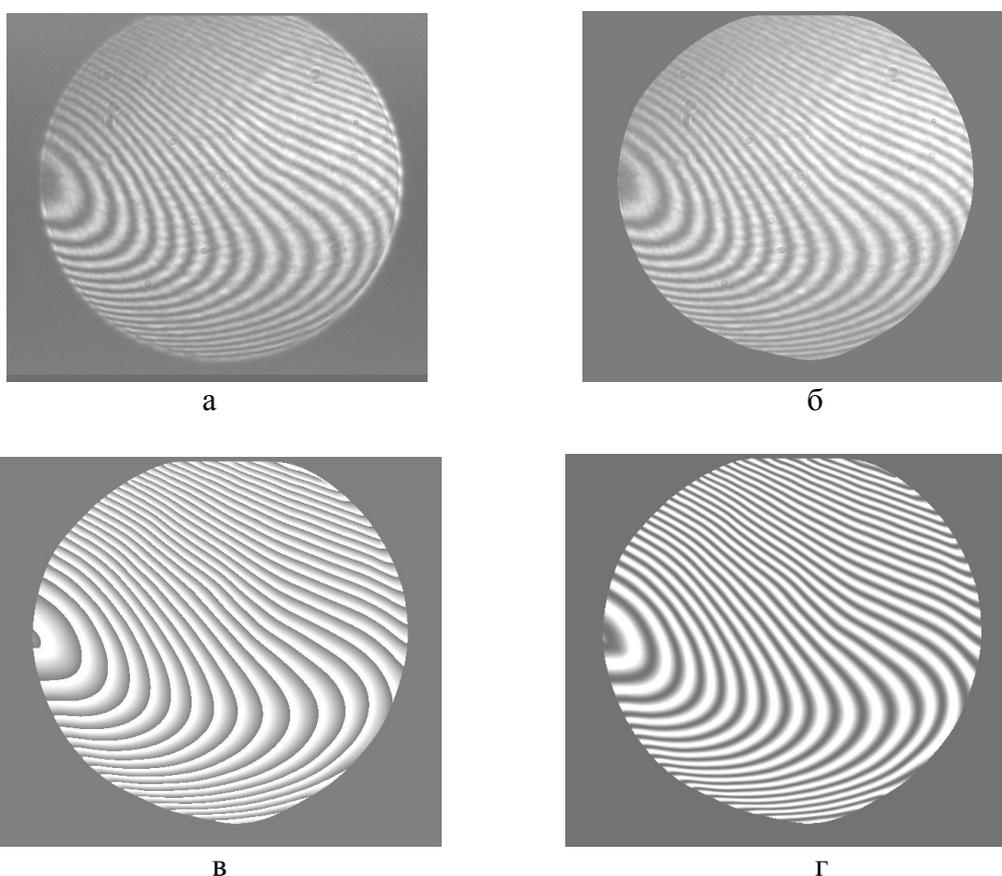


Рис. 4. Этапы обработки интерферограммы

Применение программ типа "Zebra" и аналогичных ей, вернее, интерпретационное решение обработки интерферограмм не может справиться с задачей восстановления поверхностей или волновых фронтов с большими отклонениями от референтной сферы. Возможно, конечно, применение более высоко разрешающей ПЗС-матрицы, однако матрицы с количеством пикселей значительно более миллиона пока не известны в открытом применении, и остаются другие проблемы, в частности, проблема качества проекционной оптической системы. Кроме того, увеличение числа данных в аппроксимационных задачах, как известно, приводит к обратному эффекту в смысле получения еще более сглаженной картины и увеличения степени неопределенности компонент

высшего порядка в полиномиальном разложении (это касается не только степенного базиса, но и полиномов Цернике). Имеется в виду, что интерпретационный путь решения задачи восстановления профиля сложной формы в принципе не подходит для контроля асферик высшего порядка.

Поэтому приходится разрабатывать технологии для обработки интерферограмм с большим числом полос. Такая технология обработки базируется на теории сигналов или, более конкретно, на методах обработки изображения. Эта технология представляет исключительно большой интерес в применении к дифракционному интерферометру, поскольку дает возможность восстановления фазы и формы контролируемой поверхности с высокой точностью независимо от сложности интерференционной картины.

Обработка интерферограмм производится в несколько этапов, что иллюстрируется представленными ниже изображениями (рис. 5, а–г). В предварительную обработку входят, наряду со служебными процедурами, этап неглубокой пространственной фильтрации для удаления "мусора" из интерферограммы, идентификации полос, частичное оконтуривание интерферограммы, что приводит к повышению точности расшифровки интерференционной картины до  $\lambda/10000$ .

## Заключение

В работе рассмотрены требования к точности контроля объектива для рентгеновской области спектра. Рассмотрен контроль асферических поверхностей высшего порядка в анаберрационных схемах. Показано, что применение дифракционного интерферометра позволит создавать объективы для рентгеновской фотолитографии.

## Литература

1. Кирилловский В.К., Кузнецов Д.Н., Трухин М.М. Интерферометр с дифрагированным фронтом сравнения для контроля прецизионных оптических поверхностей. Материалы Научно-технической конференции "Прикладная оптика - 98", ГОИ, ГИТМО, Санкт-Петербург, 16 - 18 декабря 1998
2. Кирилловский В.К., Кузнецов Д.Н. Оптимизация лазерного интерферометра с дифрагированным опорным волновым фронтом на основе использования осветительных пучков гауссовой структуры. // Материалы Международной конференции "Прикладная оптика-98", ГОИ, ИТМО (ТУ), СПб, 16-18 декабря 1998
3. Кирилловский В.К. Характеристики чувствительности структурометрических систем. 4-я Всесоюзная конференция "Фотометрия и ее метрологическое обеспечение". ВНИИОФИ АН СССР. 1982
4. Кирилловский В.К. Лазерный интерферометр с дифрагированным опорным фронтом. Российская научно-практическая конференция "ОПТИКА. - ФПЦ Интеграция". ИТМО (ТУ). Санкт-Петербург, 25-27 января 1999
5. Линник В.П. Простой интерферометр для исследования оптических систем. Труды ГОИ. Том X. Выпуск 95. Ленинград. 1934
6. Иванова Т.А., Кирилловский В.К. Проектирование и контроль оптики микроскопов. Л.: Машиностроение, 1984.
7. Малакара Д. Оптический производственный контроль. 1995., Машиностроение.
8. Кирилловский В.К., Вознесенский Н.Б., Трухин М.М. Интерферометры с дифракцией на точечном отверстии для контроля ошибок формы прецизионных поверхностей. Материалы Научно-технической конференции "Lamp -2003, IFMO (TU), Saint-Petersburg.
9. Кирилловский В.К., Вознесенский Н.Б.Ю, Гуров И.П., Ли Кенг Хи. Сравнительный анализ интерферограмм, полученных на интерферометре с дифракционной волной сравнения. // Материалы научно-технической конференции "Интеграция-2004". ИТМО (ТУ), Санкт-Петербург, 16-19 февраля 2004.

# КОМПЬЮТЕРНЫЙ КОНТРОЛЬ И ЮСТИРОВКА СКАНИРУЮЩИХ СИСТЕМ

К.А. Иванов

Научный руководитель – д.т.н., проф. С.А. Сухопаров

Рассмотрены задачи оптимизации процесса юстировки оптико-электронных сканирующих приборов. Предложены практические решения автоматизации юстировки цифровых оптических сканирующих систем, базирующиеся на анализе получаемого изображения, содержащего реперные элементы.

## Введение

К настоящему времени из общего объема затрат времени и средств на производство оптических приборов на сборку и юстировку приходится более 50%, причем на юстировку – большая часть затрат.

Причиной преобладания затрат на юстировку является сложность и субъективность ее приемов. Поэтому в технологии юстировки преобладают ручные трудоемкие операции. Сложность юстировки требует более высокой квалификации рабочих. В последнее время требования к точности, надежности приборов и удобству их эксплуатации повышаются. Большие затраты на юстировку приборов оправдывают поиски средств сокращения этих затрат.

Все это вызывает необходимость облегчения и удешевления юстировочных операций путем их автоматизации. Компьютерные технологии позволяют значительно усовершенствовать процесс юстировки оптико-электронных приборов.

## Постановка задачи

Юстировка оптико-электронных приборов имеет следующие характерные черты:

- выходные параметры прибора не могут быть объективно оценены и проконтролированы человеком;
- большое количество юстируемых параметров;
- высокая точность контроля;
- большие объемы и необходимость высокой скорости обработки поступающей информации.

Все вышеприведенные особенности приводят к неэффективности использования "ручной" юстировки и к идее использования компьютерной техники в процессе юстировки.

Автоматизация юстировки может иметь два направления – первое представляет собой оптимизацию допусков на этапе расчета конструкции прибора [1], предложенную проф. С.А. Сухопаровым и дополненную автором статьи [2], второе представляет собой автоматизацию самого процесса юстировки прибора [3], которое мы рассмотрим ниже.

Для достижения указанных целей автором была разработана методика, позволяющая автоматизировать процесс юстировки высокоточных сканирующих оптико-электронных систем. Она включает в себя идеологическое, программное и аппаратное обеспечение юстировочного процесса и позволяет производить:

- автоматизацию таких юстировочных операций, как фокусировка, подвижки изображения в поле зрения прибора, компенсация геометрических искажений, регулировка яркости;
- автоматический контроль функции передачи модуляции, являющейся одной из определяющих характеристик сканирующих систем, а также любых параметров передаваемого системой изображения.

В качестве объекта юстировки рассмотрим пример автоматической юстировки сложного высокоточного оптического сканера, предназначенного для получения отпечатков пальцев.

Задачей этого прибора является получение изображения предмета, расположенного в жестко заданной плоскости сканирования (см. оптическую схему на рис. 1).

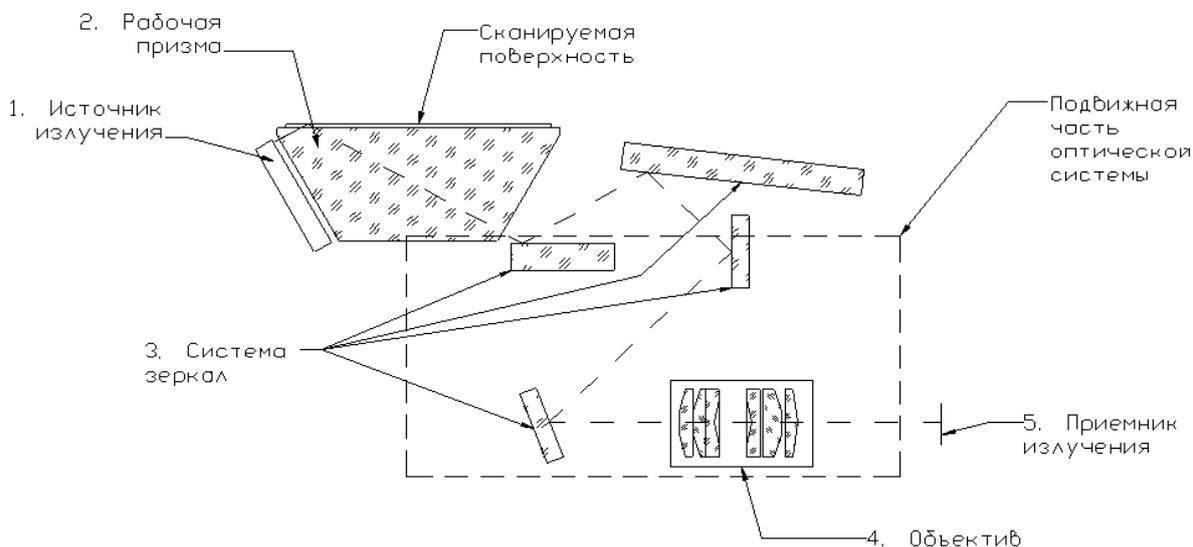


Рис.1. Оптическая схема сканера

Так как в качестве приемника излучения используется ПЗС-линейка, то для получения полноценного двухкоординатного изображения в приборе применена механическая развертка. Оптическая схема прибора разбивается на две части – подвижную и неподвижную, как показано на рисунке ниже. В результате движения подвижной части сканирующей системы на выходе прибора формируется полноценное двумерное изображение.

Для юстировки сканер необходимо подключить к специально созданному компьютеризированному стенду (рис. 2), осуществляющему моделирование работы прибора в штатном режиме. Юстировочный стенд обеспечивает двусторонний обмен данными между сканером и управляющим устройством через плату сопряжения (плату интерфейса контрольно-проверочной аппаратуры [КПА]).

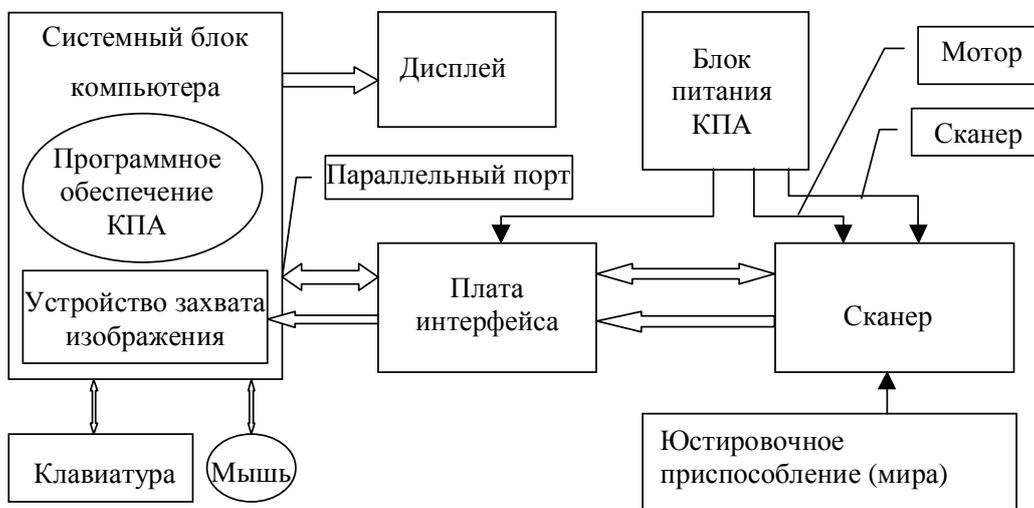


Рис. 2. Общий вид стенда контрольно-проверочной аппаратуры

Работа [1] содержит подробное описание реализации взаимодействия сканера и контрольно-проверочной аппаратуры, обеспечивающее юстировку.

Процесс юстировки базируется на анализе изображения, получаемого со сканера. Для анализа положения элементов изображения электронике необходимо задать систему координат в пространстве предмета. Это требуется для приведения ее в соответствие с системой координат самого прибора, определяемой положением ее механических элементов. Остановимся на этом вопросе подробнее.

### Задание координат в пространстве предмета

В качестве элементов "привязки" на настроечной мире предлагается использовать так называемые реперные элементы – в нашем случае это темные кружки (см. рис. 3). Три точки, представляющие собой центры темных кружков различного размера, однозначно задают систему координат в пространстве предметов (центры кружков определяются электроникой автоматически).

Так как имеются погрешности изготовления и базирования узлов и деталей, то неизбежна погрешность установки настроечного оборудования (в нашем случае настроечной миры), что приведет к невозможности для автоматики привязаться к системе координат прибора (найти реперные элементы). Чтобы этого не произошло, необходимо предусмотреть возможность для автоматики находить реперы и уметь к ним привязываться.

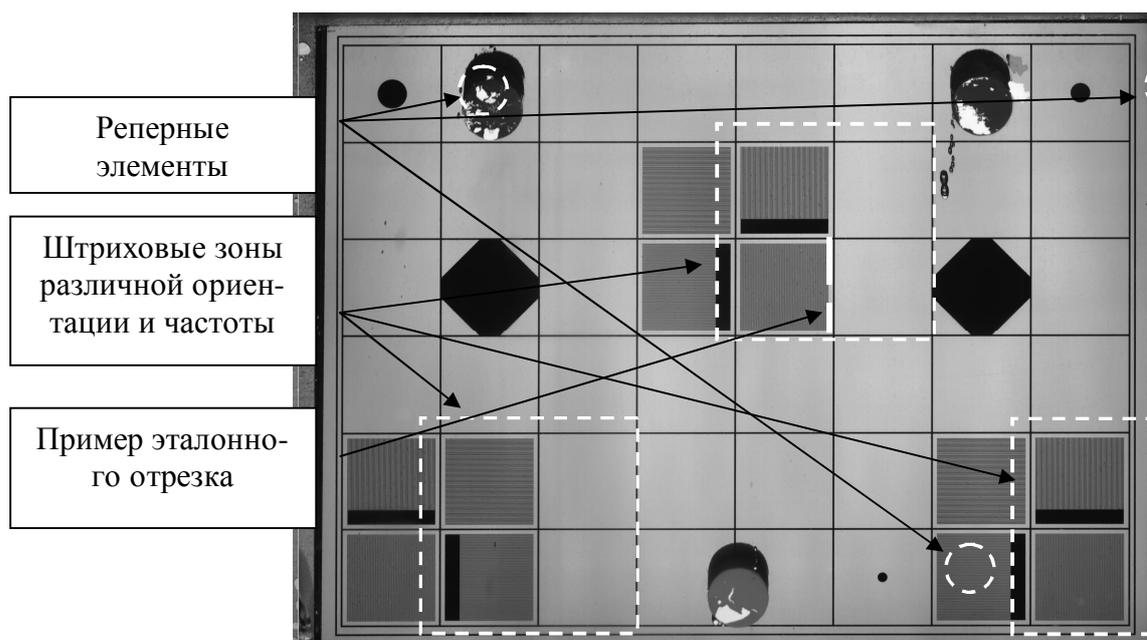


Рис. 3. Пример одной из настроечных мир (изображение, полученное на сканере).

### Определения центров реперных элементов

Чтобы программа могла "найти" необходимые ей реперные элементы, необходимо задать ей в математической модели зону поиска. Размеры зоны поиска должны превышать размеры реперного элемента, так как из-за погрешностей изготовления и базирования узлов и деталей прибора реперные элементы могут перемещаться по рабочему полю прибора. Величина отношения размеров зоны поиска к размерам реперного элемента определяется как геометрией самого реперного элемента, так и общими допусками на прибор и его элементы.

На рис. 4 приведен фрагмент изображения, полученного на сканере, отображающий один из реперных элементов (в нашем случае черный круг), входящих в настройечную миру (полное изображение миру приведено на рис. 3). Рамка из пунктирной линии отражает зону поиска, внутри которой производится поиск реперного элемента.

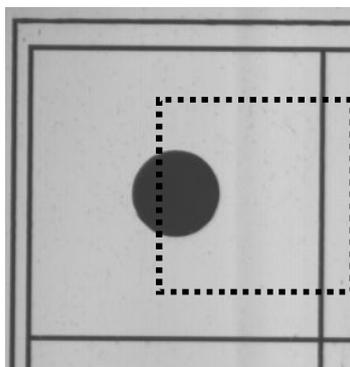


Рис. 4. Фрагмент изображения, полученного на сканере, отображающий один из реперных элементов

После того, как программа "нашла" реперный элемент (в нашем случае это круг), ей необходимо определить его центр, так как нам нужна координата точки. Существует множество методов определения центра круга, в нашем случае используется один из простейших – определение центра тяжести геометрической фигуры. Суть метода заключается в том, что для каждой точки изображения в зоне поиска вводится весовой фактор (значения которого варьируются от 0 до 255), зависящий от ее яркости. После этого вычисляется энергетический центр тяжести

$$X_C = \frac{\sum_{i=1}^n P_i \cdot x_i}{\sum_{i=1}^n P_i}, \quad Y_C = \frac{\sum_{i=1}^n P_i \cdot y_i}{\sum_{i=1}^n P_i},$$

где  $(X_i, Y_i)$  – координаты центра тяжести реперного элемента,  $(x_i, y_i)$  – координаты точки реперного элемента,  $P_i$  – весовой фактор точки реперного элемента.

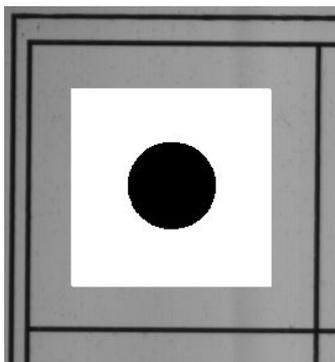


Рис. 5. Вид зоны реперного элемента после введения весового фактора.

Весовой фактор можно вычислять следующим образом. В зоне поиска реперного элемента определяются максимальное  $B_{r_H}$  и минимальное  $B_{r_L}$  значения яркости точек. Вычисляется пороговое значение (для приведения изображения к единичному

контрасту)  $Br_B = \frac{Br_H + Br_L}{2}$ , в соответствии с которым каждой точке присваивается весовой коэффициент

$$P_i = \begin{cases} 0 & \text{если } Br_i > Br_B \\ Br_B - Br_i & \text{если } Br_i \leq Br_B \end{cases}$$

После введения весового фактора и приведения картинки к единичному контрасту изображение реперного элемента приобретает вид, показанный на рис. 5:

Центр реперного элемента определяется с точностью до десятых долей пиксела, что позволяет осуществить привязку с достаточной степенью точности и избежать погрешностей измерения параметров сканирующей системы.

Используя полученную привязку, разработанное программное обеспечение [3] анализирует и корректирует:

- положение получаемого изображения в кадре;
- геометрическую точность (величину геометрических искажений по полю);
- величину функции передачи модуляции (Modulation Transfer Function, MTF) для определенных пространственных частот;
- отношение сигнал-шум (Signal-to-Noise Ratio);
- диапазон передаваемых полутонов ("градации серого");
- линейность шкалы передаваемых полутонов;
- равномерность выходного сигнала.

В работе [3] содержится детальное описание по автоматизации юстировки вышеуказанных параметров прибора. Создана математическая модель привязки эталонного элемента (миры) и параметров настраиваемого прибора. Разработаны и показаны примеры эталонов (настроечных мир), служащих для автоматизации юстировочных операций. Рассмотрены основные проблемы, связанные с возможными искажениями измеряемых параметров, такие как поворот изображения и его влияние на точность измерений.

Применение разработанного подхода к юстировке позволяет настраивать и контролировать такие параметры прибора, которые при ручной юстировке занимают много времени или попросту не могут быть осуществлены вручную (измерение отношения сигнал/шум, диапазон передаваемых полутонов, равномерность выходного сигнала) с достаточной точностью и достоверностью.

## Заключение

Широкое распространение оптико-электронных приборов и применение высокоточных специализированных измерительных и индикаторных средств способствует осуществлению автоматической компьютерной юстировки.

Разработанная концепция и метод автоматизации процесса контроля и юстировки высокоточных сканирующих приборов, основанные на анализе получаемого изображения, включающего в себя реперные элементы, позволяют значительно упростить и ускорить весь юстировочный процесс.

## Литература

1. Сухопаров С.А. Многофакторная оптимизация точностного расчета оптических приборов // Оптический журнал. 1994. №9. С. 35-38.
2. Сухопаров С.А., Иванов К.А. Компьютерная автоматизированная юстировка оптических приборов // Оптический журнал. 2001. №1. С. 76-78.
3. Иванов К.А. Компьютерная оптимизация юстировки оптических приборов // Автореферат диссертации. СПб.: СПбГУ ИТМО, 2004.

# РАЗРАБОТКА АЛГОРИТМОВ АНАЛИЗА АБЕРРАЦИЙ ОПТИЧЕСКИХ СИСТЕМ С ИСПОЛЬЗОВАНИЕМ ДИФФЕРЕНЦИАЛОВ РЕАЛЬНОГО ЛУЧА

М.Л. Гантворг

Научный руководитель – к.т.н., доц. Т.В. Иванова

В работе предлагаются алгоритмы расчета aberrаций оптических систем, которые реализованы с использованием аппарата дифференциалов лучей.

## Введение

Одной из основных процедур, которые выполняются при проектировании оптических систем, является aberrационный анализ. Задача aberrационного анализа – выполнение расчета определенного набора лучей (главные, габаритные и др.) и определение значений основных видов aberrаций (обобщенные поперечные aberrации, волновая aberrация, неизопланатизм, дефокусировка и астигматизм). При автоматизации aberrационного анализа необходимо обеспечить высокую точность и скорость вычислений. Высокая точность гарантирует получение адекватных результатов и принятие расчетчиком оптической системы соответствующих решений по aberrационной коррекции. Высокая скорость вычислений позволит повысить эффективность этапа оптимизации.

В данной работе предлагаются алгоритмы расчета aberrаций оптических систем, которые реализованы с использованием аппарата дифференциалов лучей [1]. Расчет дифференциалов лучей (или лучей, бесконечно близких к данному) широко используется в оптике наряду с расчетом действительных лучей. При этом роль оптической оси выполняет произвольный действительный луч, который описывается в этой системе двумя векторами  $s$  (вектор линейных координат) и  $q$  (ортом направления луча). Луч, бесконечно близкий к данному, характеризуется векторами  $s+ds$  и  $q+dq$ , где  $ds$  и  $dq$  – бесконечно малые дифференциалы (рис. 1). По аналогии с нулевыми лучами рассматриваются не только бесконечно малые, но и пропорционально увеличенные дифференциалы.

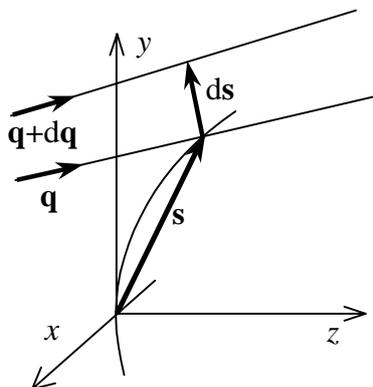


Рис. 1. Координаты действительного луча и его дифференциала

В частном случае меридионального луча дифференциал луча определяется аналогично параксиальному или нулевому лучу в системах с двумя плоскостями симметрии – двумя линейными дифференциалами  $h_x$  и  $h_y$  (координатами его пересечения с плоскостью  $Oxy$ ) и двумя угловыми дифференциалами (проекциями его орта на эту плоскость или углами  $\alpha_x$  и  $\alpha_y$ , образованными им с плоскостями  $Oyz$  и  $Oxz$ ) [1]. В данной работе предлагается отказаться от меридиональных дифференциалов и все вычисления осуществлять на основе единого аппарата дифференциалов произвольных лучей.

Для определения aberrаций системы рассчитываются:

- зрачковые дифференциалы (сагиттальный и меридиональный) – лучи, выходящие из одной предметной точки с данным, нормировка координат на входном зрачке

$$d\mathbf{s}_{pup} = \begin{pmatrix} dx_{pup} \\ dy_{pup} \\ dz_{pup} \end{pmatrix}, d\mathbf{q}_{pup} = \begin{pmatrix} X_{pup} \\ Y_{pup} \\ Z_{pup} \end{pmatrix};$$

- предметные дифференциалы (сагиттальный и меридиональный) – лучи, пересекающие входной зрачок в одной точке с данным, нормировка координат на предмете

$$d\mathbf{s}_{ob} = \begin{pmatrix} dx_{ob} \\ dy_{ob} \\ dz_{ob} \end{pmatrix}, d\mathbf{q}_{ob} = \begin{pmatrix} X_{ob} \\ Y_{ob} \\ Z_{ob} \end{pmatrix}.$$

Аберрации рассчитываются после прохождения луча и дифференциалов через систему по его координатам на поверхности изображения  $\mathbf{s}' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$ ,  $\mathbf{q}' = \begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix}$  и поверх-

ности зрачка  $\mathbf{s}'_p = \begin{pmatrix} x'_p \\ y'_p \\ z'_p \end{pmatrix}$ ,  $\mathbf{q}'_p = \begin{pmatrix} X'_p \\ Y'_p \\ Z'_p \end{pmatrix}$ .

В процессе расчета определяются следующие величины:

- обобщенные поперечные аберрации;
- обобщенные выходные зрачковые координаты;
- неизопланатизм;
- волновая аберрация;
- дефокусировка и астигматизм.

Рассмотрим алгоритмы расчета аберраций в зависимости от типа изображения.

### Расчет аберраций для изображения ближнего типа

На рис. 2 изображены (для изображения ближнего типа): последняя поверхность оптической системы, поверхность изображения и два луча – главный и неглавный.

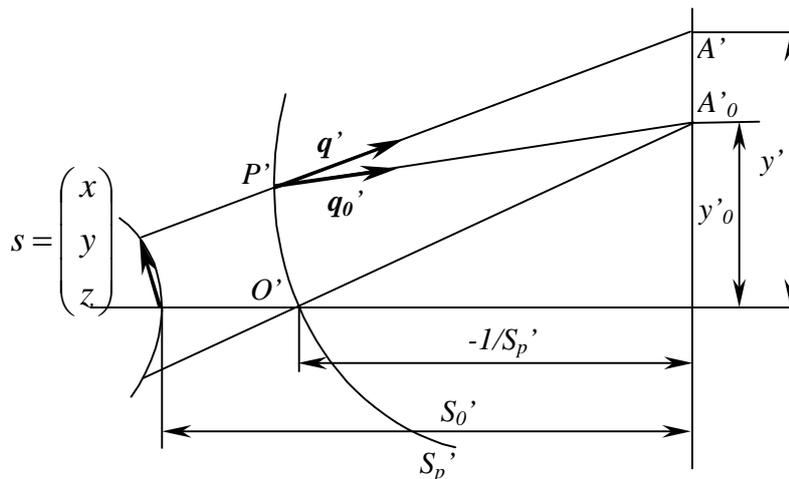


Рис. 2. Выходные координаты луча для изображения ближнего типа

Расчет аберраций для изображения ближнего типа осуществляется в следующем порядке:

1). Обобщенные поперечные aberrации измеряются в миллиметрах и определяются следующим образом:

$$\begin{aligned}\delta G' &= x' \\ \delta g' &= y' - y'_0\end{aligned}\quad (1)$$

где  $y'_0$  – обобщенная величина изображения в миллиметрах.

Для определения матрицы  $\mathbf{D}$ , которая потребуется для расчета входных лучевых координат и волновых aberrаций, вычислим определитель матрицы  $ds'$ , которая составлена из координат предметных дифференциалов:

$$\det = |ds'| = \begin{vmatrix} dx'_{ob1} & dx'_{ob2} \\ dy'_{ob1} & dy'_{ob2} \end{vmatrix}.\quad (2)$$

Если  $\det = 0$ , то плоскость изображения совпадает с плоскостью выходного луча, и необходимо задать другую плоскость изображения.

Затем формируем элементы обратной матрицы  $ds'^{-1}$ :

$$ds'^{-1} = \begin{pmatrix} dy'_{ob2} & -dx'_{ob2} \\ -dy'_{ob1} & dx'_{ob1} \end{pmatrix}.\quad (3)$$

Вычисляем матрицу  $\mathbf{D}$ :

$$\mathbf{D} = d\mathbf{q}' \cdot ds'^{-1}\quad (4)$$

2). Выходные лучевые координаты определяются следующим выражением:

$$\begin{pmatrix} p'_x \\ p'_y \end{pmatrix} = n_0 \left[ \begin{pmatrix} X' \\ Y' \end{pmatrix} - \mathbf{D} \begin{pmatrix} \delta G' \\ \delta g' \end{pmatrix} \right].\quad (5)$$

3). После расчета лучевых координат определяем значение неизопланатизма в процентах:

$$\begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix} = \left[ 1 - \frac{\begin{pmatrix} p'_x \cdot V_x \\ p'_y \cdot V_y \end{pmatrix}}{p_x} \right] \cdot 100\%.\quad (6)$$

4). Вычисление волновой aberrации осуществляется следующим образом:

$$\begin{aligned}W &= [A, A'] - \mathbf{s}'^T \mathbf{q}'; \\ W &= W - [A, A'_0] + \frac{1}{2} (\delta G' \quad \delta g') \mathbf{D} \begin{pmatrix} \delta G' \\ \delta g' \end{pmatrix},\end{aligned}\quad (7)$$

где  $[A, A']$ ,  $[A, A'_0]$  – оптические длины данного и главного лучей до поверхности изображения.

5). Затем осуществляется расчет дефокусировки и астигматизма. Для этого нам необходимо вычислить матрицу продольных дефокусировок  $d\mathbf{L}$ :

$$d\mathbf{L} = -\frac{Z'}{\det} \begin{pmatrix} dx'_{pup1} dY'_{pup2} - dx'_{pup2} dY'_{pup1} & dx'_{pup2} dX'_{pup1} - dx'_{pup1} dX'_{pup2} \\ 0 & dy'_{pup2} dX'_{pup1} - dy'_{pup1} dX'_{pup2} \end{pmatrix},\quad (8)$$

$$\det = |d\mathbf{q}'| = \begin{vmatrix} dX'_{pup1} & dX'_{pup2} \\ dY'_{pup1} & dY'_{pup2} \end{vmatrix}\quad (9)$$

$d\mathbf{q}'$  - матрица, которая составлена из координат лучевых дифференциалов.

Проведем коррекцию матрицы  $d\mathbf{L}$ :

$$\mathbf{D} = \begin{pmatrix} dl_{11}(1 - X'^2) - dl_{12} X' Y' & dl_{12}(1 - X'^2) - dl_{22} X' Y' \\ 0 & dl_{22}(1 - Y'^2) - dl_{12} X' Y' \end{pmatrix}.\quad (10)$$

Вычисление дефокусировки выполняется по формуле:

$$dz = \frac{d_{11} + d_{22}}{2} \quad (11)$$

Для борьбы с потерей значимости проведем следующие преобразования ( $k_1, k_3, k_2$  - вспомогательные величины):

$$k_1 = d_{11} - d_{22}; \quad (12)$$

$$k_2 = 2d_{12},$$

Если эти величины по модулю меньше  $10^{-8}$ , то они считаются нулевыми.

$$k_3 = \sqrt{k_1^2 + k_2^2} \quad (13)$$

Если  $k_1 < 0$ , то  $k_1, k_3, k_2$  меняют знак на противоположный.

И, наконец, вычисление астигматизма осуществляется по формуле:

$$az = -k_3 \quad (14)$$

### Расчет aberrаций для изображения дальнего типа.

На рис.3 изображены (для изображения дальнего типа): последняя поверхность оптической системы, поверхность изображения и два луча – главный и неглавный.

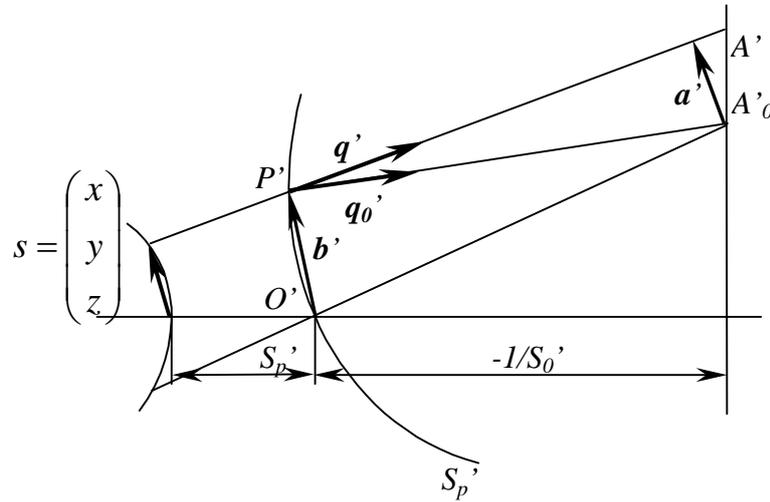


Рис. 3. Выходные координаты луча для изображения дальнего типа

Расчет aberrаций для изображения дальнего типа осуществляется в следующем порядке:

1). Для изображения дальнего типа обобщенные поперечные aberrации измеряются в радианах и определяются следующим выражением:

$$\begin{pmatrix} \delta G' \\ \delta g' \end{pmatrix} = -\frac{\mathbf{a}'}{r_p'} - \rho_p' (\mathbf{b}' + r_p' \mathbf{q}') = -\rho_p' \begin{pmatrix} x' + r_p' X' \\ y' + r_p' Y' \end{pmatrix} = -\begin{pmatrix} X' + \rho_p' x' \\ Y' + \rho_p' y' \end{pmatrix}. \quad (15)$$

Затем для расчета остальных величин необходимо выполнить пересчет дифференциалов:

$$d\mathbf{q}' = -(d\mathbf{q}' + \rho_p' ds') \quad (16)$$

Вычислим определитель матрицы  $d\mathbf{q}'$ , которая составлена из пересчитанных координат предметных дифференциалов:

$$\det = |d\mathbf{q}'| = \begin{vmatrix} dX'_{ob1} & dX'_{ob2} \\ dY'_{ob1} & dY'_{ob2} \end{vmatrix} \quad (17)$$

Если  $\det = 0$ , то плоскость изображения совпадает с плоскостью выходного зрачка, и необходимо задать другую плоскость изображения.

Матрица  $\mathbf{D}$ , которая участвует в расчете зрачковых координат и волновой абберации определяется следующим образом:

$$\mathbf{D} = ds' \cdot d\mathbf{q}'^{-1}, \quad (18)$$

где

$$d\mathbf{q}'^{-1} = \begin{pmatrix} dY'_{ob2} & -dX'_{ob2} \\ -dY'_{ob1} & dX'_{ob1} \end{pmatrix} \quad (19)$$

матрица, обратная  $d\mathbf{q}'$

2). Выходные зрачковые координаты определяются так:

$$\begin{pmatrix} p'_x \\ p'_y \end{pmatrix} = n_0 \left[ \begin{pmatrix} x' \\ y' \end{pmatrix} - \mathbf{D} \begin{pmatrix} \delta G' \\ \delta g' \end{pmatrix} \right] \quad (20)$$

3). Расчет неизопланатизма аналогичен выражению для близкого изображения:

$$\begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix} = \left[ 1 - \begin{pmatrix} p'_x \cdot V_x \\ p'_y \cdot V_y \\ p_y \end{pmatrix} \right] \cdot 100\% \quad (21)$$

4). Вычисление волновой абберации осуществляется следующим образом:

$$W = [A, P'] - [A, O'] + \frac{1}{2} (\delta G' \quad \delta g') \mathbf{D} \begin{pmatrix} \delta G' \\ \delta g' \end{pmatrix}, \quad (22)$$

где  $[A, P']$ ,  $[A, O']$  – оптические длины данного и главного лучей до поверхности выходной сферы.

5). Расчет дефокусировки и астигматизма идентичен алгоритму для близкого изображения, но в роли  $ds'$  выступает  $d\mathbf{q}'$  и наоборот:

$$d\mathbf{L} = -\frac{Z'}{\det} \begin{pmatrix} dX'_{rup1} dy'_{rup2} - dX'_{rup2} dy'_{rup1} & dX'_{rup2} dx'_{rup1} - dX'_{rup1} dx'_{rup2} \\ 0 & dY'_{rup2} dx'_{rup1} - dY'_{rup1} dx'_{rup2} \end{pmatrix}, \quad (23)$$

$$\det = |ds'| = \begin{vmatrix} dx'_{rup1} & dx'_{rup2} \\ dy'_{rup1} & dy'_{rup2} \end{vmatrix}, \quad (24)$$

$ds'$  - матрица, которая составлена из координат зрачковых дифференциалов.

Проведем коррекцию матрицы  $d\mathbf{L}$ :

$$\mathbf{D} = \begin{pmatrix} dl_{11}(1 - X'^2) - dl_{12} X' Y' & dl_{12}(1 - X'^2) - dl_{22} X' Y' \\ 0 & dl_{22}(1 - Y'^2) - dl_{12} X' Y' \end{pmatrix}. \quad (25)$$

Вычисление дефокусировки выполняется по формуле:

$$dz = \frac{d_{11} + d_{22}}{2}. \quad (26)$$

Для борьбы с потерей значимости проведем следующие преобразования ( $k_1, k_3, k_2$  - вспомогательные величины):

$$k_1 = d_{11} - d_{22}; \quad (27)$$

$$k_2 = 2d_{12},$$

Если эти величины по модулю меньше  $10^{-8}$ , то они принимаются нулевыми.

$$k_3 = \sqrt{k_1^2 + k_2^2}. \quad (28)$$

Если  $k_1 < 0$ , то  $k_1, k_3, k_2$  меняют знак на противоположный.

И, наконец, вычисление астигматизма осуществляется по формуле:

$$az = -k_3. \quad (29)$$

## **Заключение**

Представленные алгоритмы расчета аберраций построены на основе единого аппарата дифференциалов произвольного луча. Дифференциалы и реальные лучи, которые участвуют в расчете аберраций, используются в других алгоритмах, в частности при определении габаритов пучков.

Работа выполнена на кафедре прикладной и компьютерной оптики в рамках НИР №10147 "Исследование и разработка методов компьютерной оптики для моделирования, проектирования, конструирования и изготовления оптических систем".

## **Литература**

1. Родионов С.А. Автоматизация проектирования оптических систем. Л.: Машиностроение, 1982
2. Родионов С.А., Вознесенский Н.Б., Иванова Т.В. Основы оптики.
3. Родионов С.А., Пржевалинский Л.И., Шехонин А.А. Алгоритм определения габаритов пучков в программах расчета лучей через оптическую систему. / В сб. Расчет и конструирование оптических систем, 1974.
4. Родионов С.А. О расчете дифференциалов лучей через оптические системы. // Известия высших учебных заведений. Приборостроение. 1980. Т.23.
5. Родионов С.А., Пржевалинский Л.И., Шехонин А.А. О вычислении волновой аберрации при расчете лучей через оптические системы. // Известия высших учебных заведений. Приборостроение. 1974. Т.17.

# МОДЕЛИРОВАНИЕ ЛОЖНОГО ДВОЙНОГО ЛУЧЕПРЕЛОМЛЕНИЯ

И.Ю. Федорова, С.Г. Фролова, С.В. Златина, М.А. Волкова

Научный руководитель – к.т.н., доц. **А.А. Цуканов**

Произведено моделирование явления ложного двойного лучепреломления для математических моделей классического и кругового поляриметра на базе матричного метода Мюллера. Представленные результаты свидетельствуют о преимуществе схемы кругового поляриметра.

Одним из традиционных видов продукции ОАО "ЛОМО" являются поляризационные приборы, в том числе микроскопы. При их производстве большие трудности вызывает задача контроля двулучепреломления, появляющегося в процессе изготовления при возникновении напряжений в оптических деталях, а также при сборке оптических узлов поляризационных приборов. Сложность связана с тем, что контролируют двулучепреломление, регистрируя визуальное его проявление при помощи поляриметра, построенного по классической схеме со скрещенными поляризатором и анализатором.

По действию на поляризацию элементы оптической схемы поляриметра можно разделить на две группы.

К первой относятся элементы, влияющие на состояние поляризации: поляризатор, анализатор, фазовые пластинки; ко второй – такие элементы, как объектив, конденсор, состоящие из преломляющих поверхностей и не изменяющие характер поляризации. Однако преломляющие поверхности вносят разворот плоскости колебаний (векторов  $E$  и  $H$ ) [1]. Наличие поворота плоскости колебаний не позволяет получить полного гашения проходящего излучения при скрещенных поляризаторе (P) и анализаторе (A) даже при отсутствии объекта в ходе лучей. Это явление названо ложным двойным лучепреломлением, так как характер формируемой картины схож с аналогичной картиной при наличии между P и A одноосного кристалла.

Проявление двулучепреломления в контролируемом изделии выглядит как светлые (белые) пятна и полосы на темном фоне. На эту картину одновременно накладывается картина ложного двойного лучепреломления [2, 3].

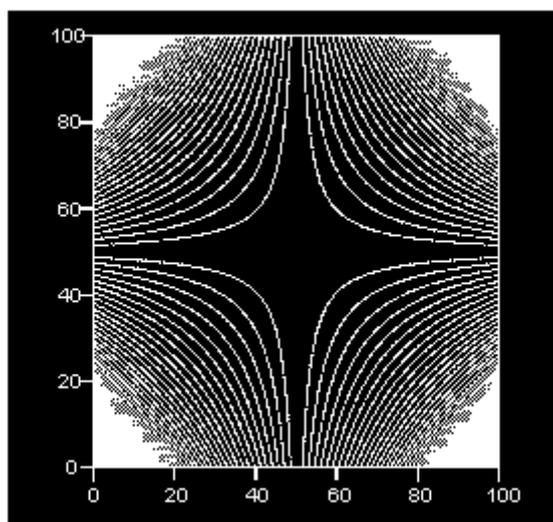


Рис. 1. Проявление ложного двойного лучепреломления

Ложное двойное лучепреломление проявляется в виде креста, показанного на рис. 1, полное гашение в котором имеет место по двум взаимно перпендикулярным направлениям, совпадающими с направлениями пропускания P и A. Если вышеуказанное просветление, вызванное двулучепреломлением, попадает на просветленный участок

картины ложного двулучепреломления, то, вероятнее всего, оно не будет замечено. Чтобы этого избежать, контролер разворачивает исследуемую деталь вокруг ее оптической оси, добиваясь попадания проявлений двулучепреломляющих участков на направление гашения указанного креста.

Однако известно [2, 4, 5], что в местах действия двулучепреломления оптическая деталь проявляет свойства одноосного кристалла, следовательно, при вращении контролируемой детали картина будет меняться (просветления могут быть погашены). Действие классического поляриметра можно выразить уравнением (1).

Для избежания искажений применяют схему кругового поляриметра [3, 4, 5], включающего в себя также две фазовые пластинки  $\lambda/4$  ( $+45^\circ$  и  $-45^\circ$ ), действие которого может быть выражено уравнением (2).

Для моделирования вышеуказанных явлений выразим действие описанных ранее систем кругового и классического поляриметров с помощью метода Мюллера, построенного на использовании 4-х мерных матриц [5]. Метод Мюллера основан на использовании физически измеримых величин (параметров интенсивности Стокса). Он позволяет описать состояние поляризации световых пучков в оптической системе (ОС), причем входящий в ОС свет может быть как поляризованным, так и неполяризованным. Использование параметров Стокса требует использования 4-х мерных матриц.

$$G_1 = AN_{i,j}TP \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (1)$$

$$G_2 = AN_3N_{i,j}TN_1P \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (2)$$

$G_1$  и  $G_2$  представляют собой значения интенсивности,

$$G_1 = 0.5 \sin^2 2\beta \sin^2(\delta/2), \quad G_2 = 0.5 \sin^2(\delta/2),$$

где  $\beta$  – угол ориентации указанного одноосного кристалла в конкретной точке,  $\delta = 2\pi\Delta/\lambda$  – разность фаз,  $\lambda$  – рабочая длина волны,  $\Delta$  – разность хода.

В выражениях (1), (2) правая столбцовая матрица-вектор Стокса определяет неполяризованный свет, входящий в поляризатор;  $P$ ,  $A$  – матрицы, описывающие действие поляризатора и анализатора соответственно;  $N_1$ ,  $N_3$  описывают действие фазовых пластинок  $\lambda/4$ ; комбинация матриц  $TN_{i,j}$  моделирует действие объекта. Главные направления элементов системы измеряем от направления пропускания поляризатора, которые принимаем за ось абсцисс ОХ. Эти матрицы определяются следующим образом:

$$P = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}; \quad (3)$$

здесь использована матрица поляризатора, пропускающего полностью колебания, параллельные оси ОХ, и гасящего колебания, параллельные оси ОУ.

$$A = \frac{1}{2} \begin{bmatrix} 1 & \cos(2\Theta) & \sin(2\Theta) & 0 \\ \cos(2\Theta) & \cos^2(2\Theta) & \sin(2\Theta)\cos(2\Theta) & 0 \\ \sin(2\Theta) & \sin(2\Theta)\cos(2\Theta) & \sin^2(2\Theta) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4)$$

Здесь  $\Theta=\pi/2$  – азимут пропускания анализатора, т.е. угол между направлением пропускания и осью ОХ.

Задание в общем виде матрицы (4) позволяет моделировать ошибки установки поляризатора и анализатора друг относительно друга.

Матрицы, описывающие действие фазовых пластинок, также целесообразно представить в общем виде:

$$N_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2(2p_1) + \sin^2(2p_1)\cos(\delta_1) & \cos(2p_1)\sin(2p_1)(1-\cos(\delta_1)) & -\sin(2p_1)\sin(\delta_1) \\ 0 & \cos(2p_1)\sin(2p_1)(1-\cos(\delta_1)) & \sin^2(2p_1) + \cos^2(2p_1)\cos(\delta_1) & \cos(2p_1)\sin(\delta_1) \\ 0 & \sin(2p_1)\sin(\delta_1) & -\cos(2p_1)\sin(\delta_1) & \cos(\delta_1) \end{bmatrix}, (5)$$

$$N_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2(2p_2) + \sin^2(2p_2)\cos(\delta_2) & \cos(2p_2)\sin(2p_2)(1-\cos(\delta_2)) & -\sin(2p_2)\sin(\delta_2) \\ 0 & \cos(2p_2)\sin(2p_2)(1-\cos(\delta_2)) & \sin^2(2p_2) + \cos^2(2p_2)\cos(\delta_2) & \cos(2p_2)\sin(\delta_2) \\ 0 & \sin(2p_2)\sin(\delta_2) & -\cos(2p_2)\sin(\delta_2) & \cos(\delta_2) \end{bmatrix}, (6)$$

где  $p_1, p_2$  – углы, определяющие ориентацию фазовых пластинок;  $\delta_1, \delta_2$  – углы фазового сдвига фазовых пластинок (в случае кругового поляризатора  $p_1=p_2=\pi/4$ ;  $\delta_1=\delta_2=\pi/2$ ).

Действие объекта определим как комбинацию (произведение) двух матриц: матрицы вращения  $T$ , которая описывает разворот плоскости колебаний за счет ложного двулучепреломления, и матрицы  $N_{i,j}$ , моделирующей двупреломляющее действие объекта, вводящее разность фаз  $\delta_{i,j}$  за счет напряжения в нем.

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\alpha_{i,j} & \sin 2\alpha_{i,j} & 0 \\ 0 & -\sin 2\alpha_{i,j} & \cos 2\alpha_{i,j} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, (7)$$

$$N_{i,j} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2(2\beta) + \sin^2(2\beta)\cos(\delta_{i,j}) & \cos(2\beta)\sin(2\beta)(1-\cos(\delta_{i,j})) & -\sin(2\beta)\sin(\delta_{i,j}) \\ 0 & \cos(2\beta)\sin(2\beta)(1-\cos(\delta_{i,j})) & \sin^2(2\beta) + \cos^2(2\beta)\cos(\delta_{i,j}) & \cos(2\beta)\sin(\delta_{i,j}) \\ 0 & \sin(2\beta)\sin(\delta_{i,j}) & -\cos(2\beta)\sin(\delta_{i,j}) & \cos(\delta_{i,j}) \end{bmatrix}. (8)$$

Значения величин  $\alpha_{i,j}$ , входящих в матрицу (7), вычисляются в результате расчета хода действительных лучей по методике из [3], а значения величин  $\delta_{i,j}$  задаем аналогично  $\beta$ . Последняя величина определяет положение одноосного кристалла, описывающего место проявления напряжения.

Матрица  $N_{i,j}$  в составе уравнений (1) и (2) позволяет моделировать действие двулучепреломления в испытуемом объекте, которое может быть визуализировано на экране монитора компьютера, например, в виде светораспределения, представленного на рис. 2: картина рис. 2а соответствует случаю классического поляризатора, а рис. 2б – круговому.

Из сравнения картин следует, что использование схемы кругового поляризатора позволяет устранить наложение картины, определяемой ложным двойным лучепреломлением, на картину, определяемую контролируемым двулучепреломлением (вызванным напряжениями). Использование описанной математической модели дает возможность описать и исследовать влияние дефектов изготовления и установки всех элементов схемы поляризатора, что в итоге позволяет оправданно назначить допуски на изготовление указанных элементов.

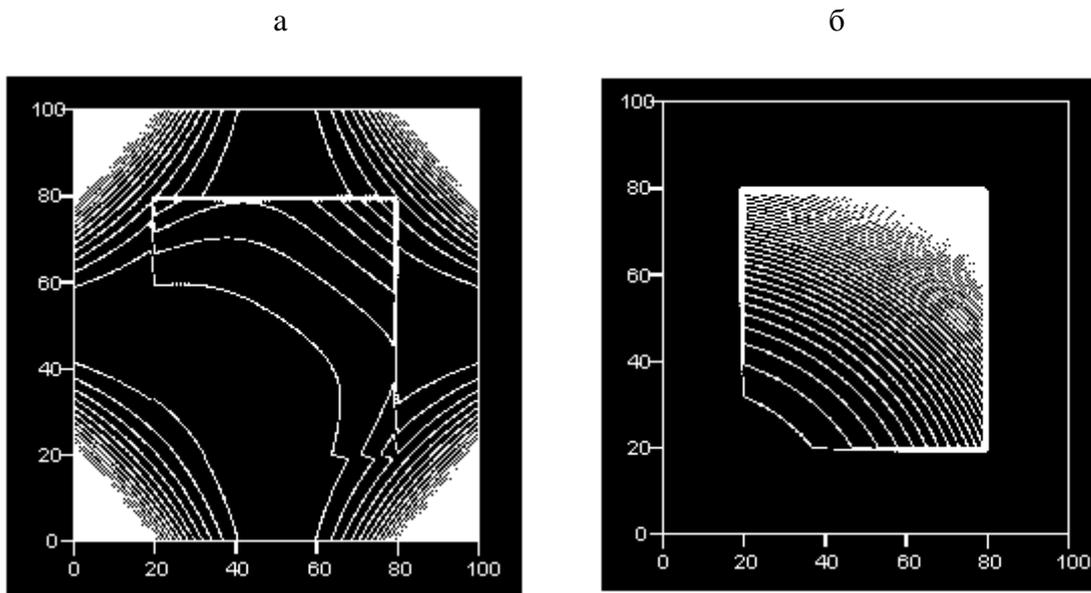


Рис. 2. Модель действия напряжений: а – для случая классической схемы поляриметра, б – для случая схемы кругового поляриметра

Описанная методика была использована при проектировании на ОАО "ЛОМО" поляриметра ПЛМ-1 для разбраковки оптических элементов (заготовок, деталей и сборок), включая микрообъективы и конденсоры поляризационных микроскопов.

В заключении следует отметить, что использование матриц Мюллера позволяет получить решение в общем виде при использовании любого источника света. Аналогичные результаты получены при моделировании на основе применения матриц Джонса, которые используются при работе со строго монохроматическим точечным источником света.

## Литература

1. Борн М., Вольф Э. Основы оптики. М., Наука, 1970. 856 с.
2. Ринне Ф., Берек М. Оптические исследования при помощи поляризационного микроскопа. ОНТИ, 1937. 332 с.
3. И. А. Дружинина, С.Н. Натаровский, Т.Ф. Селезнева и др. Ложное двойное лучепреломление в поляризационном микроскопе// оптический журнал, 2002. Т. 69. №9. С. 76-80.
4. Шерклифф У. Поляризованный свет. М., Мир, 1965. 264 с.
5. Васильев Б.И. Оптика поляризационных приборов. М. Машиностроение, 1969. 208 с.

# ОБЪЕКТИВЫ ДЛЯ ОПТИЧЕСКОЙ ЗАПИСИ И СЧИТЫВАНИЯ ИНФОРМАЦИИ.

Н. А. Лаптева

Научный руководитель – д.т.н., проф. Л.Н. Андреев

Рассмотрены оптические системы монохроматы для преобразования лазерного излучения, которые могут быть использованы в широкой спектральной области, т.е. имеющие полимонохроматическую коррекцию aberrаций. Указаны свойства этих систем и область применения. Для иллюстрации приведены результаты расчетов объективов для оптической записи и считывания информации, обладающих дифракционным качеством изображения на разных длинах волн.

## Введение

В последнее время оптическая память становится одной из перспективных альтернатив магнитной или полупроводниковой памяти. Под оптической памятью понимается такое устройство, в котором хотя бы один из видов выборки информации (запись, стирание, считывание) осуществляется с использованием лучистой энергии. Преимуществом оптической памяти является быстрый доступ к массивам информации, высокая надёжность хранения, отсутствие потребления энергии в статическом состоянии и другие. Плотность записи информации в оптическом запоминающем устройстве принципиально ограничена только дифракционным пределом применяемой оптики. [1]

Принципиальная схема устройства для оптической записи включает лазер (источник излучения), оптический диск (приёмник излучения), фокусирующий объектив, а также различные электронные и прецизионные механические устройства. Одним из основных узлов устройства является фокусирующий объектив, от числовой апертуры и качества изображения которого зависит плотность записи на оптическом диске.

Для оптической записи и считывания информации необходимы линзовые фокусирующие объективы с высокой числовой апертурой и увеличенным, по сравнению с обычными микрообъективами, рабочим расстоянием (не менее 4мм). Кружок рассеяния в этих системах, в зависимости от решаемых задач, не должен превышать 0,6...2 мкм, а качество изображения для точки на оси должно быть близким к дифракционному. Линейные поля таких объективов, как правило, не превышают 0,1 ... 0,2 мм, что вполне достаточно для слежения за информационной дорожкой. Важное требование к таким объективам - минимальная общая масса. [1]

Проведенный анализ литературы [3] показал, что за рубежом сначала велись расчеты многолинзовых систем состоящих из положительных и отрицательных линз, что приводило к жестким допускам на центрировку линз, потом при расчете объективов широко использовались однолинзовые системы с одной или двумя асферическими поверхностями высшего порядка. Такие объективы сложны в изготовлении, сложен контроль формы асферических поверхностей, поэтому большой процент от общей стоимости оптической системы записи и считывания информации составляет стоимость объектива.

В ЛИТМО в 80-е годы были рассчитаны объективы с апертурой  $A = 0.4-0.45$ , имеющие только сферические поверхности (рис.1), но они состояли из 2-3 линз.

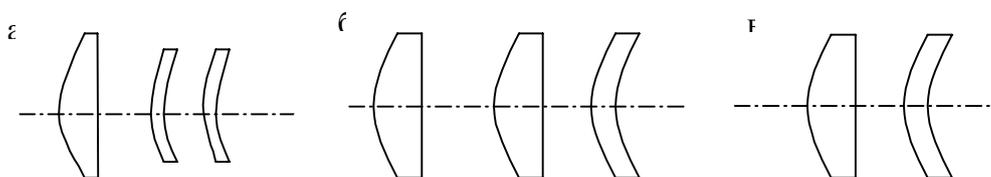


Рис. 1. Различные оптические схемы объективов со сферическими поверхностями  $A = 0,4-0,45$ :  
 а) плосковыпуклая линза и два апланатических мениска; б) две плосковыпуклых линзы и апланатический мениск; в) плосковыпуклая линза с  $n \geq 2,0$  и апланатический мениск.

Целью работы является разработка однолинзового объектива с числовой апертурой 0.35-0.45 с асферической поверхностью 2-го порядка.

Базовая схема объектива включает в себя два оптических модуля (рис.2): защитно-коррекционный элемент (оптический диск) в виде плоско-параллельной пластинки и масштабно-компенсационной части, состоящий из одной линзы. Оптический диск (плоско-параллельная пластинка, толщиной 1,2 мм) выполняет роль защитного слоя и вносит в оптическую систему положительную сферическую aberrацию.[1]



Рис. 2. Асферический однолинзовый объектив  $A = 0,4$ .

### Теоретическая часть и реализация

У объективов для записи информации учитывая их применение прежде всего должны быть исправлены сферическая aberrация ( $DS'$ ) и кома ( $Dy'_k$ ).

$$DS' = -1/2 f' \operatorname{tg}^2 y' S_I, \quad Dy'_k = -3/2 f' \operatorname{tg}^2 y' \operatorname{tg} \psi_I S_{II}.$$

Для объектива с одной асферической поверхностью второго порядка выражения для сумм Зейделя выглядят следующим образом:

$$\bar{S}_I = \bar{P} + \bar{D}\bar{C} = \bar{C}_A$$

$$\bar{S}_{II} = \bar{H}(\bar{P} + \bar{D}\bar{C}) + \bar{W} = \bar{H}\bar{P}_A + \bar{W}$$

$$\text{Где } \bar{D}\bar{C} = \frac{(\bar{b}n' - \bar{b}n)^3}{(n' - n)^2} e^2;$$

$$e - \text{эксцентриситет деформации поверхности } y^2 = 2r_0 x - (1 - e^2)x^2.$$

В таблице 1 приведены результаты расчета основных параметров  $P_A$  и  $W$  для выпукло-плоской линзы в зависимости от  $e^2$  и  $n$ .

$e^2$	0		0,5		1,0		1,5	
$n$	$\bar{W}$	$\bar{P}_A$	$\bar{W}$	$\bar{P}_A$	$\bar{W}$	$\bar{P}_A$	$\bar{W}$	$\bar{P}_A$
1,5	-0,33	2,33	-0,33	0,33	-0,33	-1,67	-0,33	-3,67
1,75	0,24	1,25	0,24	0,37	0,24	-0,52	0,24	-1,41
2,00	0,50	1,00	0,50	0,50	0,50	0	0,50	-0,50

Таблица 1. Результаты расчета основных параметров  $P_A$  и  $W$  для выпуклоплоской линзы в зависимости от  $e^2$  и  $n$ .

На рис.3 представлен график зависимости  $C_A = f(e^2, n)$ . Связь между  $C_A$  и  $e^2$  линейна, поэтому можно предварительно найти  $e^2$ , т.е. определить форму асферической поверхности. При  $e^2 = 0$  – сфера,  $e^2 = 1$  – параболоид,  $0 < e^2 < 1$  – эллипсоид,  $e^2 > 1$  – гиперboloид.

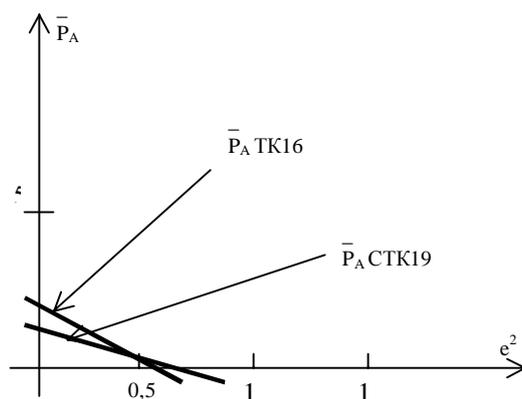


Рис. 3. Зависимости  $P_A = f(e^2, n)$  для выпуклоплоских линз из стекол ТК16 и СТК19.

По степени коррекции хроматических aberrаций оптических систем, различают: монохроматы (системы для дискретной длины волны или узкой спектральной области), ахроматы, у которых спектральная область простирается от линии F' до C', апохроматы, у которых спектральная область расширена до линии g или G и суперапохроматы (системы, у которых коррекция выполнена для видимой и УФ областей спектра).[2]

Однако, при проектировании оптических систем для преобразования лазерного излучения было обнаружено, что системы, рассчитанные на дискретную длину волны лазера имеют также вполне удовлетворительную коррекцию aberrаций при перефокусировке и для других длин волн. Такие оптические системы могут быть названы полимонохроматами.

Для иллюстрации в таблице 2 приведены результаты расчетов объектива для оптической записи и считывания информации (m – координаты лучей на выходном зрачке, N/л – волновая aberrация, z – величина характеризующая отступление от условия изопланазии).

л, мкм	m, мм	Ду', мм	N/л	z, %	Число Штреля
0,488	1,6	0,000072	0,1048	0,8272	0,79
	$1,6\sqrt{s}$	0,000799	0,0467	0,5598	
	$1,6\sqrt{S}$	0,000583	-0,0489	0,3335	
	$1,6\sqrt{j}$	-0,000056	-0,0917	0,1472	
	0	0	0	0	
0,6328	1,6	0,001593	0,1436	0,6382	0,8
	$1,6\sqrt{s}$	0,001576	0,0048	0,4152	
	$1,6\sqrt{S}$	0,000745	-0,1134	0,2354	
	$1,6\sqrt{j}$	-0,000294	-0,1370	0,0974	
	0	0	0	0	
0,800	1,6	0,002528	0,1608	0,5338	0,81
	$1,6\sqrt{s}$	0,002079	0,0041	0,3355	
	$1,6\sqrt{S}$	0,000889	-0,1139	0,1814	
	$1,6\sqrt{j}$	-0,000390	-0,1331	0,0701	
	0	0	0	0	

Таблица 2. Результаты расчетов объектива для оптической записи и считывания информации.

## Заключение

Анализ проведенных абберационных расчетов (табл.2) показывает, что система может применяться с различными лазерами, в том числе Ar ( $\lambda = 0,488$  мкм), He – Ne ( $\lambda = 0,632$  мкм) и полупроводниковым ( $\lambda = 0,800$  мкм), так как для всех перечисленных длин волн сохраняется практически дифракционное качество изображения. Это связано с высокой коррекцией сферохроматической абберации при неисправленном хроматизме положения.

## Литература

1. Андреев Л.Н. Прикладная теория аббераций. Учебное пособие. СПб. СПб ГИТМО (ТУ), 2002, - 96 с.
2. Слюсарев Г.Г. Расчет оптических систем. Л.: Машиностроение, 1975, -639 с.
3. Заявки ФРГ №3231984, № 3236666; патенты: США №4257679, № 4332442, № 4258981, № 4270842, №4.768.867, №4.729.645, Япония №63-31.766, EP №0.252.614

# ОБЪЕКТИВЫ ДЛЯ ОПТИЧЕСКОЙ ЗАПИСИ И СЧИТЫВАНИЯ ИНФОРМАЦИИ

Н. А. Лаптева

Научный руководитель – д.т.н., проф. Л.Н. Андреев

Рассмотрены оптические системы монохроматы для преобразования лазерного излучения, которые могут быть использованы в широкой спектральной области, т.е. имеющие полимонохроматическую коррекцию aberrаций. Указаны свойства этих систем и область применения. Для иллюстрации приведены результаты расчетов объективов для оптической записи и считывания информации, обладающих дифракционным качеством изображения на разных длинах волн.

## Введение

В последнее время оптическая память становится одной из перспективных альтернатив магнитной или полупроводниковой памяти. Под оптической памятью понимается такое устройство, в котором хотя бы один из видов выборки информации (запись, стирание, считывание) осуществляется с использованием лучистой энергии. Преимуществом оптической памяти является быстрый доступ к массивам информации, высокая надёжность хранения, отсутствие потребления энергии в статическом состоянии и другие. Плотность записи информации в оптическом запоминающем устройстве принципиально ограничена только дифракционным пределом применяемой оптики [1].

Принципиальная схема устройства для оптической записи включает лазер (источник излучения), оптический диск (приёмник излучения), фокусирующий объектив, а также различные электронные и прецизионные механические устройства. Одним из основных узлов устройства является фокусирующий объектив, от числовой апертуры и качества изображения которого зависит плотность записи на оптическом диске.

Для оптической записи и считывания информации необходимы линзовые фокусирующие объективы с высокой числовой апертурой и увеличенным, по сравнению с обычными микрообъективами, рабочим расстоянием (не менее 4 мм). Кружок рассеяния в этих системах, в зависимости от решаемых задач, не должен превышать 0,6–2 мкм, а качество изображения для точки на оси должно быть близким к дифракционному. Линейные поля таких объективов, как правило, не превышают 0,1–0,2 мм, что вполне достаточно для слежения за информационной дорожкой. Важное требование к таким объективам – минимальная общая масса [1].

Проведенный анализ литературы [3] показал, что за рубежом сначала велись расчеты многолинзовых систем, состоящих из положительных и отрицательных линз, что приводило к жестким допускам на центрировку линз, потом при расчете объективов широко использовались однолинзовые системы с одной или двумя асферическими поверхностями высшего порядка. Такие объективы сложны в изготовлении, сложен контроль формы асферических поверхностей, поэтому большой процент от общей стоимости оптической системы записи и считывания информации составляет стоимость объектива.

В ЛИТМО в 80-е гг. были рассчитаны объективы с апертурой  $A = 0,4-0,45$ , имеющие только сферические поверхности (рис. 1), но они состояли из 2–3 линз.

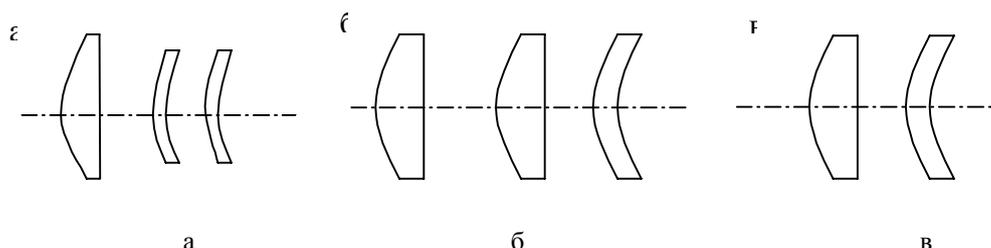


Рис. 1. Различные оптические схемы объективов со сферическими поверхностями  $A = 0,4-0,45$ : а – плосковыпуклая линза и два апланатических мениска; б – две плосковыпуклых линзы и апланатический мениск; в – плосковыпуклая линза с  $n \geq 2,0$  и апланатический мениск

Целью работы является разработка однолинзового объектива с числовой апертурой 0.35-0.45 с асферической поверхностью 2-го порядка.

Базовая схема объектива включает в себя два оптических модуля (рис.2): защитно-коррекционный элемент (оптический диск) в виде плоско-параллельной пластинки и масштабно-компенсационной части, состоящий из одной линзы. Оптический диск (плоско-параллельная пластинка, толщиной 1,2 мм) выполняет роль защитного слоя и вносит в оптическую систему положительную сферическую aberrацию.[1].



Рис. 2. Асферический однолинзовый объектив  $A = 0,4$

### Теоретическая часть и реализация

У объективов для записи информации, учитывая их применение, прежде всего должны быть исправлены сферическая aberrация ( $\Delta S'$ ) и кома ( $\Delta y'_k$ ):

$$\Delta S' = -1/2 f' \cdot \text{tg}^2 \sigma' \cdot \bar{S}_I,$$

$$\Delta y'_k = -3/2 \cdot f' \cdot \text{tg}^2 \sigma' \cdot \text{tg} \omega_1 \cdot \bar{S}_{II}.$$

Для объектива с одной асферической поверхностью второго порядка выражения для сумм Зейделя выглядят следующим образом:

$$\bar{S}_I = \bar{P} + \Delta \bar{P} = \bar{P}_A$$

$$\bar{S}_{II} = \bar{H} \cdot (\bar{P} + \Delta \bar{P}) + \bar{W} = \bar{H} \cdot \bar{P}_A + \bar{W},$$

где  $\Delta \bar{P} = \frac{(\alpha' \cdot n' - \alpha \cdot n)^3}{(n' - n)^2} e^2$ ,  $e$  – эксцентриситет деформации поверхности,

$$y^2 = 2r_0 x - (1 - e^2)x^2.$$

В табл. 1 приведены результаты расчета основных параметров  $\bar{P}_A$  и  $\bar{W}$  для выпукло-плоской линзы в зависимости от  $e^2$  и  $n$ .

$e^2$	0		0,5		1,0		1,5	
$n$	$\bar{W}$	$\bar{P}_A$	$\bar{W}$	$\bar{P}_A$	$\bar{W}$	$\bar{P}_A$	$\bar{W}$	$\bar{P}_A$
1,5	-0,33	2,33	-0,33	0,33	-0,33	-1,67	-0,33	-3,67
1,75	0,24	1,25	0,24	0,37	0,24	-0,52	0,24	-1,41
2,00	0,50	1,00	0,50	0,50	0,50	0	0,50	-0,50

Таблица 1. Результаты расчета основных параметров  $\bar{P}_A$  и  $\bar{W}$  для выпукло-плоской линзы в зависимости от  $e^2$  и  $n$

На рис. 3 представлен график зависимости  $\bar{P}_A = f(e^2, n)$ . Связь между  $\bar{P}_A$  и  $e^2$  линейна, поэтому можно предварительно найти  $e^2$ , т.е. определить форму асферической поверхности. При  $e^2 = 0$  – сфера,  $e^2 = 1$  – параболоид,  $0 < e^2 < 1$  – эллипсоид,  $e^2 > 1$  – гиперболоид.

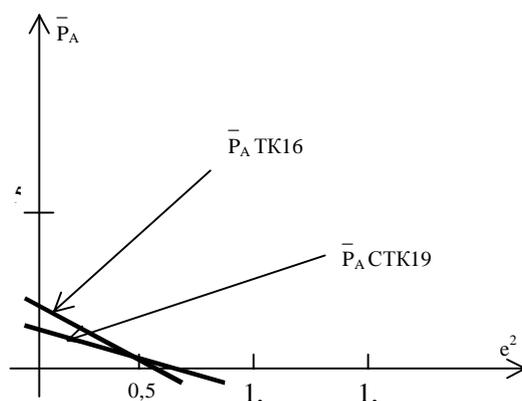


Рис. 3. Зависимости  $\bar{P}_A = f(e^2, n)$  для выпуклоплоских линз из стекол ТК16 и СТК19

По степени коррекции хроматических aberrаций оптических систем различают: монохроматы (системы для дискретной длины волны или узкой спектральной области), ахроматы, у которых спектральная область простирается от линии F' до C', апохроматы, у которых спектральная область расширена до линии g или G, и суперапохроматы (системы, у которых коррекция выполнена для видимой и УФ областей спектра) [2].

Однако при проектировании оптических систем для преобразования лазерного излучения было обнаружено, что системы, рассчитанные на дискретную длину волны лазера, имеют также вполне удовлетворительную коррекцию aberrаций при перефокусировке и для других длин волн. Такие оптические системы могут быть названы полимонохроматами. Для иллюстрации в табл. 2 приведены результаты расчетов объектива для оптической записи и считывания информации ( $m$  – координаты лучей на выходном зрачке,  $N/\lambda$  – волновая aberrация,  $\eta$  – величина, характеризующая отступление от условия изопланазии).

$\lambda$ , мкм	$m$ , мм	$\Delta y'$ , мм	$N/\lambda$	$\eta$ , %	Число Штреля
0,488	1,6	0,000072	0,1048	0,8272	0,79
	$1,6\sqrt{s}$	0,000799	0,0467	0,5598	
	$1,6\sqrt{S}$	0,000583	-0,0489	0,3335	
	$1,6\sqrt{j}$	-0,000056	-0,0917	0,1472	
	0	0	0	0	
0,6328	1,6	0,001593	0,1436	0,6382	0,8
	$1,6\sqrt{s}$	0,001576	0,0048	0,4152	
	$1,6\sqrt{S}$	0,000745	-0,1134	0,2354	
	$1,6\sqrt{j}$	-0,000294	-0,1370	0,0974	
	0	0	0	0	
0,800	1,6	0,002528	0,1608	0,5338	0,81
	$1,6\sqrt{s}$	0,002079	0,0041	0,3355	
	$1,6\sqrt{S}$	0,000889	-0,1139	0,1814	
	$1,6\sqrt{j}$	-0,000390	-0,1331	0,0701	
	0	0	0	0	

Таблица 2. Результаты расчетов объектива для оптической записи и считывания информации

### Заклучение

Анализ проведенных aberrационных расчетов (табл. 2) показывает, что система может применяться с различными лазерами, в том числе Ar ( $\lambda = 0,488$  мкм), He-Ne

( $\lambda = 0,632$  мкм) и полупроводниковым ( $\lambda = 0,800$  мкм), так как для всех перечисленных длин волн сохраняется практически дифракционное качество изображения. Это связано с высокой коррекцией сферохроматической aberrации при неисправленном хроматизме положения.

## Литература

1. Андреев Л.Н. Прикладная теория aberrаций. Учебное пособие. СПб. СПб ГИТМО (ТУ), 2002. 96 с.
2. Слюсарев Г.Г. Расчет оптических систем. Л.: Машиностроение, 1975. 639 с.
3. Заявки ФРГ №3231984, № 3236666; патенты: США №4257679, № 4332442, № 4258981, № 4270842, №4.768.867, №4.729.645, Япония №63-31.766, EP №0.252.614

# СВЕТОДИОДЫ В МИКРОСКОПИИ

А.Л. Дубовиков, М.А. Волкова, С.В. Златина

Научные руководители – к.т.н., доц. **А.А. Цуканов**,

д.т.н., проф. Г.Н. Грязин

Работа посвящена сравнительному анализу параметров различных светодиодных источников света, в том числе матриц, с точки зрения их использования в качестве основного источника света для осветительных устройств микроскопов.

## Введение

Об использовании светодиодов (СД) в системах технического зрения (СТЗ) известно [1,2]. В [1] приведены типичные характеристики излучения СД, из которых следует перспективность их использования также и в визуальных каналах, например, микроскопов. На рис. 1 приведена кривая спектра излучения белого СД и цветных СД из [1]. Эти данные говорят о возможности реализации белого свечения за счет использования триады или большего количества СД, перекрывая тем самым большую площадь цветового треугольника. Из хода кривых также следует, что спектр излучения СД захватывает ближние УФ и ИК области.

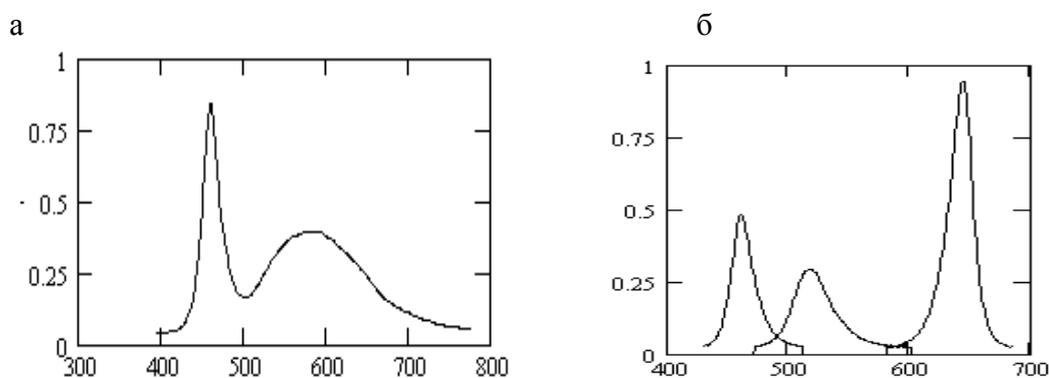


Рис. 1. а – спектр излучения белого светодиода, б – спектр излучения цветных светодиодов

По данным из [3], цветовые координаты белого СД составляют  $x = 0.29–0.36$ ,  $y = 0.27–0.34$ . При этом следует отметить, что координаты  $x$  и  $y$  для ксеноновой лампы при  $T=6100^\circ\text{K}$  равны 0.33 (это координаты источника белого цвета типа E), для ртутной лампы  $x = 0.307$ ,  $y = 0.383$ . Укажем, что в качестве эталона белого цвета приняты источники света А, В, С и Д. Ниже указаны их координаты  $x$  и  $y$  и цветовые температуры.

Источник А:  $x=0.448$ ,  $y=0.407$ ,  $T=2856^\circ\text{K}$ .

Источник В:  $x=0.348$ ,  $y=0.352$ ,  $T=4870^\circ\text{K}$ .

Источник С:  $x=0.310$ ,  $y=0.316$ ,  $T=6770^\circ\text{K}$ .

Источник Д:  $x=0.313$ ,  $y=0.329$ ,  $T=6500^\circ\text{K}$ .

Источник А определяется лампой накаливания с вольфрамовой нитью (мощностью 100 Вт), источники В и С получаются от той же лампы с помощью фильтров и имитируют прямое солнечное освещение и рассеянный дневной свет. Источник Д не стандартизован и определяет усредненное действие дневного света [4].

В табл. 1 [5] приведены координаты цвета излучения Солнца и рассеянной небом солнечной радиации. Из таблицы ясно, что понятие цвета может носить географический характер, весьма неоднозначно, субъективно и требует серьезного подхода, особенно если учесть, что эти вопросы в колориметрии еще не до конца изучены [6].

Координаты	Солнце	Небо С-Пб	Небо Алма-Ата
x	0.321	0.310	0.249
y	0.332	0.336	0.253
z	0.346	0.354	0.498

Таблица 1. Координаты цвета излучения Солнца и рассеянной небом солнечной радиации

Отметим, что цветовые координаты белого СД достаточно близки к источнику С, т.е. дневному рассеянному свету, что делает актуальным вопрос о исследовании возможности таких светодиодов в микроскопии, где требования к цветопередаче всегда высоки.

В СТЗ детектором обычно выступает фотоэлектрический приемник, например, линейный или матричный ПЗС, а в визуальном канале детектором является глаз, что определяет одно из основных отличий таких систем. Для фотоэлектрического приемника главной энергетической характеристикой излучения СД является лучистый поток, а для глаза – его яркостная характеристика. Мощность излучения современных единичного СД не превышает 1 Вт (обычно не более 150–200 мВт), а вот коэффициент светоотдачи может принимать значения от 10 до 60 лм/Вт. Это говорит о высоких яркостных характеристиках СД, сопоставимых с лампами накаливания и даже дуговыми источниками света [4]. Невысокие значения мощности излучения СД приводят к использованию в СТЗ одновременно нескольких СД, расположенных на какой-либо поверхности и образующих линейку или матрицу источников света [2]. При решении энергетической задачи это приводит к другим проблемам. Например, СД одной марки имеют довольно большой разброс по мощности излучаемого пучка и его яркостной характеристике, что при освещении близко расположенных объектов приводит к трудностям в получении равномерного освещения [7], а в случае слабо рассеивающего объекта появляются жесткие требования на геометрическое расположение СД в линейке или матрице в зависимости от режима работы и апертуры оптической системы, строящей изображение освещаемого объекта [8].

### Теоретически-экспериментальная часть

Авторами совместно с "Промсервисом" была проведена экспериментальная проверка по использованию СД в микроскопии. Использовались СД фирмы ОПТЭЛ (г. Москва). Были сняты спектральные характеристики белых и цветных СД. Рассчитаны их цветовые координаты, разработано осветительное устройство со СД для микроскопов.

На рис.1 приведена экспериментально полученная кривая спектра излучения белого СД. Пик в области  $\lambda = 470$  нм говорит о высокой цветовой температуре (8000–9000 °К) и, следовательно, о возможности более полной реализации контрастной чувствительности глаза, а также о потенциальной возможности в телеканале получить более высокий контраст на высоких пространственных частотах. Этот факт был подтвержден экспериментально. На рис. 2 приведены кривые передачи контраста изображения в телеканале микроскопа ПОЛАМ Р-312 (об 9×/0.2 190-П, прямой тубус) при использовании в качестве источника света лампы накаливания и белого СД. Изображения фиксировались на телекамеру VNC-743 фирмы ЭВС и обрабатывалось с помощью компьютера. В составе комплекса использовалась программа OSC-16, кроме того, были использованы средства MathCAD. В качестве объекта использовалась мира с пространственными частотами 455, 333, 250, 167, 91 ММ-1. Из общего изображения был выделен одинаковый фрагмент и выполнен расчет контраста изображения и построен график функции передачи контраста. Из хода кривых видим, что на средних частотах для СД ПФ выше примерно в 2 раза.

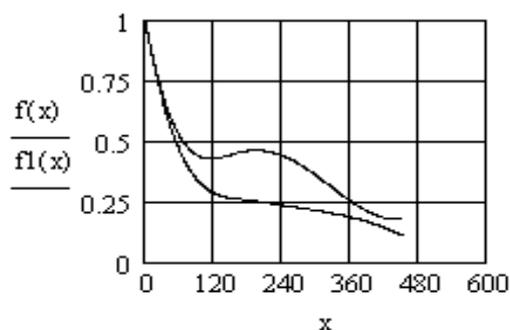


Рис. 2. Кривые передачи контраста изображения в телеканале микроскопа ПОЛАМ

В табл. 2 приведены основные характеристики СД России, Белоруссии. Достоинством СД является низкое потребление энергии, небольшие световые потоки. Значение величин излучаемой мощности СД является важным для теле-, фото- каналов, так как в отличие от глаза для фотоприемников важна не яркость, а поток излучения.

$$\Omega = 2\pi(1 - \cos \theta)$$

Производитель	Цвет свечения	длина волны	сила света	угол излучения	световой поток	яркость
		$\lambda$ [нм]	$J$ [кД]	$2\theta_{0.5}$	$\Phi = J \cdot \Omega$ [лм]	$B$ [нт] $\times 10^{-6}$
ОПТЭЛ Москва	Полноцвет	-	25	8°	0,383	25
	Красный	640	25	4°	0,096	25
	Синий	475	8	4°	0,031	8
	Зеленый	572	7	4°	0,027	7
Корвет Москва	Полноцвет	-	2	60°	1,684	2
	Красный	624	7	20°	0,668	7
	Синий	470	5	10°	0,12	5
	Зеленый	525	12	10°	0,287	12
США	Полноцвет	-	180	10°	4,304	180
	Красный	627	660	10°	15,78	660
	Синий	470	100	10°	2,391	100
	Зеленый	530	500	10°	11,955	500
Белоруссия	Полноцвет	-	6,5	15°	0,349	6,5
	Полноцвет	-	4	30°	0,856	4
	Полноцвет	-	2,5	75°	3,246	2,5
	Полноцвет	-	6,5	15°	0,349	6,5
	Полноцвет	-	5	20°	0,477	5
	Полноцвет	-	7,2	18°	0,557	7,2
	Красный	660	2,5	50°	1,472	2,5
	Синий	470	3,2	25°	0,477	3,2
	Синий	470	1,6	60°	1,347	1,6
	Синий	470	2	25°	0,298	2
	Синий	470	2	60°/30°	1,684/0,428	2
	Синий	470	3,2	25°	0,477	3,2
	Зеленый	520	6,0	25°	0,894	6,0
	ПЛАНЕТА Новгород	Синий	470	2-4	20°	0,19-0,38
Синий		470	2-4	45°	0,957-1,913	2-4

Таблица 2. Основные характеристики СД России, Белоруссии

Нами были вычислены координаты цвета ( $x, y, z$ ) для указанных СД, причем для белого СД координаты цвета вычислялись дважды с учетом и без действия светофильтра ЖС-12. В табл. 3 приведены расчетные значения координат цвета:

	Белый СД	Белый СД+ЖС-12	Красный СД	Синий СД	Зеленый СД
$x$	0.335	0.376	0.699	0.137	0.193
$y$	0.315	0.374	0.301	0.051	0.678
$z$	0.351	0.250	0.000	0.812	0.129

Таблица 3. Расчетные значения координат цвета

Результаты табл. 3 показывают, что белый СД без фильтра имеет цветовые координаты, близкие источнику типа Е, для которого все три координаты равны 0.333, а спектральная характеристика излучения представляет прямую, параллельную оси абсцисс. Отсюда следует, что белый СД как источник света в оптическом приборе может обеспечить высокое качество цветопередачи в изображении.

### Обсуждение результатов

Использование СД в качестве источника света в микроскопе интересно как с точки зрения их высоких эксплуатационных свойств (низкий нагрев, большой срок службы и пр.), так и в связи с тем, что в современных микроскопах удельный вес все более переводится с визуального на телевизионный канал. Таким образом, в микроскопе одновременно используется СТЗ и визуальный канал, что часто приводит к необходимости решения задачи баланса и целесообразности использования в осветительном устройстве матрицы СД.

В [3] приведены параметры полноцветной матрицы СД размером  $8 \times 8$  СД ( $79,5 \times 79,5$  мм). Каждый из СД состоит из четырех цветных СД RGBG. По габаритным характеристикам эта матрица для микроскопии великовата, но принципиальную задачу решает. В то же время следует ожидать появление на рынке матриц СД с более приемлемыми характеристиками. Постараемся их сформулировать.

Коэффициент светоотдачи отдельного СД желательно иметь не ниже 10 лм/Вт, апертура излучаемого пучка должна быть не более  $0,17$  ( $2\theta \leq 20^\circ$ ), шаг матрицы около 0.5 мм, количество элементов в матрице при указанном шаге и размере матрице  $30 \times 30$  мм составит 3600 штук. Экономические соображения могут, с одной стороны, скорректировать задачу (1\$ на пиксель), а с другой – привести к иным схемным решениям, сочетающим свойства матрицы СД и возможности оптических решений. На рис.3 приведены варианты, позволяющие реализовать критический способ освещения (рис.3а) и освещение по методу Келера (рис.3б).

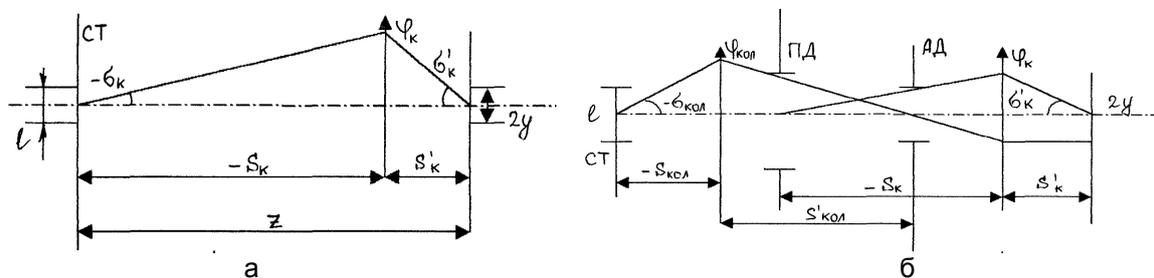


Рис. 3. Реализация различных способов освещения

### Заключение

1. Несмотря на то, что спектр излучения белого светодиода отличаются от спектров стандартных источников А, В, С, Д, по своим цветовым координатам он близок к источнику типа В, С, Д.

2. Отсутствие в его спектре ультрафиолетовой и инфракрасной составляющих, делает его применение весьма перспективным, в системах медико-биологического назначения, например, в телеканалах микроскопах.

3. Использование как одиночных светодиодов, так и линеек или матриц светодиодов позволяет решать по-новому многие задачи оплотехники, например, в микроскопии – задачи апертурного синтеза, задачи анализа движущихся изображений и пр.

## Литература

1. Волков В., Коган Л. Телевизионные системы с использованием светодиодных осветителей. // Электронные компоненты. 2002. №2.
2. Рувинова Э. Автоматизированный оптический контроль печатных плат. // Электроника НТБ, №6, 2002.
3. Тезисы докладов V Международной светотехнической конференции. Свет и прогресс. СПб. 2003.
4. Справочная книга по светотехнике. М.: Энергоатомиздат, 1983.
5. Демкина Л.В. и др. Спектральные и цветовые характеристики некоторых подстилающих поверхностей и источников света (Справочные материалы). // ОМП. 1988. №10.
6. Чуриловский В.Н. Теория оптических приборов. М.-Л.: Машиностроение, 1966.
7. Слюсарев Г.Г. Расчеты оптических систем. Л.: Машиностроение: 1979.
8. Дубовиков А.Л. // Тезисы докладов международной конференции "Оптика 2002".

# ИССЛЕДОВАНИЕ СТРУКТУРЫ ТОНКИХ ДИЭЛЕКТРИЧЕСКИХ СЛОЕВ ОПТИЧЕСКИМИ МЕТОДАМИ

Н.Н. Карасев, А.О. Шакин

Научный руководитель – д.т.н., проф. Э.С. Путилин

Работа посвящена изучению возможности использования метода нарушенного полного внутреннего отражения (НПВО) для анализа структуры тонких диэлектрических слоев, полученных методами физического осаждения. Дополнительно структура образцов исследовалась на растровом электронном микроскопе (РЭМ). Анализ картины рассеяния и определение вида индикатрисы рассеяния позволит проводить неразрушающий контроль структуры полученных покрытий.

## Введение

При расчете оптических и физико-механических характеристик покрытий необходимо знать, как связан со структурой покрытия показатель преломления и главный показатель поглощения. Поэтому необходимо изучение структуры тонких пленок из различных материалов, полученных при изменяющихся технологических параметрах (скорость осаждения, температура и качество очистки поверхности подложки) для разных методов осаждения.

Микроструктура тонких пленок обычно исследуется методами: туннельной электронной микроскопии, ОЖЕ-спектроскопии, дифракции отраженных быстрых электронов, рентгеноструктурного анализа, и т.д. [1]. К недостаткам этих методов следует отнести дорогое и сложное оборудование, необходимость изготовления специальных образцов, сложные методики определения микроструктуры, рассматривается лишь небольшой участок пленки [1].

Из оптических методов наиболее распространены эллипсометрия и спектрофотометрия. Эти методы используются для определения спектральных характеристик и определения оптических постоянных пленок. Они могут дать лишь косвенные данные о структуре покрытия. Использование эффекта нарушенного полного внутреннего отражения основано на проникновении световой волны из оптически более плотной в менее плотную среду на глубину порядка длины волны. Нарушение полного внутреннего отражения при исследовании тонких слоев связано с тем, что коэффициент отражения от границы раздела двух сред становится меньше единицы вследствие наличия поглощающих неоднородностей и напряжений в слое. Достоинствами метода НПВО являются применимость к сильно поглощающим средам. Амплитуда поля, проникшего в другую оптическую среду, затухает по экспоненциальному закону в зависимости от глубины проникновения. В работе [2] авторами был предложен контроль наличия кристаллической и аморфной фаз в прозрачных пленках методом комбинационного рассеяния света (КРС). Исследованы пленки  $TiO_2$  и смесь  $TiO_2+SiO_2$ , толщиной 83 и 20 нм соответственно, нанесенные на подложки из пирекса методом осаждения из коллоидного раствора с последующим отжигом в атмосфере кислорода при температуре 550 °С и 650 °С. Коэффициент ослабления лазерного излучения для обеих пленок составил не менее 1 Дб/см. Диаметр кристаллитов (порядка 4,5 нм), рассчитанный по рамановскому смещению частоты КРС ( $\omega=33\text{см}^{-1}$ ), совпал с результатами последующего исследования отделенной от подложки пленки методом просвечивающей электронной микроскопии.

## Экспериментальные исследования

Нами для исследований использована установка, схема которой показана на рис. 1. По рассмотренному выше методу при контроле измерение параметров производится на достаточно большой площади поверхности образца. В предложенной схеме

контроля нарушенного полного внутреннего отражения рассеянный свет регистрируется с помощью поляризационного микроскопа, поэтому наблюдаемая площадь имеет малую величину, что позволяет контролировать образцы разных размеров равномерно по всей поверхности. В качестве источника излучения использован He-Ne лазер (длина волны  $\lambda = 632,8$  нм). Диаметр зондирующего пучка  $\sim 0,5$  мм.

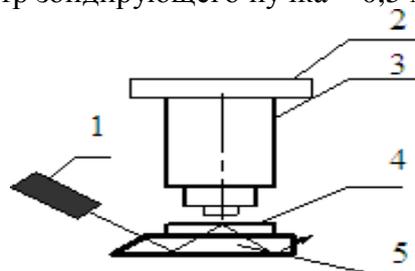


Рис 1. Схема установки: 1– источник излучения, 2 – ПЗС-камера, 3 – поляризационный микроскоп, 4 – исследуемый образец, 5 – пластина для ввода излучения

Покрытия для исследований нанесены на вакуумной установке ВУ-1А, при давлении в камере  $\sim 10^{-3}$  Па. На установке были изготовлены образцы двух видов: **1** – П В 2Н В и **2** – П 2Н В (В – четвертьволновый слой  $ZrO_2$  с показателем преломления  $n=1,92$ , Н – четвертьволновый слой  $NaCl$  с показателем преломления  $n=1,53$  на  $\lambda_0=550$  нм, стекло К8). Оптические толщины слоев одинаковы, отличны лишь скорость конденсации и угол поступления паров. В экспериментах №1 и №2 осаждение проведено при температуре  $18^\circ C$ , в эксперименте №3 –  $300^\circ C$ . Угол поступления паров  $ZrO_2$  на приемную поверхность отрицательный, если его отсчитывать от нормали к поверхности к направлению распространения потока паров, а для  $NaCl$  – положительный (положительное направление отсчета – отсчет угла от нормали к направлению распространения потока паров ведется против часовой стрелки). Диапазон изменения угла конденсации, распределение толщины пленки и скорости осаждения в зависимости от угла поступления паров сведены в табл. 1. При этом следует отметить, что максимальной скорости осаждения и толщине пленки  $ZrO_2$  соответствует минимальное значение толщины и скорости осаждения  $NaCl$ .

№	Температура поверхности подложки ( $^\circ C$ )	Диапазон изменения угла поступления паров $\alpha_1$ - $\alpha_2$ (град)		Изменение толщины пленки, min-max (нм)		Скорость роста пленки, min-max (нм/мин)	
		$ZrO_2$	$NaCl$	$ZrO_2$	$NaCl$	$ZrO_2$	$NaCl$
1	18	(0)-(-33.5)	0 - 39	63-85	128-205	21-28	45-73
2	18	(0)-(-39.6)	0-46.7	54-90	116-250	16-26	30-65
3	300	(0)-(-39.6)	0-46.7	54-90	116-250	16-26	30-65

Таблица 1 Изменение толщины и скорости роста покрытия от угла поступления паров на приемную поверхность

Для определения влияния атмосферы на свойства покрытия у полученных образцов исследованы спектральные характеристики на спектрофотометре СФ-26 в диапазоне 400–1100 нм после нанесения покрытия и через неделю. Измерения проведены относительно чистой подложки из того же материала, из которого изготовлена подложка, на которую наносилось покрытие, поэтому из полученных значений пропускания надо вычитать потери на каждой поверхности подложки.

Изменений спектральных характеристик (рис. 2) у покрытия конструкции 1 замечено не было, у покрытий конструкции 2 было замечено смещение спектральных характеристик в ИК область, возможно, это связано с десорбцией воды с поверхности подложки в пленку  $NaCl$ . Для уточнения данных, полученных методом НПВО, и по-

следующей разработки математической модели рассеяния света в покрытии свободные пленки  $ZrO_2$  исследовались методом просвечивающей электронной микроскопии (РЭМ).

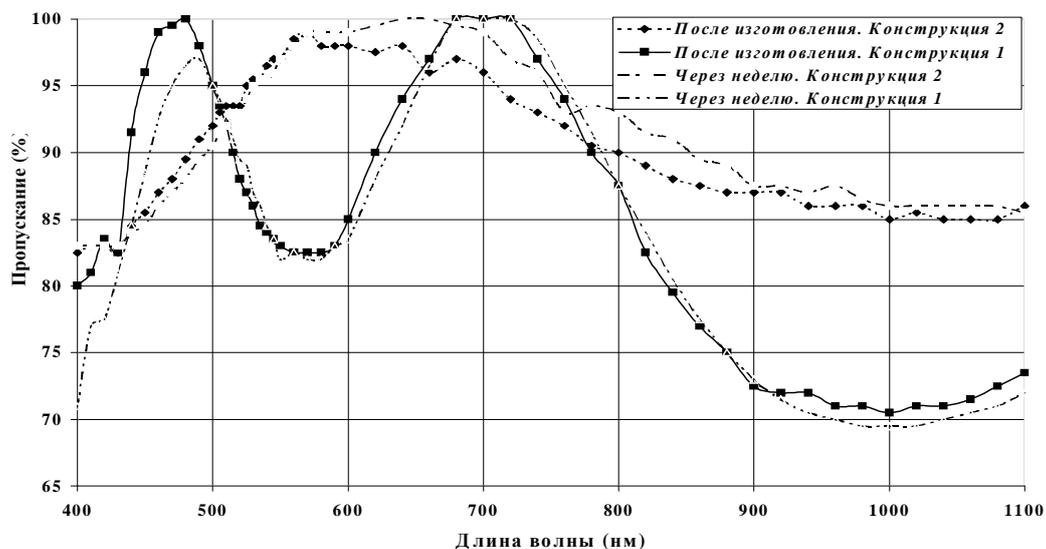


Рис. 2. Спектральные характеристики конструкции 1 и 2

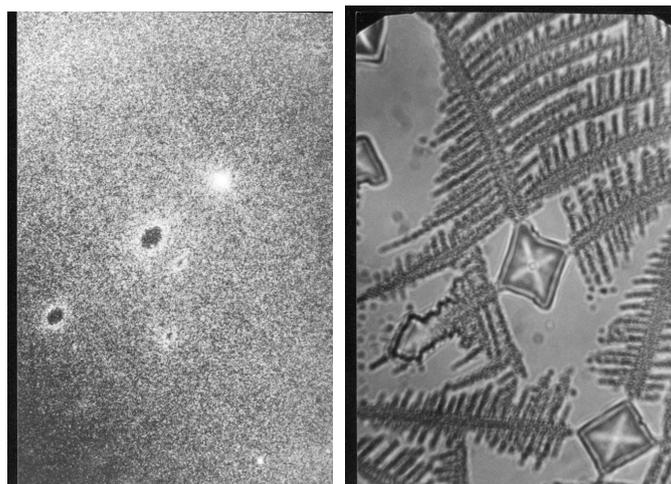


Рис. 3. Фотографии образцов: а – полученная методом НПВО, б – полученная в проходящем свете (ширина фотографии ~ 40 мкм)

На фотографиях (рис. 3), полученных методом НПВО, наблюдается область с равномерным полем рассеяния, а также несколько темных и светлых областей, локализованных в разных плоскостях по толщине пленки (при фокусировке объектива микроскопа границы одних областей становятся более резкими, у других менее). Наличие темных областей свидетельствует о том, что рассеяния в этих зонах практически не происходит, также это может свидетельствовать об отсутствии покрытия. На фотографии, полученной в проходящем свете, ясно виден рисунок травления  $NaCl$ , видны несколько кристаллитов, размер которых  $\sim 7000 \times 7000$  нм. Кристаллиты по-разному повернуты, но у всех форма косоугольного октаэдра, направленного одной из вершин по нормали к плоскости фотографии. Структуры пленок  $NaCl$  и  $ZrO_2$  взаимосвязаны друг с другом.

Анализ свободных пленок  $ZrO_2$ , полученных растворением подслоя  $NaCl$  в воде, проведен на растровом электронном микроскопе марки ЭМ-125 (рис. 4).

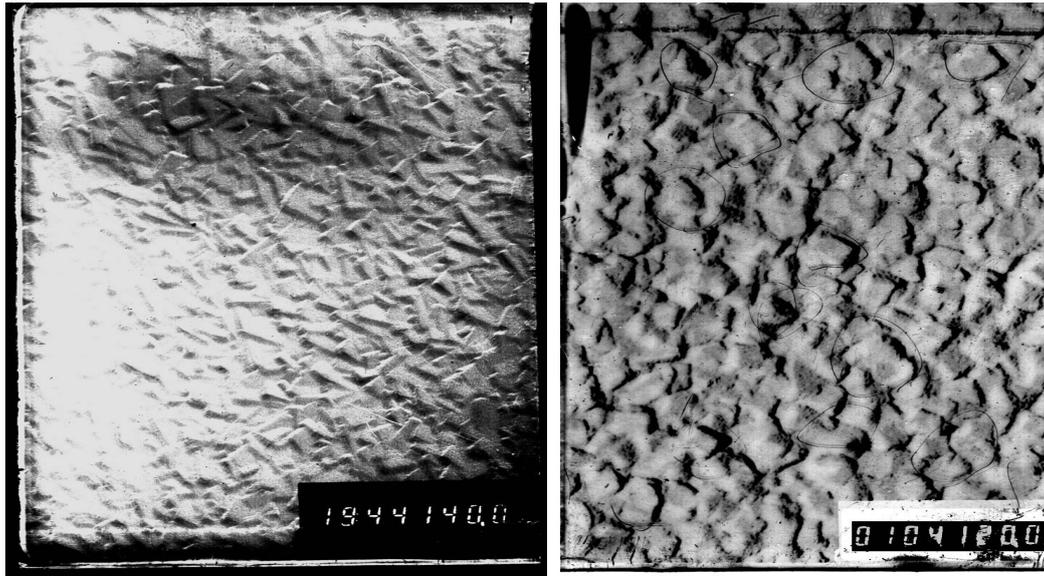


Рис. 4. Микроструктурный анализ: Изображение поверхности свободной пленки  $ZrO_2$ ; слева температура подложки  $18^\circ C$ , справа  $300^\circ C$ .

Длина фокусирующей системы  $L=500$  мм, ускоряющее напряжение 100 кВ, увеличение  $40000^{\times}$ , длина волны  $\lambda=37 \cdot 10^{-4}$  нм. Разрешение составило 5–6 нм, исследуемая площадь имеет размер  $1 \times 1$  мкм (для колец микродифракции) и  $2 \times 3$  мкм для изображения поверхности. На фотографиях (рис 4.) видна поверхность пленки  $ZrO_2$ , которая граничила с пленкой NaCl. По данным анализа картины дифракции, которые здесь не показаны, можно заключить, что пленки получились в экспериментах №1 и №2 (табл. 1) близкими к аморфным, об этом свидетельствуют размытые границы дифракционных колец, размеры кристаллитов пленки  $ZrO_2$  не превышают 5–10 нм. Из картины дифракции эксперимента №3 (табл. 1) можно заключить, что пленки также имеют структуру, близкую к аморфной, но размер кристаллитов пленки  $ZrO_2$  увеличились до 20–30 нм, об этом свидетельствуют кольца дифракции, ставшие узкими, с более яркими границами. Поверхность пленки  $ZrO_2$ , граничащая с NaCl, в экспериментах №1 и №2 имеет вид ступенек различной формы и размера (параллелепипед, куб, косой параллелепипед), развернутых преимущественно в одном направлении. Размер кристаллитов- $100 \times 80$  нм (доля 85% от общего количества). Доля кристаллитов среднего размера  $50 \times 37,5$  нм составляет 15%.

Поверхность в эксперименте №3 имеет вид ступенек, повернутых в одном направлении, форма ступенек близка к кубу, ступеньки почти не различаются по высоте. Преобладают кристаллиты среднего размера  $160 \times 120$  (95%), кристаллитов размера  $75 \times 50$  - 5%. Из проведенных экспериментов можно заключить, что увеличение температуры подложки приводит к увеличению размера кристаллитов. Средний размер кристаллитов увеличился примерно в 1,7 раза. По размеру дифракционных колец на представленной фотографии эксперимента №3 были получены данные о микроструктуре (табл. 2). Из полученных данных можно заключить, что пленка получилась моноклинной сингонии.

В дальнейшем исследовании структуры слоев  $ZrO_2$  методом НПВО получены образцы другой конструкции – П 6В – при  $\lambda_0=632,8$  нм, где В - четвертьволновый слой  $ZrO_2$  с показателем преломления  $n=1,92$ . Подложка - стекло К8. Образцы исследовались по несколько другой методике. Пленки  $ZrO_2$  сравнивались с пленками аналогичной конструкции из оптической смеси, имеющей другой показатель преломления и иную структуру, с целью сравнения характера кривых рассеяния данных покрытий. Толщина

слоя с высоким показателем преломления подбиралась, исходя из длины волны источника излучения, используемого в измерительной схеме. Изготовленная конструкция близка к симметричному планарному волноводу. Для реализации волноводного распространения излучения в слое толщина этого слоя не должна быть меньше половины длины волны распространяющегося в нем излучения, в качестве источника излучения использовался He-Ne лазер с  $\lambda=632,8$  нм.

Ориентации плоскостей(hkl)	Расстояние между плоскостями $d_{hkl}$ (нм)
100	0,51
011	0,369
111	0,319
111	0,285
002	0,213
200	0,255

Таблица 2. Структура полученного покрытия  $ZrO_2 \cdot 3H_2O$

Угол падения излучения на гипотенузную грань призмы больше критического, поэтому практически все излучение отражалось от этой грани и выходило из призмы, рассеянное излучение, прошедшее через гипотенузную грань призмы, регистрировалось с помощью ФЭУ, для исследований использовался приемник, имеющий максимальную чувствительность в данном диапазоне длин волн. Данный метод также подходит для регистрации малых величин поглощения в пленках. Исследование рассеяния проводилось в диапазоне углов  $0-90^\circ$ , за нулевое положение бралась нормаль к поверхности подложки. Сначала рассеяние определялось от чистой подложки, зафиксированной с помощью иммерсии ( $n=1,5154$ ) на поверхности гипотенузной грани призмы AP-90 (стекло К8), затем на эту подложку наносилось покрытие. Определение сигнала от чистой подложки и подложки с покрытием позволяет сравнить эти сигналы по интенсивностям и по изменению характера кривой рассеяния (индикатрисы).

На рис. 3, 4 представлены индикатрисы рассеяния, полученные от подложки и подложки с покрытием. Рассеяние определялось в нескольких точках подложки. Чтобы понять характер кривой рассеяния, измерение в одной точке проводилось многократно. На индикатрисах рассеяния, полученных от поверхности подложки с покрытием, максимум рассеяния не сдвигается, но характер кривой меняется, рассеяние регистрировалось в различных зонах. Рассеянное поверхностью покрытия излучение в разных точках примерно совпадает с интенсивностью рассеянного излучения поверхностью чистой подложки. Интенсивность в разных точках на пленке и подложке отличалась примерно в 10 раз и составила 0,001–0,09 % относительно падающего потока. Величина рассеянного сигнала сильно зависит от качества подготовки поверхности и от технологических параметров нанесения покрытия. Визуально наблюдается увеличение пятна рассеяния в некоторых направлениях, что связано с различной ориентацией поликристаллов, однако зарегистрировать его с помощью ПЗС-камеры с автоматической регулировкой усиления (АРУ) оказалось невозможным из-за больших перепадов освещенности объекта наблюдения. Как видно из полученных индикатрис рассеяния, определенных в одной точке, максимум рассеяния поверхностью чистой подложки практически не смещается и соответствует значению примерно  $70^\circ$  от нормали к поверхности грани призмы.

Характер кривой рассеяния изменяется, на ней появляются периодические перепады интенсивности, что может свидетельствовать о волноводном распространении излучения в слое. Характер кривых рассеяния на пленке и на чистой подложке в целом сохраняется, однако в отдельных точках наблюдается изменение формы кривых и рез-

кое изменение интенсивности в максимуме, что может свидетельствовать о дифракции света в царапинах на поверхности подложки.

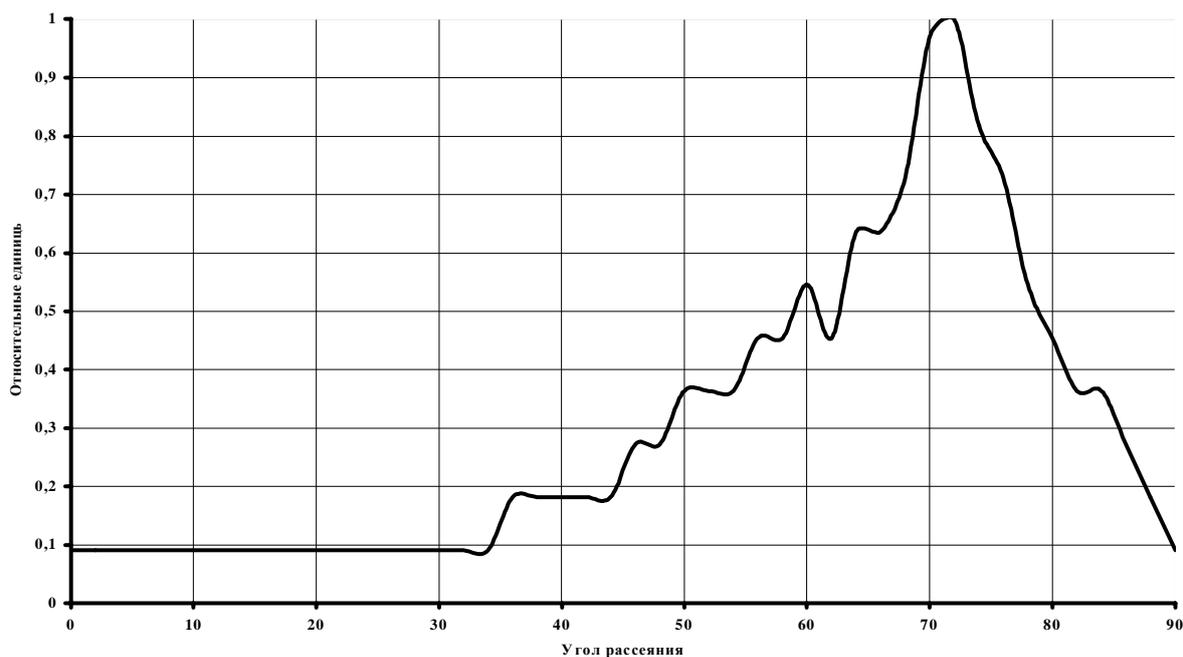


Рис. 3. Индикатриса рассеяния, зарегистрированная от поверхности чистой подложки

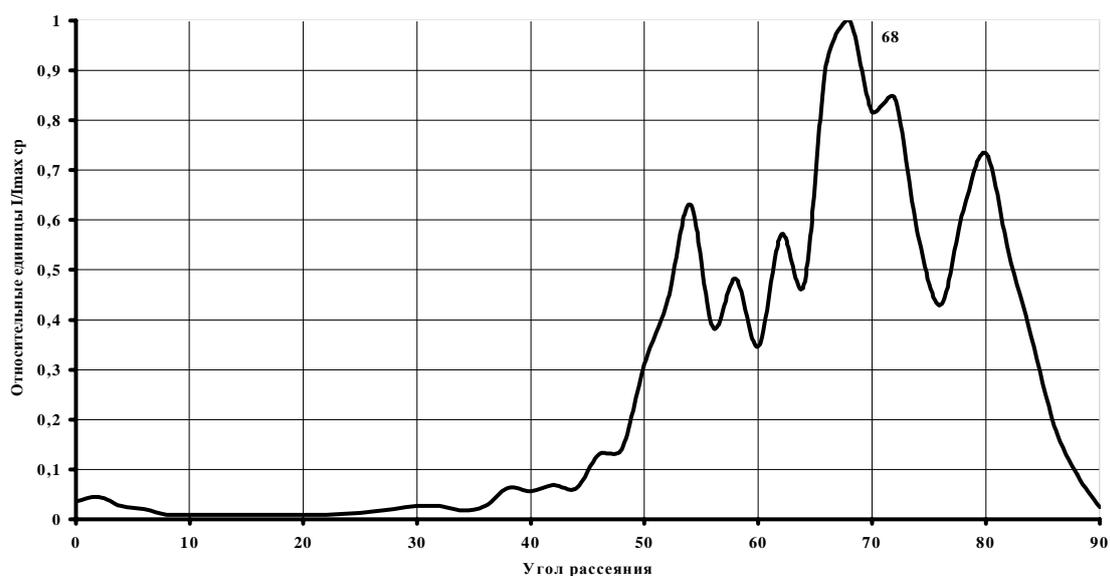


Рис. 4. Индикатриса рассеяния, зарегистрированная от поверхности покрытия

В результате работы установлено увеличение размера кристаллитов  $ZrO_2$  с ростом температуры подложки во время осаждения слоя. Показано, что изменение характера кривой рассеяния связано с изменением размера неоднородностей в данной точке. Метод позволяет проводить неразрушающий контроль качества полученных покрытий.

## Литература

1. Современные методы исследования поверхности. / Т. Вудраф., Т. Делчар. Под ред. В.И. Раховского. М.: Мир, 1989.
2. Ионно-плазменные технологии в электронном производстве. / В.Т. Барченко, Ю.А. Быстров, Е.А. Колгин. СПб: Энергоатомиздат, 2001.

## БАЛАНСИРОВКА НАГРУЗКИ В РАСПРЕДЕЛЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ

С.К. Дорожкин

Научный руководитель – к.т.н. , доц. А.А. Приблуда

Целью данной работы является определение эффективности использования специализированной системы балансировки нагрузки в распределенной вычислительной системе. Исследуется влияние системы балансировки нагрузки на использование ресурсов и производительность распределенной вычислительной системы. В качестве области исследований выбраны распределенные вычислительные системы, используемые в качестве Web-служб. Результаты экспериментов показали, что использование системы распределения нагрузки предлагаемого типа является эффективным решением.

### Введение

Построение распределенных вычислительных систем требует значительных капиталовложений, поэтому задача максимального использования возможностей такой системы, т.е. обеспечения ее максимальной производительности, является исключительно актуальной.

На сегодняшний день выделяют [1,2] три основных подхода к увеличению производительности распределенной вычислительной системы. Первый – это увеличение количества вычислительных узлов системы, второй – оптимизация алгоритмов, реализованных в программном обеспечении, используемом в распределенной системе, и третий – использование системы балансировки нагрузки, обеспечивающей равномерную загрузку вычислительных узлов распределенной системы. Все эти подходы имеют свои сильные и слабые стороны.

К недостаткам первого можно отнести то, что при обычном добавлении вычислительных узлов не удастся получить равномерного распределения нагрузки между всеми узлами распределенной системы, так как встроенные средства балансировки используют самые простые алгоритмы распределения пользовательских запросов между вычислительными узлами системы. Недостатком второго метода является то, что программные продукты, которые используются распределенной системой, имеют либо самые простейшие возможности по настройке их функционирования, либо вообще не могут быть доработаны пользователями в процессе эксплуатации [2].

### Существующие подходы к балансировке нагрузки

Использование специальной системы балансировки нагрузки позволит оптимально использовать вычислительные мощности распределенной системы и не потребует внесения изменений в программное обеспечение, которое используется на распределенной вычислительной системе. Недостатком этого метода являются задержки, которые вносит сама система балансировки, так как она становится дополнительным компонентом, который потребляет ресурсы вычислительной системы, что сказывается на производительности всей вычислительной системы в целом.

Однако использование системы балансировки нагрузки может оказаться вполне результативным, если удастся минимизировать потребление ресурсов вычислительной системы самой системой балансировки.

При разработке подходов к построению системы балансировки нагрузки самое существенное значение имеет выбор характеристик производительности вычислительной системы в целом и отдельных ее узлов. На настоящий момент существует множе-

ство тестов, которые дают оценку производительности для систем различного класса [3]. Для распределенных вычислительных систем в качестве единицы производительности имеет смысл говорить о количестве задач, выполненных за единицу времени. Это связано с тем, что распределенные системы используются для решения больших комплексных задач, и тесты, которые показывают количество простейших операций за единицу времени в одном узле или во всей их совокупности, будут некорректно характеризовать производительность всей системы в целом. За характеристику загрузки узла вычислительной системы можно принимать различные показатели [3]: это может быть загрузка процессора или памяти, а также количество операций ввода-вывода за единицу времени. На наш взгляд, наиболее эффективно загрузку вычислительного узла отражает количество процессов, находящихся в очереди на исполнение. Выбор этой характеристики связан с тем, что от количества уже ожидающих процессов напрямую зависит время ожидания обработки новых процессов.

Для разработки алгоритма балансировки нагрузки необходимо иметь информацию о порядке взаимодействия компонентов распределенной вычислительной системы в процессе ее функционирования. Общепринятым подходом для этого является моделирование. Результаты, полученные при моделировании, будут являться исходной информацией для принятия решения о разработке алгоритма балансировки нагрузки.

Средства, которыми может быть реализована система балансировки нагрузки, могут быть как аппаратными, так и программными. В случае использования программных средств один узел вычислительной системы используется для работы модуля распределения нагрузки между остальными. В случае аппаратного решения к существующей системе подключается дополнительное устройство, которое занимается сбором, обработкой информации о состоянии узлов и распределением нагрузки между узлами вычислительной системы.

Учитывая разнообразие технических средств, используемых для построения распределенных вычислительных систем, наиболее целесообразным представляется использование программных средств для реализации алгоритма балансировки нагрузки. На основе проведенных исследований был создан прототип системы балансировки нагрузки.

Под термином "система балансировки нагрузки" мы будем понимать инструментальное средство, предназначенное для распределения запросов пользователя на основе принятых критериев между узлами распределенной вычислительной системы. Пользователь при этом и не подозревает, что его запросы проходят через дополнительное звено вычислительной системы.

### **Экспериментальное исследование предлагаемого подхода к балансировке нагрузки**

Для исследований была выбрана такая область применения распределенных вычислительных систем, как Web-службы. Web-службы являются на сегодняшний момент одними из самых распространенных в Интернете и, как правило, реализованы на распределенных вычислительных системах.

Под термином "распределенная вычислительная система" в данной работе понимается кластерное решение по объединению машин в группу. Группа вычислительных узлов имеет доступ к общему дисковому массиву с данными и для конечного пользователя представляется одной виртуальным узлом. Доступ к виртуальному вычислительному узлу осуществляется через систему балансировки нагрузки.

В ходе исследований были рассмотрены различные алгоритмы распределения запросов удаленных клиентов между серверами, которые образуют единый виртуальный Web-узел [2]. После этого был выделен наиболее оптимальный с точки зрения применения к данной проблеме алгоритм, в котором решение о выборе сервера для перенаправления на него запроса принимается на основе данных, полученных с помощью

внутреннего мониторинга сервера. Критерием для принятия решения является среднее количество процессов, находящихся в очереди на исполнение. Кроме того, учитывается вычислительная мощность каждого сервера с помощью ввода так называемых весовых коэффициентов.

В качестве базы для экспериментальных исследований была выбрана платформа Sun Solaris как одна из наиболее часто используемых для построения высокопроизводительных систем. Экспериментальная распределенная вычислительная система представляла собой Web-узел, располагающийся на двух физических серверах. В качестве программного Web-сервера был выбран широко распространенный сейчас сервер "Apache".

Программная реализация состояла из двух основных частей – модуля балансировки и программы-агента, которая располагалась на каждом узле распределенной вычислительной системы. Задача программы-агента заключалась в том, чтобы собирать сведения о состоянии подконтрольного узла и передавать эти данные через определенный интервал времени модулю балансировки. Модуль балансировки принимает решение о перенаправлении запроса пользователя на конкретный узел на основе сведений, полученных от программ-агентов, а также таблицы с весовыми коэффициентами, отражающими потенциальную вычислительную мощность узла распределенной системы.

К системе, имитирующей работу Web-узла, подключался диспетчерский модуль системы балансировки нагрузки, который располагался на выделенном узле распределенной системы. Тестовая смесь задач имитировала работу удаленных пользователей с Web-узлом. Пользовательские запросы требовали не просто передачи Web-страницы от сервера клиенту, но и определенных действий на самом сервере. Таким образом, нагружался процессор узла и его подсистема ввода-вывода.

Задача, решаемая в процессе экспериментальной эксплуатации прототипа, состояла в том, чтобы оценить эффективность работы алгоритма балансировки нагрузки при увеличении количества узлов системы и определить, насколько существенны задержки, вносимые компонентами системы балансировки нагрузки. Эксперимент состоял в том, чтобы выяснить, как изменится количество обработанных запросов пользователя с увеличением количества вычислительных узлов, а также при увеличении количества соединений, устанавливаемых одновременно. Результаты эксперимента представлены на рис. 1.

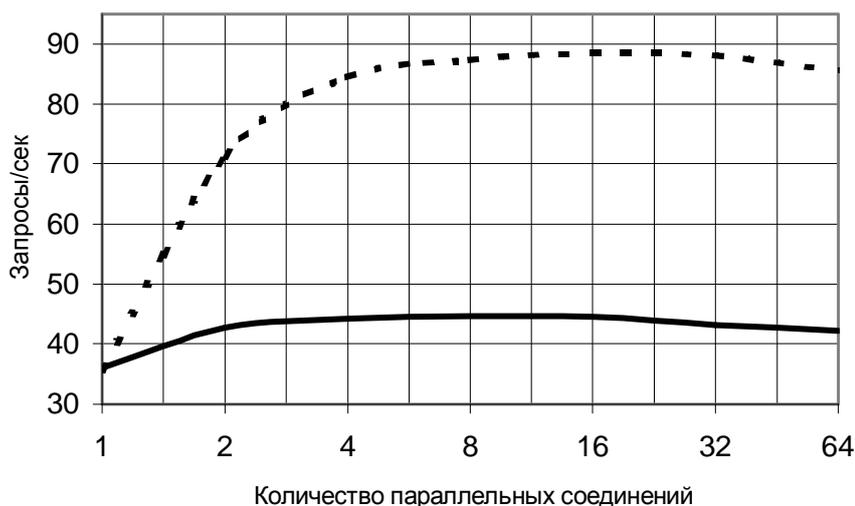


Рис. 1. Количество запросов, обработанных двумя узлами (пунктирная линия) и одним узлом (сплошная линия) в единицу времени, в зависимости от количества параллельно установленных соединений.

Результаты экспериментов показали, что использование системы балансировки нагрузки позволило практически линейно наращивать производительность вычислительной системы при увеличении количества узлов. Так, при увеличении количества вычислительных узлов в два раза количество выполненных запросов пользователей в единицу времени увеличилось в 1,98 раза. Полученные результаты позволяют сделать вывод, что алгоритм балансировки нагрузки, основывающийся на равномерном распределении запросов между узлами, является эффективным средством для балансировки нагрузки в распределенной вычислительной системе, а сама система балансировки нагрузки вносит минимальные задержки. Это позволяет сделать вывод о целесообразности использования систем балансировки нагрузки предлагаемого типа в реальных распределенных вычислительных системах.

Небольшое количество пользовательских запросов, обработанных в единицу времени, объясняется тем, что в эксперименте использовались рабочие станции, а не специализированные сервера. Главной задачей было показать потенциальные возможности подхода, а не достигнуть максимальной пропускной способности подобной распределенной системы.

Системы балансировки нагрузки могут также решать проблему доступности и надежности Web-узла, так как позволяют перенести функции по обслуживанию клиентов с одного сервера на несколько, равномерно распределяя нагрузку между ними, эффективно управляя их работой и оставаясь при этом прозрачными для конечных пользователей.

## **Заключение**

В работе изучался вопрос о целесообразности использования специальных программных средств балансировки нагрузки с целью оптимального использования имеющихся вычислительных ресурсов распределенной системы. Эксперименты показали, что использование программных решений является оправданным, задержки вносимые системой балансировки нагрузки минимальны. Число обработанных запросов с увеличением количества вычислительных запросов увеличивается практически линейно.

## **Литература**

1. Либман Л. Философия распределения нагрузки. // Журнал сетевых решений LAN. 2000. № 5.
2. Тао Чжоу. Системы балансировки нагрузки Web-серверов. // Windows 2000 Magazine. 2000. № 3.
3. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. – СПб: БХВ-Петербург, 2003.

# ПРОЕКТИРОВАНИЕ УПРАВЛЯЮЩИХ СИСТЕМ НА ОСНОВЕ ВЫСОКОУРОВНЕВЫХ ОПИСАНИЙ

И.В. Елькин

Научный руководитель – к.т.н., доц. А.Е. Платунов

В статье анализируются проблемы проектирования информационно-управляющих систем (ИУС) и современные тенденции в данной области. Рассматривается перспективная модель для высокоуровневого проектирования ИУС – модель абстрактных функциональных блоков, разрабатываемая в рамках стандарта проекта ИЕС 61499, и предлагается общий подход к проектированию ИУС на ее основе.

## Введение

Требования, предъявляемые к современным управляющим системам, постоянно возрастают, увеличивается общая сложность и масштаб таких систем. При этом условия рынка диктуют все меньшие сроки, отводимые на разработку, при меньшей стоимости. Таким образом, возникает проблема эффективного проектирования подобных управляющих систем. Подход, зачастую применяемый в настоящее время, достаточно неформален, особенно на начальной стадии проектирования, когда требования к функциональности системы выражаются на естественном языке. Он основывается на опыте и интуиции разработчика. Отсутствуют единая методология и инструменты для ведения проектирования.

Идея, лежащая в основе исследований по разработке новых подходов к проектированию, базируется на *абстракции*. Процесс проектирования следует начинать с этапа высокоуровневого анализа и проектирования (или *проектирования на системном уровне* – system level design). Данный подход основан на создании *высокоуровневых описаний* – моделей системы. *Архитектурная модель (описание) системы* высокого уровня позволяет абстрагироваться от деталей реализации системы на ранних этапах разработки и рассматривать систему как единое целое с точки зрения прикладной задачи. Такая архитектура послужит фундаментом, на основе которого в дальнейшем осуществляется низкоуровневое проектирование, реализация и совершенствование системы. Основу высокоуровневого проектирования составляют средства для создания моделей (описаний) системы – *языки описания систем*.

В статье мы проанализируем требования к языкам описания, рассмотрим перспективный язык – язык модели абстрактных функциональных блоков (ФБ) и предложим общий подход к проектированию ИУС на его основе.

## Требования к языкам описания

Высокоуровневый язык описания систем должен удовлетворять следующим требованиям:

- создание компактных и удобных для разработчика описания систем (как в графическом, так и текстовом виде) на различных уровнях,
- использование инструментальных средств для создания, редактирования, анализа, и имитационного моделирования (симуляции) описаний;
- возможность автоматизированного создания реализации системы (на языке реализации) на основе описаний высокого уровня;
- возможность повторного использования проектных решений;
- переносимость проектов между различными инструментальными средствами.

Кроме того, высокоуровневое описание также может использоваться для обмена информацией о системе между участниками проектирования в четкой и однозначной форме.

Использование языка описания систем позволяет анализировать и проверять альтернативные варианты проектных решений, что невозможно в случае непосредственного применения языков реализации системы (например, для программной составляющей – языков программирования), из-за временных затрат.

Современные управляющие системы строятся на основе микропроцессорной техники и представляют собой сложные программно-аппаратные комплексы. Соответственно, их разработка включает в себя проектирование как аппаратуры, так и программного обеспечения (ПО). Причем тенденция такова, что все большая часть усилий по разработке системы сосредоточена как раз в области ПО, в то время как проектирование аппаратуры сводится к выбору готовой платформы (например, промышленный компьютер или программируемый логический контроллер). Перенос большей части системы в область ПО позволяет получать гибкие решения за более короткий срок, за меньшие деньги, широко применять повторное использование.

Проектирование программной составляющей управляющей системы имеет много общего с ПО систем общего назначения. Однако есть ряд существенных особенностей ПО управляющих систем, что приводит к необходимости использования специальных подходов к его проектированию. Перечислим эти особенности:

1. жесткая детерминированность поведения;
2. работа в режиме реального времени;
3. повышенные требования к надежности и безопасности функционирования;
4. использование специализированных аппаратных средств;
5. как правило, распределенная архитектура системы.

Программное обеспечение управляющих систем отличается следующими характеристиками:

- параллельное исполнение,
- синхронизация,
- обработка исключительных ситуаций,
- учет временных задержек,
- использование явных состояний (автоматной модели).

Соответственно, высокоуровневые языки описания систем должны содержать средства поддержки этих характеристик. Одним из таких языков является язык модели абстрактных ФБ, который разрабатывается в рамках проекта стандарта ИЕС 61499. Рассмотрим его основные особенности подробнее.

### **Язык модели абстрактных ФБ**

Язык модели абстрактных функциональных блоков (ФБ) разрабатывается в рамках проекта стандарта ИЕС 61499 [1] и предоставляет средства для проектирования распределенных управляющих систем. Управляющее приложение (ПО), построенное в соответствии со стандартом, имеет распределенный характер, базирующийся на декомпозиции функции управления на несколько ФБ. Приложение представляет собой сеть функциональных блоков, соединенных связями по данным и событиям. При этом функциональные блоки могут быть разнесены по нескольким устройствам, объединенным линиями связи (рис. 1).

*Функциональный блок* представляет собой элемент программного обеспечения с четко определенными характеристиками – входными и выходными событиями и данными, а также поведением, задаваемым конечным автоматом функционального блока.

Каждый блок отвечает за выполнение какой-либо части управляющего алгоритма (например, управление клапаном или некоторой промышленной установкой). Функциональный блок инкапсулирует в себе прикладной алгоритм, предоставляя внешний интерфейс – совокупность входов и выходов. Алгоритмы, заложенные в функциональные блоки, теоретически могут быть описаны на любых структурных языках.

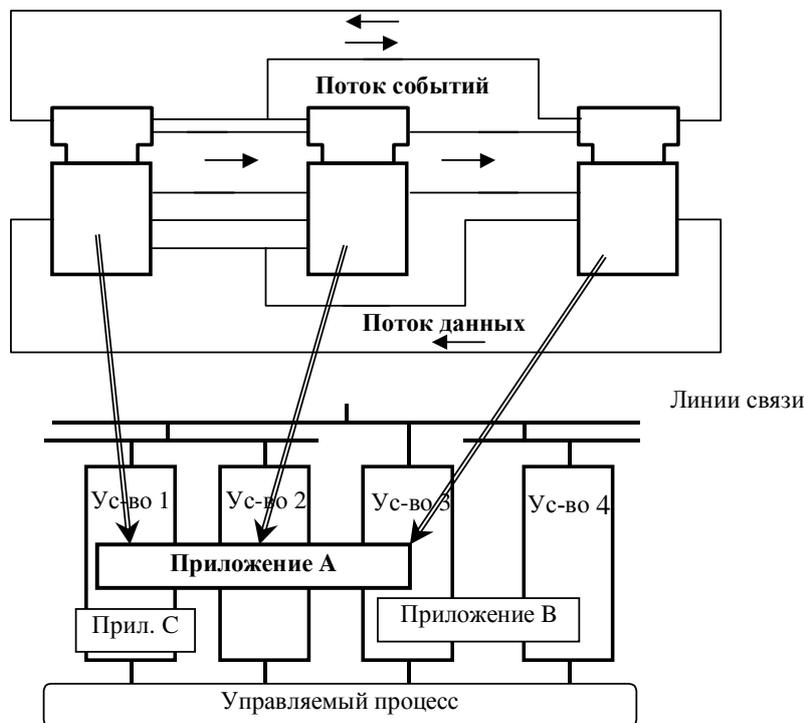


Рис. 1. Структура распределенного приложения на основе ФБ

Таким образом, здесь имеет место *объектно-ориентированный подход* к созданию приложений. Функциональные блоки являются объектами, причем существует возможность создания своих блоков и их последующего использования при проектировании, а также расширение функциональности существующих блоков. Система представляется как набор взаимодействующих объектов, каждый из которых отвечает за решение определенной задачи. Такой подход позволяет структурировать решение, уменьшить его сложность, повысить устойчивость системы при изменениях и обеспечить повторное использование компонентов системы.

Кроме того, модель абстрактных ФБ характеризуется следующими принципами.

1. *Иерархическое описание.* Модель позволяет описывать иерархическую структуру системы в виде набора подсистем и блоков, т.е. выполнять декомпозицию.
2. Поддержка построения *распределенных решений* и параллелизма работы.
3. *Функционирование, управляемое событиями.* Позволяет планировать исполнение алгоритмов управляющей системы без использования специальных механизмов синхронизации, что особенно важно в распределенных системах. Алгоритмы в отдельных ФБ запускаются на исполнение по получении событий. При этом функционирование является асинхронным, что помогает сократить задержки. Механизм событий также позволяет быстро реагировать на изменения, происходящие в системе.

4. *Использование модели конечного автомата (state-machine)*. Поведение ФБ задается автоматом с конечным числом состояний, который определяет условия перехода объекта из одного состояния в другое и набор действий, который при этом выполняется. Такая модель увеличивает предсказуемость поведения системы, облегчает отладку и позволяет поддерживать высокоскоростные, управляемые событиями алгоритмы, использующиеся в управляющих системах.

При использовании описанного подхода к проектированию важным является стандартизация формальных описаний функциональных блоков, что облегчило бы их повторное использование и распространение. С этой целью стандарт IEC 61499 определяет возможность описания функциональных блоков с использованием языка XML (Extended Markup Language) [2].

### **Проектирование на основе модели абстрактных ФБ**

В общем, проектирование системы на основе рассмотренной модели ФБ состоит в логическом описании прикладного алгоритма, заложенного в требованиях, как совокупности функциональных блоков, связанных между собой, независимо от конкретных аппаратных средств. На конечной стадии проектирования осуществляется отображение отдельных блоков приложения на вычислительные узлы управляющей сети – построение распределенного решения. К сожалению, стандарт IEC 61499 лишь описывает средство – язык абстрактных ФБ, но не затрагивает процесса проектирования на его основе.

Рассмотрим основные принципы предлагаемого подхода к проектированию, разработанного в результате анализа возможностей модели абстрактных ФБ.

В основу подхода положен *итеративный процесс разработки*, широко применяемый в настоящее время. Функциональные возможности системы в данном случае наращиваются постепенно, в течение нескольких итераций. Каждая итерация состоит из следующих этапов.

1. Сначала осуществляется *оценивание* текущего варианта проекта системы на соответствие установленным требованиям с целью *формирования требований* по улучшению системы на данной итерации (на первой итерации проектирования выполняется формирование исходных требований к системе). Если заданная степень соответствия достигнута, то процесс разработки прекращается (начинается эксплуатация системы).
2. Затем система проектируется в соответствии со сформулированными требованиями. Работа на данном этапе разделяется по процессам: *Основные процессы: Анализ и проектирование, Реализация, Тестирование, Развертывание.*
3. Третий этап заключается в *опытной эксплуатации* спроектированной на данной итерации системы для последующего оценивания.

Основные принципы, составляющие основу предлагаемого подхода:

1. использование модели прецедентов (use cases) для формулирования требований. [3];
2. использование объектно-ориентированного анализа с применением UML [4];
3. использование паттернов (шаблонов) проектирования [5];
4. визуальный подход к проектированию и тестированию (симуляция).

Все эти методы широко применяются при разработке программных систем. Они могут быть адаптированы для разработки ИУС на основе ФБ.

*Прецеденты* (use case) позволяют специфицировать поведение системы или ее части и представляют собой описания множества последовательностей действий системы и ее взаимодействия с внешней средой - *сценариев*. Прецеденты системы, как

правило, образуют структурированную систему – модель прецедентов, в которой они связаны между собой.

Вследствие схожести модели ФБ и объектно-ориентированной модели *объектно-ориентированное моделирование* с использованием UML может применяться на этапе анализа системы. В процессе анализа на основе моделирования *поведения* системы и *структуры*, реализующей это поведение, необходимо создать *набор объектов* – компонентов системы, которые реализуют системные требования.

Для *распределения обязанностей* между компонентами системы предлагается использовать *шаблоны (паттерны) проектирования* – формализованные решения типовых задач проектирования. Паттерны предлагают *каркас (framework)* – скелетную структуру, определяющую архитектуру системы, на основе которой наращивается требуемая функциональность приложения.

Сформулированный подход к проектированию основан на идее *визуального проектирования*. Разработка заключается в последовательном создании и уточнении визуальных моделей системы – сначала статических моделей структуры и поведения (UML), а затем и динамических, отражающих функционирование управляемого процесса (симуляция), а также соответствующих алгоритмов управления.

На основе перечисленных принципов была предложена организация процесса разработки РИУС в виде последовательности шагов проектирования. Данные шаги, а также *эволюция моделей* отражены на рис. 2.

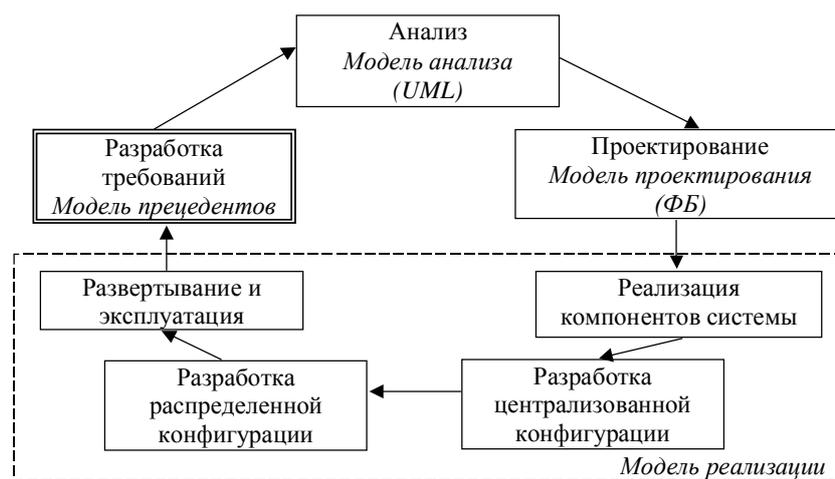


Рис. 2. Процесс проектирования на основе модели ФБ

### Формирование требований

**Шаг 1.** На данном шаге осуществляется формулирование требований к разрабатываемой системе управления в виде *модели прецедентов*.

### Анализ и проектирование

**Шаг 2.** На основе сформулированных требований в виде прецедентов необходимо создать набор компонентов системы, которые реализуют прецеденты. Эта задача решается на основе моделирования *поведения* и *структуры*, обеспечивающей это поведение, результатом которых является создание высокоуровневой *модели анализа*, описывающей архитектуру системы. Моделирование осуществляется с применением языка UML.

Для построения высокоуровневой архитектуры системы предлагается использовать многоуровневый каркас на основе паттернов. При этом система разделяется на

объекты интерфейса оператора, объекты оконечных устройств и объекты, реализующие алгоритмы управления в соответствии с решаемой задачей.

**Шаг 3.** На данном шаге осуществляется переход от абстрактных объектных моделей системы на языке UML к модели абстрактных ФБ, Такое преобразование возможно, поскольку между ними существует практически однозначное соответствие. Полученные модели объединяются в общую модель *распределенного приложения* - совокупность взаимосвязанных ФБ. Такая детальная *модель проектирования* подлежит дальнейшему уточнению уже в терминах ФБ с целью оптимизации проекта при обязательном удовлетворении всех требований к системе.

#### **Реализация**

**Шаг 4.** На данном шаге реализуются ФБ, которые входят в состав системы, спроектированной на предыдущих шагах.

**Шаг 5.** На основе модели *распределенного приложения* в совокупности с реализованными на предыдущем шаге ФБ создается прототип системы в *централизованной конфигурации* и тестируется в симуляторе. Централизованное приложение реализует только функциональные требования к управляющей системе, не принимая во внимание вопросы *топологии* системы.

**Шаг 6.** На следующем шаге осуществляется переход от централизованной конфигурации приложения к распределенной, адекватно отражающей *топологию* физической системы. Задача заключается в распределении ФБ приложения по устройствам распределенной системы. Связь между распределенными блоками осуществляется с помощью вспомогательных коммуникационных ФБ. Осуществляется тестирование в симуляторе.

**Шаг 7.** На последнем шаге распределенное приложение доставляется в реальные устройства, осуществляется конфигурирование устройств системы через инструментальные средства и проводится тестовая эксплуатация системы.

Данный переход от централизованной конфигурации к реальной системе обеспечивается моделью абстрактных ФБ. На всех этапах разработка сопровождается тестированием функциональности системы с использованием инструментальных средств (симуляция).

Сформулируем достоинства предложенного подхода.

1. Использование модели прецедентов и объектно-ориентированного моделирования удачно дополняет принципы модели ФБ и позволяет представить процесс разработки как процесс последовательного уточнения модели системы – модель прецедентов, модель анализа, модель реализации (централизованная, распределенная), реальная система.
2. Использование прецедентов и объектно-ориентированного моделирования (с применением UML) является общей практикой в настоящее время, следовательно, позволяет задействовать имеющиеся навыки специалистов и существующие инструментальные средства
3. Паттерны предоставляют скелетную структуру (каркас), определяющую базовую архитектуру будущей системы, на основе которой наращивается требуемая функциональность и предлагают проверенные решения типовых задач проектирования, а также позволяют осуществлять тестирование системы в симуляторе.
4. Кроме этого, рассмотренный подход демонстрирует возможности автоматизации отдельных задач проектирования: автоматизированный процесс перехода от модели анализа UML к модели проектирования на основе ФБ, проверка целостности и корректности описаний объектов и их соединений, автоматизированный переход от централизованной конфигурации системы к распределенной и др.

Работоспособность предложенного подхода обеспечивается использованием методов, широко применяемых в настоящее время при разработке программных систем и проверенных инженерной практикой.

### **Заключение**

В статье рассмотрен подход к проектированию информационно-управляющих систем на основе высокоуровневых описаний (моделей). Подробно проанализирован перспективный язык модели абстрактных ФБ (IEC 61499) и предложен подход к проектированию ИУС на его основе.

### **Литература**

1. IEC 61499-1 (PAS), Function Blocks for Industrial-Process Measurement and Control Systems, Part 1 – Architecture, International Electrotechnical Commission, Geneva, 2000.
2. W3C, eXtended Markup Language (XML) Specification. URL: <http://www.w3c.org/TR/1998/REC-xml-19980210>, 1998.
3. I. Jacobson, Object-Oriented Software Engineering: A use-case driven approach, Addison Wesley 1992.
4. OMG, OMG Unified Modeling Language Specification, Version 1.3, OMG, ad/99-06-08, June 1999.
5. B.P. Douglass, Real-Time UML, Addison Wesley Longman, 1998, ISBN 0-201-32579-9.

# КРИТЕРИИ ПОЛНОТЫ ТЕСТИРОВАНИЯ ПРОГРАММ

Ю.А. Яковлев

Научный руководитель – к.т.н., доц. Т.А. Павловская

В статье вводится понятие критерия тестирования, определяются и классифицируются существующие критерии. Вводятся параметры оценки трудоемкости автоматизированной проверки критериев.

## Введение

Тестирование является одним из основных методов достижения надежности программного обеспечения. Целью тестирования является обнаружение ошибок путем испытания программы на тестовых данных. Качество тестирования во многом определяется свойствами тестовых наборов. Желательно минимизировать размер тестовых наборов и, следовательно, трудоемкость, в то же время обеспечив наибольшее число найденных ошибок. Для решения этой задачи были разработаны критерии тестирования.

Тестирование может быть проведено с различной степенью полноты. Поскольку чаще всего полный перебор всех возможных значений входных данных нереален либо теоретически, либо ввиду ограниченных вычислительных ресурсов, то весьма актуальным является вопрос: какой объем тестирования следует выполнить, чтобы достигнуть достаточной степени надежности программы. Если программа протестирована недостаточно, то следует определить, какие тестовые наборы нужно добавить, чтобы повысить уровень ее тестируемости.

Таким образом, требуется:<sup>1</sup>

1. признак достаточности проведенного тестирования и возможности его прекращения;
2. мера качества проведенного тестирования, когда с каждым тестовым набором можно связать численную характеристику полноты тестирования;
3. правило выбора или создания тестовых наборов.

Для решения данной задачи используются критерии тестирования программ. С одной стороны, критерий определяется как предикат, который может для данного тестового набора и программы либо выполняться, либо не выполняться. Добавление новых тестовых данных и, следовательно, процесс тестирования могут быть остановлены, когда критерий становится истинным. С другой стороны, критерий обычно формулируется таким образом, что возможно численно определить процентную долю выполнения требований, предъявляемых критерием, при этом истинность критерия означает выполнение этих требований на 100 %.

Таким образом, критерий тестирования может быть строго определен<sup>2</sup> как функция  $C: P \times S \times T \rightarrow \{\text{true}, \text{false}\}$ , где  $P$  – множество программ,  $S$  – множество спецификаций,  $T$  – множество всех подмножеств  $D$ ,  $D$  – множество входных данных для программ из  $P$ . Выполнение критерия для программы  $p$  по отношению к спецификации  $s$  на тестовом наборе  $t$  означает  $C(p,s,t)=\text{true}$ .

## Критерии функционального тестирования

В данной статье основное внимание уделяется критериям структурного тестирования, поэтому будет рассмотрен лишь пример критерия функционального тестирования. Критерий основан на одной из наиболее часто встречающихся стратегий функционального тестирования – разбиения множества возможных входных данных на классы эквивалентностей – непересекающиеся подмножества. Критерий выполняется, если тесты выбраны из разных подмножеств.

---

<sup>1</sup> [3] стр. 336-427

<sup>2</sup> [5] стр. 1

## Критерии структурного тестирования

### Критерии потока управления

*Покрытие вершин (операторов) – C0:* критерий выполняется, если каждая вершина в управляющем графе была достигнута хотя бы один раз.

*Покрытие дуг (решений) – C1:* критерий выполняется, если каждая дуга в управляющем графе была пройдена хотя бы один раз.

*Покрытие всех путей:* критерий выполняется, если пройден каждый полный путь от начальной вершины к конечной. Применение критерия в данной формулировке связано со следующими проблемами.

- Программе, содержащей циклы, может соответствовать бесконечное множество полных путей. Поэтому применяются методы ограничения числа полных путей, требуемого критерием.
- Некоторые пути могут быть недостижимы, когда не существует такого набора тестовых данных, который приведет к прохождению по заданному пути. Поэтому применяются методы обнаружения недостижимых путей и исключения их из числа требуемых критерием.

*Покрытие всех путей длины n (Cn):* данный критерий позволяет решить проблему бесконечного числа всех полных путей, ограничивая длину требуемых путей.

Для определения нижеследующих критериев<sup>3</sup> используются понятия *решения* и *условия*. Решение представляет собой выбор одного пути перехода из нескольких возможных (например, в случае использования конструкции if-then-else). Решение может включать в себя одно или несколько условий – элементарных булевских предикатов - объединенных логическими операторами.

*Покрытие условий:* критерий выполняется, если каждое из условий в решениях приняло возможные значения (true или false) хотя бы один раз.

*Покрытие решений / условий:* критерий выполняется, если каждое из условий в решениях приняло возможные значения (true или false) хотя бы один раз, а также каждое решение приняло оба возможных исхода (true или false) хотя бы один раз, т.е. каждая дуга управляющего графа была покрыта.

*Комбинаторное покрытие условий:* критерий выполняется, если все были получены все возможные комбинации значений условий в каждом из решений.

*Модифицированный критерий покрытия решений/условий:* критерий выполняется, если каждое из условий в решениях приняло возможные значения (true или false) хотя бы один раз, а также было показано, что значение каждого условия независимо от других условий влияет на результат решения, т.е. изменение значения данного условия при фиксированных значениях остальных условий влечет за собой изменение результата решения.

### Критерии потока данных

Критерии потока данных формулируются с использованием следующих понятий:

- определение переменной (definition, def) – это оператор или выражение, меняющее значение переменной. Это может быть оператор присваивания, вызов процедуры или функции.
- использование переменной (use) - появление переменной в правой части оператора присваивания (с-use, computational use) или в составе логического предиката (p-use, predicate use).

Для данного набора тестов определим Р как множество полных путей, пройденных на этих тестах. Будем рассматривать пути от определения переменной до ее использования, которые не содержат внутри себя повторных определений переменной.

---

<sup>3</sup> [6] стр.2

Выделяют следующие критерии потока данных.

*Покрывание определений(all-defs)*: критерий выполняется, если Р включает в себя все пути от каждого определения переменной до какого-либо ее использования.

*Покрывание def-use-цепочек(all-uses)*: критерий выполняется, если Р включает в себя все пути до каждого использования переменной от какого-либо определения.

*Покрывание def – c-use цепочек (all-c-uses)*: критерий выполняется, если Р включает в себя все пути от каждого определения переменной до ее c-use-использования.

*Покрывание def – p-use цепочек (all-p-uses)*: критерий выполняется, если Р включает в себя все пути от каждого определения переменной до ее p-use-использования.

*Критерий all-c-uses/some-p-uses*: критерий выполняется, если Р включает в себя все пути от каждого определения переменной до ее c-use-использования. Если определению переменной не соответствует ни одного c-use-использования, то Р должен включать в себя путь до некоторого p-use-использования этой переменной.

*Критерий all-p-uses/some-c-uses*: критерий выполняется, если Р включает в себя все пути от каждого определения переменной до ее p-use-использования. Если определению переменной не соответствует ни одного p-use-использования, то Р должен включать в себя путь до некоторого c-use-использования этой переменной.

*Критерий all-du-path* выполняется, если Р включает в себя все пути от определения переменной до ее использования, т.е. если для конкретного определения и конкретного использования переменной существует несколько соединяющих их путей, все они должны быть покрыты.

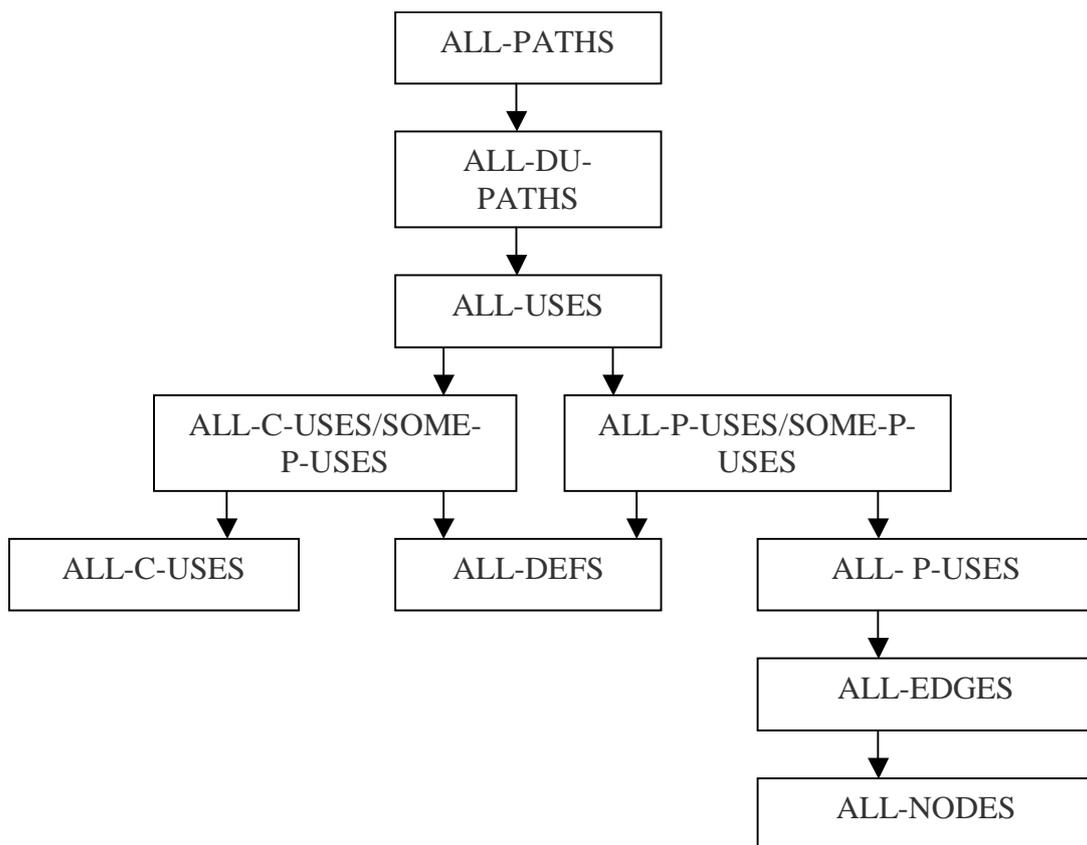


Рис. 1. Иерархия критериев тестирования

### **Мутационные критерии**

Приведенные ниже слабомутационные критерии предназначены для обнаружения неправильного использования арифметических операций и имен переменных и требу-

ют построения такого набора тестов, который позволил бы выявить все такие неточности.<sup>4</sup>

*Критерий слабой мутации арифметических операций.* Для данного критерия программы-мутанты получают из исходной программы заменой одного из арифметических операторов другим.

*Критерий слабой мутации имен переменных.* Для данного критерия программы-мутанты получают из исходной программы заменой вхождения одной переменной на другой с тем же типом и областью действия.

### **Отношение включения**

Пусть  $C1$  и  $C2$  – некоторые критерии тестирования. Критерий  $C1$  включает в себя критерий  $C2$ , если для всех программ и тестовых наборов выполнение  $C1$  подразумевает выполнение  $C2$ . Упорядочив критерии тестирования в соответствии с отношением включения, можно построить иерархию критериев, наглядно представляющую относительную "силу" и "слабость" критериев в смысле строгости требований к покрытию элементов программы.

### **Трудоемкость применения критериев**

Существующие работы по оценке и сравнению критериев в основном концентрируют внимание на способности этих критериев выявлять ошибки. Однако при необходимости реализации системы, осуществляющей проверку этих критериев, следует учитывать также и сложность вычислительных алгоритмов, необходимых для данной проверки. Эта сложность также является фактором выбора подмножества критериев, проверку которых следует реализовать.

Трудоемкость вычислений, связанных с определением истинности критерия для конкретного тестового набора, определяется:

1. сложностью выделения элементов, которые требуется покрыть, а также простой или составной структурой этих элементов.

Простой структурой характеризуются покрываемые элементы основных критериев потока управления –  $C0$  и  $C1$  (дуги и вершины). Составной структурой характеризуются покрываемые элементы критериев потока данных, поскольку включают в себя и вершину графа, содержащую определение переменной, и вершину, содержащую использование этой же переменной. Кроме того, дополнительный анализ требуется для выявления типа оператора и набора переменных, как определяемых, так и используемых в рамках оператора.

2. сложностью определения, покрывает ли определенный путь требуемые элементы.

Для большинства критериев эта сложность постоянна, поскольку после выделения требуемых фрагментов или элементов путей остается определить, являются ли они частью заданного пути.

### **Заключение**

Обзор существующих критериев тестирования показывает их многочисленность, которая приводит к проблеме выбора критериев для конкретного проекта. Эта проблема должна решаться с учетом:

1. имеющихся инструментов проверки критериев;
2. доступных ресурсов для тестирования;
3. возможности автоматизировать построение тестовых наборов.

---

<sup>4</sup> [2] стр. 53

При этом при разработке инструментов проверки критериев следует учитывать сложность необходимых алгоритмов проверки.

## Литература

1. Goodenough J.B., Gerhart S.L. Toward a theory of test data selection. - IEEE Transactions on Software Engineering, 1975, SE-2, № 3
2. Кауфман А. В., Черноножкин С.К. Критерии тестирования и система оценки полноты набора тестов // Программирование 1998, № 6
3. Zhu H., Hall P.A., May H.R. Software unit test coverage and adequacy. ACM Computing Surveys, Vol. 29, No.4, December 1997
4. Gupta N., Gupta R. Data Flow Testing // The University of Arizona, Tucson, 2002
5. Zhu H. A formal analysis of the subsume relation between software test adequacy criteria // IEEE Transactions Software Engineering, Vol. 22 No. 4 1996
6. Vilkomir S. A., Bowen P. J. Formalization of software testing criteria using the Z notation // Proc. COMPSAC 01: 25th IEEE Annual International Computer Software and Applications Conference, Chicago, Illinois, USA, 2001

# СИСТЕМА УПРАВЛЕНИЯ КАЧЕСТВОМ СЕРВИСА НА ОСНОВЕ ПОЛИТИК – РАЗРАБОТКА СЕРВЕРА КОНФИГУРИРОВАНИЯ СЕТЕВЫХ УСТРОЙСТВ

С.З. Оскотский, И.В. Григорьев

Научный руководитель – д.т.н., проф. И.П. Гуров

В работе рассматривается подход к управлению качеством сервиса, основанный на политиках. Описывается разработанная авторами архитектура системы управления качеством сервиса на основе политик. Формулируются требования к серверу конфигурирования сетевых устройств, являющемуся одним из ключевых компонентов системы управления качеством сервиса, и описываются результаты его проектирования.

## 1. Введение

В настоящее время стала очень актуальной задача управления качеством сервиса – задача обеспечения чувствительным к качеству сервиса приложениям определенных характеристик передачи информации по сети.

На сегодняшний день существуют две основные технологии управления качеством сервиса в IP-сетях: Integrated Services (IntServ) и Differentiated Services (DiffServ).

Технология IntServ индивидуально обеспечивает каждому потоку трафика определенное качество сервиса. Достигается это посредством того, что каждый маршрутизатор на пути следования пакетов резервирует для потока трафика часть своих ресурсов. Технология IntServ практически неприменима в глобальных сетях, поскольку при большом количестве потоков невозможно поддерживать на маршрутизаторе информацию о каждом проходящем через него потоке.

Технология DiffServ разбивает весь трафик на небольшое количество классов. Класс, к которому принадлежит пакет, определяется по значению специального поля – DiffServ Codepoint (DSCP) – в заголовке IP пакета. Сеть, в рамках которой значения поля DSCP имеют одинаковую интерпретацию, называется доменом дифференцированного обслуживания. В домене дифференцированного обслуживания различают граничные маршрутизаторы и внутренние маршрутизаторы. Граничные маршрутизаторы – это те маршрутизаторы, которые имеют связи с маршрутизаторами других доменов, в которых, возможно, те же значения поля DSCP имеют другую интерпретацию. Задачей граничных маршрутизаторов является маркировка входящего в домен трафика, а также отбрасывание поступающих в домен пакетов в силу каких-либо условий. Внутренние маршрутизаторы домена обрабатывают пакеты только на основании значения поля DSCP. Пакеты, имеющие одинаковое значение поля DSCP, одинаково обрабатываются внутренним маршрутизатором. Поэтому маршрутизаторы имеют дело с небольшими объемами информации, и технология DiffServ может быть использована для управления качеством сервиса в глобальных сетях.

Главный недостаток технологии DiffServ состоит в том, что индивидуальный поток трафика не получает гарантированного обеспечения качеством сервиса. При несогласованной работе маршрутизаторов домена дифференцированного обслуживания большой объем трафика других потоков, отнесенных к тому же классу, может существенно ухудшить качество сервиса, предоставляемое данному потоку. Чтобы этого не происходило, необходимы системы управления качеством сервиса, позволяющие согласованно управлять всеми маршрутизаторами домена дифференцированного обслуживания. Одна из таких систем и была разработана авторами.

## 2. Система управления качеством сервиса на основе политик

Для обеспечения всех клиентов оговоренным качеством сервиса необходимо, чтобы все маршрутизаторы подконтрольной провайдеру сети были согласованно скон-

фигурированы в соответствии с некими правилами. Подобное конфигурирование большого количества сетевых устройств может быть выполнено при помощи системы, управляющей сетью на основе так называемых политик.

Авторы принимали участие в разработке системы управления качеством сервиса на основе политик. Описываемая система является одной из реализаций идеи управления сетью на основе политик. Все ключевые требования, выдвигаемые к подобным системам, удовлетворяются описываемой системой.

Система управления качеством сервиса на основе политик дает возможность администратору сети централизованно и согласованно управлять набором маршрутизаторов, образующих вверенную ему сеть. Посредством системы управления качеством сервиса администратор задает набор политик, описывающих, как маршрутизаторами сети должен обрабатываться трафик. Набор политик может динамически изменяться.

Основных сценария использования системы управления качеством сервиса на основе политик три:

*Приписывание ролей.* Работая с системой управления качеством сервиса, администратор оперирует набором ролей. Понятие роли введено для облегчения конфигурирования портов маршрутизаторов, одинаково обрабатывающих трафик. Политика, будучи заданной один раз для роли, автоматически устанавливается на все порты, которым была присвоена эта роль.

*Редактирование политик.* Условие политики – это логическое выражение сравнения полей заголовка IP-пакета с константами. Действие политики состоит из двух частей. Первая часть действия политики – описание правил обработки трафика в зависимости от скорости его поступления. Вторая часть действия политики – отнесение трафика к одному из классов – задается как имя одного из классов трафика. В системе существует единственный набор классов трафика (и соответствующая ему дисциплина обслуживания). Каждый из классов трафика отражается на отдельную очередь на порту каждого маршрутизатора подконтрольной системе сети. Отнесение трафика к одному из классов – это фактически указание, в какую из очередей на порту маршрутизатора должен помещаться трафик.

Изменение политик может превратить весь набор политик системы в несогласованный набор политик. Установка несогласованного набора политик на маршрутизаторы может привести к непредсказуемому поведению сети. Поэтому изменяемые администратором в ходе редактирования политики не устанавливаются на сетевые устройства до момента распространения политик.

*Распространение политик.* Распространение политик – это установка политик на порты маршрутизаторов, то есть конфигурирование сетевых устройств согласно набору политик, составленному администратором. Поскольку одна и та же роль может быть приписана нескольким политикам, на один порт маршрутизатора могут оказаться установленными несколько политик. Политики устанавливаются согласно приоритетам. При поступлении в порт очередного пакета, к пакету в порядке убывания приоритетов применяются все политики, установленные на данном порту.

## **2.1. Требования к системе**

При постановке задачи разработки описываемой системы были выдвинуты следующие требования:

- конфигурирование сетевых устройств должно осуществляться посредством протокола COPS;
- в качестве хранилища политик должна использоваться LDAP-директория;
- система должна быть масштабируемой, то есть быть способной обслуживать большое количество сетевых устройств;

- система должна быть надежной, то есть функционировать (быть способной конфигурировать сетевые устройства) при выходе из строя одного или нескольких элементов системы;
- система должна предоставлять возможность централизованного управления сетевыми устройствами;
- в ходе процедуры распространения политик время, в течение которого информация в хранилищах политик несогласованна, должно быть сведено к минимуму;
- в ходе процедуры распространения политик время переконфигурации сетевого устройства должно быть минимальным;
- интерфейс пользователя для управления системой должен быть многоплатформенным.

В соответствии с выдвинутыми требованиями была разработана архитектура системы.

## 2.2. Архитектура системы

Архитектура системы представлена распределенной структурой с применением централизованного управления. Основные компоненты системы: консоль задания политик (Policy Management Tool); сервер политик (Policy Server); сервера конфигурирования сетевых устройств (Policy Decision Point); хранилища политик (Policy Repository); менеджеры хранилищ (Repository Manager).

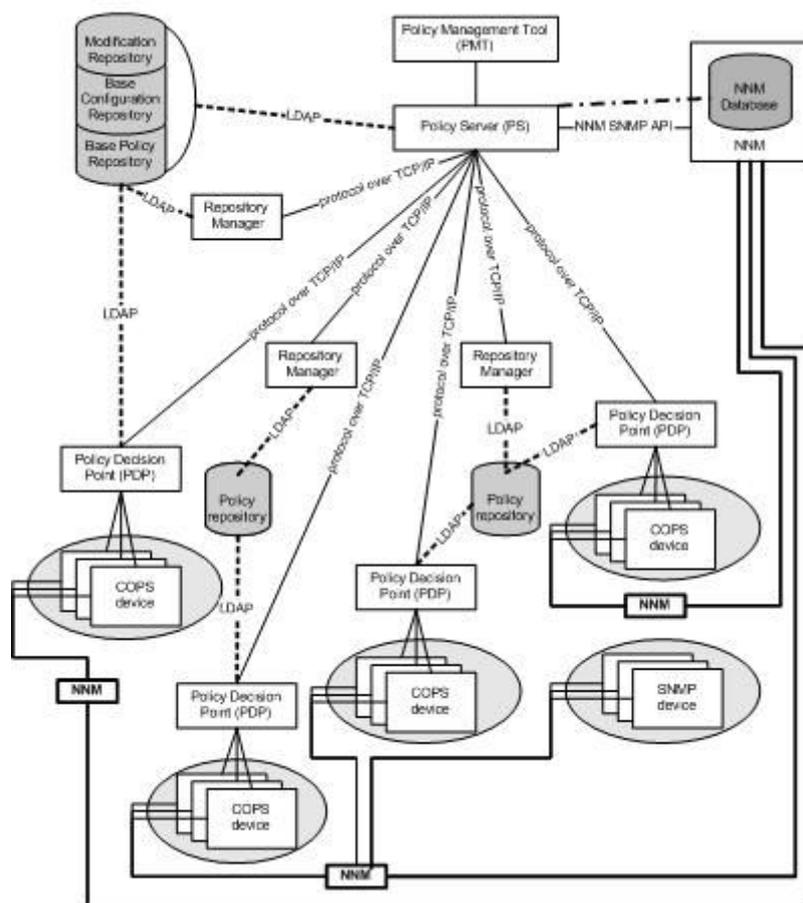


Рис. 1. Архитектура системы управления качеством сервиса на основе политик

### **2.2.1. Консоль задания политик**

Консоль задания политик (Policy Management Tool, PMT) предоставляет пользовательский интерфейс для управления всей системой. Основные функции, входящие в пользовательский интерфейс, таковы:

- приписывание ролей политикам и портам сетевых устройств;
- просмотр и редактирование политик;
- распространение политик;
- конфигурация параметров сетевых устройств, определяющих режим доступа к серверам конфигурирования сетевых устройств;
- управление конфигурацией системы – списком хранилищ политик, списком менеджеров хранилищ и списком серверов конфигурирования сетевых устройств.

Чтобы удовлетворить требование многоплатформенности пользовательского интерфейса, консоль задания политик была реализована на языке Java.

### **2.2.2. Сервер политик**

Сервер политик (Policy Server, PS) является центральным координирующим звеном системы управления качеством сервиса на основе политик. Сервером политик выполняются следующие функции:

- обработка запросов от консоли задания политик и предоставление консоли затребованной информации;
- получение, изменение и запись информации, имеющей отношение к политикам;
- управление ролями;
- распространение политик;
- изменение общей конфигурационной информации на сетевых устройствах;
- сохранение и предоставление консоли задания политик информации о событиях в системе и сообщениях, полученных от сетевых устройств;
- координация и управление хранилищами политик;
- координация и управление менеджерами хранилищ;
- координация и управление серверами конфигурирования сетевых устройств.

### **2.2.3. Сервер конфигурирования сетевых устройств**

Сервер конфигурирования сетевых устройств (Policy Decision Point, PDP) – это компонент, непосредственно отвечающий за установку политик на сетевые устройства. Сервера конфигурирования сетевых устройств взаимодействуют с сервером политик посредством специального протокола прикладного уровня, работающего над протоколом TCP.

Требования к серверу, а так же функциональность и архитектура сервера конфигурирования сетевых устройств более подробно описаны в следующем разделе.

### **2.2.4. Хранилище политик**

Хранилище политик (Policy Repository, PR) – это сервер, поддерживающий стандартный протокол LDAP, с установленной на нем схемой, позволяющей хранить информацию о политиках. Сервера конфигурирования сетевых устройств получают информацию о политиках из хранилищ. Информация в хранилищах обновляется сервером политик посредством менеджеров хранилищ.

### **2.2.5. Менеджер хранилища политик**

Менеджер хранилища политик (Repository Manager, RM) – это компонент, ответственный за обновление политик в хранилище политик. Менеджер хранилища используется сервером политик в ходе процедуры распространения политик для обновления содержимого хранилищ. Менеджеры хранилищ взаимодействуют с сервером политик посредством специального протокола прикладного уровня, работающего над протоко-

лом TCP. Менеджеры хранилищ политик введены в архитектуру системы для максимально быстрого обновления содержимого хранилищ, то есть для минимизации времени, в течение которого информация в хранилищах является несогласованной. Обновление содержимого хранилищ предпочтительнее осуществлять с менеджеров хранилищ, чем с сервера политик, потому что:

- как правило, канал связи между менеджером хранилища и хранилищем намного лучше канала связи между хранилищем и сервером политик;
- количество хранилищ может быть велико, и обновление информации всех хранилищ с сервера политик может занять длительное время, в то время как при использовании менеджеров хранилищ работы по обновлению содержимого хранилищ выполняются параллельно.

### **3. Сервер конфигурирования сетевых устройств**

Сервер конфигурирования сетевых устройств отвечает за установку политик на сетевые устройства. В ходе своей работы сервер конфигурирования взаимодействует со следующими компонентами (см. п. 2.2):

- сервер политик (взаимодействие происходит посредством специального протокола, работающего над протоколом TCP): в ходе процедуры распространения политик сервер конфигурирования получает от сервера политик команду переконфигурации. Поступление команды переконфигурации означает, что информация в хранилище политик обновилась и необходимо переконфигурировать подконтрольные серверу конфигурирования устройства;
- хранилище политик (взаимодействие происходит посредством протокола LDAP): из хранилища политик сервер конфигурирования извлекает политики, которые должны быть установлены на подконтрольные серверу устройства;
- сетевые устройства (взаимодействие происходит посредством протокола COPS): сервер конфигурирует подконтрольные ему сетевые устройства во время процедуры распространения политик или по запросу самих устройств.

Задачей сервера является установка записанных в хранилище политик на подконтрольные сетевые устройства, то есть конфигурация на портах устройств цепочек элементов управления качеством сервиса, соответствующих устанавливаемым политикам. Различные сетевые устройства (сетевые устройства от разных производителей) могут иметь отличающиеся друг от друга наборы элементов управления качеством сервиса. Кроме того, различные сетевые устройства могут иметь различные ограничения на типы соединяемых элементов и на количество элементов в цепочке. Для каждого подконтрольного сетевого устройства сервер конфигурирования должен преобразовывать извлекаемые из хранилища политики в такие цепочки элементов управления качеством сервиса, которые могут быть установлены на данное сетевое устройство.

#### **3.1. Требования к серверу**

При постановке задачи разработки сервера конфигурирования сетевых устройств были выдвинуты следующие требования:

- политики, устанавливаемые на сетевые устройства, должны быть согласованным. Выполнение этого требования осложняется тем, что LDAP-директория не поддерживает транзакций, т.е. возможна ситуация, когда сервер конфигурирования обратится к хранилищу политик в то время, когда менеджер хранилища обновляет в хранилище информацию. Если не принимать специальных мер, это может привести к тому, что политики, извлеченные сервером конфигурирования из хранилища, будут несогласованными. Часть политик будет принадлежать старому набору политик, а часть – новому набору политик;

- количество обращений к хранилищу политик должно быть минимальным;
- переконфигурация сетевых устройств должна быть максимально быстрой и минимально нагружать сеть;
- сервер конфигурирования должен быть разработан таким образом, чтобы была возможность подключения к нему модулей управления устройствами. Это необходимо для того, чтобы можно было использовать сервер конфигурирования для управления устройствами, неизвестными на момент разработки сервера (устройствами, у которых набор элементов управления качеством сервиса отличается от набора элементов управления качеством сервиса тех устройств, для управления которыми разрабатывается сервер).

В соответствии с выдвинутыми требованиями был спроектирован набор классов, образующих сервер конфигурирования сетевых устройств.

Основным сценарием работы сервера конфигурирования сетевых устройств является установка на подконтрольные устройства новых политик. Процедура установки новых политик инициируется специальной командой, поступающей с сервера политик.

### **3.2. Соответствие требованиям**

Рассмотрим соответствие сервера конфигурирования сетевых устройств требованиям, выдвинутым при постановке задачи.

#### **3.2.1. Установка согласованных политик**

Поскольку LDAP-директория не имеет поддержки транзакций, необходимо принятие специальных мер для избежания установки на сетевые устройства несогласованных политик. Перед началом считывания политик из хранилища сервер конфигурирования проверяет, не проводится ли в данный момент менеджером хранилища обновление политик, сервер конфигурирования приостанавливает свою работу на некоторое время, после чего снова проверяет согласованность записанных в хранилище политик.

Далее, сервер конфигурирования приступает к извлечению политик из хранилища. По окончании считывания политик сервер конфигурирования проверяет, не произошли ли изменения данных хранилища в течение считывания политик, и если никаких изменений не произошло, то считанные политики являются согласованными, и их можно устанавливать на сетевые устройства.

В противном случае считанные политики могут быть несогласованными. Чтобы избежать установки несогласованных политик на сетевые устройства, необходимо повторить процедуру считывания политик.

#### **3.2.2. Минимизация количества обращений к хранилищу**

Для минимизации количества обращений к хранилищу политик в архитектуру сервера конфигурирования включен специальный компонент – PolicyCache. Этот компонент хранит копию политик, записанных в хранилище. При необходимости установки политик на сетевое устройство сервер конфигурирования обращается за политиками не к хранилищу, а к компоненту PolicyCache. Компонент PolicyCache заполняется новыми политиками из хранилища только при расхождении значений версии компонента и версии хранилища.

#### **3.2.3. Быстрая переконфигурация устройств**

При установке на сетевое устройство новых политик цепочки элементов управления качеством сервиса на портах устройства целиком не заменяются. Неизменные участки цепочек элементов управления качеством сервиса не переустанавливаются. В архитектуру сервера конфигурирования включен специальный компонент – PolicyComparator, – который сравнивает цепочки управления качеством сервиса и вычисляет разницу, которую необходимо применить к установленной цепочке для того,

чтобы получить новую цепочку элементов управления качеством сервиса. Вычисленная разница передается на сетевое устройство. Такой подход позволяет максимально быстро переконфигурировать сетевые устройства.

#### **4. Заключение**

В данной работе рассматривается подход к решению проблемы обеспечения качества сервиса в IP-сетях, основанный на задании набора правил, называемых политиками. В качестве результатов работы показываются результаты проектирования сервера конфигурирования сетевых устройств, а также общей структуры системы, в разработке которой авторы принимали участие. Дальнейшей работой в данном направлении может являться реализация описанной системы управления качеством сервиса.

#### **Литература**

1. Weibin Zhao, David Olshefski, Henning Schulzrinne “Internet Quality of Service: an Overview” (Columbia University)
2. Y. Bernet, S. Blake, D. Grossman, A. Smith “An Informal Management Model for Diffserv Routers” (RFC 3290)
3. K. Nichols, S. Blake, F. Baker, D. Black “Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers” (RFC 2474)
4. S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss “An Architecture for Differentiated Services” (RFC 2475)
5. Chuck Semeria “Supporting Differentiated Service Classes: Queue Scheduling Disciplines” (White Paper of Juniper Networks, Inc.)
6. D. Durham, J. Boyle, R. Cohen, S. Herzog, R. Rajan, A. Sastry “The COPS (Common Open Policy Service) Protocol” (RFC 2748)
7. K. Chan, J. Seligson, D. Durham, S. Gai, K. McCloghrie, S. Herzog, F. Reichmeyer, R. Yavatkar, A. Smith “COPS Usage for Policy Provisioning (COPS-PR)” (RFC 3084)

# СЕМАНТИЧЕСКИЙ АНАЛИЗ ТЕХНИЧЕСКОЙ ДОКУМЕНТАЦИИ В CASE-ТЕХНОЛОГИЯХ

А.Е. Харитонова

Научный руководитель – к.т.н., доц. А.А. Приблуда

Предлагается новый подход к проектированию информационных систем, основанный на повторном использовании уже реализованных компонентов с помощью лингвистического процессора. Исследуются базовые принципы его функционирования.

В настоящее время одним из самых распространенных подходов к проектированию информационных систем является использование CASE-технологий. Важным свойством любых средств проектирования, основанных на CASE-технологиях, можно считать возможность повторного применения уже спроектированных компонентов готовых продуктов. В целях автоматизации выбора таких компонентов можно использовать семантический анализ текстовой информации о проектах.

Одним из вариантов решений поставленной задачи является способ формального описания проекта с использованием метаязыка [1]. Основная идея этого подхода состоит в том, что проект готовой системы описывается специальным метаязыком и в таком виде хранится в базе данных. Таким же образом описывается и проект создаваемой системы. В дальнейшем, применяя средства автоматизации, можно из описаний проектов, хранящихся в базе данных, и описания проекта разрабатываемой системы выделить сходные части и использовать соответствующие им проектные решения повторно.

Однако основной механизм такого подхода – описание на метаязыке – является одной из самых медленных частей системы проектирования, которая его использует. Описание на метаязыке существенно замедляет процесс проектирования, так как выполняется преимущественно разработчиком. Кроме того, при использовании этого подхода не принимается во внимание контекстная информация, что не способствует большой точности выбора при выделении сходных частей проектов. Следует заметить, что работает такой метод исключительно с проектными данными, а не с конечным продуктом разработки.

Более эффективным представляется подход с использованием лингвистического процессора для семантического анализа технической информации о проектах. При этом под проектом понимается система в целом, т.е. техническое задание, проектная информация, описанная в формате какой-либо CASE-среды, документация, программный код и т.п. Предварительная автоматизированная обработка готового проекта перед помещением его в базу данных заключается в анализе лингвистическим процессором всех текстовых данных этого проекта с учетом содержащейся в них контекстной информации. Результатом предварительной обработки является структура данных, определяющая взаимосвязь частей реализованной системы. Ядром такой структуры является семантическое описание технического задания.

При создании нового проекта техническое задание проектируемой системы обрабатывается лингвистическим процессором для получения семантического описания. Поскольку семантические описания в первом приближении представляют собой массивы объектов одного типа, то сравнение их не является трудоемкой задачей и может выполняться автоматическими средствами.

Таким образом, предлагаемый подход с использованием лингвистического процессора основывается на методе автоматизированного семантического анализа текстовой информации о проекте.

Для дальнейшей разработки предлагаемого метода были проведен ряд теоретических исследований по определению возможностей метода с точки зрения точности выбора сходных частей проектов. В их число входили:

1. точность сравнения семантических описаний. Целью этого исследования было определить насколько соответствуют результаты автоматического сравнения, сравнению аналогичных компонентов человеком;
2. вероятность ошибочного результата – вероятность того, что результаты автоматического сравнения будут полностью противоречить результатам сравнения человеком.
3. степень влияния полноты представленных на анализ данных о проектах на конечный результат работы лингвистического процессора.

Оценивались также возможности метода с точки зрения трудозатрат, сложности программной реализации и т.п.

При создании прототипа лингвистического процессора для системы автоматизированного проектирования выяснилось, что предлагаемый метод является достаточно трудоемким в программной реализации, так как алгоритмы обработки взаимокорреляции всех слов в анализируемом тексте очень сложны. Основной проблемой являлось получение семантического описания, в котором единичный элемент должен представлять собой группу слов-существительных, имеющих сходное значение [2]. Поэтому в прототипе для реализации ряда функций лингвистического процессора использовались упрощенные алгоритмы. Алгоритмы были упрощены как с математической, так и с лингвистической точки зрения. Общая структура алгоритма может быть представлена следующим образом.

1. Синтаксическо-статистическая обработка текста. На этом этапе составляется распределение существительных в обрабатываемом тексте.
2. Синтаксическая обработка. С помощью функций библиотек программы linkparser [3] в тексте определяются подлежащие, сказуемые и дополнения. Полученные результаты дают возможность сделать предположение о том, какие слова несут ключевую информацию о тексте. На этом же этапе происходит формирование концептов с единичным содержанием.
3. Семантическая обработка. Происходит расчет взаимосвязей концептов между собой, т.е. строится матрица связей. двумя из четырех различных методов [4]. На основе этой матрицы концепты объединяются.

В упрощенном алгоритме в недостаточной степени учитывалась корреляция всех используемых слов из текста. На начальных этапах разработки алгоритма сходные слова группировались вокруг какого-либо одного слова, выбираемого случайно. В результате тестирования оказалось, что в группах, сформированных таким образом, возможны противоречия. А именно, при наличии связи с основным словом концепта рассматриваемое слово могло быть недостаточно связано с остальными присоединенными или вообще противоречить им (не иметь никакой связи). Это можно показать следующим примером. Обрабатывается текст:

*“Our ultimate goal, though, is to obtain a representation of the definitions in a formalism such as conceptual graphs in order to compare definitions from various sources. This study is part of a larger project aimed at comparing definitions of medical terms in specialized sources such as medical dictionaries with those in general resources such as WordNet. In other words, our goal is to compare definitions of medical terms for health professionals and for users of consumer health applications. Although not completely unrelated to them, the task of characterizing definitions is quite different from other tasks in which definitions are involved. Characteristics of the definitions of terms, especially from several sources, represent valuable information. Among other things, this study confirmed the predominance of genus-differentia definitions for anatomical terms in both WordNet and specialized resources. This knowledge is expected to help perform the deeper analysis needed for representing the definitions in a formalism suitable to their comparison” .*

После его обработки лингвистическим процессором получены следующие результаты:



ра. Она состояла из технического задания, документации к системе и программного кода. При этом в понятие “документация” входит вся внутренняя документация системы, используемая, как правило, только разработчиками.

Несмотря на намеренное упрощение алгоритма и формата хранения проектных данных, результаты работы созданного прототипа лингвистического процессора для системы автоматизированного проектирования подтвердили теоретические исследования и позволили сделать вывод, что система даже в таком исполнении обладает достаточно большой точностью – схожие части систем используются наиболее полно. Это достигается за счет того, что при формировании основной структуры, описывающей готовую систему, используется контекстная информация из комментариев к программному коду, проектной части, непосредственно самого кода и документации к системе. К достоинствам предлагаемого подхода можно отнести то, что от пользователя системы автоматизированного проектирования с использованием лингвистического процессора не потребуется преобразовывать готовые проекты, с которыми работает система, в какой-либо формат. Кроме того, возможна работа не только с полностью готовыми проектами, но и с их частями.

Следует заметить, что прототип лингвистического процессора обрабатывает тексты пока только на английском языке, так как используемые в нём библиотеки и словари ещё не реализованы для русского языка.

## Литература

1. Брюхов Д.О., ИПИ РАН; Калиниченко Л.А., ИПИ РАН; Кравченко Д.В., ИПИ РАН; Скворцов Н.А., ИПИ РАН. Средства проектирования информационных систем, ориентированные на повторное использование компонентов // Материалы семинара Московской Секции ACM SIGMOD 29 октября 1998 г.
2. Гайдышев И. Анализ и обработка данных: специальный справочник. – СПб: Питер, 2001.
3. Link Grammar Parser. <http://www.link.cs.cmu.edu/link/>
4. Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures, Alexander Budanitsky and Graeme Hirst, Department of Computer Science University of Toronto
5. Wordnet improves Text Document Clustering, Andreas Hotho, Steffen Staab, Gerd Stumme Institute AIFB, University of Karlsruhe, Germany. 2003.

# ИСПОЛЬЗОВАНИЕ АВТОМАТНОГО ПОДХОДА ДЛЯ ПОСТРОЕНИЯ ВИЗУАЛИЗАТОРОВ

М.А. Казаков

Научный руководитель – д.т.н., проф. А.А. Шалыто

Предлагается новая технология для построения визуализаторов алгоритмов на основе формального преобразования программной реализации алгоритма в конечный автомат Мили. В статье представлено подробное описание каждого этапа предлагаемой технологии на примере "дискретной задачи о рюкзаке".

## Введение

Одним из методических приемов [1–4] при обучении программированию и дискретной математике является разработка визуализаторов классических вычислительных алгоритмов [5,6].

Визуализатор – это программа, иллюстрирующая выполнение алгоритма при определенных входных данных. Примерами визуализаторов могут служить, например, программа, занимающаяся построением графика функции, либо программа, визуально моделирующая какой-либо физический процесс. Применительно к дискретной математике и программированию визуализаторы обычно моделируют некоторые алгоритмы, давая возможность обучающемуся при помощи интуитивно понятного интерфейса проходить алгоритм шаг за шагом от начала до конца, а при необходимости и обратно.

Перечислим отличительные характеристики визуализаторов.

1. Простота использования, определяемая понятностью интерфейса. Поэтому для работы с визуализатором обычно не требуется специальная подготовка.
2. Четкость и простота представления визуализируемого процесса.
3. Компактность визуализаторов, ввиду их реализации в виде Java-апплетов. Это при необходимости упрощает передачу визуализаторов в сети Интернет, что особенно важно при дистанционном обучении.

Визуализаторы в рассматриваемой области решают следующие задачи, возникающие в процессе обучения.

1. графическое и словесное разъяснение действий алгоритма на конкретных наборах входных данных. При этом понимание алгоритма не обязательно, поскольку именно визуализатор должен объяснить действие алгоритма;
2. предоставление пользователю инструмента, реализующего данный алгоритм. При этом учащийся освобождается от необходимости выполнять шаги алгоритма, так как их автоматически выполняет визуализатор.

Визуализатор выполняет обычно следующие функции.

1. пошаговое выполнение алгоритма;
2. возможность просмотра действия алгоритма при разных наборах входных данных, в том числе и введенных пользователем;
3. возможность просмотра действия алгоритма в динамике;
4. возможность перезапуска алгоритма на текущем наборе входных данных.

Динамическая визуализация наглядно демонстрирует такую характеристику алгоритма, как трудоемкость (особенно при пошаговой демонстрации). Для некоторых алгоритмов (например, машины Тьюринга) динамический вариант демонстрации вообще представляется более естественным, чем любой набор статических иллюстраций.

До недавнего времени технология разработки визуализаторов сводилась к эвристике. При этом каждый алгоритм требовал нового подхода и осмысления реализации визуализатора. Другими словами, технология разработки визуализаторов отсутствовала, и основные успехи в этой области были педагогическими [3, 4, 7].

Основным недостатком при построении визуализаторов является то, что обычно используются только такие понятия, как "входные и выходные переменные", а понятие

"состояние" в явном виде не используется. Однако, по мнению авторов, совершенно естественным кажется, что каждый шаг визуализации необходимо проводить в соответствующем состоянии или на переходе. Поэтому при построении программ визуализации целесообразно использовать технологию автоматного программирования, базирующуюся на конечных автоматах [7].

В настоящей работе показано, что применение программирования с явным выделением состояний превращает процесс разработки логики визуализаторов в технологию.

Основным требованием к визуализатору является возможность пошаговой реализации алгоритма. Поэтому предлагаемая технология должна преобразовать исходный алгоритм в набор шагов для отображения на экране.

Визуализатор "по-крупному" предлагается строить следующим образом:

- по программе строится автомат логики визуализатора (контроллер – controller);
- выбираются визуализируемые переменные (модель – model);
- проектируется формирователь изображений и комментариев, который преобразует номер состояния и соответствующие значения визуализируемых переменных в "картинку" и поясняющий текст (представление – view).

Такая конструкция визуализатора соответствует одному из основных паттернов проектирования объектно-ориентированных программ, который обозначается аббревиатурой MVC (Model-View-Controller) [9].

Излагаемая ниже технология основана на методе построения конечного автомата по тексту программы [10]. При этом по указанному тексту строится схема алгоритма, которая на основе соответствующего кодирования преобразуется в граф перехода конечного автомата.

Исходя из изложенного, сформулируем технологию построения визуализаторов.

1. Постановка задачи.
2. Решение задачи (в словесно-математической форме).
3. Выбор визуализируемых переменных.
4. Анализ алгоритма для визуализации. Анализируется решение с целью определения того, что и как отображать на экране.
5. Алгоритм решения задачи.
6. Реализация алгоритма на выбранном языке программирования. На этом шаге производится реализация алгоритма, его отладка и проверка работоспособности.
7. Построение схемы алгоритма по программе.
8. Преобразование схемы алгоритма в граф переходов автомата Мили [10, 11].
9. Формирование набора невизуализируемых переходов.
10. Выбор интерфейса визуализатора.
11. Сопоставление иллюстраций и комментариев с состояниями автомата.
12. Архитектура программы визуализатора.
13. Программная реализация визуализатора.

Известно, что обычно создание программы разбивается на пять стадий: разработка требований, анализ, проектирование, реализация и тестирование. В данном случае к разработке требований относятся первые три пункта технологии. Анализ – пункты четыре и пять технологии. К проектированию относятся пункты с шестого по одиннадцатый. В отличие от традиционного для объектно-ориентированного программирования подхода, выбор архитектуры программы относится не к стадии проектирования, а к реализации, так как все предыдущие пункты технологии не зависят от реализации. Естественно, что к стадии реализации относится и последний пункт технологии.

Как будет показано ниже, при использовании предлагаемой технологии тестирование резко упрощается, а время написания визуализатора сокращается.

Продemonстрируем предложенную технологию на примере построения визуализатора алгоритма, использующего динамическое программирование [5], которое является одним из самых сложных разделов алгоритмики.

## 1. Постановка задачи

Рассмотрим словесную постановку задачи, которая в литературе носит название "дискретная задача о рюкзаке" [5, 12]. Приведем одну из формулировок этой задачи.

*Имеется набор из  $K$  неделимых предметов, для каждого из которых известна масса  $M_i$  (в кг), являющаяся натуральным числом. Задача состоит в том, чтобы определить, существует ли хотя бы один набор предметов, размещаемый в рюкзаке, суммарная масса которых равна  $N$ . Если такой набор существует, то требуется определить его состав.*

## 2. Решение задачи

Приведем словесную формулировку решения этой задачи.

### 2.1. Построение функции, представляющей оптимальное решение

*Построим вектор  $M(1, K)$ , каждый элемент  $M_i$  которого соответствует массе  $i$ -го предмета.*

*Введем функцию  $T(i, j)$ , значение которой равно единице, если среди предметов с первого по  $i$ -ый существует хотя бы один набор предметов, сумма масс которых равна  $j$ , и нулю – в противном случае.*

### 2.2. Рекуррентные соотношения, связывающие оптимальные значения для подзадач

*Определим начальные значения функции  $T$ :*

- $T(0, j)=0$  при  $j \geq 1$  { без предметов нельзя набрать массу  $j \geq 1$ };
- $T(i, 0)=1$  при  $i \geq 0$  { всегда можно набрать нулевую массу }.

*Определим возможные значения функции  $T$  при ненулевых значениях аргументов.*

*Существуют две возможности при выборе предметов: включить предмет с номером  $i$  в набор или не включать. Это в математической форме может быть записано следующим образом:*

- если предмет с номером  $i$  не включается в набор, то решение задачи с  $i$  предметами сводится к решению подзадачи с  $(i-1)$ -им предметом. При этом  $T(i, j)=T(i-1, j)$ ;
- если предмет с номером  $i$  включается в набор, то это уменьшает суммарную массу для  $i-1$  первых предметов на величину  $M_i$ . При этом  $T(i, j)=T(i-1, j-M_i)$ . Эта ситуация возможна только тогда, когда  $M_i \leq j$ .

*Для получения решения необходимо выбрать большее из значений рассматриваемой функции. Поэтому рекуррентные соотношения при  $i \geq 1$  и  $j \geq 1$  имеют вид:*

- $T(i, j)=T(i-1, j)$  при  $j < M_i$ ;
- $T(i, j)=\max(T(i-1, j), T(i-1, j-M_i))$  при  $j \geq M_i$ .

### 2.3. Задание исходных данных

*Пусть  $K=5$  – заданы пять предметов.*

*Эти предметы имеют следующие массы:  $M_1=4$ ;  $M_2=5$ ;  $M_3=3$ ;  $M_4=7$ ;  $M_5=6$ .*

*Суммарная масса предметов, которые должны быть размещены в рюкзаке, равна 16 ( $N=16$ ).*

### 2.4. Вычисление значений функции $T(i, j)$ (прямой ход алгоритма)

*Используя приведенные выше рекуррентные соотношения, построим таблицу значений рассматриваемой функции (табл. 1).*

*Из таблицы следует, что  $T(5, 16)=1$ , и поэтому существует хотя бы один набор предметов, удовлетворяющий постановке задачи.*

		j																
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
i	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0
	3	1	0	0	1	1	1	0	1	1	1	0	0	1	0	0	0	0
	4	1	0	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1
	5	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Таблица 1. Значения функции  $T(i, j)$

## 2.5. Определение набора предметов (обратный ход алгоритма)

Рассмотрим элементы  $T(5,16)$  и  $T(4,16)$  таблицы. Так как их значения равны, то массу в 16 кг можно набрать с помощью первых четырех предметов – пятый предмет в набор можно не включать.

Теперь рассмотрим элементы  $T(4,16)$  и  $T(3,16)$ . Их значения не равны, и поэтому набор из трех предметов не обеспечивает решение задачи. Следовательно, четвертый предмет должен быть включен в набор. Поэтому "остаток" массы, размещаемой в рюкзаке, который должен быть набран из оставшихся предметов, равен  $16 - M_4 = 16 - 7 = 9$ .

Для элементов  $T(3,9)$  и  $T(2,9)$  значения равны – третий предмет в набор не включается.

Для элементов  $T(2,9)$  и  $T(1,9)$  значения не равны – второй предмет должен быть включен в набор. При этом остаток равен  $9 - M_2 = 9 - 5 = 4$ .

И, наконец, рассмотрим элементы  $T(1,4)$  и  $T(0,4)$ . Их значения не равны – первый предмет должен быть включен в набор, а "остаток" массы для других предметов равен 0.

Таким образом, одно из решений задачи найдено – в набор включены первый, второй и четвертый предметы.

## 3. Выбор визуализируемых переменных

Перечислим переменные, которые содержат значения, необходимые для визуализации:

- $i$  – номер предмета;
- $j$  – сумма весов предметов;
- $T[i][j]$  – значения функции  $T$ ;
- $positions$  – номера предметов.

## 4. Анализ алгоритма для его визуализации

Из рассмотрения алгоритма следует, что его визуализация должна выполняться следующим образом:

- показать таблицу, заполненную начальными значениями;
- отразить прямой ход решения задачи – процесс построения таблицы значений  $T(i, j)$ ;
- отразить обратный ход решения задачи – процесс поиска набора предметов за счет "обратного" движения по указанной таблице.

## 5. Алгоритм решения задачи

Приведем алгоритм решения задачи на псевдокоде [5]:

```
1  for  $j \leftarrow 1$  to  $N$ 
2      do  $T[0, j] \leftarrow 0$ 
3  for  $i \leftarrow 0$  to  $K$ 
4      do  $T[i, 0] \leftarrow 1$ 
5  for  $i \leftarrow 1$  to  $K$ 
6      do for  $j \leftarrow 1$  to  $N$ 
7          do if  $j < M[i]$ 
8              then  $T[i, j] \leftarrow T[i-1, j]$ 
9              else  $T[i, j] \leftarrow \text{Max}(T[i-1, j], T[i-1, j-M[i]])$ 
10  $sum \leftarrow N$ 
11 if  $T[K, sum] = 1$ 
12     then
13         for  $i \leftarrow K$  downto 1
14             do if  $T[i, sum] \neq T[i-1, sum]$ 
15                 then
16                     добавить в искомый набор
17                      $sum \leftarrow sum - M[i]$ 
18                 return искомый набор
19     else return "набор не существует"
```

## 6. Реализация алгоритма на языке Java

Перепишем программу на псевдокоде с помощью языка Java.

```
/**
 * @param N суммарный вес предметов в рюкзаке
 * @param M массив весов предметов
 * @return список номеров предметов, сумма весов которых равна N,
 * либо null если это невозможно
 */
private static Integer[] solve(int N, int[] M) {
    // Переменные циклов
    int i, j;
    // Количество предметов
    int K = M.length;
    // Массив T
    int[][] T = new int[K + 1][N + 1];
    // Результирующие номера
    ArrayList positions = new ArrayList();
    // Результат работы true, если суммарный вес можно получить
    boolean result = false;

    // Определение начальных значений функции T
    for (j = 1; j <= N; j++) {
        T[0][j] = 0;
    }
    for (i = 0; i <= K; i++) {
        T[i][0] = 1;
    }

    // Построение таблицы значений функции T
    for (i = 1; i <= K; i++) {
        for (j = 1; j <= N; j++) {
            if (j >= M[i - 1]) {
                T[i][j] = Math.max(T[i - 1][j], T[i - 1][j - M[i - 1]]);
            }
            else {

```

```

        T[i][j] = T[i - 1][j];
    }
}

// Определение набора предметов
int sum = N;
if (T[K][sum] == 1) {
    // Решение существует
    for (i = K; i >= 1; i--) {
        if (T[i][sum] != T[i - 1][sum]) {
            positions.add(new Integer(i));
            sum -= M[i - 1];
        }
    }
    // Решение найдено
    result = true;
}

// Если решение найдено, то оно возвращается, иначе возвращается null
return (Integer[])(result ? positions.toArray(new Integer[0]) : null);
}

```

В этой программе операции ввода/вывода не приведены, так как их будет выполнять визуализатор.

## 7. Построение схемы алгоритма по программе

Схема алгоритма, построенного по тексту программы, представлена на рис. 1.

## 8. Преобразование схемы алгоритма в граф переходов автомата Мили

В работе [10] показано, что при программной реализации по схеме алгоритма может быть построен как автомат Мура, так и автомат Мили. В автоматах первого типа действия выполняются в вершинах, а в автоматах второго типа – на переходах. Визуализаторы могут быть построены как с использованием автоматов Мура, так и автоматов Мили. Для уменьшения числа состояний в настоящей работе применяются автоматы Мили.

В соответствии с методом, изложенным в работе [10], определим состояния, которые будет иметь автомат Мили, соответствующий этой схеме алгоритма. Для этого присвоим номера точкам, следующим за последовательно расположенными операторными вершинами (последовательность может состоять и из одной вершины). Номера присваиваются также начальной и конечным вершинам.

Количество состояний, определенных описанным методом, бывает для целей визуализации недостаточным, так как в ряде случаев необходимо показывать также условия и результаты некоторых действий.

Исходя из изложенного, для схемы алгоритма на рис. 1 выделим девять состояний автомата с номерами 0 – 6, 9, 10. Еще одно состояние с номером 7 введено для визуализации обратного хода алгоритма. И, наконец, состояние с номером 8 используется для визуализации результата выбора на обратном ходе. Таким образом, автомат визуализации будет иметь одиннадцать состояний, а его граф переходов, который строится за счет определения путей между выделенными точками на схеме алгоритма – одиннадцать вершин (рис. 2).

Как следует из графа переходов, на петлях в состояниях 1 и 2 осуществляется заполнение таблицы значений функции  $T$  начальными данными в первой строке и в первом столбце. На петлях в состоянии 4 реализуется алгоритм заполнения таблицы значений функции  $T$ . Переходы из состояния 7 в состояние 8 отображают первую половину шага алгоритма обратного хода, причем более сложный переход соответствует включению предмета с номером  $i$  в результирующий набор, а второй переход – не



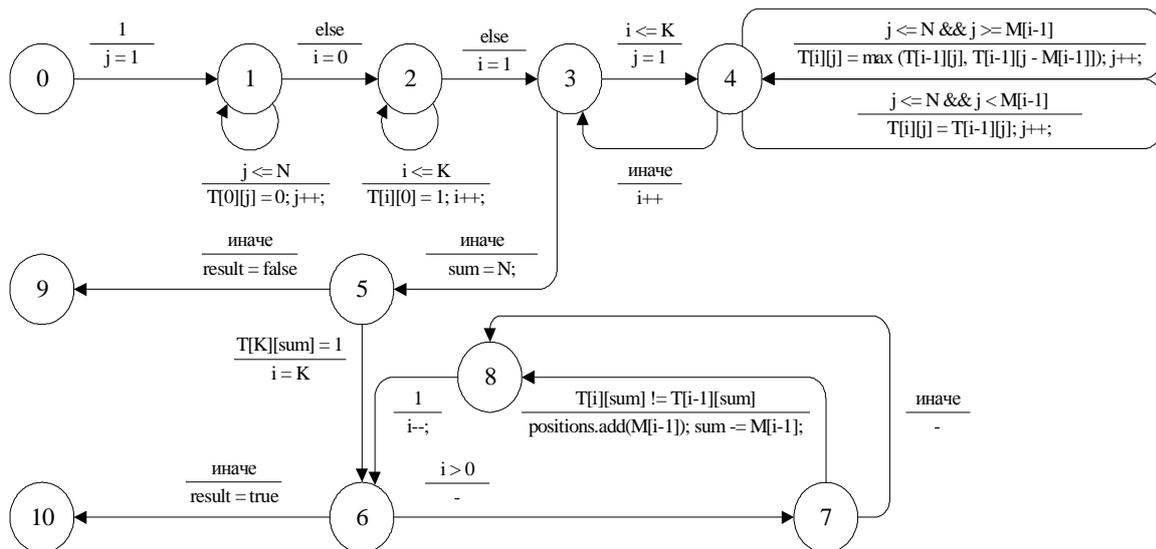


Рис. 2. Граф переходов для визуализации логики алгоритма

## 9. Формирование набора невизуализируемых переходов

Из анализа графа переходов (рис. 2) следует, что "неинтересными" (не изменяющими визуализацию) для пользователя являются следующие переходы:  $1 \rightarrow 2$ ,  $3 \rightarrow 4$ ,  $3 \rightarrow 5$ ,  $4 \rightarrow 3$ ,  $6 \rightarrow 7$ ,  $6 \rightarrow 10$ . Устранение таких переходов осуществляется за счет того, что один шаг работы визуализатора состоит из нескольких переходов автомата, реализующего его логику.

Для исключения "неинтересных" переходов шаг визуализатора реализуется следующим образом:

- 1 **while** следующий переход **in** ( $0 \rightarrow 1$ ,  $1 \rightarrow 2$ ,  $3 \rightarrow 4$ ,  $3 \rightarrow 5$ ,  $4 \rightarrow 3$ ,  $6 \rightarrow 7$ ,  $6 \rightarrow 10$ )
- 2 **do** сделать переход автомата

## 10. Выбор интерфейса визуализатора

В верхней части визуализатора (в дальнейшем называемой *иллюстрацией*) представляется следующая информация:

- таблица значений функции  $T(i, j)$ , где  $i = 0, \dots, 5$ ;  $j = 0, \dots, 16$ ;
- набор весов предметов (крайний левый столбец с цифрами 4, 5, 3, 7, 6).

В нижней части визуализатора расположена панель управления, которая содержит следующие кнопки:

- ">>" – сделать шаг алгоритма;
- "Заново" – начать алгоритм заново;
- "Авто" – перейти в автоматический режим;
- "<", ">" – изменение паузы между автоматическими шагами алгоритма.

Среднюю часть визуализатора занимает область комментариев, в которой на каждом шаге отображается описание выполняемого алгоритмом действия.

Визуализатор в исходном состоянии, которое соответствует начальному состоянию автомата, представлен на рис. 3.



Рис. 3. Начальное состояние визуализатора

## 11. Сопоставление иллюстраций и комментариев с состояниями автомата

Ввиду того, что при построении визуализатора используется автомат Мили, то будем в рассматриваемом состоянии отображать действия, которые выполняются при переходе из него. При этом серым цветом подсвечивается текущая ячейка таблицы, а черным – рабочие ячейки.

Состояние 0. Описано в предыдущем разделе.

Состояние 1. Заполнение нулями верхней строки таблицы T (рис. 4).

Состояние 2. Заполнение единицами левого столбца таблицы T (рис. 5).



Рис. 4. Состояние 1



Рис. 5. Состояние 2

Состояние 3 является "невизуализируемым", так как оно служебное.

Состояние 4 может отображаться в двух различных вариантах, потому что граф на рис. 2 имеет в этом состоянии две петли, помеченных разными условиями и соответствующими действиями. На рис. 6 приведена иллюстрация, соответствующая нижней

петле, в которой выполняется копирование значения из расположенной выше ячейки таблицы (помечена темным цветом) в ячейку, расположенную ниже (помечена серым цветом), поскольку рассматриваемый в данный момент предмет не может быть использован в наборе. Это имеет место в случае, когда сумма весов предметов для текущего столбца меньше веса предмета, соответствующего текущей строке. На рис. 7 приведена иллюстрация для верхней петли в рассматриваемом состоянии, на которой показано, что имеет место альтернатива в выборе значения, поскольку вес  $T(i, j)$  может быть получен двумя разными способами.

При использовании первого способа текущий предмет включается в набор, а для второго способа рассматриваемый предмет в набор не включается. Наличие единицы в одной из двух черных (рис. 7) ячеек соответствует одному из вариантов. Поскольку на рис. 7 выполняется соотношение  $T(1, 4) = 1$ , то набор с весом 4 может быть получен с применением одного предмета. Поэтому набор с весом 9 может быть получен из двух предметов за счет добавления второго предмета с весом 5. Следовательно,  $T(2, 9) = 1$ .



Рис. 6. Состояние 4 (нижняя петля)



Рис. 7. Состояние 4 (верхняя петля)

На этом прямой ход алгоритма завершен.

Состояние 5 отображает проверку того, что искомый суммарный вес набора предметов может быть получен, поскольку значение в правой нижней ячейке равно единице (рис. 8). Это состояние является промежуточным между прямым и обратным ходом алгоритма, на котором собственно и выполняется поиск результирующего набора.



Рис. 8. Состояние 5

Рассмотрим состояния автомата, соответствующие обратному ходу алгоритма.

Состояния 6 – 8 автомата, образующие замкнутый контур на рис. 2, порождают искомый набор предметов. Опишем, как эти состояния иллюстрируются.

Состояние 6 не изображается.

Состояние 7 отображает альтернативу первой половины шага при обратном ходе алгоритма (рис. 9) из серой ячейки в одну из черных. Вариант перехода в ячейку в том же столбце означает, что текущий предмет не включается в результирующий набор. Переход во вторую черную ячейку означает включение предмета с номером  $i$  в набор. При этом отметим, что собственно переход всегда осуществляется в ячейку, помеченную единицей. В случае если обе ячейки содержат по единице, то выбирается первый из описанных вариантов.

Состояние 8 отображает результат выбора на предыдущем шаге (рис. 10). Из левой части рисунка следует, что предмет с весом 7 включен в результирующий набор.



Рис. 9. Состояние 7



Рис. 10. Состояние 8

Для рассматриваемых исходных данных состояние 9 отсутствует.

Состояние 10 отображает искомый результат – набор предметов, реализующих суммарный вес, равный 16 (рис. 11).

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5	2	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0
3	3	1	0	0	1	1	1	0	1	1	1	0	0	1	0	0	0
7	4	1	0	0	1	1	1	0	1	1	1	1	1	1	0	1	1
6	5	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1

Конец. Искомый набор найден: [7, 5, 4]

>>   < Пауза: 250 >

Рис. 11. Состояние 10

## 12. Архитектура программы визуализатора

Во введении было признано целесообразным реализовать визуализатор на основе паттерна "Модель – Вид – Контроллер".

В данном случае "Модель" – визуализируемые переменные, которые являются глобальными (Globals), "Вид" – формирователь иллюстраций и комментариев (Drawer), а "Контроллер" – автомат (Automaton). Объединение указанных составляющих осуществляется с помощью четвертой компоненты – Applet, которая осуществляет также реализацию панели управления.

Диаграмма классов, реализующая этот паттерн для рассматриваемого примера, приведена на рис. 12.

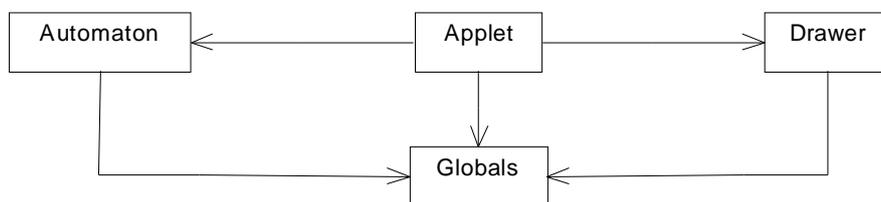


Рис. 12. Диаграмма классов визуализатора

В качестве глобальных переменных используются переменные, описанные в разд. 3. Спроектированный автомат описан в разд. 8, а иллюстрации и комментарии, "привязанные к состояниям автомата" – в разд. 10, 11.

## 13. Программная реализация визуализатора

В работе [11] было предложено переходить формально и изоморфно от графа переходов автомата к его программной реализации на основе оператора *switch*. Для рассматриваемого примера результатом преобразования графа переходов (рис. 2) является текст программы, приведенный ниже:

```

private void makeAutomationStep() {
    switch (state) {
        case 0: // Начальное состояние
            j = 1;
            state = 1;
            break;
    }
}

```

```

case 1: // Заполнение 0-й строки
    if (j <= N) {
        T[0][j] = 0;
        j++;
    } else {
        i = 0;
        state = 2;
    }
    break;

case 2: // Заполнение 0-го столбца
    if (i <= K) {
        T[i][0] = 1;
        i++;
    } else {
        i = 1;
        state = 3;
    }
    break;

case 3: // Переход к следующей строке
    if (i <= K) {
        j = 1;
        state = 4;
    } else {
        sum = N;
        state = 5;
    }
    break;

case 4: // Заполнение очередной ячейки
    if (j <= N && j >= M[i - 1]) {
        T[i][j] = Math.max(T[i - 1][j], T[i - 1][j - M[i - 1]]);
        j++;
    } else if (j <= N && j < M[i - 1]) {
        T[i][j] = T[i - 1][j];
        j++;
    } else {
        i++;
        state = 3;
    }
    break;

case 5: // Конец заполнения таблицы
    // Решение найдено
    if (T[K][sum] == 1) {
        i = K;
        state = 6; // Делать обратный ход
    } else {
        result = false;
        state = 9; // Обратный ход не требуется
    }
    break;

case 6: // Обратный ход
    if (i > 0) {
        state = 7;
    } else {
        result = true;
        state = 10;
    }
    break;

case 7: // Поиск очередного предмета
    if (T[i][sum] != T[i - 1][sum]) {
        positions.add(new Integer(i - 1));
        sum -= M[i - 1];
        state = 8;
    } else {
        state = 8;
    }
}

```

```

        break;

    case 8:
        i--;
        state = 6;
        break;
    }
}

```

Такая реализация автомата резко упрощает построение визуализатора, так как к каждому состоянию автомата (метке *case*) могут быть "привязаны" соответствующие иллюстрации и комментарии. В силу того, что реализация визуализатора выполняется с помощью указанного выше паттерна, "привязка" в данном случае осуществляется с помощью второго оператора *switch*. Таким образом, в визуализаторе используется два оператора *switch*, первый из которых реализует автомат, а второй применяется в формирователе иллюстраций и комментариев.

Отметим, что формальный подход к построению логики визуализатора привел к тому, что отладка логики отсутствовала, а небольших изменений потребовала неформализуемая часть программы, связанная с построением иллюстраций и комментариев.

Полный исходный текст программы визуализатора приведен на сайте <http://is.ifmo.ru/> в разделе "Визуализаторы".

## Заключение

Изложенный метод демонстрирует эффективность стиля программирования от состояний [13], который совместно с объектно-ориентированным программированием позволяет говорить об "объектно-ориентированном программировании с явным выделением состояний" [7].

Отметим, что для рассмотренного примера переход от классически написанной программы к автоматной увеличил размер кода незначительно – всего на 15 – 20 %.

При этом после построения автомата программа визуализатора была написана в течение двух часов, в то время как при традиционном подходе путь от классической программы до визуализатора занимает обычно несколько дней из-за "неформального движения по этому пути". Кроме того, автоматный визуализатор практически не требовал отладки.

Для других классов алгоритмов увеличение объема кода автоматной реализации по сравнению с классической может достигать нескольких раз. Однако и в этом случае время написания визуализатора на основе предлагаемого подхода остается крайне незначительным при простой отладке.

Описанный подход может быть усовершенствован. Во-первых, вместо автоматов Мили могут применяться автоматы Мура. При этом в общем случае число состояний увеличивается, однако визуализация действий в вершинах является более естественной, чем на переходах. Во-вторых, этапы технологии, связанные с проектированием и реализацией, могут быть автоматизированы.

Анализ Интернет-ресурсов в области построения визуализаторов [14–17], а также опыт их разработки и приема у студентов в рамках учебного процесса в 1999/2000 гг. на кафедре "Компьютерные технологии" СПбГУ ИТМО позволяют утверждать, что предложенный подход является "прорывом" в области построения визуализаторов.

## Литература

1. Интернет-школа программирования. <http://ips.ifmo.ru/>
2. Казаков М.А., Мельничук О.П., Парфенов В.Г. Интернет школа программирования в СПбГИТМО (ТУ). Реализация и внедрение // Всероссийская научно-методическая конференция "Телематика'2002". СПб.: 2002. – С.308-309. ([http://tm.ifmo.ru/tm2002/db/doc/get\\_thes.php?id=170](http://tm.ifmo.ru/tm2002/db/doc/get_thes.php?id=170))

3. Столяр С.Е., Осипова Т.Г., Крылов И.П., Столяр С.С. Об инструментальных средствах для курса информатики // II Всероссийская конференция "Компьютеры в образовании". СПб.: 1994.
4. Казаков М.А., Столяр С.Е. Визуализаторы алгоритмов как элемент технологии преподавания дискретной математики и программирования // Международная научно-методическая конференция "Телематика'2000". СПб.: 2000. – С.189-191.
5. Кормен Т., Лайзерсон Ч., Ривест Р. Алгоритмы. Построение и анализ. М.: МЦМНО, 2000. – 960 с.
6. Кнут Д. Искусство программирования. Т. 1: Основные алгоритмы. М.: Вильямс, 2001. - 712 с.
7. Rling, G., Naps, T. L. A Testbed for Pedagogical Requirements in Algorithm Visualizations. // 7th Annual SIGCSE/SIGCUE Conference on Innovation and Technology in Computer Science Education (ITiCSE'02). <http://nibbler.tk.informatik.tu-darmstadt.de/publications/2002/TestbedPedReqinAV.pdf>
8. Шалыто А.А. Технология автоматного программирования // Мир ПК. 2003 № 10. С.74-78. <http://is.ifmo.ru>, раздел "Статьи".
9. Гамма Э., Хэлм Р., Джонсон Р., Влиссидес Дж. Приемы объектно-ориентированного проектирования. Паттерны проектирования. СПб.: Питер, 2001. – 325 с.
10. Шалыто А.А., Туккель Н.И. Преобразование итеративных алгоритмов в автоматные // Программирование. 2002. № 5. – С.12-26. <http://is.ifmo.ru>, раздел "Статьи".
11. Шалыто А.А. SWITCH-технология. Алгоритмизация и программирование задач логического управления. СПб.: Наука, 1998. – 628 с.
12. Седжвик Р. Фундаментальные алгоритмы на C++. Анализ. Структуры данных. Сортировка. Поиск. М.: DiaSoft, 2001. – 687 с.
13. Непейвода Н.Н., Скопин И.Н. Основания программирования. Ижевск: Научно-издат. центр "Регулярная хаотическая динамика", 2003. – 896 с.
14. Complete Collection of Algorithm Animations. <http://www.cs.hope.edu/~algnim/ccaa/>
15. Interactive Data Structure Visualizations <http://www.student.seas.gwu.edu/~idsv/idsv.html>
16. Jawa examples <http://www.cs.duke.edu/csed/jawaa2/examples.html>
17. Sorting Algorithms <http://www.cs.rit.edu/~atk/Java/Sorting/sorting.html>

# СИСТЕМА АВТОМАТИЧЕСКОГО ТЕСТИРОВАНИЯ ПРОГРАММНЫХ РЕШЕНИЙ И ПРОВЕДЕНИЯ СОРЕВНОВАНИЙ В РЕЖИМЕ ON-LINE

М.А. Казаков

Научный руководитель – д.т.н., проф. А.А. Шалыто

Система PCMS2 web (Programming Contest Management System v.2. for web) является частью проекта PCMS2, реализуемого на кафедре компьютерных технологий Санкт-Петербургского государственного университета информационных технологий, механики и оптики (СПбГУ ИТМО). Реализации и концепциям, лежащим в основе этой системы, и посвящена данная статья. Основная часть изложенного материала касается пользовательской модели, используемой в рамках системы, и особенностям поведения системы в рамках Интернет-школы программирования (IPS – Internet Programming School).

## Введение

В настоящее время широкое распространение получили технологии дистанционного обучения (ДО), разработано множество систем ДО в различных областях знания, существует классификация этих систем и методики их создания.

Первая цель настоящего исследования заключается в реализации и внедрении новых идей в рамках системы дистанционного обучения "Интернет-школа программирования". Школа работает на технической базе кафедры компьютерных технологий и пользуется успехом не только у абитуриентов и школьников. Некоторые средние общеобразовательные школы проявили интерес к нашей системе дистанционного обучения и собираются использовать ее в своем учебном процессе.

Вторая цель настоящей работы заключается в разработке автономной системы автоматического проведения олимпиад. Разработанный для реализации системы дистанционного обучения подход оказался достаточно универсальным. Это позволило автору объединить две системы на базе одной – PCMS2 web. Основное отличие данной системы проведения олимпиад от уже существующих в том, что она предоставляет персональный подход к каждому пользователю, что позволяет использовать ее не только в качестве системы управления соревнованиями, но и в качестве обучающей системы для студентов и школьников, желающих приобрести навыки программирования.

Обе упомянутые системы совместно функционируют на технической базе кафедры компьютерных технологий СПбГУ ИТМО и зарекомендовали себя с положительной стороны. В настоящей статье описан единый подход, использованный для построения двух разных систем на базе единого ядра. Единство двух систем обеспечивается единством концепции построения системы PCMS2 web.

## Общие понятия

В настоящее время тема дистанционного обучения и тестирования обсуждается довольно широко. На текущий момент имеется целый ряд методик дистанционного обучения в различных областях науки и технологиях. Прежде чем обратиться к существу вопроса, остановимся на используемых терминах. Многие из них имеют более узкий смысл, чем обычно. Это связано в первую очередь с тем, что данная работа посвящена именно программированию и обучению программированию.

**Дистанционное обучение** – новая организация образовательного процесса, базирующаяся на принципе самостоятельного обучения студента. Среда обучения характеризуется тем, что учащиеся в основном, а часто и совсем, отдалены от преподавателя в пространстве и (или) во времени, в то же время они имеют возможность в любой момент поддерживать диалог с помощью средств телекоммуникации. В рамках данной работы под термином "дистанционное обучение" подразумевается обучение, в котором связь между преподавателем и учеником осуществляется посредством Интернета, и

рассматривается суженный круг систем дистанционного обучения – системы, организованные на базе технологии WWW, если отдельно не отмечено обратное.

**Дистанционное тестирование** – аналогично дистанционному обучению, является новой формой определения уровня накопленных (остаточных) знаний. Характеризуется тем, что тестируемые учащиеся отдалены от преподавателя в пространстве и (или) во времени.

**Интернет школа программирования** (Internet Programming School(IPS)) – это реализация основных концепций дистанционного обучения, совмещенных с накопленным опытом обучения программированию в автономной дистанционной обучающей системе.

**Урок** – в рамках Интернет-школы программирования это набор, состоящий из следующих элементов: текст урока, набор упражнений, набор задач для самостоятельного решения и автоматической проверки, набор визуализаторов.

**Визуализатор** – это программа, иллюстрирующая в нашем случае действие алгоритма на определенные входные данные, в общем случае – действия любого объекта на окружение.

**Автоматическая система проверки** – это система, обеспечивающая проверку исходных кодов программ учащихся без участия человека. При необходимости система может выдавать осмысленные комментарии к ошибкам, допущенным учащимися.

**Олимпиада** – в рамках данной работы это соревнование по программированию, характеризуемое определенными правилами проведения, набором задач, сроками и местом проведения. Правила проведения олимпиады регламентируют порядок регистрации участников, порядок проведения соревнований, правила подсчета результатов. Результатом олимпиады является итоговая таблица результатов, характеризующая некоторый рейтинг участников по определенной шкале. В рамках данной работы рассматриваются олимпиады, проводимые по правилам американской ассоциации ACM (Association for Computing Machinery). Подробные правила, а также другую полезную информацию можно почерпнуть из [10].

**Непосредственное участие в олимпиаде** – присутствие в рамках зала для участников олимпиады. Характеризуется наиболее строгим контролем со стороны жюри олимпиады за ходом участия. При более строгом контроле над соблюдением правил соревнований также характеризуется и более значимыми призами, и другими положительными аспектами для победителей. Обычно используется как антоним дистанционного участия.

**Дистанционное участие в олимпиадах** – участие в олимпиаде посредством средств телекоммуникаций. Обычно характеризуется отсутствием строгого контроля над соблюдением правил соревнований. Очень редко, если не оговорено обратное, такое участие зачитывается как официальный результат. Исключение обычно составляют чисто дистанционные олимпиады, когда непосредственных соревнований не проводится.

За последние пять лет соревнования по программированию стали очень популярны в России. Стоит только отметить, что в последние годы более 300 команд со всего бывшего СССР собираются, чтобы сразиться за возможность поехать на финальные соревнования и получить желанный кубок мира. Понятно, что даже 300 команд по всей России – это всего лишь 900 человек. Но если учесть постоянно растущий интерес к программированию и высоким технологиям среди молодежи, то получается, что требуется иметь достаточно высокую квалификацию, чтобы выйти на уровень хотя бы России. А как быть остальным студентам, у которых нет возможности подготовиться к соревнованиям? К счастью, в настоящее время мировая сеть Интернет доступна почти каждому студенту. Такая возможность позволила развиваться целому направлению – дистанционное проведение олимпиад через Интернет. Отличительная черта таких олимпи-

ад особенно актуальна для России. Участники не должны передвигаться между городами, они могут участвовать в олимпиадах, сидя дома, в институте или в школе, и решать те же задачи, что и остальные участники.

СПбГУ ИТМО уже в течение шести лет проводит полуфинальные соревнования по северо-восточному европейскому региону в рамках командного чемпионата мира по программированию по правилам ACM. Кроме того, налаженный процесс подготовки победителей олимпиад позволил за эти годы накопить большой архив задач и специальных методик подготовки. Одна из важнейших составляющих тренировок – участие в реальных соревнованиях.

В настоящее время на кафедре компьютерных технологий разрабатывается уникальная система проверки задач (PCMS2 – Programming Contest Management System v. 2). Основной задачей этой системы является именно проведение олимпиад. Но в силу своей универсальности автором данной работы совместно с автором и разработчиками системы PCMS2 был разработан специальный интерфейс, позволяющий использовать ядро системы PCMS2 для проверки решений, приходящих через Интернет. Очевидной возможностью, которую открыл данный подход, стало создание системы управления дистанционными соревнованиями. Поскольку система, так же как и Интернет-школа программирования, носит учебный характер, то было решено создать единую инфраструктуру двух систем, чтобы ученики школы могли опробовать свои знания в соревнованиях. Безусловно, предполагается и обратный ход – для участников соревнований, не обладающих необходимыми навыками, предоставляется возможность повысить свою квалификацию в Интернет школе. Единая инфраструктура двух таких систем позволяет свободно пользоваться обеими системами без дублирования информации и функциональности.

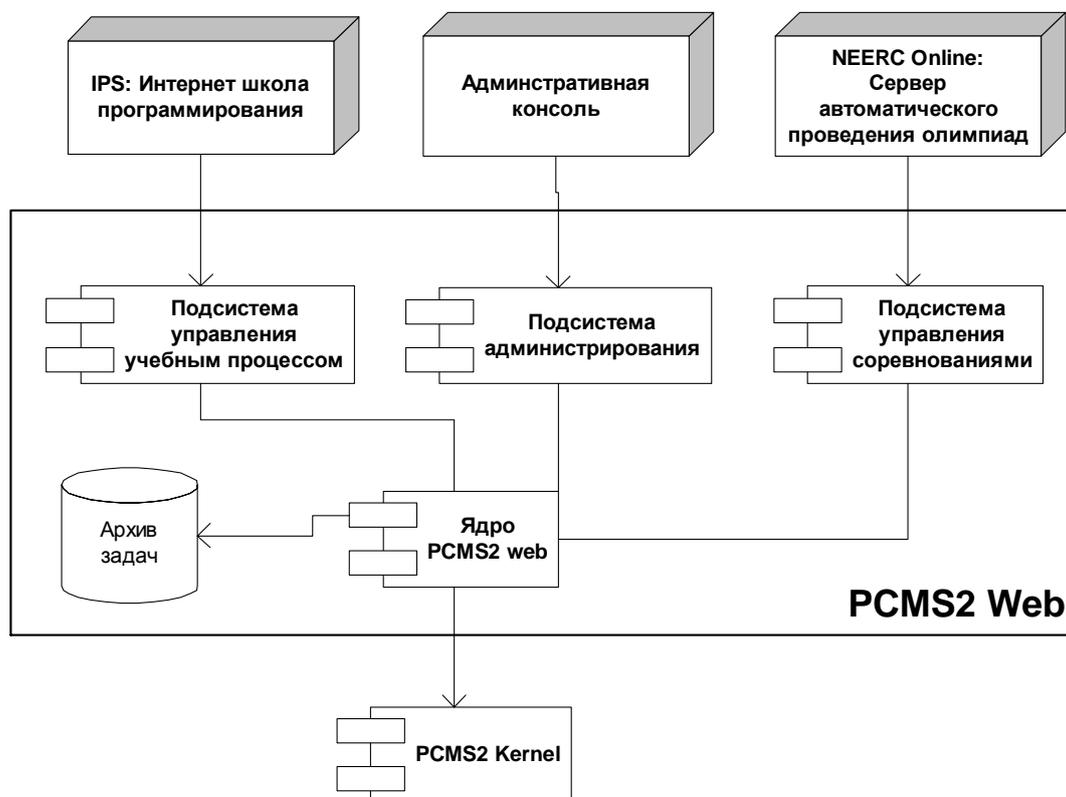


Рис. 1 Общая схема PCMS2 Web

Целью создания такой системы проведения олимпиад является не только проведение олимпиад как таковых, но и предоставление доступа к уникальному архиву задач с возможностью проверить свои способности в решении любой из них. В учебных це-

лях предлагается также участи в "олимпиадах из архива". Это основная отличительная черта данной системы от уже существующих систем. Таким образом, первый взгляд на разрабатываемую систему дает нам приблизительно следующую картину (рис. 1).

### Концепции PCMS2 web и IPS

В этой главе рассмотрим основные понятия и концепции, лежащие в основе как методической, так и технической частей сервера Интернет-школы программирования. Далее рассмотрим основные сведения об олимпиадах по программированию. После изложения необходимых сведений об олимпиадах и автоматическом тестировании уровня знаний учащихся школы мы сформулируем основные концепции системы PCMS2 web, необходимые для ее создания. Иллюстрация на рис. 2 дает общую структуру взаимосвязей между различными концепциями, задействованными при написании данной работы.

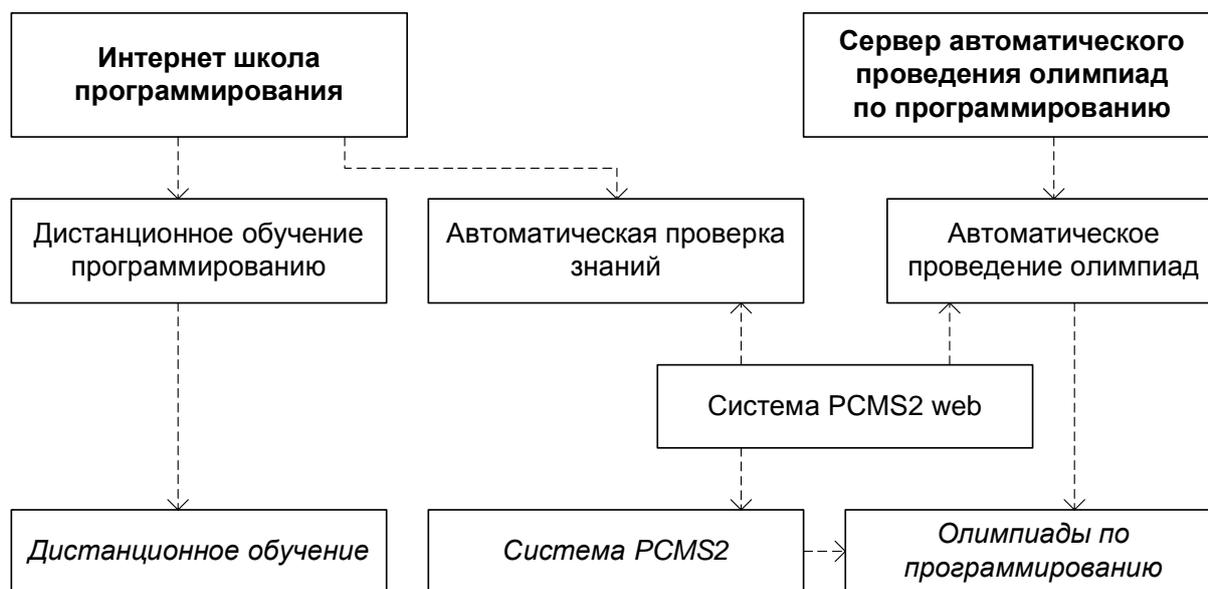


Рис. 2. Структура концептуальных зависимостей

Рассмотрим более подробно зависимости на этой иллюстрации. Как видно из обозначений, "Интернет-школа программирования" и "Сервер автоматического проведения олимпиад по программированию" являются главными концепциями. Они объединяют в себе более частные концепции. Не останавливаясь подробно на каждой из указанных частей, поясним лишь назначение каждой из них:

- **Интернет-школа программирования** объединяет в себе все концепции, заложенные в проект Интернет-школы программирования. Содержит в себе два основных раздела: "Автоматическая проверка знаний" и "Дистанционное обучение программированию".
- **Дистанционное обучение программированию** – это набор идей и концепций, связанных с особенностями дистанционного обучения программированию. Основывается на общих концепциях дистанционного обучения.
- **Автоматическая проверка знаний** – набор идей, лежащих в основе проверки знаний посредством проверки практических навыков программирования.
- **Автоматическое проведение олимпиад** – на базе основной концепции проведения олимпиад по программированию был выведен ряд ограничений и расширений, положивших основу для концепции автоматического проведения олимпиад.

- Система **PCMS2 web** взяла в основу идеи, заложенные в PCMS2. Путем внесения специфики, связанной с тем, что система PCMS2 web работает исключительно на web, были сформулированы основные идеи ее построения.
- Система **PCMS2** – это автоматическая система проведения олимпиад, поэтому основы ее построения положили хорошую базу для создания системы PCMS2 web. Кроме того, поскольку между PCMS2 и PCMS2 web необходимо налаживать общий интерфейс для интеграции, сходство идей, заложенных в основу обеих систем, помогло сделать организацию такого интерфейса максимально простой.

Существует достаточно большое количество источников, описывающих концепции дистанционного обучения. Например, описание концепций дистанционного обучения, которое выходит за рамки настоящей статьи, можно найти в соответствующих источниках ([1, 7]). В рамках данной работы рассмотрим лишь особенности дистанционного обучения программированию, используемых в интернет-школе программирования. Кроме того, не будем описывать концепции, лежащие в основе проведения олимпиад по программированию, поскольку существует множество источников, описывающих цели и задачи олимпиад по программированию, а также их правила. Рекомендуем обратиться к источнику [10], содержащему самую оперативную информацию относительно олимпиад по программированию в рамках правил ACM.

### **Обзор испытаний PCMS2 web**

Система PCMS2 web в настоящее время функционирует в течение полутора лет в автономном режиме. За время работы системы не возникало никаких исключительных ситуаций, приводящих к остановке системы. Те немногие временные сбои, которые появлялись для пользователей, были обусловлены перебоями в подаче питания. Два раза система останавливалась, поскольку на кафедре компьютерных технологий отключали электричество на сроки, превышающие аккумуляторную емкость источников бесперебойного питания. Кроме того, возникновение таких ситуаций привело к решению поставить источники бесперебойного питания не только для серверной части, т.е. непосредственно того персонального компьютера, на котором работают все компоненты PCMS2 web, но и для компьютеров, на которых располагается система PCMS2. Остальные остановки системы обуславливались ошибками в проверочной системе PCMS2. Ошибки в данный момент устранены, и система действительно стала полностью автономной. Но не сразу система стала работать настолько стабильно. Рассмотрим основные этапы тестирования и проверки системы в реальных условиях.

**Полуфинальные и четвертьфинальные соревнования чемпионата мира по программированию 2000 года.** В этом году были проведены испытания системы IPS. Надо отметить, что эти соревнования лишь подтвердили печальный вывод: первая версия системы довольно плохо спроектирована и не может работать в автономном режиме, поскольку сбои случаются постоянно. Она, как и предыдущие системы проверки решений задач, требует постоянного наблюдения со стороны администратора. Единственным преимуществом этой системы была интеграция с Web, что позволяло в самом примитивном виде обеспечить доступ к соревнованиям пользователям, желающим участвовать дистанционно в соревнованиях. Результаты этих испытаний были плачевны. Аналогичные выводы были сделаны и по поводу используемой системы автоматического тестирования APPEP и ее приложения для проведения локальных соревнований PCMS. По результатам этих испытаний было решено объединить усилия разработчиков систем для создания архитектуры единой системы, объединяющей в себе функциональность всех предыдущих систем.

**Полуфинальные и четвертьфинальные соревнования чемпионата мира по программированию 2001 года.** К этим соревнованиям был готов работающий прототип системы PCMS2 и системы PCMS2 web. К сожалению, не удалось добиться созда-

ния единой системы, и было решено разбить ее на две тесно связанные части. Слишком велики оказались различия бизнес-моделей Интернет-школы программирования и дистанционных соревнований и олимпиад, проводимых непосредственно в локальной сети. Итак, первая часть выполняла действия по проведению соревнований через локальную сеть, а вторая система (PCMS2 web) организовывала соревнования через Интернет. Надо признать, что сходство в архитектуре этих двух систем все-таки были достаточно высоко для того, чтобы создать интерфейс взаимодействия между PCMS2 web и ядром системы PCMS2. Это позволило проводить одновременно соревнования через Интернет и локально на одних задачах и в одно и то же время, формируя общие таблицы результатов. Таким образом, была реализована идея проведения соревнований через Интернет. Система PCMS2 web показала себя с лучшей стороны. За время соревнований она работала достаточно стабильно, и проблем у пользователей не возникало. Таким образом, была реализована первая часть функциональности PCMS2 web. Надо отметить, что, несмотря на то, что функциональность по обслуживанию Интернет-школы программирования имела больший приоритет, было решено не упускать возможности испытать систему в стрессовых условиях, поэтому реализация IPS была отложена на более поздний срок.

**Работа в автономном режиме.** Сразу по окончании соревнований в течение декабря 2001 года была завершена реализация интерфейса PCMS2 web для Интернет-школы программирования. Это позволило 25 декабря произвести пробный запуск школы, а к концу января запустить школу в автономном режиме. Таким образом, к началу февраля 2002 года была выделена техника для размещения PCMS2 (для функционирования этой системы с точки зрения безопасности необходимо иметь не менее 2 персональных компьютеров), и Интернет-школа программирования начала работать. К этому же времени была доработана система автоматического проведения олимпиад. Она содержала ряд мелких ошибок. Это было обусловлено тем, что в феврале 2002 года начались тренировки сборной команды ИТМО перед поездкой на финальные соревнования чемпионата мира по программированию. Тренировки могли послужить идеальным испытанием для реальных возможностей системы. Система оправдала возложенные на нее надежды. Многие тренировки работали в режиме реального времени в рамках системы PCMS2 web. Часть тренировок была выложена чуть позднее, но даже в таких тренировках пользователи могли участвовать в режиме эмуляции и соревноваться с реальными командами, чьи результаты имеются в PCMS2 web. Все тренировки выложены и в настоящее время и открыты для участия. Однако следует отметить, что, поскольку основную часть пользователей составляли команды и студенты, заинтересованные в участии в олимпиадах по программированию, то после финала чемпионата мира популярность сервера несколько упала.

Итак, что же произошло со школой в период автономной работы системы? Школа постоянно функционирует и пользуется популярностью. Конечно, надо отметить, что, несмотря на то, что пользователей, работающих в Интернет-школе программирования больше, чем участвующих в олимпиадах, участники олимпиад являются более активной частью аудитории. Это объясняется в первую очередь тем, что участие в работе Интернет-школы – это процесс обучения, который является более размеренным, в то время как участие в олимпиадах является действием динамичным, тем более, что участники соревнований ограничены во времени при решении задач.

Добавим еще несколько слов о полуфинальных соревнованиях в рамках чемпионата мира по программированию в 2002 и 2003 году. В эти годы система PCMS2 web сыграла очень важную роль в проведении соревнований. В 2002 году было принято решение апробировать систему на четвертьфинальных соревнованиях в Узбекистане. Во время этих соревнований все участники одновременно начали соревнования на сервере в СПбГУ ИТМО. Соревнования прошли успешно, и уже в 2003 году проводилось

одновременно два четвертьфинальных этапа чемпионата мира – в Узбекистане и в Киргизстане. Обе олимпиады проводились одновременно на базе системы PCMS2 web.

В заключение раздела приведем интересную статистику системы за время работы. Отметим, что, поскольку система едина, приводится сводная статистика по двум системам.

- Количество зарегистрированных пользователей – около 2800. Из них около 20% составляет активная аудитория. Под активной аудиторией мы подразумеваем тех пользователей, которые решили как минимум одну задачу полностью и сделали не менее 10 подходов.
- Количество опубликованных задач – около 350.
- Количество опубликованных соревнований – 85.
- Количество подходов – 20000, из них 4000 являются правильными решениями задач.

### Перспективы развития

Итак, система заработала. А что же дальше? Остановимся на дальнейшей деятельности команды, которую уже начали собирать, в рамках разработки системы PCMS2 web.

**Административная консоль** является новым продуктом и только вошла в стадию разработки. Эта новая подсистема будет содержать функциональность для удобной конфигурации системы, пополнения набора задач, просмотра статистики и так далее. Функциональность этой подсистемы почти не ограничена. В настоящее время реализована лишь базовая функциональность этой подсистемы.

**Рабочее место куратора.** Эта новая функциональность будет реализована в рамках подсистемы Административной консоли и будет являться консолью преподавателя Интернет-школы программирования с ограниченными правами. В функциональность этой консоли будет входить следующее:

- обратная связь с учениками интернет школы программирования;
- формирование учебных классов и просмотр статистики и другой полезной информации, контроль над успеваемостью;
- проведение тестирований для закрытой группы учащихся;
- ручная проверка решений. Эта функциональность, на первый взгляд, кажется шагом назад, однако в образовательных целях очень полезно иногда посмотреть на исходный код и явно указать на ошибку в алгоритме.

Сформировав такое рабочее место, мы сможем реализовать совместные проекты по проведению факультативных занятий для школьников с различными общеобразовательными школами. Уже существуют школы, проявившие интерес к нашей Интернет-школе, и мы готовы идти им навстречу.

**Проведение закрытых соревнований.** В настоящее время пользователи могут участвовать в соревнованиях лишь в общем потоке, т.е. таблица результатов является сводной по все пользователям. Это вызывает нарекания у пользователей и стимулирует не совсем чистых пользователей создавать множество сессий для "загрязнения" таблицы результатов. Новая функциональность будет позволять пользователю создавать персональные соревнования и предоставлять доступ к ним только определенной группе лиц. Это позволит использовать наш сервер как сервер для проведения тренировок или других внутренних соревнований. В свою очередь, мы можем предоставить сервис, схожий с тем, что предоставляет система [11], т.е. по присланным конфигурациям задач проводить соревнования, в обмен на этом мы будем размещать задачи в нашем архиве задач.

Безусловно, предполагается создание и развития раздела архива задач, где каждый желающий сможет решить любую задачу из архива, не открывая олимпиадную сессию. Архив задач будет снабжен системой поиска и другими полезными возможностями.

**Белые страницы.** Этот раздел системы будет позволять пользователям не только персонально работать с системой, но и общаться, обмениваться идеями, больше узнавать друг друга. Такой подход к расширению возможностей в общении может потенциально привести к формированию новой группы общения по интересам. Такая тенденция скажется очень положительно – ведь сервер NEERC является популярным среди студентов-программистов.

Кроме перечисленных выше глобальных возможностей, безусловно, существуют недочеты и другие небольшие ошибки. Все эти ошибки требуют исправления. Поэтому разработка и совершенствование системы требует еще довольно длительного времени и ресурсов. В связи с этим надеемся в ближайший год привлечь к работе над системой еще нескольких студентов с младших курсов.

## Заключение

В заключении отметим основные полученные результаты. Задачей работы была разработка, реализация и внедрение новой системы PCMS2 web с целью создания автоматической системы тестирования практических навыков в программировании в рамках Интернет-школы программирования. Второй задачей работы была разработка сервера автоматического проведения олимпиад на базе той же системы PCMS2 web. Эти задачи решены. В процессе разработки производились постоянные испытания системы и апробация методик дистанционного обучения как на абитуриентах, так и на студентах кафедры КТ разных уровней подготовленности. Методики показали себя с самой лучшей стороны. Система получила популярность, особенно в период подготовки команд к финальным соревнованиям по программированию 2002 и 2003 года. За время работы система показала надежность, стабильность. Удобство работы с сервером автоматического проведения олимпиад и с Интернет-школой программирования подтверждается многочисленными положительными отзывами пользователей.

Автор не рассматривает работу как полностью завершенную. Существуют большие перспективы развития системы PCMS2 web как со стороны олимпиад, так и со стороны дистанционного обучения. Поскольку технология обучения через образование является перспективной, то данная работа может служить хорошей базой для проработки большого числа передовых идей.

Работа в этой области продолжается и в настоящее время. Поскольку количество уже имеющихся сейчас в проработке идей достаточно велико, то предполагается и дальнейшее расширение, и усовершенствование системы PCMS2 web, а вместе с ней и серверов, работающих на ее основе.

## Литература

1. Васильев В.Н., Елизаров Р.А., Парфенов В.Г., Столяр С.Е. Организация дистанционного обучения программистов. // Телематика'98. Всероссийская научно-методическая конференция. - Санкт-Петербург, 8-11 июня 1998 г.: Тез. докл. - С. 172-173.
2. Васильев В.Н., Казаков М.А., Парфенов В.Г., Столяр С.Е. Проблемы подготовки профессиональных программистов и дистанционное обучение в Интернет-школе СПбГИТМО(ТУ). // Дистанционное обучение. Проблемы и перспективы взаимодействия вузов Санкт-Петербурга с регионами России. Материалы II межрегиональной научно-практической конференции. – СПб, 15-18 марта, 1999. - С.45-48.
3. Казаков М.А., Мельничук О.П., Парфенов В.Г. Интернет школа программирования в СПбГИТМО(ТУ). Реализация и внедрение. // Телематика 2002. Всероссийская научно-методическая конференция. – Санкт-Петербург, 3-6 июня 2002 г.: Тез. докл. - С.308-309.

4. Казаков М.А., Осипова Т.Г., Парфенов В.Г., Столяр С.Е. Интернет-школа программирования в СПбГИТМО. // Телематика'99. Всероссийская научно-методическая конференция. - Санкт-Петербург, 7-10 июня 1999 г.: Тез. докл. - С.165-166.
5. Казаков М.А., Столяр С.Е. Визуализаторы алгоритмов как элемент технологии преподавания дискретной математики и программирования. // Телематика'2000. Международная научно-методическая конференция. - Санкт-Петербург, 29 мая - 1 июня 2000 г.: Тез. докл. - С.189-191.
6. Командный чемпионат мира по программированию ACM 2001/2002. Северо-Восточный Европейский регион / Под ред. Проф. В.Н. Васильева и проф. В.Г. Парфенова – Санкт-Петербург: СПбГИТМО(ТУ), 2001. – 144 с.
7. Подготовка и проведение учебных курсов в заочно-дистанционной форме обучения. / Под редакцией профессора И.А. ЦИКИНА. – Изд-во СПбГТУ, 2000
8. Столяр С.Е. Интернет-школа программирования в СПбГИТМО: позади 1 семестр.// Компьютерные технологии в образовании, 1999, 1-2.
9. Столяр С.Е., Казаков М.А. Подготовка тестов как часть технологии автоматизированного тестирования. // XXX науч.-техн. конф. проф.- преподавательского состава 25-28 января 1999 года. Тез. докл. - СПб: СПбГИТМО(ТУ), 1999. - С.105.
10. The ACM-ICPC International Collegiate Programming Contest Web Site sponsored by IBM (<http://icpc.baylor.edu/>)
11. Valladolid Online Judge - Contest Hosting Service (<http://acm.uva.es/contest/>)

# АВТОМАТИЧЕСКИЙ ГЕНЕРАТОР ТЕСТОВЫХ ЗАДАНИЙ

С.В. Степанюк

Научный руководитель – к.ф.-м.н, доц. А.В. Норин

Изучается проблема компьютерного моделирования задач математических тестов. Разработан метод синтеза тестовых заданий. На основе этого метода построены открытые тесты по математике с автоматической генерацией вопросов.

## Введение

Несколько лет назад началась очередная реформа образования. Многие высшие учебные заведения, а также школы начали опробовать новый вид контроля знаний: тестирование. Первые же попытки выявили ряд трудностей: реализация удобного интерфейса, возможность угадать ответ, большой объем методической работы, сложность технического представления и т.д.

После ряда трансформаций наиболее широкое распространение получили два типа тестов: открытые (ОТ) и закрытые тесты (ЗТ). В ОТ ответ нужно вводить, а в ЗТ – выбирать из предлагаемых вариантов [1]. Каждому виду тестов присущи свои достоинства и недостатки. Основным недостатком ОТ является анализатор поля ввода ответов. Устранить этот недостаток без потерь в методической части заданий на сегодняшний день нельзя. Преимущества здесь имеют только такие дисциплины, как физика и математика, так как вопросы по ним имеют однозначные ответы (зачастую числовые). ЗТ обладают рядом других недостатков:

- вероятность, что испытуемый угадает ответ, относительно большая (15–25 %);
- задания в группе тестирующихся могут повторяться.

Конечно, увеличение количества вариантов одной задачи, а также составление правдоподобных ответов уменьшает эти недостатки. К сожалению, следствием этого является большой объем методической и технической работы. Одним из путей решения вышеуказанных противоречий может стать автоматическая генерация тестовых заданий. Возможность переложить часть работы на машину существенно уменьшит время разработки теста.

Вышеуказанный подход был реализован несколькими университетами и центрами тестирования [2,3]. Системы тестирования с генерацией вопросов успешно применяются в учебном процессе. Тесты КубГТУ доступны в сети Интернет по адресу [www.kubstu.ru](http://www.kubstu.ru). Недостатком существующих разработок является генерация вопросов до момента тестирования. Необходимо хранить множество заданий, заботится о их защите от компьютерного взлома и т.д. Также используется закрытая форма тестов, что уменьшает значение генерации вопросов.

Рассмотрим указанную ситуацию на примере конкретной дисциплины математики. Благодаря своей жесткой логической структуре она позволяет сравнительно легко осуществить не только генерацию заданий, но и создание на ее основе открытых тестов. Связано это с тем, что в большинстве задач ответом будет число. Также для математики может быть осуществлена и еще одна идея – использование при вычислениях независимых программных продуктов. Так, задачи по математике можно решать, не разрабатывая решения на языках программирования, а с помощью уже созданных математических пакетов (MathCad, Maple, Математика). Это не только облегчает программирование теста, но и существенно повышает уровень сложности разрабатываемых заданий.

Нами была реализована на практике система автоматической генерации тестовых заданий с построением на ее основе открытых тестов по математике. Часть вычислений производилось с помощью пакета MathCad 8.0.

## Метод синтеза тестовых вопросов

Ограничимся рассмотрением в качестве предметной области только математики. Первая задача, которую следует решить, – это отбор задач для генерирования. Останемся только на тех задачах, ответы в которых являются числами. Из предполагаемого раздела дисциплины выбираются основные задачи, характерные именно для него. Задачи следует отбирать так, чтобы они наиболее полно отражали структуру курса.

Далее проводится вывод общего вида задачи. Происходит это путем использования набора параметров. Задача представляется с помощью неопределенных коэффициентов. Далее в соответствии с алгоритмом решения той или иной задачи вычисляется ответ (также в общем виде). Таким образом, задача подготовлена для последующей генерации.

Необходимо позаботиться о значениях, которые могут принимать параметры. Диапазон и сами значения необходимо вычисляются заранее. Связано это с со следующими обстоятельствами:

- для некоторых задач есть ограничения на возможные значения параметров;
- часть коэффициентов может быть связана различными условиями;
- диапазон значений также не может быть неограниченным.

Приведем конкретный пример. Пусть по условию задачи необходимо решить квадратное уравнение с целыми, действительными корнями. Общий вид задачи выглядит следующим образом:  $ax^2 + bx + c = 0$ . Мы имеем три неопределенных коэффициента. При подготовке задачи к генерации изменим ее общий вид на более пригодный:  $(dx + m) \cdot (kx + n) = 0$ . Если перемножить содержимое скобок, то мы возвратимся к изначальному виду задачи. Этот вид более оптимизирован для синтеза вопросов. Варьируя коэффициенты  $m$  и  $n$ , мы получаем действительные корни. Нет необходимости в дополнительных проверках: наличия положительного дискриминанта. При условии, что параметры  $d$  и  $k$  принимают единичные значения, получаем в задачах целочисленные решения.

Отдельно следует рассмотреть диапазоны значений параметров. Конечно, все параметры должны быть целыми числами. В противном случае мы не получим целых корней в решении уравнений. Также необходимо учитывать некоторую специфику задачи. Коэффициент  $c$  есть перемножение параметров  $m$  и  $n$ , следовательно, при больших значениях  $m$  и  $n$  мы столкнемся с проблемой несоответствия задач одному уровню сложности. Некоторые вопросы будут требовать больше времени на их решения только из-за громоздкости расчетов. И действительно, при  $m$  и  $n$ , равных 9 мы получаем  $c = 81$ . Если мы проводим тестирование в 8–9 классе школы, то появление значений параметров больше 100 нежелательно: трудности в решении задач будут связаны не с разделом, по которому проводится тестирование, а с вычислениями. Таким образом, оптимальным диапазоном будут принимаемые значения в пределах от  $-9$  до  $9$ .

Нулевое значение коэффициента нас не удовлетворяет (оно изменяет общий вид задачи), и оно отбрасывается. Таким образом, для данной задачи мы варьируем два параметра. Каждый параметр может принимать 18 значений. Вычислим общее количество возможных задач по формуле (из теории вероятностей)

$$N = m_1 \cdot m_2 \cdot \dots \cdot m_k$$

Здесь  $N$  – общее количество вопросов,  $m_1, m_2, \dots, m_k$  – число значений, которые может принимать каждый параметр. Подставляя значения для каждого из параметров, получим:

$$N = 18 \cdot 18 = 324.$$

Таким образом, при использовании двух параметров можно получить 324 различные задачи. Соответственно, при использовании большего количества параметров ста-

новится возможным увеличение количества возможных задач до тысяч и даже миллионов. По сравнению с обычными тестами, в которых только 30–40 различных вариантов одной задачи, тесты-генераторы обладают громадным потенциалом заданий.

### Описание технической части разработки

Пакет тестирующих программ был разработан в среде Delphi 3.0 и представляет собой гибкую систему по созданию и использованию тестов. Это становится возможным благодаря применению вычислительных мощностей математического пакета MathCad.

Выбор среды программирования был обусловлен следующим:

- широкие возможности для визуального программирования;
- относительная легкость создания небольших программ (в частности – тестов);
- удобство интерфейса;
- поддержка OLE (Object Linking and Embedding).

Пакет MathCad был выбран по следующим причинам:

- дружелюбный интерфейс;
- поддержка OLE;
- наличие команд управления средой.

Здесь необходимо отметить, что такие пакеты, как Maple или "Математика" имеют большие вычислительные мощности. Однако необходимо учесть некоторую специфику применения пакета: обязательна поддержка OLE, наличие команд управления пакетом. Из-за подобных ограничений нет возможности использовать абсолютно любой программный продукт.

Процесс технической реализации начинается с того, что на форму теста необходимо поместить специальный объект OleContainer. Это компонент, позволяющий обмениваться данными между независимыми приложениями. Список программных продуктов, которые могут быть подключены в том или ином виде, предоставляется операционной системой. Все эти объекты установлены на данном компьютере. В него входят все пакеты, поддерживающие технологию OLE. При нажатие правой кнопкой мыши на OleContainer появляется меню, предлагающее:

- Edit (начать редактировать уже существующий объект);
- Open (открыть зарегистрированный объект);
- Insert Object (Добавить новый объект);
- и т.д.

После выбора пункта Insert Object появляется окно со списком объектов, которые могут быть подключены к данной программе (рис. 1).

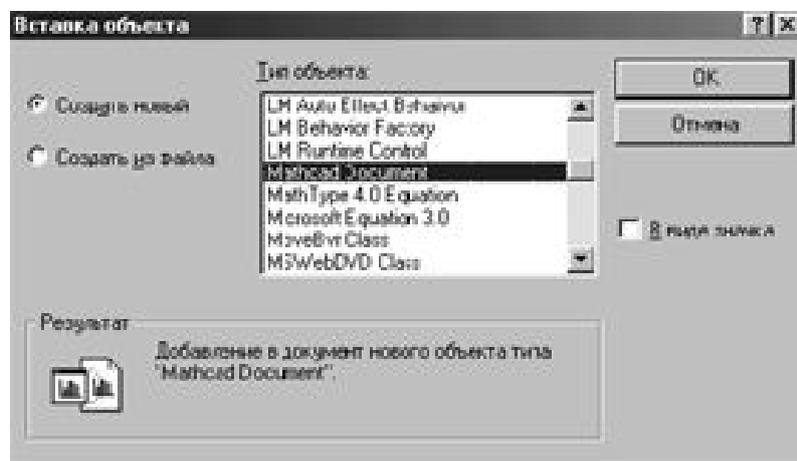


Рис. 1. Окно со списком подключаемых объектов

Из предлагаемых объектов выбираем MathCad Document и нажимаем ОК. Затем загружается MathCad, в котором открывается новый документ. С этого момента начинается методическая разработка задания: в появившемся окне Вы программируете имеющимися средствами задачу. На рис 2. предлагается конкретный пример одного из вопросов. Представлена задача из раздела "Пределы" – неопределенность вида единица в степени бесконечность.

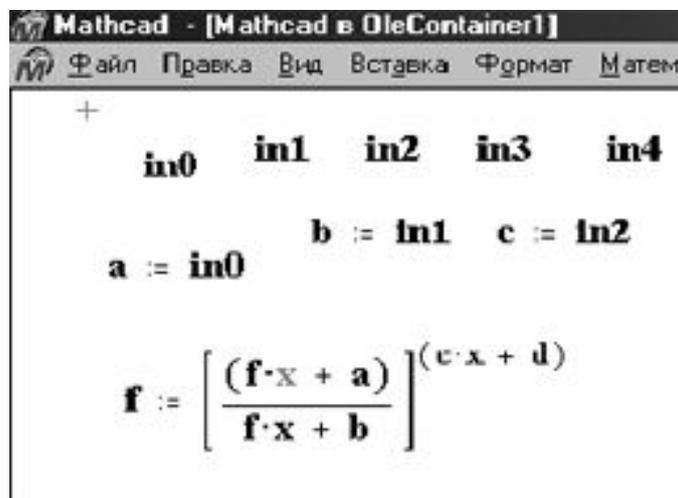


Рис. 2. Пример вопросов

Рассмотрим подробнее структуру документа. В начале любого подобного документа должны стоять входные переменные (входные параметры) in0, in1, ... и т.д. Это зарегистрированные системой переменные. Всего их только 10 (с нулевой по девятую). Через каждую может передаваться в MathCad комплексное число. Далее для удобства эти переменные переобозначаются через буквенные параметры: a, b, c, и т.д. Затем выписывается решение задачи в общем виде. Полученный ответ записывается в переменные вида out0, out1 ... out9. Каждая выходная переменная также может содержать комплексное число.

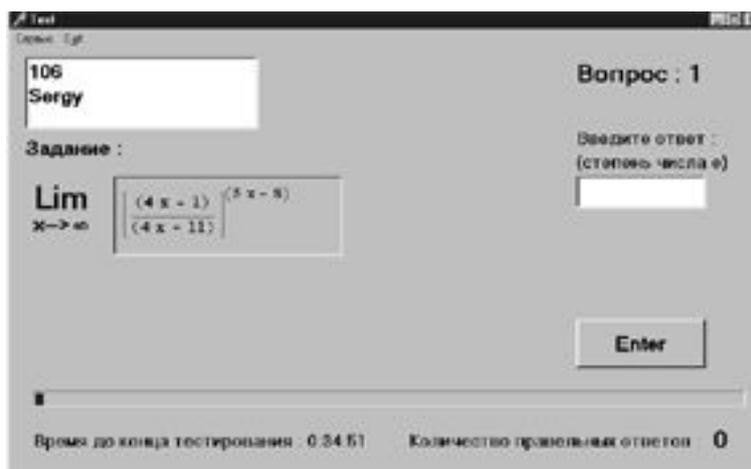


Рис. 3. Внешний вид теста с задачей

После завершения программирования MathCad документа необходимо его сохранить. Для этого нажимаем Menu и выбираем "Сохранить в OleContainer". Закрываем MathCad и приступаем к обработке OleContainer. Чтобы переслать данные в OleContainer, используется команда SetComplex (in0, param1, param2). Данная команда переда-

ет комплексное число  $\text{param1}+i*\text{param2}$  в переменную  $\text{in0}$ . После передачи параметров посылаем команду `Recalculate`, при помощи которой пересчитывается документ. Чтобы получить из `MathCad` документа ответ, воспользуемся командой `GetComplex (out0, param1, param2)`. Команда считывает значение переменной  $\text{out0}$  и записывает его в переменные  $\text{param1}$ ,  $\text{param2}$ .  $\text{Param1}$  содержит действительную часть ответа, а  $\text{param2}$  – комплексную. Также можно использовать `MathCad` и для вывода задачи на экран. Внешний вид теста с задачей представлен на рис. 3.

Итак, осуществляется взаимодействие приложений (Тест – `MathCad`). Тест пересылает начальные данные в `MathCad` (в заранее запрограммированную задачу), где они обчисляются и затем, получив конкретный ответ, пересылаются обратно, а представление задач через набор параметров позволяет автоматически генерировать их условие.

## Описание методической части разработки

В настоящее время нами разработаны следующие разделы:

### 1. "Пределы"

- неопределенность вида бесконечность/бесконечность (многочлены одинаковых степеней при условии, что аргумент стремится к бесконечности)

$$\lim_{x \rightarrow \infty} \frac{a \cdot x^n + b \cdot x^{n-1} + \dots + c}{a_1 \cdot x^n + b_1 \cdot x^{n-1} + \dots + c_1};$$

- неопределенность вида бесконечность/бесконечность (многочлены разных степеней

при условии, что аргумент стремится к бесконечности)  $\lim_{x \rightarrow \infty} \frac{a \cdot x^m + b \cdot x^{m-1} + \dots + c}{a_1 \cdot x^n + b_1 \cdot x^{n-1} + \dots + c_1};$

- неопределенность вида 0/0 (аргумент стремится к корню числителя и знаменателя (многочлены третьей степени с действительными корнями))

$$\lim_{x \rightarrow k} \frac{ax^3 + bx^2 + cx + d}{a_1x^3 + b_1x^2 + c_1x + d_1};$$

- неопределенность вида 0/0 (бесконечно малая функция в числителе)

$$\lim_{x \rightarrow 0} \frac{f(x)}{x}, \quad f(x) \xrightarrow{x \rightarrow 0} 0;$$

- неопределенность вида 0/0 (отношение бесконечно малых функций)

$$\lim_{x \rightarrow 0} \frac{f(x)}{g(x)}, \quad f(x) \xrightarrow{x \rightarrow 0} 0, \quad g(x) \xrightarrow{x \rightarrow 0} 0;$$

- неопределенность вида разность бесконечностей

$$\lim_{x \rightarrow \infty} \left( \sqrt{ax^2 + bx + c} - \sqrt{a_1x^2 + b_1x + c_1} \right);$$

- неопределенность вида 0/0  $\lim_{x \rightarrow c} \frac{\sqrt{ax+b-d}}{x-c};$

- неопределенность вида 0/0 (в знаменателе бесконечно малая функция)

$$\lim_{x \rightarrow 0} \frac{\sqrt{ax+b-d}}{f(x)}, \quad f(x) \xrightarrow{x \rightarrow 0} 0;$$

- неопределенность вида 0/0:  $\lim_{x \rightarrow 0} \frac{\sqrt[n]{(x+1)^{n^2}} - \sqrt[n^3]{(x+1)^{n^4}}}{x^{n^5}};$

- неопределенность вида бесконечность / бесконечность  $\lim_{x \rightarrow \infty} \frac{\sqrt[n^1]{x^{n^2} + a} \pm \sqrt[n^3]{x^{n^4} + b}}{\sqrt[n^5]{x^{n^6} + c} \pm \sqrt[n^7]{x^{n^8} + d}};$

- неопределенность вида единица в степени бесконечность  $\lim_{x \rightarrow \infty} \left( \frac{ax+b}{ax+c} \right)^{dx+h}$ .

## 2. "Точки разрыва"

- определить и классифицировать точки разрыва;
- определить значения параметров, при которых функция непрерывна.

## 3. "Производные"

- производная от суммы двух функций;
- производная от произведения двух функций;
- производная от частного двух функций;
- производная от сложной функции.

Каждая задача из раздела "Производные" имеет достаточно индивидуальный характер, и их представление в общем виде имеет не только технические, но и методические трудности. В связи с этим обстоятельством представлен только базовый уровень задач. Также необходимо отметить, что из-за сложности обработки символьных результатов все значения производных вычисляются в точке.

## 4. "Алгебра"

- решить систему методом Крамера (система 3-х уравнений);
- вычислить определитель 4-го порядка;
- действия с матрицами;
- определить ранг матрицы;
- решить систему методом Гаусса (система 4-х уравнений);
- вычислить обратную матрицу;
- определить собственные числа матрицы;
- вычислить алгебраическое дополнение к определителю 4-го порядка;
- разложить заданный вектор по базису;
- определить матрицу перехода к новому базису.

Материалы соответствуют программе первого курса первого семестра технического вуза. Разработанные задачи наиболее полно представляют программу первого семестра по математическому анализу и алгебре.

## Заключение

В работе исследован метод синтеза тестовых заданий. На его основе разработаны открытые тесты по математике с автоматической генерацией вопросов в процессе тестирования. Также рассмотрено применение математических пакетов для расчета и вывода задач на экран.

## Литература

1. Аванесов В.С. Форма тестовых заданий. – М., 1991.
2. Грушевский С.П., Лесниченко Н.В., Миронов Я.П., Усатиков С.В., Федоренко Д.Н. Автоматизированная система генерации тестовых заданий по курсу высшей математике // Педагогические нововведения в высшей школе, часть 2: Информатизация образования и вопросы экранной культуры. Краснодар: КубГТУ, 1998. С.41–42.
3. Солдатова Л., Шевченко И. Автоматическая генерация задач. // Тезисы докладов международной научно-производственной конференции "Информационные технологии в образовании, управлении и промышленности", ч.1.-Комсомольск-на-Амуре, Гос. Техн. Ун-т, 1996, с. 58-60.

# **ПОДХОДЫ К СОЗДАНИЮ И ПЕРЕНОСУ ЭЛЕКТРОННЫХ УЧЕБНЫХ ПОСОБИЙ НА КАРМАННЫЕ ПЕРСОНАЛЬНЫЕ КОМПЬЮТЕРЫ**

**Р.В. Койнов**

**Научный руководитель – д.т.н., проф. С.К. Стафеев**

Рассматривается возможность использования карманных персональных компьютеров в системах дистанционного обучения. Проведен анализ существующих образовательных ресурсов и программного обеспечения, которое можно использовать для создания новых учебных пособий. Создан инструментарий, позволяющий создавать новые и переносить уже имеющиеся образовательные ресурсы на платформу карманных персональных компьютеров.

## **Введение**

В силу известных социально-экономических условий в настоящее время в России процесс информатизации образования развивается явно недостаточными темпами. Однако можно утверждать, что даже полное оснащение системы образования персональными компьютерами не решает проблемы, так как не позволяет преподавателям, студентам и администрации иметь доступ к информации в удобное время и в удобном месте. Это утверждение относится и к системам дистанционного обучения (СДО). В то же время получение образования в удобном месте в удобное время и темпе является основным принципом и своего рода девизом СДО.

Задача развития и становления систем СДО в полной мере может быть решена с помощью применения карманных персональных компьютеров (КПК). КПК – это компьютеры, предназначенные для личного использования в различных целях, выполненные на базе микропроцессоров, размеры и вес которых позволяет без вреда для здоровья транспортировать и работать на них.

На сегодняшний день образовательные ресурсы для КПК представлены довольно небольшой группой различных научных калькуляторов, сборников констант, электронных справочников. В силу особенностей построения учебного материала – использование таблиц, формул, рисунков – меньше всего ресурсов создано по предметам естественнонаучного цикла (физика, математика и т.д.). В настоящей работе рассматривается возможность создания инструментария, который позволит восполнить недостаток образовательных программ для КПК путем создания новых и переноса уже имеющихся ресурсов СДО. Производится сравнительный анализ имеющихся программ и технологий, которые могут быть использованы для работы с созданными ресурсами.

## **Карманные персональные компьютеры**

Современные карманные персональные компьютеры представляют практически все возможности компьютера класса "ноутбук", при этом стоят значительно дешевле. В таких ПК за счет упрощений, связанных с заменой "мыши" на сенсорный дисплей, отказа от электромеханических винчестеров, дисководов, CD-ROM'ов, а также в результате увеличения памяти ПЗУ до 8–16 Мбайт и ОЗУ до 6–32 Мбайт удается реализовать все основные возможности ОС (текстовый редактор, электронные таблицы, переводчики, дневники, калькулятор, СУБД, ГИС, демонстрационные средства, приложения и устройства в части математической обработки и графики, для работы в сети Internet, спутниковой радио- и пейджинговой связи и т.п.). Кроме того, удается обеспечить совместимость с ОС Windows настольных ПК и интерактивный обмен файлами с ними и периферией (принтерами, сканерами, модемами, цифровым фото и т.д.) – последовательный проводной или через инфракрасные порты. Мини-компьютеры КПК-технологий снабжаются вставляемыми в них флеш-картами емкостью 8–128 Мбайт, способными нести полную библиотечную электронную поддержку учащегося по курсу,

ряду дисциплин, дипломному проектированию и т.п. Наряду с сенсорным управлением рабочего стола, в некоторых миниатюрных версиях КПК предусмотрена многоязычная виртуальная клавиатура, иницируемая с дисплея и управляемая пером. Энергопотребление миникомпьютеров КПК-технологий столь невелико, что от двух пальчиковых батареек или аккумуляторов формата АА обеспечивается автономная работа персонального компьютера-помощника в течение 8–20 часов. Первоначальная цена изделия с монохромным дисплеем без дополнительных устройств и программных приложений составляет от 100 до 250 долларов, что делает их приобретение и использование относительно реальным для значительной части учащихся и учреждений образования. Понятно, что цветные многофункциональные КПК стоят дороже – от 300 долларов.

Существующие карты расширения позволяют объединить в единое целое сотовый телефон и карманный компьютер. В результате по сотовой связи можно передавать факсы и данные или обмениваться сообщениями электронной почты через Internet. Фактически это прообраз интегрированного коммуникационно-вычислительного устройства ближайшего будущего, оптимального для обучающегося в системе ДО.

Массовое внедрение КПК в образовательную индустрию должно сопровождаться разработкой соответствующих Soft-приложений, созданием сериалов многомодульных электронных учебников, задачников, справочников, баз данных, а также научно-образовательных методик использования КПК-технологий в учебном процессе на базе централизованного и персонифицированного информационного библиотечного обеспечения технологий образования

Прежде чем приступить к созданию электронных учебных пособий (ЭУП) для КПК, было принято решение произвести изучение соответствующего сегмента рынка, чтобы определить уровень необходимых затрат для создания мультиплатформенного стандарта данных пригодного для использования в качестве основы для ЭУП. В результате произведенных исследований были получены следующие результаты (рис. 1.).

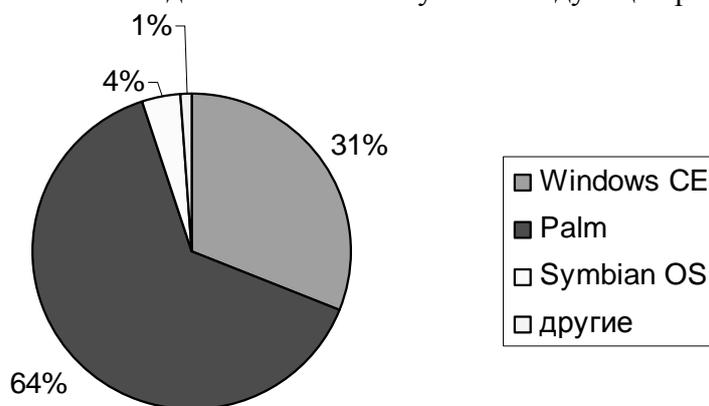


Рис. 1. Рыночные доли ОС на рынке КПК

В результате предварительных рассмотрений было принято решение не разрабатывать новый мультиплатформенный стандарт отображения информации, а создать механизм, который бы позволил с минимальными усилиями использовать один из имеющихся стандартов данных на обеих "главенствующих" ОС (PalmOS и Windows CE и ее более поздние реализации) для создания ЭУП.

Операционная система Windows CE и ее поздние версии являются облегченным вариантом обычной ОС корпорации Microsoft. Т.е. у создателя ЭУП не возникнет никаких проблем при выборе средств реализации, подойдет стандартный, уже имеющийся инструментарий и средства просмотра.

Кардинально другим образом обстоит дело с операционной системой PalmOS (только для версий ниже 5, последние версии этой ОС по своим мультимедиа возмож-

ностям и совместимости приблизились к Windows). Приведем общие сведения об этой ОС.

- PalmOS – однозадачная операционная система.
- Поддерживает графический интерфейс (диалоги, кнопки, списки, шрифты и т.д.).
- Приложения PalmOS – одно-поточные, событийно ориентированные.
- В PalmOS есть своя файловая система.
- Пользователь в PalmOS не работает ни с папками, ни с файлами. Пользователь "видит" только приложения и документы, с которыми работает данное приложение.
- В PalmOS нет таких понятий, как: "загрузка операционной системы" и "завершение работы". Вместо этого PalmOS "умеет делать" легкий рестарт (Reset) и полный рестарт (Hard Reset).
- PalmOS состоит из модулей (Manager) и библиотек (Library), например, Memory Manager – модуль управления памятью, Data Manager – модуль управления файловой системой, Библиотека TCP/IP. Модуль – это неотъемлемая часть PalmOS, а библиотеки можно добавлять и удалять.
- Каждый модуль и библиотека предоставляет документированное API для приложений.
- PalmOS автоматически управляет питанием. Например, PalmOS может полностью остановить процессор, при этом приложение ничего не заметит.
- PalmOS является ОС с микроядром (Micro Kernel OS), основное качество такого типа ОС – это гарантированно малое время реакции на любые внешние события. (микроядро AMX разработано фирмой Kadak)
- PalmOS поддерживает только один процессор - Motorola DragonBall 68K.
- PalmOS поддерживает 1-, 2-, 4- 8 и 16-битовые цвета и оттенки серого
- PalmOS способна работать после предварительной обработки с файлами следующих форматов: HTML, Java, Flash (только Sony Clie).

### **Выбор инструментария для создания электронных учебных пособий**

Перед началом работ по выбору технологии представления ЭУП на платформе КПК были сформулированы требования, которым должна удовлетворять разработка.

- *Возможность отображения графики.* Необходимым условием любого ЭУП является наличие различных графических материалов, которые позволяют обеспечить наглядность представляемой информации. Поэтому возможность просмотра графики является одним из основных критериев при выборе стандарта.
- *Возможность отображения таблиц.* Очень удобным инструментом структурирования информации являются таблицы. Большая часть учебных дисциплин естественнонаучного цикла активно использует таблицы для представления различных данных, что сделало их наличие необходимым критерием.
- *Открытость.* Открытость стандарта для сторонних разработчиков позволит дать возможность большому числу людей принять участие в разработке новых ЭУП, что в свою очередь значительным образом расширит диапазон применения технологии.
- *Легкость и простота создания ресурсов.* Одним из критериев при создании технологии является легкость освоения работы с ней, что позволит сделать доступной работу с ней не только для профессионалов, но и для рядовых преподавателей, которые смогут в короткие сроки перенести свои ЭУП на платформу КПК или создавать новые.
- *Возможность создания переходов по документу.* Наличие гипертекстовой структуры ЭУП является де-факто стандартом для них, следовательно, также должно присутствовать и в реализации для платформы КПК.

- *Возможность создания интерактивных моделей.* Современные КПК являются мультимедийными, что делает возможным создание и перенос на их платформу различных интерактивных моделей. Это позволит использовать КПК в образовательном процессе в полной мере.

Основной акцент при разработке технологий был сделан на платформу Palm в силу того, что у Windows CE нет кардинальных проблем с переносом ПО, разработанным для desktop-версий этой операционной системы, а также того, что платформа Palm занимает более 60 % рынка КПК в России.

К сожалению, в силу особенностей операционной системы PalmOS, оказалось невозможным совместить в одном типе ЭУП все описанные выше требования, поэтому было принято решение о разделении разрабатываемой технологии на 2 уровня:

- текстовые ЭУП с поддержкой отображения графики, таблиц и гипертекстовой структуры документа;
- мультимедийные ЭУП.

Для PalmOS существует большой спектр различных программ, предназначенных для отображения текстовой информации, но не все из них удовлетворяли выдвигаемым условиям, среди которых были:

- возможность отображения таблиц;
- возможность отображения формул;
- возможность корректного отображения больших изображений.

Перед тем как приступить к разработке технологии создания ЭУП первого уровня для платформы Palm, были проведены исследования программ, которые больше всего распространены среди пользователей КПК этой ОС, на их соответствие указанным выше условиям. Рассматривались три программы – ReachReader, iSilo и Plucker. В дополнении к сравнению по выполнению того или иного условия была также проведена проверка на силу "сжатия" (компрессии) исходного документа для представления на экране КПК. В качестве тестовых образцов были взяты два html-документа, содержащие гипертекстовое форматирование, графику и таблицы. Полученные результаты представлены на рис. 2 и в табл. 1.

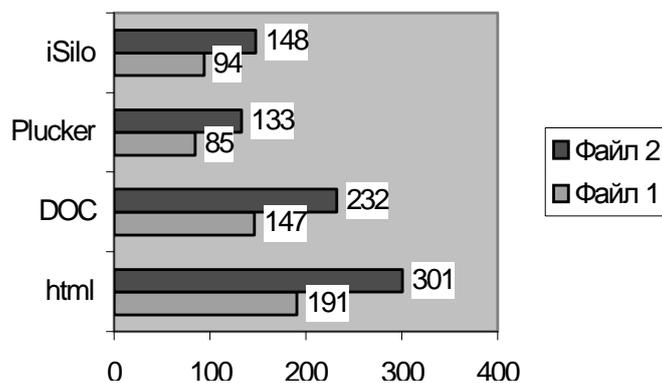


Рис. 2. Степень сжатия различными программами

	ReachReader	iSilo	Plucker
Графика	-	+	+
Таблицы	-	+	+
Аналог для платформы WindowsCE	-	+	+
Лицензия	Free	SoftWare	GPL
Корректное отображение графики/таблиц	-	-/-	+/-
Метод сжатия	Doc	Soft	zLib

Таблица 1. Сравнительная оценка различных программ для PalmOS

На основании проведенных исследований был сделан вывод, что Plucker является наиболее оптимальной основой для создания технологии переноса ЭУП первого уровня с desktop на платформу Palm.

В настоящий момент только две технологии могут быть использованы для разработки мультимедийных ЭУП второго уровня. Это Java и Macromedia Flash. Аналоги этих технологий существуют для обеих платформ КПК.

Был произведен сравнительный анализ эффективности использования той или иной технологии для создания мультимедийных ЭУП, в результате чего были сделан выбор в сторону Flash-технологий по следующим причинам:

- идентичность визуализации на всех поддерживаемых платформах. К сожалению, Java-технологии не смогли обеспечить достаточной цветовой совместимости приложений запущенных на различных моделях КПК;
- простота разработки графических и анимированных объектов, отсутствие необходимости навыков программирования для создания красивых наглядных эффектов;
- качественное масштабирование. Flash является векторной графикой, поэтому возникает минимум проблем при переходе с одного расширения экрана на другой;
- простое встраивание звука. Для Java эта возможность реализована не так хорошо в силу ее ориентированности на чистое программирование;
- малое количество памяти, требуемой для хранения. Отсутствие необходимости каких-либо библиотек для поддержания работоспособности ЭУП значительным образом сокращает объем памяти, требуемый моделью на Flash.

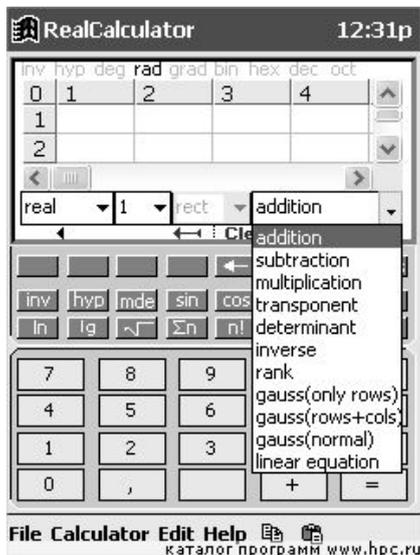
Перечислим задачи, которые ставились при разработке технологии создания мультимедийных ЭУП второго уровня

- Создание библиотеки готовых типовых объектов, которые можно использовать при создании различных интерактивных демонстраций.
- Создание шаблонов ЭУП, которые могут быть использованы в качестве основы для разработок различного рода (текст с графикой, анимация, интерактивная демонстрация, видеофрагмент).
- Разработка механизма, позволяющего оптимизировать отображение ресурсов для проигрывателя 5 версии, так как в настоящий момент для платформы Palm доступен только Flash-плеер 5 версии, в котором отсутствуют некоторые возможности (в том числе работа с видео) Flash более поздних версий.

## **Практическая реализация**

В ходе практической реализации ставилась задача разработки инструментария, который бы позволил создавать и переносить на КПК различные электронные образовательные ресурсы.

За основу инструментария для создания ЭУП 1 уровня был взят уже имеющийся транслятор ("парсер") для программы Plucker. Выбор был обусловлен не только результатами сравнения с другими программами аналогичной направленности, но и тем, что исходные коды парсера находятся в свободном доступе и написаны на языке Python. Доработка программы позволила сделать возможным достаточно корректное отображение таблиц, которые встречаются в документах. В настоящий момент реализованы две формы программы – с графической оболочкой типа Explorer и консольный вариант. Второй вариант позволяет производить более точную настройку процесса перевода ресурсов на КПК.



а



б

Рис. 3. Примеры существующих образовательных программ для КПК: а – реализация программного калькулятора для платформы Windows CE; б – реализация программного калькулятора для платформы Palm OS

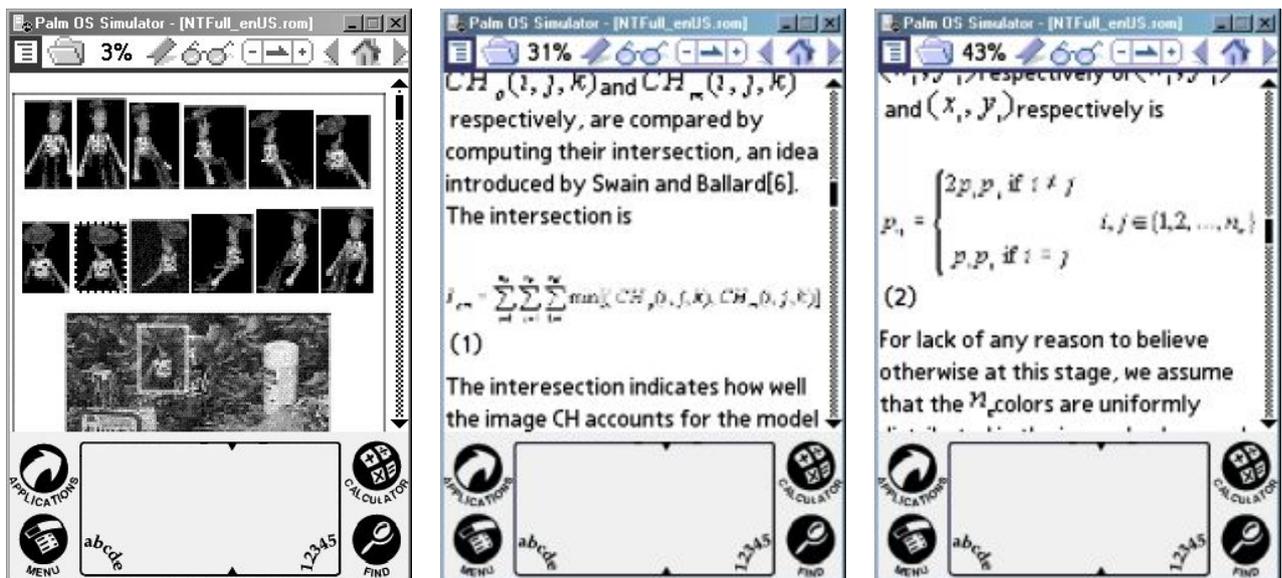


Рис. 4. Пример отображения статьи, переведенной на КПК с помощью разработанных технологических методик (скриншоты сделаны с использованием эмулятора PalmOS 5 версии).

В перспективе рассматривается возможность реализации внешнего редактора ресурсов со встроенным графическим редактором формул, который бы позволил решить проблему отображения сложных математических объектов. В настоящий момент эта функция частично реализована с помощью редактора Microsoft Equation.

В качестве основы для создания ЭУП 2 уровня была взята технология Flash. Основными целями при этом являлись:

- создание библиотеки стандартных типовых объектов, которые можно было бы использовать при создании виртуальных лабораторий;

- создание шаблонов ЭУП, которые могут быть использованы в качестве основы для разработок различного рода (текст с графикой, анимация, интерактивная демонстрация, видеофрагмент);

Полная реализация поставленных целей позволит создавать (и переносить) образовательные ресурсы на платформу КПК, причем этот процесс будет максимально упрощен путем автоматизации выполнения типовых операций

## **Заключение**

На сегодняшний день сложилась ситуация, когда технологические возможности КПК позволяют использовать их в качестве инструмента образования, однако сами образовательные ресурсы для КПК представлены довольно небольшой группой различных научных калькуляторов, сборников констант, электронных справочников. В силу особенностей построения учебного материала меньше всего ресурсов создано по предметам естественнонаучного цикла (физика, математика и т.д.). В результате проведенных исследований были получены следующие результаты.

- Разработана технология и инструментарий для переноса и корректного отображения технических (учебных) документов на платформы КПК различных операционных систем;
- Создан инструментарий для создания и перевода с desktop интерактивных демонстраций на платформу КПК.

Разработанные технологии и методики позволят в относительно короткие сроки произвести перенос необходимых ЭУП на платформу КПК и расширить диапазон применения СДО, не ограничивая его только сферой Internet.

# ПРОЦЕССНЫЙ ПОДХОД В УПРАВЛЕНИИ ИНФОРМАЦИОННЫМИ ПОТОКАМИ ПРИ ВНЕДРЕНИИ КОРПОРАТИВНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Д.А. Заколдаев

Научный руководитель – к.т.н., проф. Н.Д. Фролов

В работе рассматривается технология консолидации функциональной и процессной систем управления на основе транзакционных данных корпоративной информационной системы на этапе проектирования ее архитектуры и внедрения на предприятии.

## Введение

В современных условиях российскому предприятию, в первую очередь, необходима эффективная технология управления, нацеленная не на управление сложившимися на предприятии структурными подразделениями (функциями), а на управление деловыми процессами (бизнес-процессами), способными производить товары и услуги в точном соответствии с требованиями потребителей.

Традиционное функциональное управление – это перечень задач, которые решает предприятие для достижения поставленных целей. Безусловно, в рамках предприятия можно выделить множество функций. Традиционно на предприятии каждой функции или ее части соответствует структурное подразделение. Управление функцией понимается как управление структурным подразделением. *Функция отвечает на вопрос – что делать.* В настоящее время менее изучен процессный подход в управлении предприятием, процессный подход – это распределение ресурсов, ориентированное на выполнение процесса, а не функции.

Принципиальная схема функциональной модели управления приведена на рис. 1.

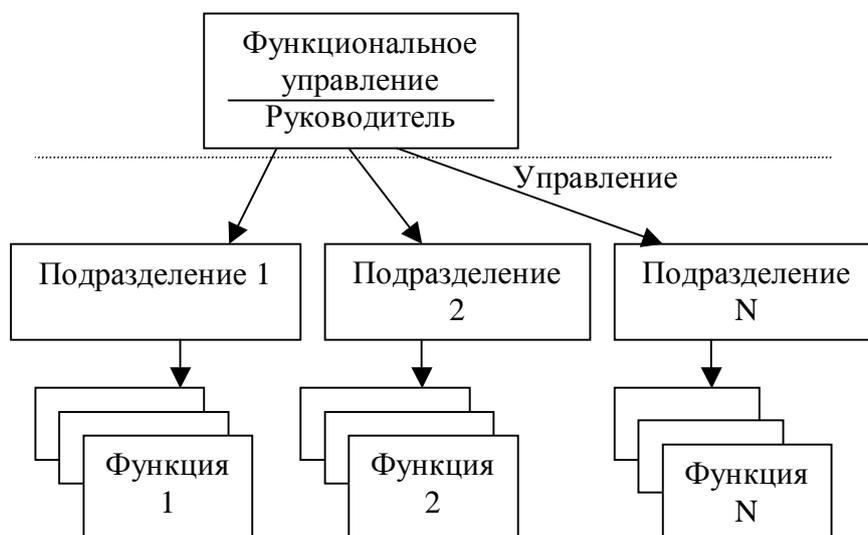


Рис. 1. Структурная модель функциональной системы управления

Процесс отвечает на вопрос – как делать. Система процессного управления позволяет оптимизировать систему общего корпоративного управления, сделать ее прозрачной для руководства и способной гибко реагировать на изменения внешней среды. Система процессного управления регламентирует:

- порядок планирования целей и деятельности;
- взаимодействие между процессами и подразделениями предприятия;
- ответственность и полномочия должностных лиц, в том числе владельцев процессов;

- порядок работы и действий в нестандартных ситуациях;
- порядок и формы отчетности перед высшим руководством;
- систему показателей, характеризующих результативность и эффективность деятельности предприятия в целом и его процессов;
- порядок рассмотрения результатов деятельности и принятие управленческих решений по устранению отклонений и достижению плановых показателей.

Принципиальная схема процессной системы управления приведена на рис. 2.

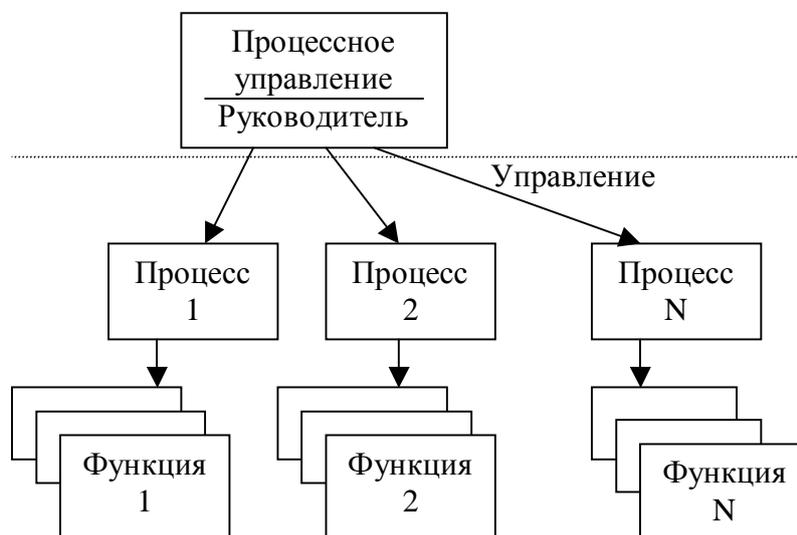


Рис. 2. Структурная модель процессной системы управления

Внедрение на предприятии системы управления бизнес-процессами (далее СУБП) в первую очередь подразумевает работу по описанию и регламентации бизнес-процессов, в рамках которой:

- проводится распределение ответственности за результаты работ, входящих в состав процессов;
- определяется система взаимодействия процессов между собой и с внешними поставщиками и потребителями;
- определяется перечень документации, необходимой для функционирования процессов (инструкции, порядки, положения, методики, должностные инструкции и т.д.);
- составляется график разработки и внедрения этой документации;
- устанавливаются показатели деятельности процессов, способы и формы сбора информации и порядок отчетности перед руководителями;
- определяются границы показателей, характеризующие нормальное течение процессов;
- устанавливаются критерии, по которым начинается работы по устранению причин отклонения.

Система процессного управления позволяет получить и использовать систему показателей и критериев оценки эффективности управления на каждом этапе производственной/управленческой цепочки. Система показателей, построенная в рамках процессного управления, структурируется по четырем направлениям:

- показатели результата деятельности отдельных процессов и предприятия в целом (достижение запланированных результатов – по объему, качеству, номенклатуре и срокам);
- показатели эффективности деятельности отдельных процессов и предприятия в целом (отношение полученных результатов к затратам времени, финансовых и других ресурсов);

- показатели продуктов, производимых процессами предприятия;
- показатели удовлетворенности клиентов результатами работы.

Система процессного управления предусматривает двухступенчатую систему показателей:

а) показатели, по которым владелец процесса оценивает результативность и эффективность своего процесса и работ, входящих в его состав;

б) показатели, по которым владелец процесса отчитывается перед высшим руководством о результатах деятельности процесса. В число процессов, которые существуют в организации, входит процесс управления организацией. Владелец этого процесса является Генеральный директор. Управление деятельностью предприятия производится на основе отчетных показателей, которые владельцы процессов передают высшему руководству.

Внедрение процессного подхода и построение системы менеджмента качества гарантирует четко определенный порядок и ответственность за разработку, согласование, утверждение и ведение документации, соответствие ее действующей нормативной базе РФ.

Система управления бизнес-процессами обеспечивает возможность построения на предприятии системы менеджмента качества в соответствии с требованиями международного стандарта ISO 9001:2000. Разработанная и внедренная СУБП обеспечивает реализацию в организации процессного подхода в соответствии с требованиями последней версии международного стандарта ISO 9001:2000 и получение сертификата ISO 9001:2000.

В настоящее время в промышленно развитых странах признан факт, что предприятие является системой связанных между собой пяти основных групп бизнес-процессов:

- производство продукции и услуг;
- сбор, обработка и документирование информации о реализации бизнес-процессов предприятия;
- анализ информации и принятие качественных управленческих решений;
- управление ресурсами производственных бизнес-процессов;
- постоянное улучшение системы управления производством.

Это видение нашло отражение в международном стандарте ISO 9001 версии 2000 года.

### **Объединение функциональной и процессной систем управления**

Суть процессного подхода заключается в том, что процессы должны взаимодействовать друг с другом по принципу "*клиент - исполнитель*". Рассмотрим применение этого тезиса к производственному предприятию.

При проведении интервью руководителей подразделений и служб были сформулированы следующие основные функции предприятия:

1. маркетинг рынка готовой продукции,
2. формирование плана производства,
3. закупка сырья и материалов,
4. приемка сырья и материалов,
5. хранение сырья и материалов,
6. входной контроль сырья и материалов,
7. выдача в производство сырья и материалов,
8. производство продукции,
9. контроль качества продукции,
10. приемка готовой продукции,
11. хранение готовой продукции,

12. отпуск готовой продукции,
13. ППР.

Посредством применения вышеуказанного подхода получаем карту основных процессов предприятия.

1. планирование производства,
2. логистика сырья и материалов,
3. входной параметрический контроль материалов и ПКИ,
4. технологический процесс,
5. контроль качества готовой продукции,
6. логистика готовой продукции,
7. обслуживание и ремонт.

Таким образом, процесс – это набор функций, направленных на выполнение совокупности последовательных действий для достижения поставленной задачи (цели). В свете этого становится очевидным, что функции определяют процесс, а процесс определяет функции.

Недостаток процессного подхода заключается в том, что во многих случаях для его внедрения требуется революционный реинжиниринг системы управления. Такой способ приемлем не во всех случаях, особенно для действующих предприятий. Для решения данной проблемы предлагается методика объединения функциональной и процессной систем управления.

Практически внедрение новых технологий управления возможно только при полной информационной управляемости. Под управлением понимается процесс планирования, организации, мотивации и контроля, необходимый для того, чтобы сформулировать и достичь целей организации. Для обеспечения информационной управляемости необходима автоматизация управления, т.е. в данном контексте – применение технических средств, частично или полностью освобождающих человека от непосредственного участия в процессах получения, преобразования, передачи и использования информации.

Перечислим задачи разработчиков информационной системы.

- формализация процесса,
- определение набора функций процесса,
- регистрация показателей в области пересечения процесса и функций.

Определение "глубины диффузии" функции в процесс для регулирования степени использования функциональной или процессной систем управления (под областью "диффузии" функции в процесс понимается область формализации задачи).

Основой процессного управления является принятие решений, основанных на аналитических данных, поэтому большое значение для создания процессного управления имеет информационная система.

Посредством применения методики разработки и внедрения КИС, сущностью которой является консолидация двух подходов (на уровне информационной среды), обеспечивается улучшение показателей при переходе предприятия на процессную систему управления с сохранением функциональных свойств. Это выражено в: снижении рисков при формировании аппаратной структуры, гибкости настройки, снижении временных показателей при внедрении системы управления, наличие этапа опытной эксплуатации разрабатываемой системы для отладки системы управления.

### **Методика внедрения процессного управления информационными потоками на предприятии с функциональной системой управления**

1. При предпроектном обследовании:
  - анализ области формализации задачи;
  - определение состава функций;

- формализация основных бизнес-процессов предприятия (рекомендуется методология SADT, стандарты IDEF);
- анализ основных подпроцессов;
- определение точки корреляции процессов и подпроцессов;
- формирование процессов, нуждающихся в реинжиниринге;
- создание сети процессов;
- анализ принятой на предприятии отчетности;
- формализация документооборота процессов.

2. При формировании технических требований на информационную систему на основе полученных данных становится возможным произвести следующие действия, направленные на управление информационными потоками при разработке архитектуры КИС:

- оптимизация сети процессов дифференцированной по функциям;
- определение предельных точек диффузии функции в процесс (в контексте информационной среды);
- создание нормативно-справочной информации;
- классификация документооборота процессов;
- разработка диктатурной системы маршрутизации документооборота;
- формализация и оптимизация форм отчетности, отвечающей требованиям процессной технологии управления;
- определение функционального состава КИС;
- оптимизация последовательности внедрения компонентов КИС под процессную систему.

Методика объединения разрозненных функциональных задач в сеть процессов организации посредством внедряемой информационной системы позволяет добиться последовательного итерационного реинжиниринга системы управления предприятием путем формирования консолидированной аналитической базы данных, характеризующей реализацию процессов организации.

## **Выводы**

Применение методики внедрения процессного управления информационными потоками на предприятии с функциональной системой управления позволяет произвести определение процессов, оптимизацию количества функций, формализовать набор функций процесса, провести регистрацию контрольных показателей выполнения функции в рамках процесса, осуществлять контроль по процессам в рамках существующей структуры предприятия, точную формализацию задачи при разработке информационной системы посредством определения граничных условий функции и процесса.

В заключение необходимо отметить, что концепция бизнес-процессов и процессный подход являются лишь инструментами повышения эффективности компании. Для эффективного их использования необходимо знать показания к применению, области применения, а также другие инструменты, существующие для решения подобных задач.

## **Литература**

1. Реинжиниринг бизнеса: Реинжиниринг организаций и информационные технологии. М.: Финансы и статистика, 1997.
2. Управление производством на базе стандарта MRP II. – СПб: Питер, 2002.
3. Моделирование бизнес-процессов – М.: Диалог-мифи, 2003.

# ПРОБЛЕМА И ВЫБОР СПОСОБА ПРОТИВОДЕЙСТВИЯ ВРЕДНОСНЫМ ПРОГРАММАМ, ИСПОЛЬЗУЮЩИМ ИНТЕРПРЕТАТОРЫ

К. Н. Шилов

Научный руководитель – д.т.н., проф. А.Ю. Щеглов

Настоящая работа является анализом существующего положения дел в сфере противодействия вредоносным программам, в частности, использующим интерпретируемые средства программирования. На сегодняшний день вредоносные программы, использующие командные интерпретаторы, "...являются самой большой проблемой современной вирусологии." (Е. Касперский). Результатом проведенного исследования является предложение подхода к построению системы противодействия, позволяющего упростить реализацию антивирусного программного обеспечения.

## Введение

Для создания первых компьютерных вредоносных программ необходим был опыт программирования на языке ассемблера или довольно сложных языках высокого уровня, таких как C/C++, что требовало высокой квалификации программиста. В середине 90-х годов XX века, с внедрением в один из наиболее популярных программных продуктов Microsoft Office эффективного языка автоматизации приложений Visual Basic for Applications, ситуация резко изменилась. Люди, называющие себя вирусмейкерами, получили прекрасный инструмент для реализации своих идей довольно тривиальными методами. В настоящее время доступность создания макро-вирусов привела к тому, что вредоносные макро-программы, по данным разных источников, составляют от 20% до 70% из почти 70000 известных на сегодня компьютерных вирусов.

Для учреждений характерным является организация электронного документооборота при административном ограничении полномочий пользователя по установке программного обеспечения и копирования файлов в компьютерные системы. В этом случае интерпретируемые вирусы и программные закладки являются основным источником опасности.

Тип вирусов	Наносимый ущерб (%%)
Скрипт-вирус	4
DOS-вирус	7
Windows-вирус	10
Троянский конь Backdoor	13
Интернет-червь	14
Троянский конь PSW	22
Макро-вирус	30

Таблица 1. Самые распространенные типы вирусов

## Условия существования вредоносных макро-программ

Чем совершеннее программное средство, тем уязвимее становится оно для компьютерных злоумышленников. В Windows заложены два концептуальных решения, взаимно дополняющие друг друга – СОМ (модель компонентных объектов) и автоматизация. Технология СОМ определяет стандартные механизмы и правила предоставления одними программными продуктами функций другим, не обладающим такими возможностями. Разработка технологии СОМ стала неизбежной в условиях появления проектов, содержащих огромное количество строк исходного кода. Главным достоинством концепции СОМ является возможность разбиения проекта на самостоятельные, логически завершенные части. Под автоматизацией понимается предоставление приложением своих возможностей другим. Автоматизация, построенная на технологии

COM, затронула не только пользовательские программные продукты, но проникла в операционную систему и стала неотъемлемой ее частью. Ядром автоматизации Windows является Visual Basic. В то же время в пользовательских приложениях используется разновидность VB – Visual Basic for Application. Пакеты офисных продуктов Microsoft Office, начиная с 6 версии, также имеют встроенные средства автоматизации. Данные инструментальные средства представляют собой мощнейший командный процессор. Благодаря COM обеспечивается доступ ко всем внешним командам и программам, которые обладают дополнительными возможностями.

На сегодняшний день проблема противодействия вредоносным программам лежит в самой технологии, отказаться от использования которой не представляется возможным, так как произойдет возврат на уровень начала 90-х годов.

Одно из лидирующих мест по количеству созданных вредоносных программ занимает Microsoft Word, входящий в состав офисного пакета Microsoft Office. Для существования вредоносной программы, написанной на макро-языке, необходимо наличие следующих возможностей:

- привязки программы на макро-языке к конкретному файлу;
- копирования макро-программ из одного файла в другой;
- возможность получения управления макро-программой без вмешательства пользователя.

Данным условиям удовлетворяют редакторы из пакета Microsoft Office. Системы из этого пакета содержат в себе макро-языки: Word - Word Basic, Excel, Office (включая Word, Excel и Access) - Visual Basic for Applications. При этом:

- макро-программы находятся внутри файла (Word, Excel, Office);
- макро-язык позволяет копировать файлы или перемещать макро-программы в служебные файлы системы и редактируемые файлы;
- при работе с файлом при определенных условиях (открытие, закрытие и т.д.) вызываются макро-программы, которые имеют стандартные имена.

## **Подходы к построению антивирусной защиты**

Сегодня используется несколько подходов, которые лежат в основе построения модулей антивирусной защиты. К их числу относятся:

- сканеры и мониторы;
- поведенческие блокираторы.

Принцип работы сканеров и мониторов основан на знании характерных фрагментов (сигнатур) конкретных вирусов. Однако, чтобы подготовить сигнатуру, нужно получить экземпляр вируса, проанализировать его, выпустить средства для распознавания и лечения, а затем распространить обновление по тем же каналам Internet, по которым распространяется и сам вирус. Кроме того, появились полиморфные макро-вирусы, которые умеют модифицировать свой код, что не позволяет распознавать их с помощью сигнатур. Поэтому традиционные средства часто не успевают блокировать распространение новых вредоносных программ. С определенной погрешностью (в окрестности 3%) эффективность стационарных антивирусных средств (к ним относятся модули, использующие антивирусные базы данных) можно оценить в 90%. Любой компьютер более всего уязвим в момент между появлением нового вируса и установкой обновления антивирусного ПО, защищающего от этого вируса. К безусловным плюсам сегодняшнего дня относятся скорость доставки обновлений на удаленные компьютеры.

Сканеры и мониторы не решают главной проблемы – защиты от неизвестных вирусов. Данного недостатка лишены так называемые поведенческие блокираторы. Они функционируют как резидентные программы, которые перехватывают различные

события и в случае "подозрительных" действий запрещают эти действия или запрашивают разрешение у пользователя.

Блокираторы совершают не поиск уникального программного кода, сходного с сигнатурой, принадлежащей вредоносной программе, а отслеживает и нейтрализует вредоносные программы по их характерным действиям. Но, как любой технологии, поведенческий блокиратор свойственны достоинства и недостатки. Достоинством является независимость от базы сигнатур и, как следствие, возможность предотвращения проникновения неизвестного вирусного кода. Недостатком является не способность идентификации и последующего лечения зараженного объекта. Работа поведенческого анализатора (блокиратора) основана на проверке корректности запрашиваемого действия интерпретатором. Отличительной особенностью использования блокираторов в интерпретируемых языках программирования состоит в том, что количество макрособытий ограничено.

К недостаткам поведенческого блокиратора можно отнести:

- необходимость иметь перечень всех возможных макрокоманд, которые могут быть затребованы на выполнение интерпретатором при анализе программного кода;
- полное отсутствие документации на API-интерфейсы обращения к интерпретатору VBA;
- зависимость конкретной версии поведенческого блокиратора от используемой версии командного интерпретатора, в частности Visual Basic for Applications.

Хотя поведенческий анализатор обладает перечисленными выше недостатками, он по-прежнему остается наиболее эффективным средством раннего предупреждения проникновения нового вредоносного кода в систему.

Предлагаемый метод повышения эффективности раннего предупреждения активности вредоносного кода основан на реализации дополнительных средств контроля выполнения макрокоманд не на этапе передачи очередной команды в интерпретатор, а при вызове системных функций.

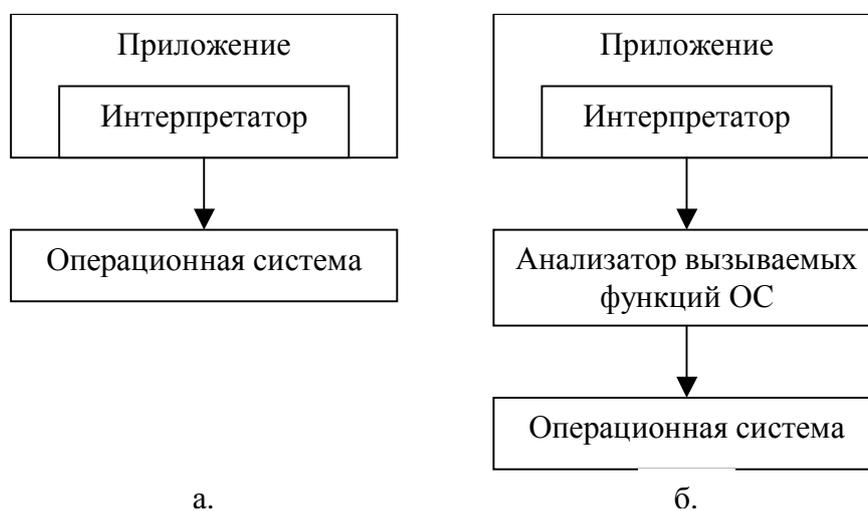


Рис 1. Схемы вызовов функций ОС из приложения – без анализатора (а) и с внедренным (б) анализатором вызываемых функций ОС

Таким образом, выполнение приложения, использующего автоматизацию средствами интерпретатора VBA, осуществляется в замкнутой среде, что уменьшает вероятность воздействия вредоносных программ на другое пользовательское программное обеспечение, на операционную систему в целом, на обрабатываемые и хранящиеся данные. Реализация данного подхода осуществима через замену динамической библиотеки интерпретатора VBA на библиотеку, содержащую анализатор вызываемых функций ОС. Ссылка на библиотеку VBA находится в системном реестре ОС, и поэтому не

представляет технических трудностей, с точки зрения программной реализации, осуществить перехват загрузки оригинального программного модуля VBA в вызывающее приложение и выполнить необходимые действия по инициализации контроля вызовов функций ОС. Таким образом, возможно построение дополнительных средств защиты, не зависящих от конкретных версий командных интерпретаторов, что решает одну из проблем по реализации подобного программного обеспечения.

## **Заключение**

В работе проведен анализ существующего положения дел в сфере противодействия вредоносным программам, использующим интерпретируемые средства программирования. Выявлены присущие недостатки существующих решений и предложен метод, позволяющий упростить программную реализацию.

## **Литература**

1. Касперский Е. Компьютерные вирусы: что это такое и как с ними бороться?
2. Полиморфные макро-вирусы, <http://securitylab.ru/33903.html>.
3. Коржов В. "Песочница" для вируса // Computerworld. 2001. № 9.
4. Эффективные антивирусные технологии, <http://www.bytemag.ru/Article.asp?ID=1550>

# ОБЪЕДИНЕНИЕ РЕЛЯЦИОННОГО И МНОГОМЕРНОГО ПРЕДСТАВЛЕНИЯ ДАННЫХ НА ПРИМЕРЕ ORACLE 9i

А.К. Дорожкин

Научный руководитель – к.т.н., проф. В.В. Кириллов

Изучается новый подход к представлению данных, предложенный фирмой Oracle. Получено формальное описание подхода, выведен критерий эффективного применения данной формы представления. Произведена на практике оценка эффективности применения подхода по сравнению с уже существующими.

## Введение

С момента появления первых реализаций технологии OLAP (сам подход возник еще в начале 70-х, но реализован он впервые был только в 90-е) программные комплексы, осуществляющие анализ данных на основании многомерного представления, выпускались как отдельные продукты. Эти программные комплексы имели возможность извлекать данные из одной или нескольких различных СУБД, а также плоских файлов. Первым, кто сломал данную традицию, была корпорация Oracle, сделавшая попытку объединить в одном ядре как реляционное, так и многомерное представление данных. О том, как и насколько удачно это осуществляется в настоящее время, и пойдет речь в работе.

Изначально в рамках продуктов, предоставляемых фирмой Oracle, многомерный анализ осуществлялся с помощью сервера Oracle Express. Первые версии Express не были собственной разработкой фирмы Oracle. Они создавались компанией IRI Multidimensional и назывались IRI Express Server. Лишь с 1995 г. данный продукт был приобретен корпорацией Oracle и получил название Oracle Express Server. Линейка продуктов Oracle, связанных с проведением многомерного анализа, хорошо известна специалистам, так что перейдем к рассмотрению нововведений, появившихся в версии Oracle 9i.

Вышедшая в 2001 г. версия Oracle 9i включала в себя целый ряд усовершенствований, среди которых было заявленное ранее объединение реляционного и многомерного представления данных в одном ядре, а также единый метод описания различных форм представления данных на базе Common Warehouse Metamodel (CWM). На данный момент времени актуальной является версия Oracle 9i Release 2. На ее примере и рассмотрим реализацию подобного объединения.

## 1. Формальное описание различных представлений данных

Для оценки эффективности новой модели рассмотрим три различных способа хранения данных:

1. Реляционное представление данных на основе схемы "звезда". Такое представление по сравнению с остальными характеризуется эффективным использованием дискового пространства и низкой производительностью при извлечении данных из-за отсутствия хранения обобщенных показателей, которые используются в первую очередь в аналитических приложениях.

2. Многомерное представление данных, в первую очередь, нацелено на эффективное извлечение данных за счет неоптимального использования дискового пространства и избыточности данных.

3. Новая форма представления – реляционные кубы, предназначенные для достижения преимуществ двух вышеперечисленных подходов – эффективного использования дискового пространства и приемлемой производительности.

Эффективность использования дискового пространства можно определить, исходя из принципов положенных в основу хранения данных того или иного подхода. Намного сложнее задача определения производительности, так как она зависит от многих факторов, в том числе и типа клиентского приложения. Поэтому в настоящей работе более

производительным способом хранения данных будем считать тот способ, который при равной степени детализации данных занимает меньший объем на диске:

$$\frac{C_{det} + C_{agg}}{V} \rightarrow \min, \quad (1)$$

где  $C_{det}$  – объем детальных данных, получаемый из оперативных источников,  $C_{agg}$  – объем обобщенных данных, получаемый в результате агрегации и  $V$  – объем данных, занимаемый базой данных на диске.

1. Объем данных, занимаемый реляционным представлением данных в схеме "звезда", можно оценить как

$$C_{star} = \left( \sum_{i=1}^n C_{dim i} \right) + C_{fact}, \quad (2)$$

где  $C_{dim i}$  – объем данных в  $i$ -ом измерении,  $C_{det}$  – объем данных в таблице фактов,  $n$  – количество измерений. Исходя принципов хранения данных, принятых в концепциях реляционных баз данных,  $C_{dim i}$  и  $C_{det}$  определяются как

$$C_{dim} = \sum_{j=1}^t \sum_{k=1}^m r_{jk}, \quad (3)$$

где  $t$  – количество записей в таблице,  $m$  – количество полей,  $r_{jk}$  – размер  $k$ -ого поля  $j$ -ой записи (так как размер поля варьируется для некоторых типов данных переменной длины, например строки).

2. Объем данных, занимаемый многомерным представлением данных, можно оценить как

$$C_{cube} = \sum_{i=1}^v \left( \prod_{j=1}^{n_i} d_{ij} \right) t_i, \quad (4)$$

где  $v$  – количество показателей в многомерном кубе,  $n_i$  – количество измерений, по которым индексируется  $i$ -ый показатель,  $d_{ij}$  – количество различных значений в  $j$ -ом измерении,  $t_i$  – размер типа данных показателя.

Как видно из выражения (4), основным фактором, влияющим на размер многомерной базы данных, является количество различных значений в измерении. Однако на практике переменные индексируются далеко не по всем значениям измерений, поэтому для эффективного управления памятью используют измерения двух типов: разреженные (sparse) и плотные (dense). Использование того или иного типа измерения непосредственно влияет на параметр  $d_{ij}$ . Так, в [4] приводится рекомендация использовать плотные измерения только в тех случаях, когда переменная индексируется как минимум по 80% значений данного измерения.

Из выражений (3) и (4) видно, что реляционное представление данных в схеме звезда является более компактным, так как вторая операция выражения (3) является суммированием, а не умножением как в выражении (4). Однако нет четкой закономерности между  $t$  и  $v$ , а также  $m$  и  $n_i$  при переходе от реляционного представления данных к многомерному. Дело в том, что характер этого перехода целиком зависит от данных, хранимых в таблицах. В общем случае размер, занимаемый многомерным кубом, можно выразить через параметры реляционной модели данных, следующим образом:

$$C_{cube} = \left( \prod_{i=1}^{m-v} d_i \right) * \sum_{j=1}^v t_j + \sum_{i=1}^{m-v} d_i r_i \quad (5)$$

где  $d_i$  – количество различных значений в  $i$ -ом поле,  $m$  – количество полей таблицы фактов,  $v$  – количество полей, содержащих значения показателей,  $t_j$  – размер типа данных  $j$ -ого показателя,  $r_k$  – размер типа данных  $k$ -ого измерения. В выражении (5) вторая сумма отвечает за объем данных, занимаемый измерениями, и обычно на практике составляет незначительную часть объема всего куба.

Задачи многомерного анализа в основном связаны с отображением данных на разных уровнях детализации, а вычисление агрегатов "на лету", как происходит в случае использования реляционных данных в схеме "звезда", при больших объемах данных приводит к значительному увеличению времени отклика системы. Поэтому для увеличения производительности таких систем в Oracle 9i была предложена новая форма представления данных, позволяющая, используя все преимущества реляционного подхода, увеличить эффективность поддержки обобщенных данных. Для этого в реляционных кубах создается набор материализованных представлений для различных срезов куба, предназначенных для хранения агрегированных данных. К этим представлениям в конечном итоге и обращается пользователь.

При таком подходе к общему объему данных, хранящемуся в схеме "звезда" (выражение (3)) добавляется еще агрегированные данные, общее число которых, исходя из [3], равно

$$C_{agg} = \prod_{i=1}^n (d_i + 1) - \prod_{i=1}^n d_i \quad (6),$$

где  $n$  – количество измерений,  $d_i$  – количество различных значений в  $i$ -ом измерении. Таким образом, объединив выражения (3) и (6), получим объем реляционного куба:

$$C_{rel} = \sum_{j=1}^t \sum_{k=1}^m r_{jk} + \sum_j^m \left( \prod_{i=1}^n (d_i + 1) - \prod_{i=1}^n d_i \right) r_j \quad (7)$$

## 2. Результаты экспериментов

Для определения критерия эффективности применения нового типа представления данных был проведен эксперимент по определению реальных объемов данных в различных формах представления. В качестве платформы использовался Oracle 9i для реляционных форм представления данных – схема "звезда" и реляционные кубы – и Oracle Express 6 для многомерных представлений данных – разреженных и плотных измерений. В качестве исходных данных использовалась модель данных с тремя измерениями и одним показателем целого типа (рис. 1). Количество значений в каждом измерении равно 1024, число уровней агрегации – 2. В качестве эксперимента были созданы три куба и различной степенью заполнения данными.

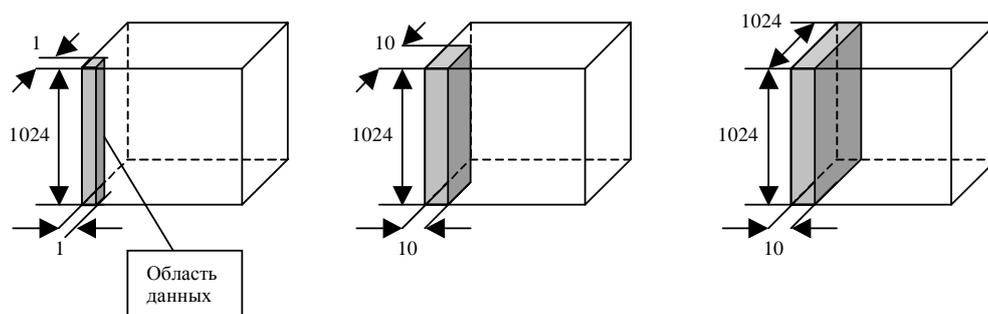


Рис. 1. Заполнение тестовых кубов

В табл. 1 и на рис. 2 приведены результаты эксперимента для четырех форм представления данных, в зависимости от формы представления и объема загруженных данных в мегабайтах.

Форма \ Объем	1024	102400	10485760
Схема "звезда"	0,191	4	528
Реляционный куб	0,389	22	1310
Многомерный куб (разреженный)	2	13	1056
Многомерный куб (плотный)	2,7	23	172

Таблица 1. Результаты эксперимента

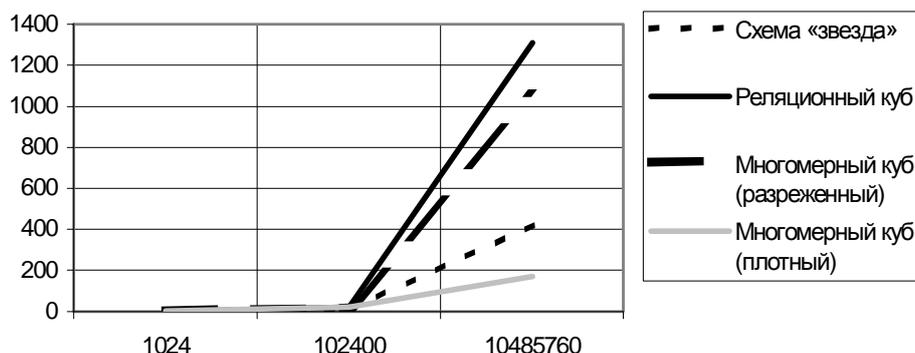


Рис. 2. Результаты эксперимента

Как видно из рис. 1, объем данных, занимаемый реляционным кубом, несколько больше, чем у многомерного разреженного. При таком масштабе видна общая закономерность роста базы данных для различных форм представления и совсем не видна область небольших объемов данных, при которых применение реляционных кубов себя оправдывает.

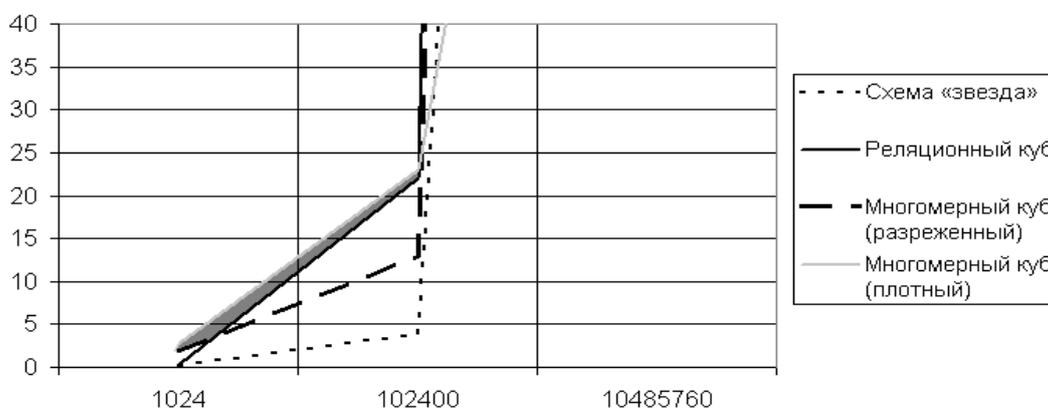


Рис.3 Область эффективного применения

На рис.3 закрашена область эффективного применения реляционных кубов. Используя выражения (7) и (4), можно вывести условие эффективного использования подобной формы представления данных  $C_{rel} \leq C_{cube}$  в общем виде:

$$\sum_{j=1}^t \sum_{k=1}^m r_{jk} + \sum_j \left( \prod_{i=1}^n (d_i + 1) - \prod_{i=1}^n d_i \right) r_j \leq \sum_{i=1}^v \left( \prod_{j=1}^{n_i} d_{ij} \right) t_i \quad (8)$$

## Заключение

В работе проведено исследование формы представления данных, объединяющей достоинства многомерного и реляционного подхода, предложенной фирмой Oracle. Получено выражение критерия эффективности применения данной формы. Описан проведенный эксперимент, наглядно показывающий область эффективного применения данного подхода.

## Литература

1. Архипенков С.Я. Хранилища данных. М.: Диалог – МИФИ, 2002.
2. Сахаров А.А., Принципы проектирования и использования многомерных баз данных. // Системы управления базами данных. 1996. № 03.
3. Хрусталёв Е.М., Алеф Консалтинг & Софт Агрегация данных в OLAP-кубах
4. Oracle Express: Database design and performance guide.

# УПРАВЛЕНИЕ ПАМЯТЬЮ КАК ОСНОВОПОЛАГАЮЩИЙ ФАКТОР БЫСТРОДЕЙСТВИЯ ПРИЛОЖЕНИЙ

С. В. Антонов, Д. В. Соловьев

Научный руководитель – к.т.н., доц. М.Ю. Юдин

В статье рассматриваются основные механизмы управления памятью, предоставляемые операционной системой, а также указываются области их рационального применения.

## Введение

Несмотря на всю важность проблемы рационального использования ресурсов системы, в настоящее время лишь немногие разработчики уделяют этому вопросу достаточно внимания. С приходом мощных библиотек компиляторов и высокоуровневых языков программирования (Java, С# и т.д.) вся работа по распределению ресурсов легла на плечи этих библиотек и операционную систему.

Однако очевидно, что, так как ни операционная система, ни runtime библиотека не знают особенностей программы, то вероятность оптимального выделения ресурсов невелика. Выходом из этой ситуации может быть собственный механизм выделения памяти. Подобный механизм описан в [1]. Однако его реализация целесообразна только в очень больших проектах, поэтому в данной статье речь идет о том, как оптимизировать управление памятью, не прибегая к созданию мощных надстроек.

## Реализация управления памятью

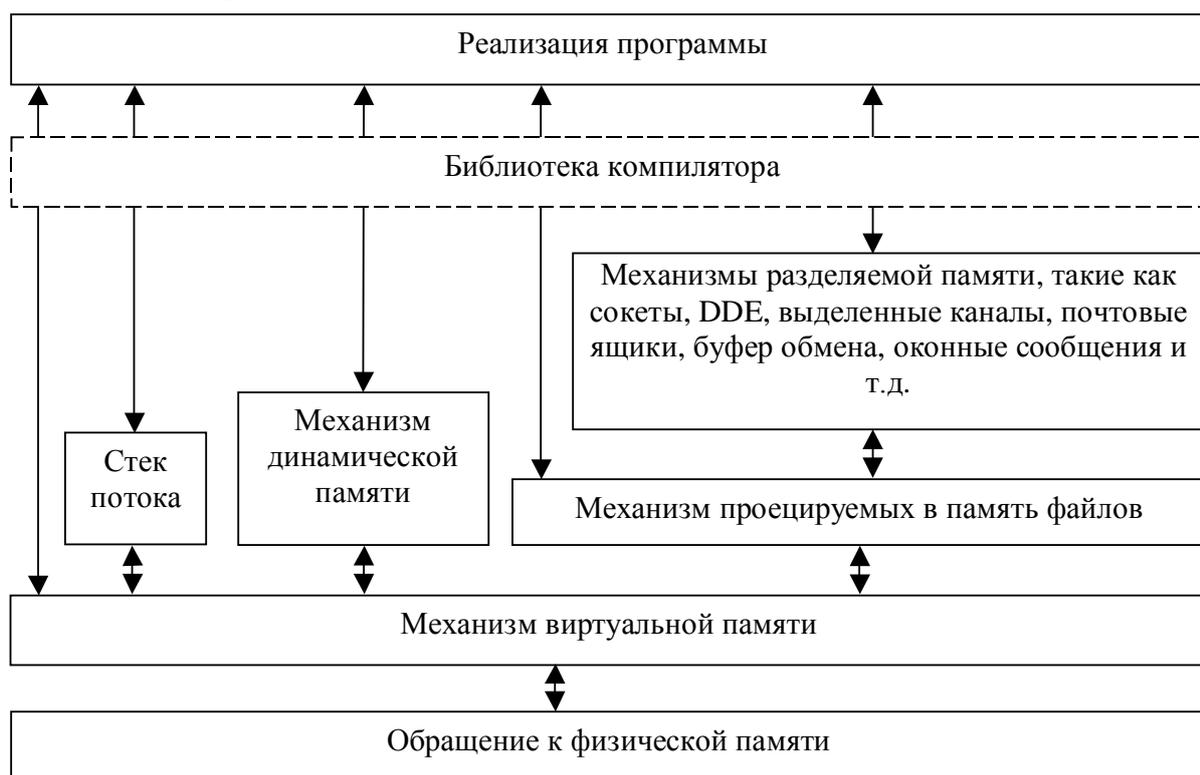


Рис. 1. Основные механизмы распределения памяти, доступные программе

В первую очередь встает вопрос о том какой механизм памяти использовать для той или иной задачи. На рис. 1 схематично показаны основные механизмы управления памятью, предоставляемые программе.

Как видно из рис.1, в основании находится механизм виртуальной памяти. В ОС Windows в его обязанности входит передача физической памяти, резервирование виртуальной памяти, проверка прав доступа к страницам и ряд других функций.

Что касается стека, то в первом приближении может показаться, что он не зависит от механизма виртуальной памяти, так как при обращении к стеку программа просто вызывает соответствующие команды процессора. Однако для того, чтобы программа могла обратиться к какому-либо региону стека, операционная система должна зарезервировать данный регион и передать ему физическую память.

Также важно отметить, что обычно в стеке размещаются только данные небольшого размера, такие как числа или структуры. Если же создается локальная строковая переменная, то в стеке будет размещен только указатель на нее, а для хранения самих данных будет задействована динамическая память.

Что же касается самой динамической памяти, то в ней рекомендуется размещать как можно меньше данных и на непродолжительный срок, так как эта область наиболее подвержена фрагментации. Снижению фрагментации способствует возможность создания дополнительных пулов памяти для разных типов данных, предоставляемая операционной системой. Подробнее об этом см. [2], глава 18.

Следующим базовым механизмом, основанным на виртуальной памяти, является механизм проецируемых в память файлов (memory mapped files (MMF)). Это очень мощный механизм, применяемый для чтения и редактирования файлов, а также создания совместно используемых областей памяти. На MMF основаны многие механизмы операционной системы, например, загрузка программ в память, передача текста из окна в буфер и т.д. Принцип действия MMF заключается в том, что при загрузке файла в память в адресном пространстве процесса резервируется блок, достаточный для размещения всего файла целиком, а в оперативной памяти выделяется только блок, необходимый в данный момент. Таким образом, программа может сканировать файлы очень больших размеров, а именно до 18 Экзабайт ( $18 \cdot 10^{18}$  байт), при этом выделяя сравнительно немного физической памяти.

После рассмотрения основных механизмов, предоставляемых операционной системой, возникает вопрос о том, для каких целей какой механизм выбрать и как наиболее оптимально его реализовать. При этом не следует забывать, что важно не только быстродействие алгоритма, но и его надежность.

Одной из основных проблем быстродействия является фрагментация, в частности, нерациональное распределение строковых данных в памяти. Дело в том, что при выполнении многих строковых операций, которыми изобилуют программы, данные многократно записываются и удаляются из динамической памяти, причем отнюдь не последовательно. В результате этого свободное пространство и данные распределены хаотично. Если в такой ситуации попытаться выделить память, то система может просто не найти свободного блока достаточного размера. Именно по этой причине операционные системы семейства Windows 9x не могли работать продолжительное время, и их приходилось перезагружать.

Самым простым методом снижения фрагментации является расположение данных по принципу "большие в начале, меньшие в конце". Также важно размещать данные и код, который их использует, в одном месте. Что же касается самого кода, то его желательно реализовывать таким образом, чтобы та часть кода, выполнение которой маловероятно, находилась отдельно от основного кода и, вероятно, не была загружена в память.

Важно отметить, что не все части программы нуждаются в оптимизации. Так, повышение быстродействия процедур инициализации и обратных им нецелесообразно, так как они выполняются только один раз, причем очень важно, чтобы они работали корректно. Не имеет смысла оптимизация блоков с множеством тяжелых функций (тяжелыми являются функции ядра, так как на переход в режим ядра требуется более 1000 тактов, функции GDI, функции синхронного чтения или записи на диск и некоторые другие).

## Заключение

Наиболее оптимальное сочетание производительности и надежности (а так же наглядности) кода программы обеспечивает построение алгоритма, наилучшим образом отражающего суть решаемой задачи, и его оптимизация на этапе доработки с учетом тонкостей среды, в которой он будет выполняться.

## Литература

1. Распределение виртуальной памяти в среде Win32. Управление малыми блоками
2. Рихтер Дж. WINDOWS для профессионалов: Создание эффективных Win32-приложений с учетом специфики 64-разрядной версии Windows. / Пер, англ. - 4-е изд. – СПб; Питер; М.: Издательско-торговый дом «Русская Редакция», 2001. – 752 с.; ил.
3. Цикл статей «Драйверы режима ядра». <http://www.wasm.ru>
4. А. Мешков, Ю. Тихомиров, VISUAL C++ и MFC. Программирование для WINDOWS NT и WINDOWS 95. В двух томах.
5. П. Дейтел, Х. Дейтел. Как программировать на C++: Пер. с англ. – М.: ЗАО «Издательство БИНОМ» 2000 г. – 1024 с.: ил.
6. Румянцев П.В. Работа с файлами и Win32 API. М.: Горячая линия – Телеком, 2000. – 200 с., ил.

# НОВЫЙ ПРОТОКОЛ IPv6: ПРЕИМУЩЕСТВА, МОДЕЛИ ПЕРЕХОДА

Д.В. Соловьёв, С.В. Антонов

Научный руководитель – к.т.н., доц. М.Ю. Юдин

В работе кратко рассматриваются основные положения протокола IPv6 и модели перехода на эту технологию.

## Введение

Internet в последнее время в нашей жизни стал таким же необходимым и неотъемлемым атрибутом, как телевизор или стиральная машина. И уже почти не осталось людей, которые не представляют что это такое. Кибер-пространство увеличивается в размере, каждый день принимая в свои ряды тысячи новых пользователей. Интернет вошёл в различные области жизни человека: личную и бизнес. Чёткая структурированность и определённая специфика населения Сети позволяют более четко проводить маркетинговые кампании, существенно удешевлять расходы на рекламу и экономить на персонале, принимающем заказы. Набирает обороты и развлекательная часть сети, назревает бум ролевых онлайн-игр. Множество этих фактов являются предпосылками для внедрения нового протокола. Кроме того, телефонные компании, а именно Nokia, уже представила новую технологию "MobileIPv6", основанную соответственно на IPv6. Производители сетевого оборудования и мобильных телефонов в Европе уже готовятся к производству и выпуску мобильного телефона, поддерживающего IPv6. Уже можно говорить о так называемом "тотальном IP", когда свой IP-адрес будут иметь сотовые телефоны, стиральные машины и т.д.

## Причины перехода на новый стандарт

- Основными причинами, подтолкнувшими на разработку нового протокола, стали:
- дефицит адресного пространства, особенно сильно это испытывают Китай, Япония и некоторые другие страны. IPv4 использует 32-битную адресацию, в то время как IPv6 128-битную;
  - появление большого числа новых приложений, например мультимедийных, видеоконференций и Internet-телевидения, которые, в свою очередь, чувствительны к задержкам пакетов"
  - внедрение Сети во все отрасли промышленности, а также в новые страны;
  - новый уровень шифрования и безопасности. Новый протокол предоставляет две встроенные опции, которые обеспечивают безопасность коммуникаций. Они могут использоваться каждая по отдельности или обе вместе. Первый механизм, называемый "Authentication Header", является расширением заголовка и обеспечивает аутентификацию и целостность данных. Второй – "IPv6 Encapsulated Security Header" – также реализуется с помощью расширения заголовка. Он отвечает за конфиденциальность IPv6-дейтаграмм;
  - возможность автоконфигурации IP-сетей. Цель новой функции – сделать так, чтобы хост автоматически определял всю информацию, необходимую для подключения к Internet. Минимальными требованиями здесь являются способность хоста сгенерировать один уникальный IP-адрес и обнаружить (получить адрес), по крайней мере, один маршрутизатор;
  - потребность в широковещании, которая реализуется в IPv6 за счет новой схемы адресации. Пакет может отсылаться как одному пользователю, группе пользователей, включенных в одну физическую подсеть, так и набору компьютеров в различных сетях;

- расширяемость. Заключается в возможности в дальнейшем развивать новый протокол, без резкого перехода и огромных затрат на новое оборудование [2,3].

### **Возможные схемы перехода на новый протокол**

Естественно, переход на новый протокол по принципу "все сразу" неосуществим ввиду огромных технических и, соответственно, денежных затрат. Поэтому предполагается несколько схем перехода:

*Модель двойного стека (Dual stack model).* В модели двойного стека все узлы IPv6, хосты или маршрутизаторы имеют двойной стек. Таким образом, связь с узлами IPv4 осуществляется с помощью стека IPv4, а связь с миром IPv6 осуществляется с помощью стека IPv6. Ограничением этого подхода является необходимость назначения IPv4 адреса каждому новому устройству, способному работать по IPv6.

*IPv6-over-IPv4 туннелирование.* Эта технология может использоваться на этапе внедрения IPv6 инфраструктуры. Здесь используется существующая инфраструктура IPv4 для передачи IPv6 пакетов. IPv4/IPv6 узлы могут инкапсулировать IPv6 пакеты внутри IPv4 пакетов и передавать их по маршрутизационной топологии IPv4.

*SOCKS64.* Инструмент SOCKS Gateway представляет собой систему, которая принимает SOCKS-соединения от хостов IPv4 и ретранслирует их IPv4 или IPv6 узлам. Этот механизм в большей мере предназначен для сайтов, которые уже используют SOCKS-клиентов и SOCKS-сервер. Для таких сайтов SOCKS Gateway обеспечивает простой способ реализации возможности связать хосты IPv4 с хостами IPv6. Не требуется никаких модификаций в DNS и отображений адресов. Этот принцип может также использоваться, чтобы позволить соединить узлы IPv6 с узлами IPv4, узлы IPv4 через сети IPv6 и узлы IPv6 через сети IPv4. Последние два случая имеют сходство с методами туннелирования без возможных проблем с фрагментацией или с количеством промежуточных узлов.

*SIIT (Stateless IP/ICMP Translation Algorithm).* Протокол SIIT описывает метод трансляции между IPv6 и IPv4. Трансляция ограничивается заголовками IP пакетов. В работе не описывается метод присваивания временного IPv4 адреса узлу IPv6. Транслятор работает в бесконтекстном режиме (без сохранения состояния), что означает, что трансляции должен подвергаться каждый пакет.

*NAT-PT (Network Address Translation - Protocol Translation).* NAT-PT относится к связи между хостами, работающими только по протоколу IPv6, и хостами, работающими только по протоколу IPv4. Связь реализуется с помощью специального устройства, которое выполняет трансляцию между IPv4 и IPv6 адресами и хранит состояние в течение продолжительности сеанса связи. Устройство NAT-PT может также включать DNS-шлюз прикладного уровня для того, чтобы сделать возможной трансляцию DNS запросов и ответов между IPv4 и IPv6 узлами.

*BIS (Bump-In-The-Stack).* Модель Bump-In-The-Stack позволяет не способным к работе с IPv6 приложениям, выполняющимся на IPv4 хосте, взаимодействовать с хостами, работающими только по протоколу IPv6. К стеку IPv4 добавляются три модуля, которые находятся между прикладным и сетевым уровнями – расширение к резольверу (name resolver), отображатель адресов (address mapper) и транслятор. Основная идея заключается в том, что, когда приложению IPv4 требуется связь с хостом, работающим только по протоколу IPv6, IPv6 адрес этого хоста отображается на IPv4 адрес из пула, локального для хостов с двойным стеком. Пакет IPv4, формирующийся для взаимодействия, транслируется в пакет IPv6 в соответствии с SIIT. Можно рассматривать Bump-In-The-Stack как конкретную реализацию NAT-PT внутри IP стека хоста.

*Транслятор транспортного уровня (Transport Relay Translator).* Транслятор транспортного уровня (Transport Relay Translator), определенный в RFC "An IPv6-to-IPv4 transport relay translator", дает возможность организации прямой связи между хос-

тами IPv6 и хостами IPv4. Этот механизм чем-то напоминает NAT-PT, но выполняет трансляцию на транспортном уровне, а не на уровне IP. На сайте должен быть выделен маршрутизатор для трансляции {UDP,TCP}/IPv6 в {UDP,TCP}/IPv4 и обратно. Кроме того, должен быть DNS сервер, который может отображать IPv4 адреса на IPv6 адреса. На хостах IPv6 и хостах IPv4 не требуется никаких модификаций. Для обеспечения масштабируемости на сайте могут быть инсталлированы несколько выделенных устройств с несколькими фиктивными IPv6 префиксами. Трафик UDP может ретранслироваться тем же самым методом, что и трафик TCP.

*DSTM (Dual Stack Transition Mechanism)*. Механизм перехода на базе двойного стека DSTM, описанный в Internet-Draft "Dual Stack Transition Mechanism (DSTM)", представляет собой механизм, который позволяет узлу с двойным стеком, у которого стек IPv4 разрешен, но еще не сконфигурирован, временно получать IPv4 адрес для связи с приложениями, работающими только по IPv4. Основная идея заключается в том, что когда IPv4/IPv6 хосту требуется IPv4 адрес, он запрашивает его на время связи с DSTM сервером. Связь с DSTM сервером осуществляется поверх IPv6. При отсутствии внутренней IPv4 инфраструктуры маршрутизации хост с двойным стеком будет инкапсулировать IPv4 пакеты в пакеты IPv6 до конечной точки туннеля, которая будет их декапсулировать и вводить в инфраструктуру IPv4. Эта инкапсуляция выполняется с помощью виртуального интерфейса DTI. Связь, иницируемая из узла IPv4 к узлу DSTM, стек IPv4 которого еще не сконфигурирован, в данный момент не поддерживается, но будет определена в будущей работе.

*ISATAP (Intra-Site Automatic Tunnel Addressing Protocol)*. Этот простой механизм позволяет инкрементно внедрять IPv6 внутри IPv4 сетей. ISATAP позволяет двухстековым узлам, которые не имеют прямого соединения с IPv6 маршрутизатором, автоматически туннелировать пакеты к точке входа в IPv6 сеть через IPv4 инфраструктуру. [1,4]

## **Заключение**

Новый протокол позволит присвоить IP-адреса множеству устройств – от бытовых приборов до автомобилей, систем безопасности и идентификационных карт, обеспечит возможность дистанционного управления всем этим оборудованием. Кроме этого, он позволит существенно "расширить" сеть Internet, обеспечит поддержку новых приложений и служб, коммуникационных технологий и телефонных (сотовых) сетей 3G, что в конечном итоге приведет к еще более существенному росту массовости и популярности Сети, доступности ее услуг.

## **Литература**

1. Паркер Т., Сиян К. TCP/IP. Для профессионалов. 3-е изд. 2003.
2. С. Фейт. TCP/IP. Архитектура, протоколы, реализация (включая IP версии 6 и IP Security). 2000.
3. Стивенс У. Р. Протоколы TCP/IP. Практическое руководство. 2002.
4. Кларк Э. IPv6 - адресное пространство навсегда? 1999.

# РАСПРЕДЕЛЕНИЕ ВИРТУАЛЬНОЙ ПАМЯТИ В СРЕДЕ WIN32. УПРАВЛЕНИЕ МАЛЫМИ БЛОКАМИ

С. В. Антонов

Научный руководитель – к.т.н., доц. М. Ю. Юдин

Основным предметом статьи является рассмотрение эффективного управления памятью средствами приложения.

## Введение

Методы, которыми операционная система организует работу с памятью, подробно описаны в [1]. В большинстве случаев при необходимости многократного выделения сравнительно небольших областей памяти приложения используют динамическую память, а именно стандартную кучу<sup>1</sup> процесса. В малых и средних проектах, если не предполагается обработка больших объемов информации, на работе с памятью внимание вообще не заостряется, и все управление памятью передается библиотечным классам, которые в основном используют ту же стандартную кучу.

Основным недостатком данного подхода является то, что вследствие частого выделения блоков разной величины куча сильно фрагментируется. Разумеется, современные операционные системы борются с фрагментацией. Однако ее полное предотвращение весьма затруднительно, так как для этого требуется дополнительное место и время, причем система не знает, блоки какого объема потребуются приложению в дальнейшем. На рис. 1 проиллюстрирована типичная ситуация.



Рис. 1 Фрагментация памяти при нахождении данных разных типов в одном блоке

Для начала предположим, что куча полностью занята двумя объектами некоторого типа по 100 байт каждый и строками по 200 байт. Далее из кучи удаляется 2 объекта и высвобождается по 100 байт памяти. Однако, если теперь произвести попытку записи строки размером 200 байт, такая попытка будет неудачной. Это произойдет потому, что данные должны быть записаны единым блоком и не могут быть разделены на несколько частей<sup>2</sup>. Такое явление называется фрагментацией свободного пространства.

<sup>1</sup> Куча является зарезервированной областью адресного пространства, из которой происходит выделение физической памяти, когда её запрашивает программа.

<sup>2</sup> Так сделано, потому что иначе надо было бы обеспечить механизм, который отслеживал бы окончание фрагмента данных и обеспечивал переход на следующий фрагмент. Однако при его применении возможны значительные накладные расходы.

## Решение проблемы фрагментации

Для предотвращения фрагментации в современные версии ОС Windows была добавлена возможность создания дополнительных куч. Соответственно, для предотвращения описанной ситуации можно создавать отдельные кучи для каждого типа данных. Однако при создании куч для большого количества объектов появляется множество областей разного размера, что, в конечном счете, приводит к фрагментации пространства между кучами, учитывая тот факт, что размер кучи может меняться.

В данной работе предлагается следующий выход из этой ситуации. Так как внутри куч пространство непрерывно, а вне их пространство фрагментарно (рис. 2), то способом решения данной проблемы является централизация выделения памяти.

Адресное пространство процесса (область динамической памяти)

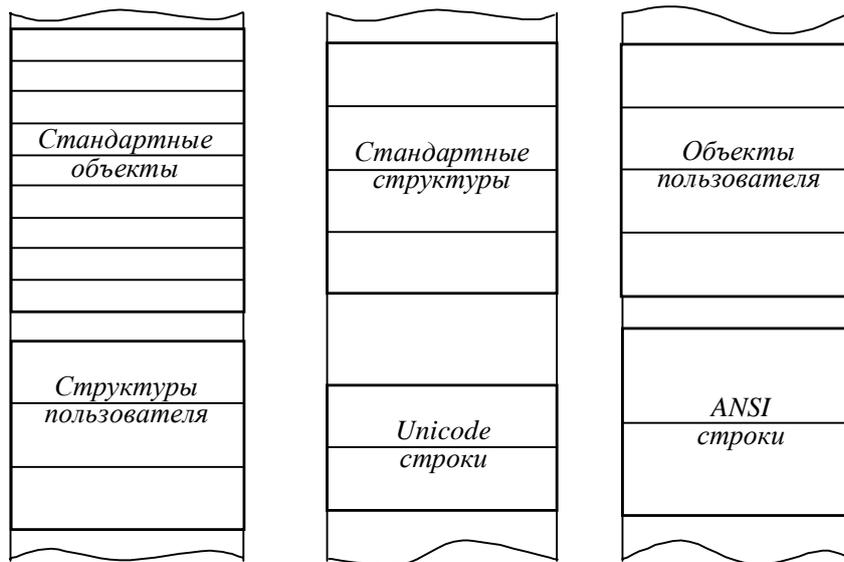


Рис. 2. Фрагментация при распределении куч с данными разных типов

Такой подход заключается в том, что первоначально при помощи механизма виртуальной памяти программа выделяет большой блок в адресном пространстве процесса, а затем распределяет его между частями кода, требующими небольших блоков памяти. В первом приближении такой подход может показаться неоправданным, однако, когда число областей с данными различных типов действительно велико, фрагментация становится чувствительной. Программу или часть программы, следящей за памятью, будем называть сервером памяти.

Структура сервера памяти имеет вид, представленный на рис. 3.

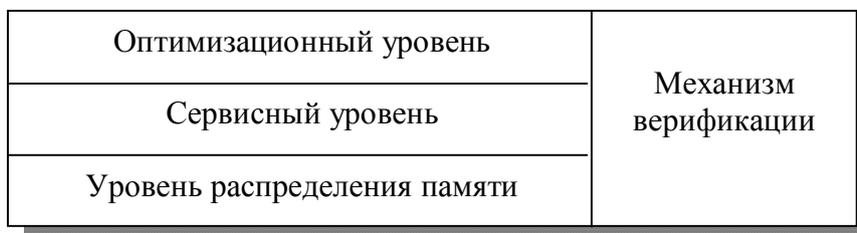


Рис. 3. Схема сервера управления памятью

Уровень распределения памяти, реализованный на основе низкоуровневых функций операционной системы, выполняет резервирование блоков и передачу им физической памяти. Также он отвечает за независимость сервера памяти от операционной системы.

Сервисный уровень, в свою очередь, распределяет память внутри выделенного региона, предотвращая ее фрагментацию и размещая блоки в соответствии с частотой доступа, их размером и рядом других параметров, тем самым обеспечивая их оптимальное расположение.

В задачи оптимизационного уровня входит наиболее оптимальное связывание сервера памяти с текущим проектом. Как следствие, его реализация зависит от конкретного проекта и учитывает все его особенности.

Механизм верификации отвечает за корректное связывание уровней и осуществляет проверку данных, передаваемых между ними. Кроме того, поскольку механизм верификации имеет доступ ко всем уровням сервера, он наилучшим образом подходит для обработки исключительных ситуаций, связанных с выделением памяти.

Преимуществами централизованного управления памятью являются:

- отсутствие фрагментации адресного пространства;
- снижение фрагментации физической памяти;
- независимость механизма выделения памяти от операционной системы;
- централизованное слежение за выполнением задач;
- снижение времени выделения памяти.

К недостаткам такого подхода можно отнести следующее:

- необходимость самостоятельной реализации множества библиотечных функций;
- затраты, связанные с загрузкой и нахождением сервера в оперативной памяти.

Как можно заметить, недостатка всего два, но они довольно существенны, в особенности первый. По этой причине данный подход целесообразен только в больших проектах, где рациональное распределение памяти действительно важно. Но, разумеется, переписывать для каждого проекта все функции runtime библиотеки, отвечающие за выделение памяти, слишком трудоемко и нерационально.

По этой причине сервер разделен на независимые уровни, и, изменив только оптимизационный уровень, сервер можно подключить к другому проекту. Причем важно, что не обязательно реализовывать универсальный механизм, работающий со всеми стандартными типами данных. Наиболее перспективными являются две реализации. В первой задачей механизма является распределение строковых данных, ведь именно они в первую очередь размещаются в динамической памяти. Вторая реализация подразумевает распределение большого числа специфических объектов, особенно если они велики.

## **Заключение**

Рассмотренный подход вполне оправдан, если программа активно использует динамическую память, а время ее работы достаточно велико. Дополнительную информацию по этой теме можно найти в [2], где подробно описываются режимы работы сервера памяти и рассматриваются причины необходимости его создания.

## **Литература**

1. Дж. Рихтер, WINDOWS для профессионалов: Создание эффективных Win32-приложений с учетом специфики 64-разрядной версии Windows. / Пер, англ. - 4-е изд. - СПб; Питер; М.: Издательско-торговый дом "Русская Редакция", 2001. - 752 с.; ил.;
2. Кашинский Д.В. Цикл статей «Управление памятью» на сайте <http://www.wasm.ru>
3. Walter O., Programming the Microsoft Driver Model, Microsoft Press, 1999
4. А. Мешков, Ю. Тихомиров, VISUAL C++ и MFC. Программирование для WINDOWS NT и WINDOWS 95. В двух томах.

# ПРОЕКТИРОВАНИЕ МАТЕМАТИЧЕСКОГО ОБЕСПЕЧЕНИЯ САПР СИСТЕМ АВТОМАТИЗАЦИИ ПРОИЗВОДСТВА ОПТИЧЕСКИХ МАТЕРИАЛОВ НА ОСНОВЕ НЕЙРОСЕТЕВЫХ КОМПОНЕНТОВ

К.В. Богданов

Научный руководитель – д.т.н., проф. А.Г. Коробейников

Проанализирована структура задач проектирования математического обеспечения САПР систем автоматизации производства оптических материалов (МО САПР САПОМ). Предложена схема и единообразное алгоритмическое обеспечение системы автоматизированного проектирования МО САПР САПОМ.

## Введение

Задача проектирования математического обеспечения сложных САПР является одной из ключевых проблем автоматизации проектирования. Эти задачи многоплановы и, как правило, требуют разнородного математико-алгоритмического обеспечения, что как затрудняет проектирование МО САПР САПОМ, так и удлиняет во времени сами проектные работы.

В данной работе предложено построение единой интегрированной системы проектирования МО САПР САПОМ на основе нейросетевых компонентов.

## 1. Структура задач МО САПР САПОМ

Рассмотрим структуру задач МО САПР САПОМ с точки зрения характерных особенностей и сложности решения.

### 1. Идентификация технологических процессов и систем

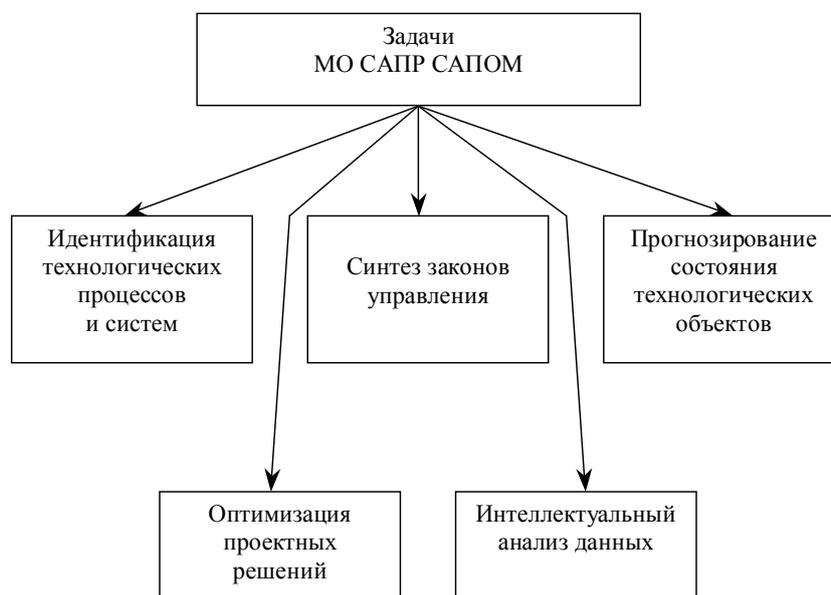


Рис. 1. Структура задач МО САПР САПОМ

При построении автоматизированных технологических комплексов (АТК) производства оптических материалов необходимо получение большого количества моделей технологических процессов (ТП) и технологических систем (ТС) в целом. Получение таких моделей в аналитической форме затруднено сложностью фазовых физико-химических процессов [5] и отсутствием подходящего математического аппарата. В данной ситуации требуется применение специальных методов идентификации.

## **2. Прогнозирование состояния технологических объектов и систем**

Задача имеет большое значение. Поведение технологических объектов и систем, особенно на опасных производствах и на установках со сложным характером протекания ТП, нуждается в прогнозировании. В частности, в последнее время выделяются два круга таких задач. Это задачи прогнозирования качества выходного продукта (что крайне актуально для ТП со сложными физико-химическими и фазовыми превращениями, к которым относятся и ТП производства оптических материалов) и задачи анализа предкатастрофной динамики (что актуально для наладки и поддержания работоспособности сложных АТК оптического производства, особенно непрерывного цикла).

Задача прогнозирования состояния во многом схожа с задачей идентификации, однако имеет свои отличия. В частности, для сложных систем с большим количеством параметров прогнозирование чрезвычайно затруднено. Вместе с тем именно для таких систем её решение представляет наибольший интерес.

## **3. Синтез законов управления**

Необходимым условием для этого является получение в явном или неявном виде модели объекта управления (см. задачи 1 и 2). Как мы уже отмечали, для сложных фазовых ТП с неочевидной динамикой параметров крайне сложно получить аналитическое описание процесса, а значит – и синтезировать соответствующий управляющий регулятор.

## **4. Оптимизация проектных решений**

Традиционной задачей МО САПР являются оптимизационные задачи. Они могут носить как изошрёпный характер, так и иметь достаточно простую структуру. Однако для оптимизации проектных решений сложных АТК оптического производства даже простые задачи имеют очень большую размерность. В условиях применения классических методов и алгоритмов оптимизации это приводит к неприемлемым затратам вычислительных мощностей или времени. Даже решение элементарных по структуре, "школярских" задач оптимизации (к примеру, всего семейства транспортных) может потребовать значительного машинного времени, что отрицательно скажется на скорости проектирования АТК в целом.

## **5. Интеллектуальный анализ данных**

Сложные САПР опираются на базы типовых решений и типовых элементов. Однако, помимо этого, при проектировании нередко приходится отбрасывать неявные или трудно формализуемые закономерности в параметрах проектируемого объекта. Безусловно, то или иное усечение входных параметров проектирования всегда можно правдоподобно обосновать. Однако можно ли при этом поручиться за то, что эти параметры не повлияют нештатным образом на объект в ходе эксплуатации? Эта опасность тем выше, чем более сложные ТС мы проектируем. Поэтому одной из важнейших задач МО САПР САПОМ является наработка алгоритмов для определения неявных корреляций между параметрами проектируемого объекта (в т.ч. в иерархии его подсистем, что особенно актуально для АТК).

## **2. Нейронные сети как единый инструмент создания МО САПР САПОМ**

Разнообразие задач МО САПР САПОМ требует для их решения достаточно широкого набора методов, моделей и алгоритмов из различных областей прикладной математики. Всё вместе это сильно затрудняет реализацию интегрированных систем проектирования МО САПР САПОМ и в конечном итоге увеличивает время на разработку самих САПОМ.

Нашей задачей было показать возможность построения единообразного математического аппарата для системы проектирования МО САПР САПОМ. В качестве такого аппарата были предложены нейронные сети.

Теоретически аппарат нейронных сетей неплохо разработан, существует большой пласт работ по их практическому внедрению. Так, нейронные сети идеально приспособлены для идентификационных задач – в тех условиях, когда проектировщик накопил достаточно представительную выборку параметров "вход-выход".

Тесно связанный с идентификационными задачами класс задач нейросетевого управления тоже получил развитие. Так, адаптивные нейроконтроллеры широко применяются для управления динамическими объектами.

Более подробно процесс создания нейросетевых систем управления описан, например, в [2, 3, 8]. Была предложена схема адаптивного управления для процессов выращивания оптического кристалла и варки оптического стекла [4].

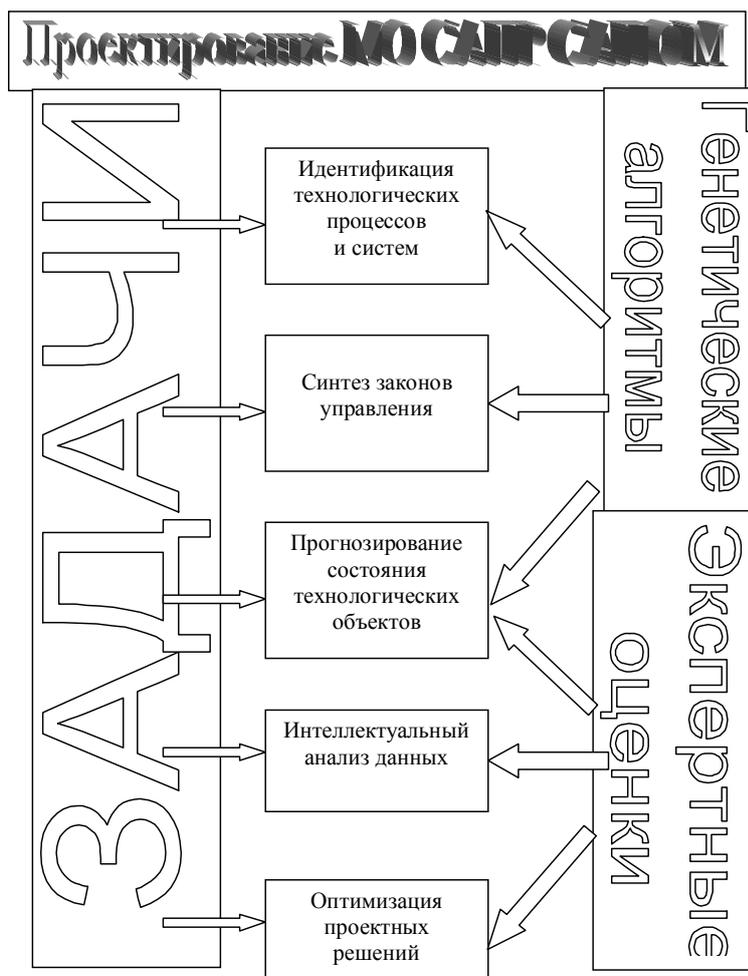


Рис. 2. Методы проектирования МО САПР САПОМ

Для оптимизационных задач применимы рекуррентные нейронные сети Хопфилда [1]. В процессе эволюции хопфилдовской сети уменьшается так называемая "функция энергии", переход которой к минимуму означает перевод всей системы к искомому оптимальному значению [9].

Более подробно практическое решение задач оптимизации с помощью нейронных сетей и проблематика конструирования нейросетевых оптимизирующих систем рассмотрены, например, в [6, 7, 9].

Прогнозные и экспертные нейронные сети имеют наиболее простую архитектуру. Как правило, это многослойные перцептроны [9].

Проектирование и обучение нейронной сети для решения той или иной задачи МО САПР САПОМ – задача для отдельного исследования. Подробно часть из этих методов рассмотрена, например, в [1,2,6,7].

## **Заключение**

Проанализирована структура задач МО САПР САПОМ и основные сложности на пути его проектирования. Предложена единая система проектирования МО САПР САПОМ на основе аппарата теории нейронных сетей.

## **Литература**

1. Hopfield J.J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities, in Proc. National Academy of Sciences, USA 1982.
2. Neural networks for control / Miller W.T., Sutton R.S., Werbos P. J., Eds., The MIT Press, 1990.
3. Антонов В.Н., Терехов В.А., Тюкин И.Ю. Адаптивное управление в технических системах. – СПб, Изд-во Санкт-Петербургского университета, 2001.
4. Богданов К.В. К применению нейросетевых алгоритмов при автоматизации технологических процессов оптического производства. // В кн. "Диагностика и функциональный контроль качества оптических материалов. Научный сборник под ред. д.т.н. проф. Ю.А.Гатчина и д.т.н. проф. В.Л. Ткалич". – СПб.: СПбГУ ИТМО, 2004
5. Гатчин Ю.А. Принципы построения и методы проектирования интегрированной системы автоматизированного производства оптических материалов. // Автореф. докторской диссертации, СПб, СПбГИТМО(ТУ), 2001.
6. Ефимов В.В. Нейроподобные сети в бортовых информационно-управляющих комплексах летательных аппаратов. Решение оптимизационных задач. – СПб., 1996.
7. Назаров А.В., Лоскутов А.И. Нейросетевые алгоритмы прогнозирования и оптимизации систем. – СПб., Наука и техника, 2003.
8. Терехов В.А., Ефимов Д.В., Тюкин И.Ю. Нейросетевые системы управления. – М.: Высшая школа, 2002.
9. Терехов С.А. Лекции по теории и приложениям искусственных нейронных сетей. – Снежинск: ВНИИТФ, 1994.

# ОПТИМИЗАЦИЯ ПРОЕКТНЫХ РЕШЕНИЙ В САПР НА ОСНОВЕ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ

К.В. Богданов

Научный руководитель – д.т.н., проф. А.Г. Коробейников

Рассмотрены показания к решению транспортной задачи в САПР автоматизированных технологических комплексов (АТК). Предложено использование для решения оптимизационных задач САПР АТК искусственных нейронных сетей (ИНС) Хопфилда. Исследована временная емкость стандартных алгоритмов в сравнении с предложенным методом.

## Введение

Задача автоматизации оптимального проектирования сложных АТК требует большого объема оптимизационных процедур. Нередко это достаточно известные задачи (транспортные задачи или задачи размещения ресурсов, разновидности задачи коммивояжера, задача целераспределения, задача "китайского почтальона" [3]). Однако вычислительная и временная емкость таких алгоритмов в случаях, типичных для проектирования и оптимизации АТК, очень велика. Это затрудняет использование классических методов оптимизации. Для решения группы оптимизационных задач в САПР АТК было предложено использование искусственных нейронных сетей (ИНС) Хопфилда, показавших свою способность к эволюции в сторону минимума энергии [1, 6, 7]. Особенности нейронных сетей позволяют создавать системы с распараллеленными вычислениями [2, 4, 8], что является перспективным, но пока недостаточно изученным направлением в автоматизации проектирования АТК. В данной работе исследуется решение оптимизационной транспортной задачи с помощью ИНС Хопфилда.

## 1. Формулировка транспортной задачи

В наиболее общем виде транспортная задача может быть сформулирована как нахождение матрицы  $U = \left\| u_{i,j} \right\|_{n \times m}$ , называемой планом, где

$n$  – количество входов модели ("производств" в классической трактовке);

$m$  – количество выходов модели ("складов" в классической трактовке);

$u$  – количество искусственных путей ("единиц продукции" в классической трактовке).

Сутью задачи является минимизация выражения

$$\sum_{i=1}^n \sum_{j=1}^m u_{ij} \Delta t_{ij} \rightarrow \min \quad (1)$$

где

$$\Delta t = \left\| \Delta t_{ij} \right\|_{n \times m} = \left\| \begin{array}{c} \Delta t_{11} \cdots \Delta t_{1m} \\ \cdots \cdots \cdots \\ \Delta t_{n1} \cdots \Delta t_{nm} \end{array} \right\| \quad (2)$$

– матрица весов искусственных путей (транспортных издержек).

Искомые количества искусственных путей  $u$  должны удовлетворять следующим условиям:

$$\begin{cases} u_{ij} \geq 0 \\ \sum_{j=1}^m u_{ij} = c_i, \quad i = [1, n] \\ \sum_{i=1}^n u_{ij} = c_j, \quad j = [1, m] \end{cases} \quad (3)$$

где первое – условие неотрицательности количества искусственных путей, второе и третье – тождественность количества исходящих/входящих путей балансным числам модели. [3]

## 2. Решение транспортной задачи с помощью рекуррентной ИНС Хопфилда

Рассмотрим решение транспортной задачи с помощью рекуррентной ИНС Хопфилда. Будем считать время системы непрерывным.

Определим структуру ИНС, решающей задачу (2) в условиях ограничений (3). Пусть задана нейронная сеть, порождающая матрицу размерностью  $n \times m \times c_{ij}$ ,  $i = [1, n]$ ,  $j = [1, m]$ ,  $k = [1, c_{ij}]$ , где  $c_{ij} = \min(c_i, c_j)$ . Динамика такой ИНС может быть описана следующей системой дифференциальных уравнений [3]:

$$\frac{dU_{ijk}}{dt} = -\frac{U_{ijk}}{\tau_{ijk}} + \sum_{\mu=1}^n \sum_{v=1}^m \sum_{\chi=1}^{c_{ij}} T_{ijk,\mu\nu\chi} V_{\mu\nu\chi} - I_{ijk} \quad , \quad (4)$$

где

$$V_{ijk} = g(U_{ijk}); V_{ijk}(t_0) = g(U_{ijk}(t_0)) = V_{ijk}^{(0)} \quad , \quad (5)$$

$T$  – матрица синаптических коэффициентов,  $V$  – выходной сигнал нейрона,  $I$  – начальное смещение.

Выбор активационной функции нейронов зависит от характерных особенностей конкретной задачи. Обычно наилучшие результаты дает сигмоидная функция (рис. 1) [2–4, 6]:

$$g(U_{ijk}) = \frac{1}{1 + e^{-\lambda U_{ijk}}} \quad , \quad (6)$$

где  $\lambda$  – константа, определяющая вид сигмоиды. Ценным свойством сигмоидных активационных функций является их способность к усилению слабых сигналов одновременно с предотвращением насыщения от сильных сигналов. Это связано с особенностями наклона кривой.

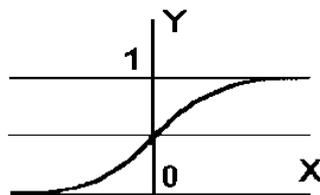


Рис.1. Сигмоидная активационная функция искусственного нейрона

Теперь поставим каждой целочисленной переменной  $u_{ij}$  в соответствие выходной сигнал  $V_{ijk}$   $ijk$ -го нейрона:

$$u_{ijk} = k \Leftrightarrow \sum_{k=1}^{c_{ij}} V_{ijk} = k \quad . \quad (7)$$

Опираясь на (7), запишем (1) и (3) в нейросетевом базисе. Целевая функция (1) представима в виде

$$\phi(V) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{c_{ij}} V_{ijk} \Delta t_{ij} \rightarrow 0. \quad (8)$$

Второе условие из системы (3) требует, чтобы в  $i$ -м слое возбуждалось не более  $c_i$  нейронов:

$$\sum_{j=1}^m \sum_{k=1}^{c_{ij}} V_{ijk} = c_i, \quad \text{где } V_{ijk} \in \{0,1\} \quad \forall i, j, k. \quad (9)$$

Такое же ограничение можно записать и для третьего условия системы (3):

$$\sum_{i=1}^n \sum_{k=1}^{c_{ij}} V_{ijk} = c_j, \quad \text{где } V_{ijk} \in \{0,1\} \quad \forall i, j, k. \quad (10)$$

Теперь соберем "функцию энергии" ИНС, минимизация которой и приведет нас к искомому решению задачи (8) с ограничениями (9) и (10), т.е. к решению исходной транспортной задачи. Будем строить функцию в виде [5]

$$E^0(V) = E_\phi^0(V) + E_\psi^0(V) + E_\varphi^0(V), \quad (11)$$

где первое слагаемое оптимизирует функцию времени и с точностью до константы  $A > 0$  однозначно определяется следующим образом:

$$E_\phi^0(V) = A \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{c_{ij}} V_{ijk} \Delta t_{ij}. \quad (12)$$

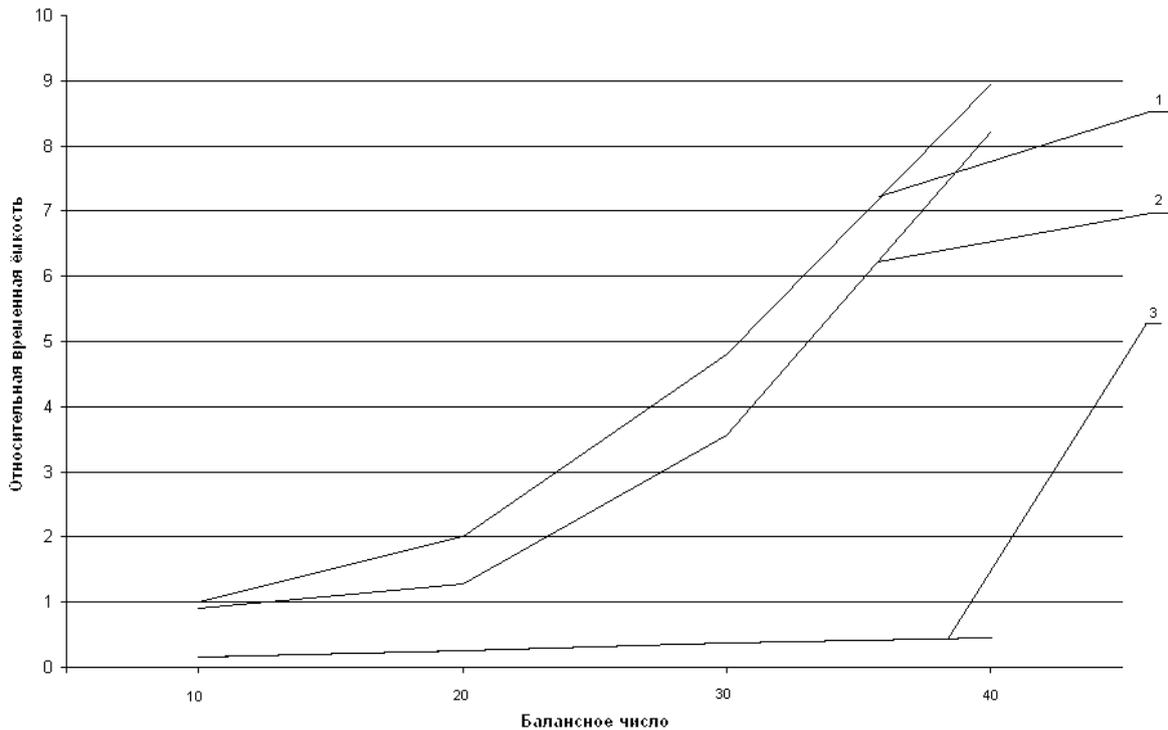


Рис. 2. Относительная временная емкость алгоритмов решения транспортной задачи (1 – метод простого перебора, 2 – венгерский метод, 3 – метод с использованием оптимизирующей ИНС Хопфилда)

Второе слагаемое обеспечивает выполнение ограничений и имеет следующий вид [5]:

$$E_\psi^0(V) = B \sum_{i=1}^n \left( \sum_{j=1}^m \sum_{k=1}^{c_{ij}} V_{ijk} - c_i \right)^2 + C \sum_{j=1}^m \left( \sum_{i=1}^n \sum_{k=1}^{c_{ij}} V_{ijk} - c_j \right)^2, \quad (13)$$

где  $B$  и  $C$  – неотрицательные константы. Выражение (13) принимает минимальное (и равное нулю) значение тогда и только тогда, когда выполняются условия (9) и (10).

Третье слагаемое связано с требованием бинарности выходных сигналов нейронов, оно минимизируется на участках, где выходные сигналы нейронов принимают значения 0 и 1 [5]:

$$E_{\varphi}^0(V) = D \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{c_{ij}} (v_{ijk} (v_{ijk} - 1)). \quad (14)$$

Экспериментальное исследование показало высокую эффективность ИНС Хопфилда в качестве средства решения транспортной задачи. На рис. 2 приведены графики относительной временной емкости двух стандартных решений задачи типовой транспортной задачи (в качестве иллюстрации выбраны метод простого перебора и венгерский метод) и ее решения с помощью ИНС Хопфилда (для различных балансных чисел модели). В качестве вычислительного средства использовался компьютер IBM PC с процессором Intel Pentium4 2800. Относительная временная емкость нормирована на время работы алгоритма простого перебора при балансном числе 10.

Заметим, что ИНС Хопфилда в первую очередь дает существенный выигрыш времени на решении задач большой размерности, что является показанием к применению подобных методов при оптимизации проектных решений в САПР сложных АТК.

## Заключение

В ходе работы рассмотрены основные показания к решению оптимизационной транспортной задачи с помощью ИНС Хопфилда. Транспортная задача интерпретирована в нейросетевом базисе. Практически обоснована эффективность подобного решения.

## Литература

1. Hopfield J.J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities, in Proc. National Academy of Sciences, USA 1982.
2. Нейроинформатика / Горбань А.Н., Дунин-Барковский В.Л., Кирдин А.Н. и др. – Новосибирск: Наука (Сибирское отделение), 1998.
3. Ефимов В.В. Нейроподобные сети в бортовых информационно-управляющих комплексах летательных аппаратов. Решение оптимизационных задач. – СПб., 1996.
4. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. – М.: Горячая Линия – Телеком, 2001.
5. Назаров А.В., Лоскутов А.И. Нейросетевые алгоритмы прогнозирования и оптимизации систем. – СПб., Наука и техника, 2003.
6. Терехов С.А. Лекции по теории и приложениям искусственных нейронных сетей. – Снежинск: ВНИИТФ, 1994.
7. Терехов С.А. Типовые задачи для информационного моделирования с использованием нейронных сетей. – Снежинск, НИИТФ, 2000.
8. Уоссермен Ф. Нейрокомпьютерная техника. Теория и практика. – М., 1992.

# МЕТОД ЭФФЕКТИВНОГО МОДЕЛИРОВАНИЯ ВЫБОРКИ ИСТОЧНИКОВ ДВОИЧНОЙ ИНФОРМАЦИИ

В.В. Семенюк

Научный руководитель – д.т.н., проф. В.Н. Васильев

Рассматривается новый высокоэффективный метод адаптивного кодирования выборки источника двоичной информации без памяти. Он предложен в качестве альтернативы наиболее часто используемому на практике методу адаптивного кодирования с периодическим масштабированием счетчиков появления символов.

## Постановка задачи

В последнее время арифметическое кодирование [1–4] все чаще вытесняет другие методы генерации кода переменной длины из реализаций различных методов экономного кодирования информации. На данном этапе мы обладаем необходимыми вычислительными ресурсами, что позволяет применять на практике сложные вероятностные модели, являющиеся основой для получения арифметического кода. Наиболее яркими примерами использования арифметического кодирования являются всевозможные методы кодирования коэффициентов дискретного косинусного преобразования, а также коэффициентов частотных диапазонов вейвлет-преобразования [5–8].

Во многих случаях вероятностное моделирование сводится к моделированию процессов порождения независимой двоичной выборки. Это означает, что достаточно часто мы можем ограничиться решением проблем, связанных с обработкой выхода информационных источников с двоичным алфавитом, состоящим из символов "0" и "1", не обладающих памятью. В подобной постановке задача моделирования может быть эффективно решена на практике. Так, разработчиками компании IBM был предложен алгоритм, позволяющий осуществлять высокопроизводительное моделирование выборки источников двоичной информации и генерацию кода на основе построенных моделей (так называемый *Q-coder* [9]). К сожалению, предложенная схема кодирования не обладает универсальностью и имеет некоторые ограничения, приводящие к потерям в эффективности кодирования. Чаще всего при разработке нового алгоритма экономного кодирования мы вынуждены прибегать к использованию упрощенных, но, в то же время, более гибких решений. Данные решения подчас обладают большей вычислительной сложностью по сравнению с хорошо оптимизированным по скорости алгоритмом *Q-coder*, однако они позволяют осуществлять как более точное моделирование, так и более эффективную генерацию кода.

## Стандартный метод моделирования

Рассмотрим схему, обычно применяемую на практике. (Существуют также иные методы моделирования. Один из наиболее удачных альтернативных методов изложен в [10].) Как известно, для генерации кода на основе арифметического кодирования необходимо знать вероятность появления кодируемого символа и формальное расположение соответствующего вероятностного интервала в базовом интервале вероятностей  $[0, 1)$  (см. [1]). На практике вероятность заменяется частотой появления символа, а длина базового интервала оказывается равной суммарной частоте появления всех символов алфавита. Таким образом, для случая двоичного алфавита нам достаточно хранить частоту появления символа "0" ( $f_0$ ) и суммарную частоту появления символов "0" и "1" ( $f_\Sigma$ ). Генерация арифметического кода сводится к интервальному проектированию, во время которого осуществляется деление внутреннего интервала арифметического кодера на суммарную частоту  $f_\Sigma$  с последующим домножением результата деления на частоту появления нулевого символа  $f_0$ . Как видно, недостаток такого решения заключа-

ется в использовании ресурсоемких операций умножения и деления, причем операция деления с вычислительной точки зрения является наиболее сложной операцией. Увеличение производительности в данном случае достижимо за счет использования таблиц с результатами деления, однако такое решение сопряжено с наложением серьезных ограничений на максимальную величину  $f_{\Sigma}$ .

Другая проблема общепринятой на практике схемы арифметического кодирования заключается в методе поддержания и обновления вероятностной модели. С появлением очередного символа суммарная частота  $f_{\Sigma}$ , а в случае, когда появляется символ "0", также частота  $f_0$  должны быть увеличены на некоторую величину  $\delta$ . При этом необходимо помнить, что счетчики появления символов, которые содержат значения  $f_0$  и  $f_{\Sigma}$ , имеют некоторые предельные значения, которые могут зависеть как от особенностей реализации интервального проектирования (например, упомянутой выше), так и от объема памяти, отводимого под тот или иной счетчик. Стандартным решением проблемы переполнения счетчиков является их периодическое масштабирование: если  $f_{\Sigma}$  достигает некоторого предельного значения  $f_{\max}$ , значения  $f_0$  и  $f_{\Sigma}$  делятся на константу, в качестве которой чаще всего выступает число 2 (минимальное масштабирующее число, деление на которое можно заменить операцией сдвига).

Для формализации вышесказанного обозначим через  $f_0^{(i)}$  и  $f_{\Sigma}^{(i)}$  значения соответствующих счетчиков после  $i$ -го масштабирования. Через  $\Delta f_0^{(i)}$  обозначим количество появлений символа "0" во временном промежутке между  $(i-1)$ -м и  $i$ -м масштабированиями. Тогда для всех  $i$ , отличных от нуля, справедливы формулы:

$$f_0^{(i)} = \frac{f_0^{(i-1)} + \delta \Delta f_0^{(i)}}{2}, \quad (1)$$

$$f_0^{(i)} = \frac{f_0^{(0)}}{2^i} + \delta \sum_{j=1}^i \frac{\Delta f_0^{(j)}}{2^{i-j+1}}, \quad (2)$$

$$f_{\Sigma}^{(i)} = \frac{f_{\max}}{2}.$$

В формулах фигурируют три параметра, значение которых необходимо установить перед началом кодирования:  $f_0^{(0)}$  – начальное значение счетчика появления символа "0",  $\delta$  – величина изменения значения счетчиков при появлении очередного символа,  $f_{\max}$  – предельное значение счетчика суммарной частоты появления символов. Есть также четвертый параметр, подлежащий установке, – начальное значение счетчика суммарной частоты появления символов  $f_{\Sigma}^{(0)}$ . Изменение указанных четырех параметров влияет на эффективность и производительность кодирования. Значение  $f_{\max}$ , с одной стороны, определяет точность представления счетчиков: чем больше величина  $f_{\max}$ , тем точнее мы можем описать вероятностные характеристики кодируемой информации. С другой стороны, совокупность значений  $f_{\max}$  и  $\delta$  влияет на частоту масштабирования, от которой напрямую зависит такой немаловажный фактор, как скорость адаптации модели к вероятностным особенностям обрабатываемой информации: с уменьшением отношения  $f_{\max} / \delta$  скорость адаптации возрастает. При этом само по себе значение  $\delta$  характеризует степень влияния начальной статистики, представленной параметрами  $f_0^{(0)}$  и  $f_{\Sigma}^{(0)}$ , на результаты моделирования на начальном этапе работы модели: увеличение  $\delta$  уменьшает это влияние.

Практика показывает, что тщательной настройкой указанных параметров можно добиться заметного повышения эффективности кодирования или изменения его произ-

водительности (скорость работы алгоритма зависит от частоты масштабирования). Однако такая настройка, как правило, сильно затруднена. Как видно, приведенные параметры в достаточной мере взаимосвязаны между собой, поэтому изменение значения одного из них требует изменения значений других параметров, причем результат подчас непредсказуем (все во многом зависит от особенностей самой обрабатываемой информации). Кроме того, как уже отмечалось, конкретная реализация интервального проектирования вносит в настройку определенные ограничения.

Еще одним недостатком рассмотренной модели является не очень удачный метод адаптации. Вызывает нарекание тот факт, что адаптация происходит не плавно, а скачками, что обуславливается периодическим масштабированием, осуществляемым после обработки очередной, достаточно большой порции информации. Так как время масштабирования никак не связано с моментами изменений вероятностных особенностей обрабатываемой информации, адаптация оказывается не достаточно эффективной.

### Новый метод моделирования

В данной работе предлагается несколько иной способ моделирования выборки источника двоичной информации, который является более удачным решением по сравнению методом, изложенным выше. Основное отличие заключается в том, что мы отказываемся от частотного подхода и работаем исключительно с вероятностными оценками. Если выше описывался алгоритм изменения значений счетчиков частот появления символов  $f_0$  и  $f_\Sigma$ , в данном случае мы предлагаем метод динамического изменения в процессе кодирования оценки вероятности появления символа "0" ( $r_0$ ).

Предположим, что выборка источника некоторым образом поделена на небольшие порции  $\Delta f_\Sigma^{(i)}$ . Долю символов "0" в каждой из этих порций обозначим через  $\Delta f_0^{(i)}$  (Здесь мы сознательно выбираем ранее использованное обозначение, так как обозначаемый объект имеет прежний смысл.). Введем параметр адаптации для учета степени вклада  $i$ -й порции информации при вычислении очередной вероятностной оценки ( $\alpha^{(i)}$ ). (Подразумевается, что  $\alpha^{(i)}$  меняется в интервале  $[0, 1]$ .) Используя данные обозначения, новый подход можно формализовать так:

$$r_0^{(i)} = (1 - \alpha^{(i)}) r_0^{(i-1)} + \alpha^{(i)} \frac{\Delta f_0^{(i)}}{\Delta f_\Sigma^{(i)}}, \quad (3)$$

$$r_0^{(i)} = \left( \prod_{j=1}^i (1 - \alpha^{(j)}) \right) r_0^{(0)} + \sum_{j=1}^i \left( \prod_{k=j+1}^i (1 - \alpha^{(k)}) \right) \alpha^{(j)} \frac{\Delta f_0^{(j)}}{\Delta f_\Sigma^{(j)}}. \quad (4)$$

Здесь  $r_0^{(i)}$  - оценка вероятности появления символа "0" после обработки  $i$ -й порции информации.

Если положить  $\alpha^{(i)} = 1/2$  и  $f_\Sigma^{(i)} = f_{\max}/2\delta$ , формулы (3) и (4) превращаются в формулы (1) и (2). Таким образом, рассмотренный вначале подход является частным случаем предлагаемого. Однако это не означает, что новый метод обладает свойствами старого. Как будет показано ниже, новый метод моделирования на практике реализуется несколько иначе, поэтому ему не присущи некоторые недостатки рассмотренного выше метода.

### Практическая реализация

Основная задача при практической реализации предлагаемого метода – обеспечение хранения и обработки оценки вероятности появления символа "0"  $r_0^{(i)}$  и параметра

адаптации  $\alpha^{(i)}$  с достаточной степенью точности. Вместо вещественных  $r_0^{(i)}$  и  $\alpha^{(i)}$  мы будем работать с целочисленными значениями  $R_0^{(i)} = [r_0^{(i)} B_r]$  и  $A^{(i)} = [\alpha^{(i)} B_\alpha]$  (здесь  $B_r$  и  $B_\alpha$  – целочисленные основания для представления вещественных  $r_0^{(i)}$  и  $\alpha^{(i)}$  в целочисленной форме). Домножив обе части формулы (3) на  $B_r$ , а правую часть дополнительно домножив и поделив на  $B_\alpha$ , получим:

$$r_0^{(i)} B_r = \frac{(B_\alpha - \alpha^{(i)} B_\alpha) r_0^{(i-1)} B_r + \alpha^{(i)} B_\alpha \frac{\Delta f_0^{(i)}}{\Delta f_\Sigma^{(i)}} B_r}{B_\alpha}.$$

Отсюда получаем целочисленное приближение

$$R_0^{(i)} = \left[ \frac{(B_\alpha - A^{(i)}) R_0^{(i-1)} + A^{(i)} \frac{\Delta f_0^{(i)}}{\Delta f_\Sigma^{(i)}} B_r}{B_\alpha} \right]. \quad (5)$$

Для удобства вычислений выбираем величину порции выборки источника и основания целочисленного представления кратными степени двойки:  $\Delta f_\Sigma^{(i)} = 2^{S_\Sigma^{(i)}}$ ,  $B_r = 2^{S_r}$ ,  $B_\alpha = 2^{S_\alpha}$ . Тогда формула (5) может быть записана в следующей форме (знаки "<<" и ">>" обозначают операции левого и правого битового сдвига):

$$R_0^{(i)} = \left( (2^{S_\alpha} - A^{(i)}) R_0^{(i-1)} + (A^{(i)} \Delta f_0^{(i)} \ll (S_r - S_\Sigma^{(i)})) + 2^{S_\alpha - 1} \right) \gg S_\alpha. \quad (6)$$

Для того, чтобы использовать результаты, полученные при применении формулы (6), в арифметическом кодировании, в качестве частоты появления символа "0" надо брать значение  $R_0^{(i)}$ , а в качестве суммарной частоты появления символов - значение  $2^{S_r}$ . Здесь налицо одно из преимуществ рассматриваемой схемы: при осуществлении интервального проектирования операция деления может быть заменена операцией сдвига на величину  $S_r$ .

### Важный частный случай

Рассмотрим частный случай, когда все порции выборки состоят ровно из одного символа ( $\Delta f_\Sigma^{(i)} = 1$ ), а параметр адаптации не меняется с течением времени ( $A^{(i)} = A$ ). Формула (6) приобретает вид:

$$R_0^{(i)} = \begin{cases} \left( (2^{S_\alpha} - A) R_0^{(i-1)} + 2^{S_\alpha - 1} \right) \gg S_\alpha & \text{при } \Delta f_0^{(i)} = 0 \\ \left( (2^{S_\alpha} - A) R_0^{(i-1)} + (A \ll S_r) + 2^{S_\alpha - 1} \right) \gg S_\alpha & \text{при } \Delta f_0^{(i)} = 1 \end{cases}. \quad (7)$$

При осуществлении моделирования на основе формулы (7) подразумевается, что величины сдвигов  $S_r$  и  $S_\alpha$ , определяющие основание целочисленного представления, изначально фиксируются с учетом особенностей программной реализации алгоритма. Таким образом, параметрами, подлежащими установке в расчете на конкретный тип информации, являются:  $R_0^{(0)}$  – начальная оценка вероятности появления символа "0" и  $A$  – параметр адаптации. По сравнению с рассмотренным выше методом нам удалось сократить количество свободных параметров с четырех до двух, причем эти два параметра являются независимыми, что упрощает процесс их настройки. Достоинством предлагаемой схемы также является очень плавная адаптация модели, которая производится после обработки каждого символа.

Попытаемся оценить производительность нового метода моделирования. При установленном параметре  $A$  переменными в формуле (7) оказываются только величины  $R_0^{(i-1)}$  и  $\Delta f_0^{(i)}$ . Если мы предварительно вычислим значения  $2^{S_\alpha - 1}$ ,  $(2^{S_\alpha} - A)$  и  $(A \ll S_r) + 2^{S_\alpha - 1}$ , каждая итерация сведется к выполнению одной операции умножения, одной операции сложения и одной операции сдвига. Заметим, что здесь можно избежать проверки каких бы то ни было условий, так как выбор одной из двух альтернатив  $\Delta f_0^{(i)} = 0$  или  $\Delta f_0^{(i)} = 1$ , может быть сведен к чтению из таблицы нужного значения слагаемого, фигурирующего в формуле (7):  $2^{S_\alpha - 1}$  или  $(A \ll S_r) + 2^{S_\alpha - 1}$ .

### Практическая эффективность

Для оценки практической эффективности нового метода мы использовали его, а также описанный ранее метод в схеме кодирования коэффициентов частотных диапазонов дискретного вейвлет-преобразования. Значение  $S_r$  было выбрано равным 16, а значение  $S_\alpha$  – равным 8 (точность представления вероятности, таким образом, составляет 1/65536, а точность представления параметра адаптации – 1/256). Для моделей обоих описанных типов были очень тщательно подобраны универсальные параметры (универсальные на множестве тестовых данных), определяющие начальную статистику и характеристики адаптации. В результате предложенный метод оказался более эффективным по сравнению со стандартным методом и более быстрым, даже несмотря на то, что при реализации стандартного метода нам удалось избежать операций деления путем использования заранее подготовленных таблиц результатов вычислений. Эффективность кодирования коэффициентов увеличилась на 0.5%, а скорость возросла более чем на 10% (была использована вычислительная платформа на основе процессора Pentium III). Учитывая тот факт, что кодирование коэффициентов не исчерпывалось кодированием только двоичной выборки, увеличение показателей в процентном соотношении оказывается даже более существенным.

### Литература

1. Семенюк В. В. Экономное кодирование дискретной информации. СПб: СПб ГИТМО (ТУ), 2001.
2. Witten I. H., Neal R. M., Cleary J. G. Arithmetic coding for data compression // CACM. 1987. Vol. 30, N. 6. P. 520–540.
3. Howard P. G., Vitter J. S. Practical implementations of arithmetic coding // Storer A. Image and text compression. Kluwer Academic Publishers, Massachusetts, USA, 1992. P. 85–112.
4. Howard P. G., Vitter J. S. Arithmetic coding for data compression // Proc. IEEE. 1994. Vol. 82, N. 6. P. 857–865.
5. Marpe D., Cycon H. L. Efficient pre-coding techniques for wavelet-based image compression // Proc. PCS'97 (ITG Fachbericht). Berlin, Germany, 1997. P. 45–50.
6. Said A. and Pearlman W. A. A new fast and efficient image codec based on set partitioning in hierarchical trees // IEEE Trans. on Circuits and Systems for Video Technology. 1996. Vol. 6, N. 3. P. 243–250.
7. ISO/IEC / JTC 1/SC 29/WG 10 Information technology – Digital compression and coding of continuous-tone still images, ISO/IEC 10918.
8. ISO/IEC / JTC 1/SC 29/WG 1 Information technology – JPEG 2000 image coding system, ISO/IEC 15444.
9. Pennebaker W. B., Mitchell J. L., Langdon G. G., Arps R. B. An overview of the basic principles of the Q-Coder adaptive binary arithmetic coder // IBM J. Res Develop. 1988. Vol. 32, N. 6. P. 717–726.
10. Рябко Б.Я., Фионов А.Н. Эффективный метод адаптивного арифметического кодирования для источников с большими алфавитами // Проблемы передачи информации. 1999. Т.35, №4. С. 95–108.

# ЭФФЕКТИВНЫЙ АЛГОРИТМ ЭКОНОМНОГО КОДИРОВАНИЯ КОЭФФИЦИЕНТОВ ДИСКРЕТНОГО КОСИНУСНОГО ПРЕОБРАЗОВАНИЯ

В.В. Семенюк

Научный руководитель – д.т.н., проф. В.Н. Васильев

Рассматривается новый алгоритм экономного кодирования коэффициентов дискретного косинусного преобразования, построенный на принципах контекстно-зависимого вероятностного моделирования и являющийся эффективной заменой общепринятому алгоритмическому решению.

## Введение

Несмотря на появление и стандартизацию новых высокоэффективных технологий экономного представления видеoinформации (стандарты JPEG2000 [1] и H.264/MPEG4 AVC [2]), разработанные ранее стандарты JPEG [3] и MPEG [4–6] до сих пор являются наиболее распространенными на практике стандартами видеокodирования, так как они удачно сочетают в себе простоту и эффективность. Как известно, оба стандарта описывают технологию кодирования, основанную на использовании дискретного косинусного преобразования. Данная технология является наиболее важной частью метода обработки информации; во многом именно за счет ее применения удается достигнуть высокой эффективности кодирования.

Технология включает в себя три основных этапа: блочное дискретное косинусное преобразование, квантование коэффициентов, полученных в результате преобразования, последующее экономное кодирование квантованных коэффициентов. Если механизм работы алгоритма на первых двух этапах в достаточной мере отлажен и вряд ли подлежит существенному изменению, третий этап – кодирование квантованных коэффициентов – представляет собой проблему, имеющую далеко не однозначное решение. В существующих стандартах описываются упрощенные схемы кодирования [3–6], обладающие низкой вычислительной сложностью, что делает возможным их применение в очень широком спектре приложений. В данной работе предлагается более сложный алгоритм, основанный на использовании контекстно-зависимого моделирования и генерации кода на основе арифметического кодирования. Алгоритм превосходит по эффективности стандартные схемы кодирования и при этом допускает быстрые практические реализации.

Следует отметить, что излагаемый подход идейно не является новым. В литературе можно встретить описания алгоритмов, использующих контекстно-зависимое моделирование коэффициентов дискретного косинусного преобразования (см., например, [7]). Оригинальность рассматриваемого в работе алгоритма заключается в деталях реализации, благодаря которым и удается добиться достаточно высокой эффективности кодирования.

## Описание алгоритма

Исходными данными для алгоритма служат блоки квантованных коэффициентов, полученных в результате дискретного косинусного преобразования произвольного размера  $N \times N$  (в стандартах JPEG и MPEG  $N = 8$ ). В процессе обработки блоков собирается статистическая информация о распределении значений коэффициентов на различных позициях и в различном контекстном окружении. На основе собранной статистики осуществляется генерация кода. В качестве метода генерации кода используется арифметическое кодирование [8–10].

Кодирование коэффициентов в блоке производится в зигзагообразном порядке, как показано на рисунке. Обработка каждого коэффициента может включать до двух

этапов: если коэффициент не может быть обработан на первом этапе, он будет обработан на втором этапе.

Выбор метода кодирования обуславливается особенностями распределения значений коэффициентов в блоке. На практике чаще всего встречаются нулевые коэффициенты и, реже, коэффициенты, единичные по абсолютному значению. Коэффициенты, абсолютные значения которых превышают единицу, встречаются довольно редко, поэтому сбор статистической информации для таких значений затруднен. Можно говорить о малой репрезентативности статистической информации, собираемой для больших абсолютных значений. Исходя из этого, мы вынуждены осуществлять моделирование различных значений коэффициентов по-разному: для маленьких абсолютных значений моделирование должно быть более точным.

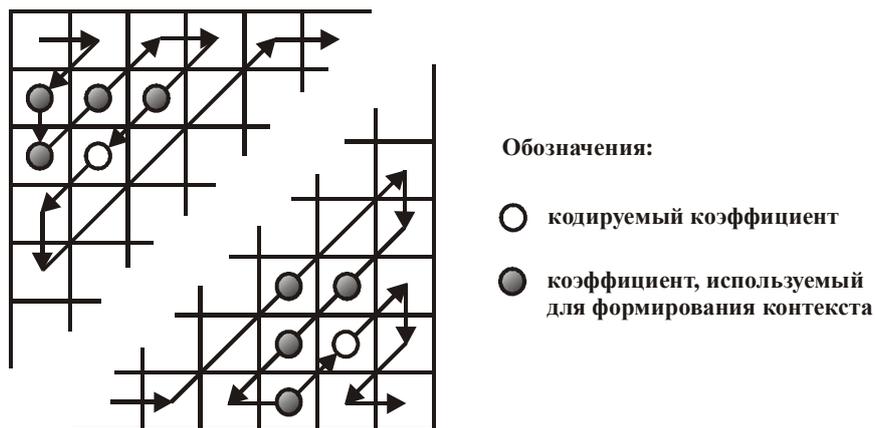


Рис. 1. Формирование контекстов коэффициентов дискретного косинусного преобразования

Другим фактором, обуславливающим выбор метода кодирования, является уменьшение абсолютных значений коэффициентов при обходе блока в зигзагообразном порядке из левого верхнего угла в правый нижний угол. Как правило, при таком обходе, начиная с некоторого момента, все коэффициенты имеют нулевое значение.

Для значения коэффициента  $v$  мы рассматриваем три альтернативы:

- $C_0: v = 0$ ,
- $C_1: |v| = 1$ ,
- $C_2: |v| > 1$ .

В дополнение первая альтернатива разделяется на две подальтернативы:

- $C_0^a: v = 0$  и хотя бы один коэффициент, который следует за данным коэффициентом при зигзагообразном обходе блока, отличен от нуля;
- $C_0^b: v = 0$  и все коэффициенты, которые в следуют за данным коэффициентом при зигзагообразном обходе блока, нулевые.

Таким образом, каждый обрабатываемый коэффициент соответствует одной из четырех ситуаций:  $C_0^a$ ,  $C_0^b$ ,  $C_1$ ,  $C_2$ .

На первом этапе кодирования коэффициента кодируется номер той ситуации, которой этот коэффициент соответствует (реально кодируется число от 0 до 3). Если это ситуация  $C_0^a$  или  $C_0^b$ , кодирование коэффициента на этом заканчивается, так как коэффициент нулевой. Более того, в ситуации  $C_0^b$  заканчивается кодирование не только данного коэффициента, но и всего блока, так как все необработанные коэффициенты тоже равны нулю.

Ситуация  $C_1$  требует дополнительного кодирования знака коэффициента. Было установлено, что знаки распределены достаточно случайно, поэтому для кодирования знака используется информационная модель, включающая два символа, каждый из которых встречается с вероятностью  $1/2$ : символ с индексом 0 соответствует отрицательному значению коэффициента, а символ с индексом 1 – положительному.

Наиболее сложной является ситуация  $C_2$ . Помимо знака, который кодируется точно так же, как и в ситуации  $C_1$ , нам необходимо закодировать абсолютное значение остатка  $|v| - 1$ . Кодирование абсолютного значения остатка осуществляется на втором этапе кодирования.

В ситуации  $C_2$  каждый коэффициент по своему абсолютному значению принадлежит одному из интервалов: [2, 2], [3, 3], [4, 4], [5, 5], [6, 6], [7, 7], [8, 8], [9, 9], [10, 11], [12, 15], [16, 23], [24, 39], [40, 71], [72, 135], [136, 263], [264, 519], [520, 1031], [1032, 2055] (подразумевается, что коэффициент по абсолютному значению не превышает числа  $2^{11} = 2048$ ). Кодирование остатка осуществляется следующим образом: сначала кодируется номер интервала (номера номеруются с нуля), куда попадает абсолютное значение остатка, а затем, если это требуется, кодируется смещение абсолютного значения остатка относительно левой границы интервала (заметим, что в подавляющем большинстве случаев интервал будет состоять из одного символа, поэтому смещение будет кодировать в очень редких случаях). Кодирование смещения осуществляется с использованием информационной модели, которая включает в себя  $2^M$  символов, встречающихся с одинаковыми вероятностями. Это становится возможным благодаря тому, что длины всех интервалов представляют собой степени двойки.

Рассмотрим пример. Пусть мы хотим закодировать коэффициент со значением  $v = -30$ . Сначала кодируется номер ситуации  $C_2$  – 3. Затем кодируется знак - нулевой символ. После этого кодируется номер интервала [24, 39] – 11 и, наконец, смещение в интервале – 6 с использованием модели с 16 символами.

Перейдем к рассмотрению особенностей построения статистических моделей. Будем называть модели, используемые на первом этапе кодирования, *первичными*, а на втором – *вторичными*. Каждая первичная модель содержит частоты появления номеров случаев, а каждая вторичная модель – номеров интервалов. Данные частоты используются в качестве параметров для генерации кода на основе арифметического кодирования. Обновление моделей сводится к увеличению значения частоты появления номера ситуации/интервала сразу после обработки соответствующей ситуации/интервала. Для хранения частот заводятся счетчики. Чтобы избежать переполнения счетчиков и одновременно повысить степень адаптивности модели к меняющимся характеристикам обрабатываемой информации значения счетчиков делятся пополам при достижении суммарной частоты всех счетчиков некоторого наперед заданного порога.

Первичная и вторичная модели хранятся в двух массивах. Индексирование массива первичных моделей осуществляется по позициям и контекстам (для каждой конкретной позиции в блоке при фиксированном контексте в массиве хранится ровно одна первичная модель). Массив вторичных контекстов индексируется только по позициям (для каждой конкретной позиции в массиве хранится ровно одна вторичная модель). Контекст формируется как комбинация четырех двоичных значений (всего  $2^4 = 16$  контекстов), каждое из которых характеризует свойство одного из четырех коэффициентов, соседствующих в плоскости блока коэффициентов с кодируемым коэффициентом, быть равным нулю (нулевое двоичное значение соответствует нулевому коэффициенту). В качестве соседних коэффициентов берутся уже обработанные коэффициенты блока: коэффициент, находящийся слева от кодируемого коэффициента; коэффициент, находящийся слева и сверху от кодируемого коэффициента; коэффициент, нахо-

дящийся сверху от кодируемого коэффициента и, в зависимости от направления зигзагообразного обхода в данном месте, либо коэффициент, находящийся справа и сверху от кодируемого коэффициента, либо коэффициент, находящийся слева и снизу от кодируемого коэффициента (см. рис. 1). Если какой-либо коэффициент, участвующий в формировании контекста, лежит за границей блока, соответствующее двоичное значение считается равным нулю.

Существенное повышение эффективности алгоритма достигается за счет установки удачных начальных значений частот и параметров адаптации моделей. Данные установки должны быть тщательно подобраны из расчета на получение оптимального результата для большинства обрабатываемых данных. Установка параметров модели также позволяет учесть одно важное свойство: ситуация  $C_0^b$  не может возникнуть вслед за ситуацией  $C_0^a$  (иначе ситуация  $C_0^a$  в действительности была бы ситуацией  $C_0^b$ ). Для учета этого свойства из первичной модели, которая соответствует контексту, имеющему нулевое двоичное значение на позиции, непосредственно предшествующей при зигзагообразном обходе данной позиции (той, которой соответствует рассматриваемая модель), исключается номер ситуации  $C_0^b$ . Исключение осуществляется путем обнуления значения соответствующего счетчика в модели.

### Практические результаты

Рассмотренный алгоритм был опробован в рамках стандартной схемы кодирования MPEG4 [6]. Эффективность данного алгоритма сравнивалась с эффективностью алгоритма, описанного в стандарте. Были взяты стандартные тестовые видеопоследовательности "Coastguard", "Foreman" и "Table Tennis". Каждая видеопоследовательность содержит 300 фреймов. В схеме MPEG4 были задействованы только I- и P-фреймы. Для фреймов каждого из указанных типов поддерживалась отдельная статистика, которая обновлялась на каждом I-фрейме (интервал между I-фреймами был равен 30 фреймам). Параметр квантования  $Q$  устанавливался равным 6 и 12; тип квантования – "MPEG" (расшифровка указанных терминов приведена в [6]).

Название	Размер	$Q$	$\Delta_1$	$\Delta_2$
"Coastguard"	176x144	6	8.4%	9.7%
"Coastguard"	176x144	12	8.9%	13.0%
"Foreman"	176x144	6	8.4%	10.4%
"Foreman"	176x144	12	7.6%	12.4%
"Coastguard"	352x288	6	11.0%	12.4%
"Coastguard"	352x288	12	12.1%	16.5%
"Table Tennis"	352x288	6	7.4%	8.9%
"Table Tennis"	352x288	12	5.5%	8.6%

Таблица 1. Повышение эффективности схемы MPEG4

В табл. 1 приведены результаты в процентах, характеризующие повышение эффективности кодирования схемы MPEG4 в целом ( $\Delta_1$ ) и отдельно кодирования квантованных коэффициентов, полученных в результате применения дискретного косинусного преобразования ( $\Delta_2$ ). Как видно, нам удалось повысить эффективность схемы MPEG4 в среднем примерно на 10%. Эксперименты по оценке производительности показали, что вычислительная сложность алгоритма является вполне приемлемой для работы схемы в реальном времени на большинстве современных вычислительных сис-

тем. Высокая скорость достигается во многом благодаря обработке ситуации  $C_0^b$ : в подавляющем большинстве случаев ненулевые коэффициенты сосредоточены в левом верхнем углу блока, что соответствует началу зигзагообразного обхода, поэтому кодирование блока почти всегда сводится к обработке небольшого числа коэффициентов.

## Заключение

Рассмотренный алгоритм может быть использован как при получении эффективных представлений неподвижных изображений, так и при кодировании потоков видеоданных. Он является частью новой высокоэффективной технологии кодирования видеоинформации VSS *Video Codec* компании "Vanguard Software Solutions Inc." (<http://www.vsofts.com/codec/vsscodec.html>). Автор данной работы принимал непосредственное участие в разработке этой технологии, в частности, им был предложен представленный в работе алгоритм.

## Литература

1. ISO/IEC / JTC 1/SC 29/WG 1 Information technology – JPEG 2000 image coding system, ISO/IEC 15444.
2. ISO/IEC JTC 1/SC 29/WG 11 and ITU-T SG 16 Q.6 Study of Final Committee Draft of Joint Video Specification, ITU-T Rec. H.264 / ISO/IEC 14496-10 AVC, JVT-F100, Dec., 2002.
3. ISO/IEC / JTC 1/SC 29/WG 10 Information technology – Digital compression and coding of continuous-tone still images, ISO/IEC 10918.
4. ISO/IEC / JTC 1/SC 29/WG 11 Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s, ISO/IEC 11172.
5. ISO/IEC / JTC 1/SC 29/WG 11 Information technology – Generic coding of moving pictures and associated audio information, ISO/IEC 13818.
6. ISO/IEC / JTC 1/SC 29/WG 11 Information technology – Coding of audio-visual objects, ISO/IEC 14496.
7. Умняшкин С. В. Использование контекстного арифметического кодирования для повышения сжатия данных по схеме JPEG // Известия вузов. Электроника. 2001. №3. С. 96–98.
8. Witten I. H., Neal R. M., Cleary J. G. Arithmetic coding for data compression // CACM. 1987. Vol. 30, N. 6. P. 520–540.
9. Howard P. G., Vitter J. S. Practical implementations of arithmetic coding // Storer A. Image and text compression. Kluwer Academic Publishers, Massachusetts, USA, 1992. P. 85–112.
10. Howard P. G., Vitter J. S. Arithmetic coding for data compression // Proc. IEEE. 1994. Vol. 82, N. 6. P. 857–865.

# ЭФФЕКТИВНЫЙ АЛГОРИТМ ЭКОНОМНОГО КОДИРОВАНИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ДИСКРЕТНОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

В.В. Семенюк

Научный руководитель – д.т.н., проф. В.Н. Васильев

Рассматривается новый алгоритм экономного кодирования изображений на основе дискретного вейвлет-преобразования. В отличие от большинства разработок, где основное внимание уделяется методам квантования, высокая эффективность предлагаемого решения, обусловливается применением сложных методов контекстно-зависимого вероятностного моделирования

## Введение

Кодирование на основе дискретного вейвлет-преобразования является на сегодняшний день наиболее эффективным способом кодирования изображений с потерями. Неудивительно, что новый стандарт JPEG2000 [1] основан именно на данной технологии. Помимо алгоритма, описанного в стандарте, существует большое количество других схем вейвлет-кодирования. Эти схемы различаются типом вейвлет-преобразования, методом квантования и способом кодирования квантованных коэффициентов частотных диапазонов в вейвлет-базисе. Очень часто алгоритмы реализуются таким образом, что можно легко изменить вейвлет-преобразование и метод квантования, оставив неизменным способ кодирования коэффициентов. Таким образом, способ кодирования является по сути ядром алгоритма, тогда как вейвлет-преобразование и квантование есть лишь его опциональные части. (Следует отметить, что иногда метод квантования и способ кодирования неразрывно связаны, что делает квантование также частью ядра алгоритма.) Поэтому основные усилия целесообразно направлять именно на создание эффективного способа кодирования.

Наиболее удачные схемы кодирования квантованных коэффициентов частотных диапазонов в вейвлет-базисе основаны на применении так называемых *встроенных нуль-деревьев*. Подход был впервые описан в [2] и получил свое развитие в [3]. Для повышения эффективности в [3] предлагается использовать арифметическое кодирование в рамках контекстно-свободной вероятностной модели. В более поздних работах описываются контекстно-зависимые методы кодирования, наиболее эффективный из которых рассмотрен в [4]. В предлагаемой схеме применяются стандартные вейвлет-преобразования и простейшее скалярное квантование. Мы решили пойти по тому же пути, выбрав одно из используемых в [4] стандартных вейвлет-преобразований и идентичный метод квантования. При этом мы существенно изменили способ кодирования, добившись увеличения эффективности кодирования, не выходя за рамки идеологии встроенных нуль-деревьев.

## Описание алгоритма

Исходное изображение подвергается ставшей уже де-факто стандартной многоуровневой диадической вейвлет-декомпозиции. На каждом уровне разрешения  $i$ , за исключением нулевого, прямоугольный блок коэффициентов  $LL_i$  путем применения горизонтальной и вертикальной фильтрации преобразуется в совокупность четырех прямоугольных блоков:  $LL_{i-1}$  – результат низкочастотной фильтрации в горизонтальном и вертикальном направлениях;  $LH_{i-1}$  – результат низкочастотной фильтрации в горизонтальном направлении и высокочастотной фильтрации в вертикальном направлении;  $HL_{i-1}$  – результат высокочастотной фильтрации в горизонтальном направлении и низкочастотной фильтрации в вертикальном направлении;  $HH_{i-1}$  – результат высокочас-

точной фильтрации в горизонтальном и вертикальном направлениях (см. рис. 1). Размеры получающихся в результате декомпозиции блоков в два раза меньше размеров исходных блоков. В процессе декомпозиции обрабатываемое изображение выступает в качестве блока коэффициентов при максимальном разрешении  $LL_{K-1}$  (здесь  $K$  – количество уровней разрешения; количество уровней вейвлет-декомпозиции всегда на единицу меньше данного количества). Вейвлет-декомпозиция осуществляется с использованием лифтинг-схемы [5]. В качестве фильтра берется широко используемый на практике биортогональный фильтр 9/7 [6].

Коэффициенты блоков, полученных в результате вейвлет-преобразования, подвергаются скалярному квантованию с мертвой зоной в окрестности нуля. Для всех блоков используется один и тот же параметр квантования  $q$ . Величина мертвой зоны составляет  $1.5q$ .

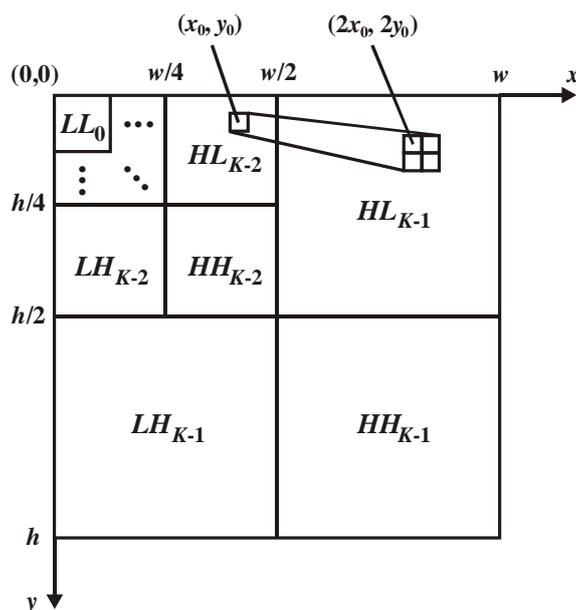


Рис. 1. Общая схема многоуровневой диадической вейвлет-декомпозиции

Кодирование квантованных коэффициентов может включать до трех этапов. На первом этапе кодируется номер старшего ненулевого двоичного разряда коэффициента. Если этот номер превышает единицу, на втором этапе кодируются все двоичные разряды коэффициента, кроме старшего. Последним этапом является кодирование знака. Оно необходимо только в случае, когда коэффициент отличен от нуля. Так как на практике значение коэффициента очень часто равно нулю, кодирование в большинстве случаев завершается на первом этапе.

При кодировании номера старшего ненулевого разряда коэффициента учитывается зависимость между значениями коэффициентов, находящихся на одних и тех же позициях в блоках, полученных с использованием идентичного типа фильтрации, но соответствующих разным разрешениям. Речь идет о зависимости между абсолютным значением коэффициента с координатами  $(x_0, y_0)$ , принадлежащего одному из блоков  $HL_i$ ,  $LH_i$  или  $HH_i$ , и абсолютными значениями коэффициентов с координатами  $(2x_0, 2y_0)$ ,  $(2x_0 + 1, 2y_0)$ ,  $(2x_0, 2y_0 + 1)$  и  $(2x_0 + 1, 2y_0 + 1)$ , принадлежащих соответствующему блоку большего разрешения:  $HL_{i+1}$ ,  $LH_{i+1}$  или  $HH_{i+1}$ . Как правило,

$$|I(x_0, y_0)| > \max\{|I(2x_0, 2y_0)|, |I(2x_0 + 1, 2y_0)|, |I(2x_0, 2y_0 + 1)|, |I(2x_0 + 1, 2y_0 + 1)|\},$$

где  $I(x, y)$  – значение коэффициента с координатами  $(x, y)$ . Отсюда автоматически следует, что коэффициент с координатами  $(x_0, y_0)$ , не принадлежащий блоку  $LL_0$ , с

большой долей вероятности превосходит по своему абсолютному значению все коэффициенты, координаты которых получаются из координат  $x_0$  и  $y_0$  путем многократного применения функций  $f(x) = 2x$  и  $f(x) = 2x + 1$ . Таким образом, каждый коэффициент, принадлежащий одному из блоков  $HL_0$ ,  $LH_0$  или  $HH_0$ , является корнем своеобразного дерева коэффициентов со степенью ветвления 4, обладающего тем свойством, что коэффициент, соответствующий любому узлу дерева, в большинстве случаев превосходит по своему абсолютному значению коэффициенты, соответствующие его дочерним узлам. Распространяя данный вывод на номера старших ненулевых разрядов коэффициентов, можно сделать заключение о том, что номер старшего ненулевого разряда коэффициента, соответствующего узлу дерева, как правило, не меньше номеров старших ненулевых разрядов коэффициентов, соответствующих его дочерним узлам. Ниже приведена схема кодирования, учитывающая указанную особенность.

Первоначально выбирается максимальный номер ненулевого разряда и осуществляется обработка блоков, во время которой кодируются абсолютные значения коэффициентов с номером старшего ненулевого разряда, совпадающим с номером, выбранным для текущего кодового прохода. В дальнейшем всегда берется номер, на единицу меньший предыдущего, и делается попытка обработать коэффициенты, еще не обработанные во время предыдущих проходов. Блоки всегда обрабатываются в следующем порядке:  $LL_0$ ,  $HL_0$ ,  $LH_0$ ,  $HH_0$ ,  $HL_1$ ,  $LH_1$ ,  $HH_1$ ,  $HL_2$ ,  $LH_2$ ,  $HH_2$ , ... ,  $HL_{K-1}$ ,  $LH_{K-1}$ ,  $HH_{K-1}$ . При этом разные блоки обрабатываются по-разному. По способу обработки их можно разделить на 4 группы:

- 1-я группа: блок  $LL_0$ ;
- 2-я группа: блоки  $HL_0$ ,  $LH_0$ ,  $HH_0$ ;
- 3-я группа: блоки  $HL_1$ ,  $LH_1$ ,  $HH_1$ , ... ,  $HL_{K-2}$ ,  $LH_{K-2}$ ,  $HH_{K-2}$ ;
- 4-я группа: блоки  $HL_{K-1}$ ,  $LH_{K-1}$ ,  $HH_{K-1}$ .

При обработке блока  $LL_0$  для каждого еще не обработанного коэффициента осуществляется кодирование факта совпадения номера его старшего ненулевого разряда с номером, выбранным для текущего прохода (вариант E1). Реально кодируется одно из двух значений: 0 соответствует несовпадению номеров, а 1 – совпадению. В случае, когда номера совпадают, производится кодирование оставшихся двоичных разрядов коэффициента (вариант E5).

Коэффициенты, принадлежащие блокам 2-й группы, обрабатываются несколько иначе. Сначала кодируется факт совпадения выбранного для текущего прохода номера разряда с номером максимального старшего разряда среди всех коэффициентов, принадлежащих дереву, корень которого соответствует кодируемому коэффициенту (вариант E1). Если номера совпадают, кодируется факт совпадения выбранного для текущего прохода номера разряда с номером максимального старшего разряда среди всех коэффициентов, принадлежащих дереву, корень которого соответствует кодируемому коэффициенту, исключая сам коэффициент (вариант E3). Несовпадение номеров в последнем случае означает, что номер старшего ненулевого разряда кодируемого коэффициента равен номеру разряда, выбранному для текущего прохода, т.е. на этапе декодирования определение этого номера возможно без какой-либо дополнительной информации. Если же данные номера совпадают, без такой информации не обойтись. В ситуации, когда номера совпадают, дополнительно осуществляется кодирование разности между номером разряда, выбранным для текущего прохода, и номером старшего ненулевого разряда данного коэффициента (вариант E4). Как и прежде, кодирование

абсолютного значения коэффициента завершается кодированием всех его оставшихся двоичных разрядов (вариант E5).

Способ кодирования коэффициентов, принадлежащих блокам 3-й группы, очень похож на способ кодирования коэффициентов, принадлежащих блокам 2-й группы, за одним исключением: при обработке блоков 3-й группы для каждого четырех смежных коэффициентов с координатами  $(2x, 2y)$ ,  $(2x+1, 2y)$ ,  $(2x, 2y+1)$  и  $(2x+1, 2y+1)$  нам может быть заранее известно, что номера старших ненулевых разрядов всех четырех коэффициентов меньше номера разряда, выбранного для текущего прохода. Это становится возможным благодаря кодированию по рассмотренному ранее варианту E3 блока с тем же типом фильтрации, но соответствующего меньшему на 1 разрешению. Когда возникает подобная ситуация, кодирование всех четырех коэффициентов на данном проходе не осуществляется. Если же такая ситуация не возникает, есть большая вероятность того, что один из указанных четырех коэффициентов будет закодирован на данном проходе, поэтому в данном случае целесообразно вместо варианта кодирования E1, использовать особый вариант кодирования E2.

Блоки 4-й группы обрабатываются практически по той же схеме, что и блоки третьей группы. Отличие заключается в том, что кодирование по варианту E3 в данном случае оказывается ненужным, так как коэффициенты, принадлежащие блокам данной группы, всегда соответствуют листовым узлам деревьев коэффициентов.

Выбор порядка кодирования коэффициентов обуславливается удобством реализации. Коэффициенты блоков 1-й и 2-й групп кодируются в строчном порядке. При обработке блоков 3-й и 4-й групп коэффициенты обрабатываются по четыре (четверку образуют смежные коэффициенты с координатами  $(2x, 2y)$ ,  $(2x+1, 2y)$ ,  $(2x, 2y+1)$  и  $(2x+1, 2y+1)$ ), при этом выбор очередной четверки совпадает с построчным обходом блока с тем же типом фильтрации, но соответствующего меньшему на 1 разрешению.

После кодирования абсолютных значений коэффициентов кодируются их знаки (вариант E6). Кодирование знака также представляет собой кодирование двоичного значения: 0 соответствует отрицательному числу, а 1 – положительному.

В общем случае генерация кода осуществляется с применением контекстно-зависимого арифметического кодирования [7–9]. Для каждого упомянутого выше варианта кодирования поддерживается набор арифметических моделей, представляющих собой совокупность частот появления различных символов на выходе кодируемого источника. Частоты являются входными данными для генератора арифметического кода. Выбор модели, на основе которой осуществляется генерация кода, во всех вариантах, кроме E4 и E5, зависит от текущего двумерного пространственного контекста кодируемого элемента. (Вариант E5 подразумевает несколько иной способ моделирования, о чем будет сказано ниже. Кодирование по варианту E4 является контекстно-свободным, т.е. в данном случае используется одна модель, которая не зависит от контекста.) Для различения контекстов вводится численная величина *значение контекста*. Формула, по которой определяется значение контекста, зависит от конкретного варианта кодирования.

Введем ряд обозначений. Для коэффициента с координатами  $(x, y)$  введем функцию  $S1(x, y)$ , которая будет указывать, какое бинарное значение было закодировано в последний раз, когда данный коэффициент обрабатывался по варианту E1 или по варианту E2. Введем аналогичную функцию  $S2(x, y)$  для кодирования по варианту E3. Также определим функцию  $SIGN(x, y)$ , описывающую результат обработки знака по варианту E6: если знак коэффициента с координатами  $(x, y)$  уже был закодирован, то в случае отрицательного коэффициента значение функции полагается равным 1, а в случае положительного коэффициента – 2. Если при вычислении значения каждой из указанных трех функций коэффициент на момент вычисления еще не обрабатывался по

соответствующем варианту или координаты коэффициента выходят за допустимые пределы, значение соответствующей функции считается равным нулю.

Используя данные функции, можно легко выписать значения контекстов для кодирования по вариантам E1, E2, E3 и E6. Значение контекста для позиции с координатами  $(x, y)$  при кодировании по варианту E1 вычисляется по формуле:

$$C1(x, y) = 6 \cdot S1(x-1, y) + 3 \cdot S1(x, y-1) + S1(x+1, y) + S1(x, y+1).$$

При кодировании по варианту E2 значение контекста зависит от того, какую позицию занимает обрабатываемый коэффициент в четверке смежных коэффициентов, имеющих координаты  $(2x, 2y)$ ,  $(2x+1, 2y)$ ,  $(2x, 2y+1)$  и  $(2x+1, 2y+1)$ , и от выполнения или невыполнения условия

$$S1(2x, 2y) + S1(2x+1, 2y) + S1(2x, 2y+1) + S1(2x+1, 2y+1) = 0$$

(1)

Если условие (1) не выполнено, значение контекста  $C2(2x, 2y)$  берется равным

$$6 \cdot S1(2x-1, 2y) + 3 \cdot S1(2x, 2y-1) + S1(2x+1, 2y) + S1(2x, 2y+1);$$

если же условие выполняется, то равным

$$6 \cdot S1(2x-1, 2y) + 3 \cdot S1(2x, 2y-1) + S1(2x+2, 2y+1) + S1(2x+1, 2y+2).$$

После обработки коэффициента с координатами  $(2x, 2y)$  условие (1) снова проверяется на выполнение. Если оно не выполнено, то значение контекста  $C2(2x+1, 2y)$  берется равным

$$6 \cdot S1(2x, 2y) + 3 \cdot S1(2x+1, 2y-1) + S1(2x+2, 2y) + S1(2x+1, 2y+1),$$

в противном случае – равным

$$6 \cdot S1(2x+2, 2y) + 3 \cdot S1(2x+1, 2y-1) + S1(2x-1, 2y+1) + S1(2x, 2y+2).$$

После обработки коэффициента с координатами  $(2x+1, 2y)$  условие (1) вновь проверяется, и в случае его невыполнения значение контекста  $C2(2x, 2y+1)$  полагается равным

$$6 \cdot S1(2x-1, 2y+1) + 3 \cdot S1(2x, 2y) + S1(2x+1, 2y+1) + S1(2x, 2y+2),$$

а в случае выполнения – равным

$$2 \cdot S1(2x-1, 2y+1) + S1(2x+2, 2y+1).$$

После обработки коэффициента с координатами  $(2x, 2y+1)$  условие (1) проверяется последний раз. Если оно не выполнено, значение контекста  $C2(2x+1, 2y+1)$  берется равным

$$6 \cdot S1(2x, 2y+1) + 3 \cdot S1(2x+1, 2y) + S1(2x+2, 2y+1) + S1(2x+1, 2y+2).$$

Если условие выполнено, кодирование по варианту E2 вообще не осуществляется, так как в этом нет необходимости (известно, что кодируемое значение равно 1).

Значение контекста, используемого для кодирования коэффициента с координатами  $(x, y)$  по варианту E3, вычисляется по формуле

$$C3(x, y) = 6 \cdot S2(x-1, y) + 3 \cdot S2(x, y-1) + S2(x+1, y) + S2(x, y+1),$$

а значение соответствующего контекста, используемого при кодировании знака (вариант E6), – по формуле

$$C6(x, y) = 3 \cdot SIGN(x-1, y) + SIGN(x, y-1).$$

Вычисление значения контекста в случае кодирования по варианту E5 (кодирование оставшихся двоичных разрядов абсолютного значения коэффициента) не подразумевает использование какой бы то ни было информации о соседних коэффициентах. Обозначим через  $M$  номер старшего ненулевого разряда коэффициента, а через  $N$  – номер кодируемого двоичного разряда (в данных обозначениях 1-й номер соответствует младшему разряду). Значение контекста определяется по формуле

$$C5(M, N) = 0 + 1 + 2 + \dots + (M-2) + N.$$

## Практические результаты

Для получения численных результатов, характеризующих эффективность предложенного алгоритма, мы применили его к стандартным тестовым изображениям "Barbara", "Goldhill" и "Lena". В каждом случае использовались два варианта кодирования: с 4 и 5 уровнями вейвлет-декомпозиции. В табл. 1 указаны численные значения качества восстановленных изображений в dB (пиковое соотношение сигнал-шум) при эффективностях кодирования 0.25, 0.5 и 1 бит на точку изображения (bpp).

Изображение	Количество уровней декомпозиции	0.25 bpp	0.5 bpp	1 bpp
"Barbara"	4	28.35	32.30	37.33
"Barbara"	5	28.39	32.33	37.36
"Goldhill"	4	30.79	33.51	37.03
"Goldhill"	5	30.83	33.55	37.06
"Lena"	4	34.42	37.50	40.60
"Lena"	5	34.50	37.56	40.65

Таблица 1. Эффективность предложенного алгоритма

Отметим, что по сравнению с результатами, приведенными в [4], нам удалось увеличить эффективность в среднем на 0.05–0.1 dB.

Производительность алгоритма сложно оценить безотносительно к конкретной вычислительной системе и конкретному изображению, тем не менее, использование встроенных нуль-деревьев гарантирует невысокую вычислительную сложность предлагаемой схемы кодирования. Эксперименты показывают, что на современных вычислительных системах алгоритм позволяет обрабатывать в реальном времени не только неподвижные изображения, но и видеопотоки.

## Литература

1. ISO/IEC / JTC 1/SC 29/WG 1 Information technology – JPEG 2000 image coding system, ISO/IEC 15444.
2. Shapiro J. M. Embedded image coding using zerotrees of wavelets coefficients. // IEEE Trans. on Signal Processing. 1993. Vol. 41, N.12. P. 3445–3462.
3. Said A. and Pearlman W. A. A new fast and efficient image codec based on set partitioning in hierarchical trees // IEEE Trans. on Circuits and Systems for Video Technology. 1996. Vol. 6, N. 3. P. 243–250.
4. Marpe D., Cycon H. L. Efficient pre-coding techniques for wavelet-based image compression // Proc. PCS'97 (ITG Fachbericht). Berlin, Germany, 1997. P. 45–50.
5. Calderbank R., Daubechies I., Sweldens W., Yeo B. L. Wavelet transforms that map integers to integers // J. Applied and Computational Harmonic Analysis (ACHA). 1998. Vol. 5, N. 3. P. 332–369.
6. Antonini M., Barlaud M., Mathieu P., Daubechies I. Image coding using wavelet transform // IEEE Trans. on Image Processing. 1992. Vol. 1, N. 2. P. 205–220.
7. Witten I. H., Neal R. M., Cleary J. G. Arithmetic coding for data compression // CACM. 1987. Vol. 30, N. 6. P. 520–540.
8. Howard P. G., Vitter J. S. Practical implementations of arithmetic coding // Storer A. Image and text compression. Kluwer Academic Publishers, Massachusetts, USA, 1992. P. 85–112.
9. Howard P. G., Vitter J. S. Arithmetic coding for data compression // Proc. IEEE. 1994. Vol. 82, N. 6. P. 857–865.

# АНАЛИЗ ВЫЧИСЛИТЕЛЬНОГО ПРОЦЕССА НЕЛИНЕЙНОЙ ДИСКРЕТНОЙ ФИЛЬТРАЦИИ КАЛМАНА С ИСПОЛЬЗОВАНИЕМ ПРОЦЕССОРА INTEL PENTIUM 4

М.А. Таратин, А.С. Захаров

Научный руководитель – д.т.н., проф. И.П. Гуров

Представлены аналитические выражения и структурная схема, определяющие алгоритм нелинейной дискретной фильтрации Калмана. Получена оценка производительности фильтрации при восстановлении параметров интерференционного сигнала. Указаны возможные пути повышения скорости вычислений.

## Введение

В настоящее время возрастает потребность в обработке информационных сигналов в реальном времени поступления данных. Одной из актуальных задач является динамическое восстановление параметров интерференционных сигналов малой когерентности в оптической когерентной томографии. Возможным подходом к решению этой задачи является использование нелинейной дискретной фильтрации Калмана [1].

Известны публикации [2], в которых рассматриваются особенности применения дискретного нелинейного фильтра Калмана для восстановления параметров интерференционного сигнала, однако в этих работах не приводится оценок скорости выполнения вычислений.

Для проведения вычислений в процессе фильтрации могут быть использованы специализированные процессоры. Такой подход позволяет обеспечить максимальное быстродействие операций обработки и необходимую точность вычислений, но обладает существенными недостатками: высокой стоимостью и длительностью реализации.

Другим возможным подходом является использование для вычислений более доступных персональных компьютеров. В последнее время они достигли достаточно высокой производительности даже при работе с числами с плавающей точкой и могут быть использованы для решения сложных вычислительных задач.

Целью данной работы является оценка максимальной производительности нелинейной дискретной фильтрации Калмана при восстановлении параметров интерференционного сигнала на однопроцессорном компьютере на базе Intel Pentium 4.

## Нелинейный дискретный фильтр Калмана

Фильтр Калмана определяется уравнением наблюдения и уравнением состояния. При дискретной фильтрации они имеют вид, соответственно:

$$s(k) = h(\theta(k)) + n(k), \quad (1)$$

$$\theta(k) = \mathbf{F}(k-1)\theta(k-1) + w(k), \quad (2)$$

где  $h(\theta)$  – заданная функция наблюдения,  $\mathbf{F}(k)$  – матрица перехода,  $\theta$  – вектор параметров,  $n$  – шум наблюдения,  $w$  – шум системы. Результирующее уравнение для оценивания параметра  $\theta$  дискретным фильтром Калмана имеет рекуррентный характер:

$$\theta(k) = \mathbf{F}(k-1)\theta(k-1) + \mathbf{C}(k)\sigma_{\theta}^{-2}(k)\sigma_n^{-2}(k) \times [s(k) - h(\mathbf{F}(k-1)\theta(k-1))] \quad (3)$$

где  $\sigma_n(k)$  – дисперсия шума наблюдения,  $\sigma_{\theta}^2(k)$  – апостериорная дисперсия параметра

$$\theta(k), \quad \mathbf{C}(k) = h'_{\theta}(\theta(k)). \quad (4)$$

Структурная схема фильтра показана на рис. 1.

Фильтр Калмана, располагая оценкой  $\theta(k-1)$  на предыдущем шаге и основываясь на уравнении (2), прогнозирует оценочное значение  $\mathbf{F}(k-1)\theta(k-1)$  на  $k$ -ый шаг. При получении  $k$ -го наблюдения  $s(k)$  прогнозируемая оценка дополнительно корректируется на значение, пропорциональное невязке (так называемому обновляющему процессу), т.е. отклонению прогноза  $h(\mathbf{F}(k-1)\theta(k-1))$  от полученного значения  $s(k)$ . Коэффициент пропорциональности  $\mathbf{C}(k)\sigma_0^2(k)\sigma_n^{-2}$ , регулирующий "вес" невязки для  $\theta(k)$  по сравнению с прогнозом  $\mathbf{F}(k-1)\theta(k-1)$ , называют коэффициентом усиления фильтра Калмана.

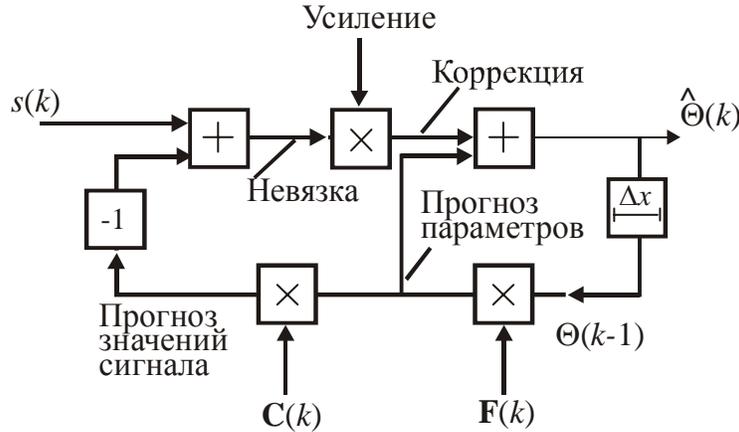


Рис. 1. Структурная схема фильтра Калмана

Коэффициент усиления, помимо значения известной функции  $\mathbf{C}(k)$ , определяется также дисперсией  $\sigma_n^2$  шума наблюдения и апостериорной дисперсией  $\sigma_0^2(k)$  параметра  $\theta(k)$ . Можно показать, что  $\sigma_0^2(k)$  совпадает с квадратом разности оценки и истинного значения  $\theta(k)$ , усредненным по всем возможным реализациям шума наблюдения, и, следовательно, характеризует точность оценки на  $k$ -ом шаге.

Далее рассматриваем случай интерферометрической системы. Для неё вектор параметров состоит из четырех компонентов:

$$\theta = (s_0, A, \varphi, f)^T, \quad (5)$$

где  $s_0(k)$  – фоновая составляющая,  $A(k)$  – огибающая сигнала полос,  $f$  – локальная частота полос,  $\varphi$  – фаза полос.

Можно показать, что:

$$f(k) = f(k-1) + P_f(k)\{s(k) - A(k-1)\cos\varphi(k/k-1)\}, \quad (6)$$

$$\varphi(k) = \varphi(k-1) + 2\pi f(k-1)\Delta x + P_\varphi(k)\{s(k) - s_0(k-1) - A(k-1)\cos\varphi(k/k-1)\}, \quad (7)$$

$$A(k) = A(k-1) + P_A(k)\{s(k) - s_0(k-1) - A(k-1)\cos\varphi(k/k-1)\}, \quad (8)$$

$$s_0(k) = s_0(k-1) + P_{s_0}(k)\{s(k) - s_0(k-1) - A(k-1)\cos\varphi(k/k-1)\}, \quad (9)$$

где  $s(k)$  – дискретные отсчеты сигнала,  $k=1, 2, \dots, K$ , соответствующие дискретным значениям независимой переменной  $x_k = k\Delta x$ ,  $\Delta x$  – шаг дискретизации, значение которого для упрощения вида формул принят равным единице,

$$P_A(k) = \tilde{\sigma}_A^2 \cos\varphi(k/k-1) / \tilde{\sigma}^2, \quad (10)$$

$$P_\varphi(k) = -\tilde{\sigma}_\varphi^2 A(k-1) \sin\varphi(k/k-1) / \tilde{\sigma}^2, \quad (11)$$

$$P_f(k) = -2\pi\tilde{\sigma}_f^2 A(k-1) \sin \varphi(k/k-1)/\tilde{\sigma}^2, \quad (12)$$

$$P_{s_0}(k) = \tilde{\sigma}_{s_0}^2 / \tilde{\sigma}^2, \quad (13)$$

$\tilde{\sigma}_A^2, \tilde{\sigma}_\phi^2, \tilde{\sigma}_f^2, \tilde{\sigma}_{s_0}^2$  - дисперсии ошибки предсказания (параметры фильтрации),

$$\begin{aligned} \tilde{\sigma}^2 = & \tilde{\sigma}_A^2 \cos^2 \varphi(k/k-1) + \tilde{\sigma}_\phi^2 A^2(k-1) \sin^2 \varphi(k/k-1) \\ & + 4\pi^2 \tilde{\sigma}_f^2 A^2(k-1) \sin^2 \varphi(k/k-1) + \tilde{\sigma}_{s_0}^2 + \sigma_n^2 \end{aligned} \quad (14)$$

$\sigma_n^2$  - дисперсия шума наблюдения.

## Оценка производительности вычислений

Оценим максимальную производительность дискретного нелинейного фильтра Калмана для интерферометрического сигнала с учетом особенностей [3] микроархитектуры процессора Intel Pentium 4.

Прежде всего, необходимо получить оценку количества операций, выполняемых на каждом шаге фильтрации. Для этого исключим дублирующиеся вычисления. Результаты изображены на рис. 2.

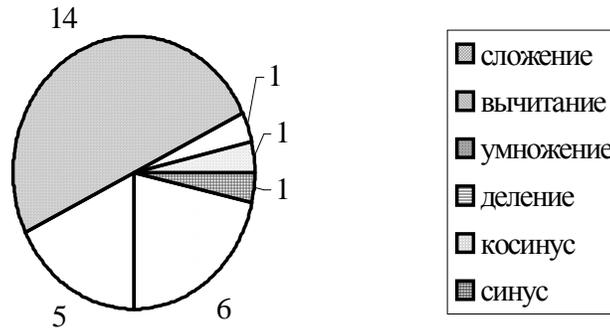


Рис. 2. Количество вычислительных операций на одном шаге фильтрации

Входными данными для каждого шага фильтрации является значение сигнала и оценка значений параметров с предыдущего шага. Значение сигнала должно эффективно кэшироваться при последовательном чтении данных из массива. Значения оценки параметров должны остаться в кэш с предыдущего шага, поэтому время выполнения операций чтения и записи этих значений будет занимать пренебрежимо малое время. Кроме этого, благодаря небольшому количеству выполняемых операций на каждом шаге фильтрации, невелико и время выполнения операций чтения и сохранения данных, используемых в теле цикла, так как эти данные тоже должны находиться в кэше. Поэтому можно предположить, что при выполнении каждого шага фильтрации операции будут производиться только с данными из регистров процессора или кэш (первого или второго уровня). В связи с тем, что операции с памятью в Intel Pentium 4 вынесены в отдельный операционный блок и могут выполняться параллельно с другими операциями, можно считать, что время их выполнения не окажет влияния на общее время выполнения одного шага. Таким образом, для оценки максимального быстродействия достаточно оценить время выполнения вычислительных операций и не учитывать время выполнения операций чтения и записи данных.

В качестве базового типа вычислений будем рассматривать четырехбайтовый тип с плавающей точкой (8 бит порядок, 24 бита мантисса).

Все операции, за исключением операций вычисления значений синуса и косинуса, можно выполнять с помощью набора векторных инструкций Streaming SIMD 2 Extensions (SSE). Благодаря тому, что размер представления числа составляет четыре байта, возможна работа со значениями, упакованными в вектора размерностью четыре. Это позволит значительно повысить производительность операций.

Оценка времени выполнения [4] каждой из используемых инструкций SSE приведена в табл. 1. Здесь и далее под временем выполнения предполагается количество тактов процессора перед выполнением следующей операции того же типа.

Сложение ADDPS	2
Вычитание SUBPS	2
Умножение MULPS	2
Деление DIVPS	39

Таблица 1. Время выполнения операций (в тактах процессора)

Из-за того, что в наборе инструкций SSE отсутствуют тригонометрические операции, для вычисления значений функций синус и косинус приходится использовать блок Floating Point Unit (FPU). Благодаря тому, что операнды у этих команд одинаковы, инструкции FSIN и FCOS можно заменить на FSINCOS. Время выполнения этой операции на процессоре Intel Pentium 4 составляет 140 тактов.

Предположим, что при выполнении множества шагов фильтрации возможно изменить последовательность инструкций таким образом, чтобы загрузка всех выполняющих операции устройств процессора была оптимальной. Таким образом, минимальное время  $T$  выполнения одного шага фильтрации (выраженное количеством тактов процессора) будет определяться общим временем выполнения операций умножения, деления и вычисления синуса и косинуса

$$T = \frac{((14 + 5 + 6) \times 2 + 39 \times 1)}{4} + 140 \approx 162. \quad (15)$$

Для последовательности интерферометрических данных форматом 300×300 элементов обработка одного кадра требует 162×300×300 тактов, что при тактовой частоте процессора 3 ГГц позволяет обрабатывать 205 кадров в секунду.

## Заключение

Проведен анализ вычислительного процесса нелинейной дискретной фильтрации Калмана. Получены оценки вычислительной сложности алгоритма фильтрации при вычислениях в числах с плавающей точкой с одинарной точностью.

Поскольку длина волны равна 800 нм, шаг перемещения опорного отражателя должен быть не более 200 нм. Таким образом, для исследования образца на глубину 0,2 мм необходимо обработать 1000 кадров. Оценка времени исследования образца составляет 5 сек при приемлемом времени исследования 1 сек. Следовательно, необходима дальнейшая оптимизация фильтрации.

Определяющим фактором, снижающим скорость фильтрации, является время выполнения операций вычисления значений синус и косинус. Возможные пути повышения про-

изводительности – замена этих инструкций на более быстрые аналоги, возможно, используя инструкции SSE/SSE2, или использование массивов предварительно рассчитанных значений синуса и косинуса.

## Литература

1. Kalman R.E. A new approach to linear filtering and prediction problems //Trans. ASME, J. Basic Eng. 1960. V.82. P. 35-45.
2. Alarousu E., Gurov I., Hast J., Mullula R., Prikari T., Zakharov A. Optical coherence tomography evaluation of internal random structure of wood fiber tissue. In: Sixth International Conference of Quality Control by Artificial Vision, Tobin K.W.Jr., Meriaudiau M., eds. //Proc. SPIE. 2003. Vol. 5132. pp. 149-160.
3. Glenn Hinton, Dave Sager, Mike Upton, Darrell Boggs, Doug Carmean, Alan Kyker, Patrice Roussel. The Microarchitecture of the Pentium 4 Processor. // Intel Technology Journal Q1, 2001. pp. 2-11.
4. Intel Pentium 4 and Intel Xeon Processor Optimization, Reference Manual. Intel Corporation // Order Number: 248966-05 pp.c1-c16.

# ИСПОЛЬЗОВАНИЕ ДИСКРЕТНЫХ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЙ ДЛЯ СЖАТИЯ ПОЛУТОНОВЫХ ИЗОБРАЖЕНИЙ

М.В. Гришин

Научные руководители – д.т.н., проф. А.А Ожиганов, д.т.н., проф. А.Ю. Тропченко

Рассматриваются новые методы сжатия изображений на основе дискретных ортогональных вейвлет-преобразований. Приведены сравнительные характеристики рассмотренных методов с методом JPEG сжатия.

## Введение

В настоящее время существует достаточно много алгоритмов сжатия изображений. Они осуществляют сжатие либо в пространственной, либо в частотной областях изображения. В пространственной области нельзя добиться достаточно большого коэффициента сжатия, поэтому в настоящее время большинство алгоритмов сжатия работают в частотной области изображения.

Процесс сжатия изображения в частотной области можно представить четырьмя стадиями:

1. применение обратимых дискретных ортогональных преобразований к изображению;
2. квантование полученных частотных коэффициентов;
3. выбор наиболее значимых частотных коэффициентов;
4. вторичное сжатие выбранных коэффициентов, например арифметическим или статистическим алгоритмом сжатия.

В работе в качестве обратимого дискретного ортогонального преобразования, выбрано вейвлет-преобразование. В алгоритмах использовались два вида преобразования: преобразование Хаара и вейвлет-преобразование (5,3).

Выбор вейвлет-преобразования обусловлен тем, что его коэффициенты обладают следующими свойствами:

1. вейвлет-коэффициенты одновременно локализованы как в пространственной, так и в частотной областях;
2. если данный вейвлет-коэффициент большой либо маленький, то смежный с ним вейвлет коэффициент также большой, либо маленький;
3. масштаб значений вейвлет-коэффициентов сохраняется от уровня к уровню преобразования [1].

## Квантование вейвлет-коэффициентов

В отличие от JPEG кодирования, при квантовании вейвлет-коэффициентов используется однородное квантование. Это значит, что либо все коэффициенты из всех субполос квантуются на одно и то же число, либо для каждой из субполос выбирается свой коэффициент квантования, но для всех частотных коэффициентов данной субполосы он остается постоянным. Выбор различных коэффициентов квантования для разных субполос является более предпочтительным, т.к. позволяет "тонко" регулировать степень сжатия и качество восстановленного изображения.

Следует отметить, что параметры квантования выбирается экспериментально и являются одними из входных параметров для всех рассматриваемых алгоритмов сжатия.

## Методы обхода вейвлет-коэффициентов

Для качественного сжатия вейвлет-коэффициентов необходим учет взаимосвязей между коэффициентами из различных субполос разложения. В высокочастотных субполосах обычно имеются большие области с нулевой или малой энергией. Области с

высокой энергией повторяют от субполосы к субполосе свои очертания и местоположение. Медленно изменяющиеся, гладкие области исходного изображения хорошо описывают низкочастотные (НЧ) вейвлет-базисы, что приводит к "упаковке" энергии в малом числе коэффициентов НЧ области.

Наиболее очевидным в данном случае методом обхода вейвлет-коэффициентов является древовидный обход.

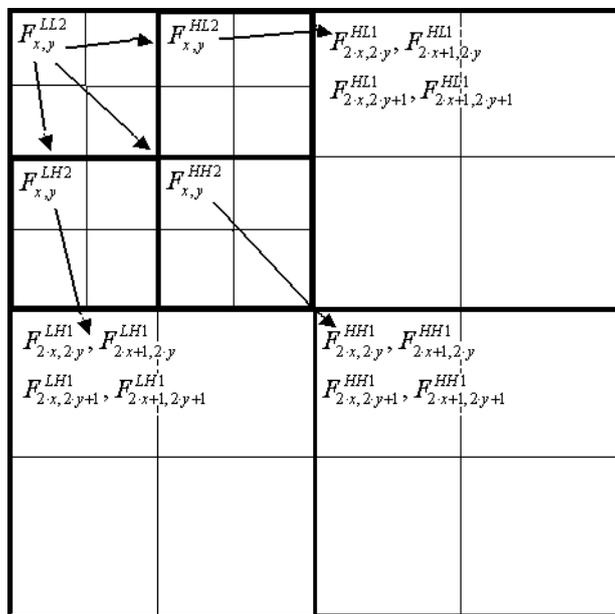


Рис. 1. Древовидный обход вейвлет-коэффициентов

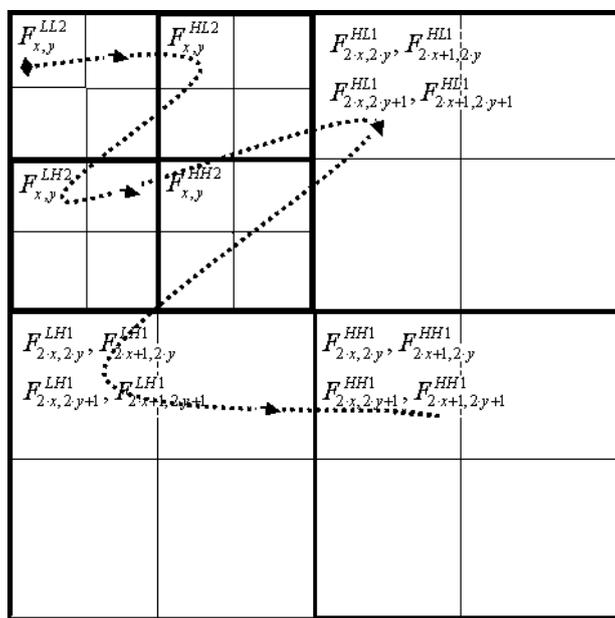


Рис. 2. ZigZag обход вейвлет коэффициентов (метод "1")

Впервые идея нуль-дерева была предложена А.Льюисом и Г. Ноулесом [2]. В их алгоритме применялась древовидная структура данных для описания вейвлет коэффициентов (рис.1). Корневой узел дерева представляет коэффициент масштабирующей функции в самой НЧ области и имеет три отпрыска. Узлы дерева соответствуют вейвлет-коэффициентам масштаба, равного их высоте в дереве. Каждый из узлов имеет четыре отпрыска, соответствующих вейвлет коэффициентам следующего уровня и того



Второй метод ZigZag обхода основан на переупорядочивании вейвлет-коэффициентов в блоке 4×4 (рис. 2). Здесь коэффициент  $F_{x,y}^{LL2}$  помещается в позиции (0,0) блока 4×4,  $F_{x,y}^{HL2}$  – в позиции (0,1),  $F_{x,y}^{LH2}$  – в позиции (1,0),  $F_{x,y}^{HH2}$  – в позиции (1,1), и т.д. Затем расставленные в блоке коэффициенты обходятся в ZigZag порядке, как и при JPEG сжатии. В итоге для каждого коэффициента из субполосы LL2 получается один вектор из 16 элементов, который кодируется в дальнейшем на основе адаптивного или разрядно-срезового алгоритмов.

## Реализованные методы кодирования

В работе были разработаны и реализованы следующие методы сжатия: "Метод нуль-дерева", "Метод нуль-дерева с кодовой книгой", "Адаптивное кодирование вейвлет-коэффициентов, упорядоченных в ZigZag порядке", "Кодирование вейвлет-коэффициентов, упорядоченных в ZigZag порядке, на основе кодовой книги", "Разрядно-срезовой алгоритм". Количественной характеристикой качества восстановленного изображения, по отношению к оригиналу, являлись: MSE – средняя квадратичная ошибка, SNR – соотношение сигнал/шум, PSNR – соотношение пика сигнала к показателю SNR.

### Метод нуль-дерева

Кодирование нульдеревом основано на том, что, если коэффициент мал, его отпрыски на дереве зачастую тоже малы. Это объясняется тем, что значимые коэффициенты возникают вблизи контуров и текстур, которые локальны. Нетрудно увидеть, что этот подход является разновидностью предсказания. Это предсказание можно свести к минимуму, предположив, что, если какой-либо коэффициент незначим (т.е. он меньше по модулю некоторого порога Thr), то все его потомки также будут незначимыми. Дерево или субдерево, которое содержит только незначимые коэффициенты, называется нуль-деревом [2].

Выбранные таким образом вейвлет-коэффициенты затем сжимаются с помощью вторичного алгоритма сжатия (ZIP). Сравнение данного метода сжатия с другими методами приведено в табл. 1.

### Метод нуль-дерева с кодовой книгой

Данный метод является модификацией метода кодирования нуль-деревом. Здесь сжатие ветвей, не являющихся нуль-деревом, происходит с использованием кодовой книги.

В отличие от классического кодирования, в данном методе используются несколько другие ветви вейвлет-коэффициентов. Поскольку субполоса LL2 содержит низкочастотные коэффициенты, то крайне нежелательно их изменять или терять при сжатии. Поэтому коэффициенты субполосы LL2 сохраняются без изменений. Далее рассматриваются оставшиеся ветки вейвлет-коэффициентов (рис.1):

$$F_{x,y}^{HL2}, F_{2-x,2-y}^{HL1}, F_{2-x+1,2-y}^{HL1}, F_{2-x,2-y+1}^{HL1}, F_{2-x+1,2-y+1}^{HL1}$$

$$F_{x,y}^{LH2}, F_{2-x,2-y}^{LH1}, F_{2-x+1,2-y}^{LH1}, F_{2-x,2-y+1}^{LH1}, F_{2-x+1,2-y+1}^{LH1}$$

$$F_{x,y}^{HH2}, F_{2-x,2-y}^{HH1}, F_{2-x+1,2-y}^{HH1}, F_{2-x,2-y+1}^{HH1}, F_{2-x+1,2-y+1}^{HH1}$$

эти ветки образуют вектора из 5 элементов.

Рассмотрим процесс кодирования таких ветвей на примере одной из ветвей:

$$F_{x,y}^{HL2}, F_{2-x,2-y}^{HL1}, F_{2-x+1,2-y}^{HL1}, F_{2-x,2-y+1}^{HL1}, F_{2-x+1,2-y+1}^{HL1}$$

В начальном состоянии есть одна "чистая" кодовая книга.

Если данная ветвь является нуль-деревом, т.е. (в рассматриваемом примере) все коэффициенты субполосы НН1 меньше некоего порога  $Thr$ , то на приемную сторону передается значение -1, что приведет при восстановлении к обнулению всех коэффициентов субполосы НН1 данной ветви и установлению  $F_{x,y}^{HL2} = Thr$ .

Метод сжатия	До вторичного сжатия	После вторичного сжатия (zip)
Исходный файл BMP	227k	137k
JPEG	-----	25k
Метод нуль-дерева	67k	38k
Метод нуль-дерева с кодовой книгой	61k	22k
Адаптивное кодирование вейвлет-коэффициентов, упорядоченных в ZigZag порядке	82k	45k
Кодирование вейвлет-коэффициентов, упорядоченных в ZigZag порядке на основе кодовой книги	41k	28k
Разрядно-срезовой алгоритм (сохраняются 8 бит исходных коэффициентов)	150k	73k
Разрядно-срезовой алгоритм (сохраняются 7 бит исходных коэффициентов)	140k	59k
Разрядно-срезовой алгоритм (сохраняются 6 бит исходных коэффициентов)	113k	47k

Таблица 1. Сравнение реализованных алгоритмов по степени сжатия изображения

Метод сжатия	MSE	SNR	PSNR
JPEG	9,5	22.7	28.6
Метод нуль-дерева	10,1	19.3	25.0
Метод нуль-дерева с кодовой книгой	9,3	22.7	28.7
Адаптивное кодирование вейвлет-коэффициентов, упорядоченных в ZigZag порядке	19,4	16.5	22.3
Кодирование вейвлет-коэффициентов, упорядоченных в ZigZag порядке на основе кодовой книги	24,5	14.4	20.3
Разрядно-срезовой алгоритм (сохраняются 8 бит исходных коэффициентов)	5,8	26.9	32.7
Разрядно-срезовой алгоритм (сохраняются 7 бит исходных коэффициентов)	10,2	21.9	27.8
Разрядно-срезовой алгоритм (сохраняются 6 бит исходных коэффициентов)	13,3	19.7	25.6

Таблица 2. Сравнение реализованных алгоритмов по качеству восстановленного изображения

Если данная ветвь не является нуль-деревом, то процесс кодирования организуется следующим образом. Первая найденная ветвь, не являющаяся нуль-деревом, помещается в кодовую книгу, а на приемную сторону передается код данной ветви (поскольку ветвь была первой, то передается код 0). В дальнейшем любая найденная ветвь сравнивается со всеми ветвями, уже находящимися в кодовой книге. Пусть  $S$  – текущий размер кодовой книги. Если все коэффициенты данной ветви  $Bv_i$  отличаются от

какой-то ветви  $Br_n$  с кодом  $C$ , находящейся в кодовой книге, не более чем на порог  $Thr$ , то эти ветви считаются идентичными, и на приемную сторону передается код  $C$ . Если ни одна ветвь в словаре не совпала с кодируемой ветвью, то в словаре заводится новая запись для ветви  $Br_i$  и передается новый код  $S+1$ . Чтобы коды всех ветвей были однобайтовыми, размер кодовой книги был ограничен 250-ю значениями. Таким образом, если размер книги превысит 250, то будет заведена новая кодовая книга, и дальнейшая работа будет производиться только с ней. Эксперименты показали, что такой подход к кодированию вейвлет-ветвей очень эффективен. В большинстве случаев наблюдался большой процент попадания кодируемых ветвей в кодовую книгу. Но из-за того, что многие ненулевые ветви заменяются аналогом, а не оригиналом, на некоторых изображениях могут появляться визуальные артефакты, ухудшающие качество восстановленного изображения.

Сравнение данного метода сжатия с другими методами приведено в табл. 1, 2.

### **Кодирования вейвлет-коэффициентов, упорядоченных в ZigZag порядке**

После двухуровневого вейвлет-преобразования и квантования вейвлет-коэффициентов для каждого коэффициента из субполосы LL2 формируется вектор из 16 соответствующих коэффициентов, упорядоченных ZigZag методом "1" или "2".

Кодирование таких векторов осуществлялось тремя разными способами, что позволяет выделить три различных метода сжатия.

#### ***Адаптивное кодирование***

В данном методе, по аналогии с кодированием нуль-деревом, предпринимается попытка адаптивно ограничить длину передаваемого вектора на приемную сторону. В отличие от кодирования нуль-деревом, здесь нельзя однозначно предполагать, что, если данный узел меньше порога, то оставшиеся коэффициенты незначимы, и ими можно пренебречь. Чтобы не допустить потерю значимых коэффициентов, вводятся следующие правила.

1. Длина сохраняемого вектора не может быть меньше 3. Это правило принудительно заставляет сохранять и передавать на приемную сторону низкочастотные и средне-частотные коэффициенты, чтобы избежать сильного искажения при восстановлении изображения.
2. Передачу вектора можно прекратить, если выполняются следующие условия:
  - данный коэффициент  $br[i]$  меньше порога  $Thr_1$ ,  $0 < br[i] < Thr_1$ , где  $i$  – порядковый номер коэффициента в векторе [0;15];
  - модуль коэффициента следующего за  $i$ , меньше порога  $Thr_2$ ,  $abs(br[i+1]) < Thr_2$ . Пороги  $Thr_1$  и  $Thr_2$  выбираются экспериментально и являются параметрами метода. Обычно  $Thr_1$  выбирается близким к 0, а  $Thr_2$  – равным нескольким единицам или десяткам. Увеличение порогов приводит к уменьшению объема сжатых данных, но и к увеличению вносимых искажений при восстановлении изображения.

Сравнение данного метода сжатия с другими методами приведено в табл. 1, 2.

Несмотря на то, что получаемый размер сжатого файла больше, чем при сжатии методом нуль-дерева с кодовой книгой, необходимо при этом отметить, что визуальное качество восстановленного изображения намного лучше по сравнению с другими методами. Кроме того, метод позволяет "тонко" регулировать процесс сжатия, как на уровне квантования, так и при выборе длины передаваемого вектора.

#### ***Кодирование с кодовой книгой***

Данный метод аналогичен методу сжатия нуль-деревом с кодовой книгой, за исключением того, что вводится жесткий размер длины  $L$  кодируемого вектора. Чем дли-

на  $L$  меньше, тем меньше размер сжатых данных, но хуже качество восстановленного изображения. Эксперименты показали, что при  $L < 14$  восстановленное изображение претерпевает сильные искажения.

Сравнение данного метода сжатия с другими методами приведено в табл. 1, 2.

### Разрядно-срезовый алгоритм кодирования

Разрядно-срезовый алгоритм кодирования применяется к векторам, полученным с помощью ZigZag сканирования для каждого коэффициента из субполосы LL2.

Суть алгоритма состоит в следующем. Каждый из коэффициентов представляется в двоичном виде. Коэффициенты размещаются друг под другом, образуя двоичную матрицу  $8 \times 16$ .

	Исходное значение		Двоичное значение							
			0	1	2	3	4	5	6	7
1	28	=	0	0	0	1	1	1	0	0
2	0	=	0	0	0	0	0	0	0	0
3	-8	=	1	1	1	1	1	0	0	0
4	0	=	0	0	0	0	0	0	0	0
5	1	=	0	0	0	0	0	0	0	1
6	0	=	0	0	0	0	0	0	0	0
7	0	=	0	0	0	0	0	0	0	0
8	0	=	0	0	0	0	0	0	0	0
Новый П/Н			↓	↓	↓	↓	↓	↓	↓	↓
			32	32	32	-96	-96	-128	0	8
			1	3	5	7	9	11	13	15
9	-2	=	1	1	1	1	1	1	1	0
10	0	=	0	0	0	0	0	0	0	0
11	-1	=	1	1	1	1	1	1	1	1
12	0	=	0	0	0	0	0	0	0	0
13	1	=	0	0	0	0	0	0	0	1
14	0	=	0	0	0	0	0	0	0	0
15	0	=	0	0	0	0	0	0	0	0
16	0	=	0	0	0	0	0	0	0	0
Новый П/Н			↓	↓	↓	↓	↓	↓	↓	↓
			-96	-96	-96	-96	-96	-96	-96	40
			2	4	6	8	10	12	14	16

Таблица 3. Алгоритм разрядно-срезового кодирования

Рассмотрим пример такого преобразования для конкретного набора коэффициентов (табл. 3).

Допустим, что кодируемый вектор коэффициентов имеет вид  $IshV[i]$ , где  $i \in [0;15]$ : 28, 0, -8, 0, 1, 0, 0, 0, -2, 0, -1, 0, 1, 0, 0, 0

Расположим эти коэффициенты друг под другом, одновременно переводя их в двоичную форму.

Теперь разобьем двоичную таблицу пополам и рассматриваем получившиеся таблицы по столбцам с 0 по 7. Поскольку в данном столбце по 8 строк, содержащих двоичные значения, то переведем их в десятичную форму. Упорядочим полученные числа в порядке, соответствующим значению поля *Новый П/Н*, тем самым получим новый

набор из 16 чисел, но уже содержащих меньшее число различных значений. В результате имеем вектор  $PlaneV[i]$ , где  $i \in [0;15]$ :

32, -96, | 32, -96, | 32, -96, | -96, -96, | -96, -96, | -128, -96, | 0, -96, | 8, 40

Данный вектор обладает одним хорошим свойством: его последние элементы составлены из наименее значимых бит исходных коэффициентов, и поэтому их можно не передавать на приемную сторону, что уменьшит объем хранимых данных. Качество восстановленного изображения определяется тем, сколькими наименьшими битами пренебрегается при сжатии. Поскольку значения  $PlaneV[i]$  могут повторяться от вектора к вектору, то наиболее эффективно кодировать их с помощью словаря.

Сравнение данного метода сжатия с другими методами приведено в табл. 1, 2..

## Заключение

В работе рассмотрены новые методы сжатия изображений на основе дискретного вейвлет-преобразования. Показаны возможные способы обхода частотных вейвлет-коэффициентов. Приведены сравнительные характеристики реализованных методов со стандартным методом сжатия изображений JPEG.

## Литература

1. Дьяконов В.П. От теории к практике. Вейвлеты. М.: Солон-Р, 2002.
2. Воробьев В.И., Грибунин В.Г. Теория и практика вейвлет преобразований. СПб: Военный университет связи; 2000.
3. F. Keissarian; M.F. Daem. Block pattern coding of HVS-based on wavelets for image compression" // SPEE Volume 4472.

# МОДЕЛИРОВАНИЕ ТРАФИКА В КОМПЬЮТЕРНЫХ СЕТЯХ НА ОСНОВЕ МАРКОВСКИХ ПРОЦЕССОВ

М.П. Шалаев

Научный руководитель – д.т.н., проф. И.П. Гуров

Статья посвящена разработке модели состояния компьютерной сети при помощи динамического анализа стохастических потоков трафика и объективной оценки характеристик сети на основе строгих критериев качества.

## Введение

Развитие информационного сообщества в последние годы привело к возникновению новых и расширению уже существующих сетевых инфраструктур, что связано как с непосредственной заинтересованностью капитала, так и с инициативой правящих кругов различных стран, решающих стратегические задачи. Создаются глобальные транснациональные структуры, объединяющие различные национальные научно-исследовательские сети, которые довольно скоро обнаруживают недостаточность финансирования при постоянно растущем спросе на скорость обмена информацией. При этом механизмы рынка требуют, чтобы политика использования ресурсов подобных сетей не противоречила общим принципам использования услуг Интернет.

При решении подобных проблем прежде всего необходима организация системы мониторинга и анализа трафика магистральных сетей и крупных провайдеров Интернет-услуг.

Использование таких систем позволяет проводить различный количественный и качественный анализ работы сети. К сожалению, при исследовании процессов в информационных сетях обычно получают оценки характеристик сети на достаточно коротком промежутке времени либо для небольших сетевых объединений, что не дает ответа на вопрос о характере функционирования глобально распределенной сети.

Вследствие влияния различных факторов на режим работы информационной сети, природа сетевых взаимодействий является случайной, и корректный количественный анализ характеристик исследуемой системы возможен на основе использования статистического подхода и аппарата теории случайных процессов [1–4].

На сегодняшний день наибольший интерес и значимость представляет задача исследования характеристик сетевого трафика с целью анализа состояния сети для решения одной из важнейших задач сетевого администрирования, которой является мониторинг маршрутизаторов и другого оборудования магистральных сетей с целью выявления аномальных режимов функционирования системы или сбоев в работе сети. Данную задачу можно решать с помощью сбора и анализа различной статистической информации по IP-трафику [5, 6], проходящему через тот или иной интерфейс сетевого устройства. При этом аномалии в характеристиках трафика могут представлять собой, например, необоснованный рост или падение интенсивности трафика, изменения в стационарном характере трафика, чрезмерное повышение интенсивности использования отдельных частей сети и т.д. Выявление и распознавание аномальных режимов функционирования сети администраторами часто основывается на так называемых "ad hoc", т.е. специальных, субъективных методах, которые основаны на опыте долговременной работы в области управления сетью.

Задачей настоящей работы является разработка метода оперативной диагностики состояния сети на основе динамического анализа стохастических потоков трафика с привлечением математического аппарата теории случайных процессов и, следовательно, обеспечивающего объективную оценку характеристик информационной сети на основе строгих критериев качества.

## Математический аппарат

Рассмотрение современных теоретических подходов к формализации анализа работы информационных сетей показывает целесообразность и продуктивность использования математического аппарата теории марковских случайных процессов [7, 8]. В отличие от сугубо теоретических подходов к исследованию свойств сетевого трафика, цель данной работы заключается в математической формализации системы, позволяющей на практике проводить оперативную диагностику режима работы сети и определять, нормально ли функционирует сеть.

Одним из возможных подходов к обнаружению сетевых аномалий является сравнение наблюдаемого режима работы системы с так называемым нормальным режимом функционирования, который определяется совокупностью характеристик рассматриваемой сети в течение определенного промежутка времени, соответствующей функционированию системы без существенных отклонений от нормы.

Рассмотрим систему, имеющую конечное число  $K$  всех возможных состояний  $\vartheta_1, \vartheta_2, \dots, \vartheta_k, \dots, \vartheta_K$  [9, 10]. Под состоянием системы будем понимать определенную совокупность параметров, описывающих систему. Пусть система шаг за шагом (в моменты времени  $t_0 < t_1 < t_2 < \dots$ ) случайным образом меняет свое состояние, т.е. имеют место переходы  $\theta_0 \rightarrow \theta_1 \rightarrow \theta_2 \rightarrow \dots$ , где  $\theta_n = \theta(t_n)$  состояние системы через  $n$  шагов, а  $\theta_0 = \theta(t_0)$  – начальное состояние системы.

Полное вероятностное описание поведения рассматриваемой системы определяется совместными конечномерными вероятностями  $P(\theta_0, \theta_1, \dots, \theta_n)$  при всех  $n$ . На основании теоремы умножения вероятностей можем написать

$$\begin{aligned} P(\vartheta_{k_0}, \vartheta_{k_1}, \dots, \vartheta_{k_n}) &= P\{\theta_0 = \vartheta_{k_0}, \theta_1 = \vartheta_{k_1}, \dots, \theta_n = \vartheta_{k_n}\} = \\ &= P\{\theta_0 = \vartheta_{k_0}, \theta_1 = \vartheta_{k_1}, \dots, \theta_{n-1} = \vartheta_{k_{n-1}}\} P\{\theta_n = \vartheta_{k_n} \mid \theta_0 = \vartheta_{k_0}, \theta_1 = \vartheta_{k_1}, \dots, \theta_{n-1} = \vartheta_{k_{n-1}}\}, \end{aligned} \quad (1)$$

где  $\vartheta_{k_\mu}$  – одно из возможных состояний  $\theta_\mu$ .

Расписывая аналогично вероятность  $P\{\theta_0 = \vartheta_{k_0}, \theta_1 = \vartheta_{k_1}, \dots, \theta_{n-1} = \vartheta_{k_{n-1}}\}$ , получаем

$$P(\theta_0, \theta_1, \theta_2, \dots, \theta_{n-1}, \theta_n) = P_0(\theta_0) \prod_{\mu=1}^n \pi_\mu(\theta_\mu \mid \theta_0, \dots, \theta_{\mu-1}), \quad (2)$$

где  $P_0(\theta_0)$  – вероятность начального состояния (в момент времени  $t_0$ ), а

$$\pi_\mu(\vartheta_{k_\mu} \mid \vartheta_{k_0}, \dots, \vartheta_{k_{\mu-1}}) = P\{\theta_\mu = \vartheta_{k_\mu} \mid \theta_0 = \vartheta_{k_0}, \dots, \theta_{\mu-1} = \vartheta_{k_{\mu-1}}\} \quad (3)$$

– условная вероятность, т.е. вероятность перехода на  $\mu$ -м шаге,  $1 \leq \mu \leq n$ .

Следовательно, для вычисления совместных вероятностей  $P(\theta_0, \theta_1, \dots, \theta_n)$  нужно задать начальное состояние системы и указать правило осуществления смены состояний, позволяющее вычислить вероятности перехода  $\{\pi_\mu\}$ . К определению простой цепи Маркова приводит случай, когда вероятность состояния параметра  $\theta_n$  в момент времени  $t_n$  зависит лишь от того, в каком состоянии находилась система в непосредственно предшествующий ему момент времени  $t_{n-1}$ , и не зависит от того, в каких состояниях находилась система в более ранние моменты времени  $t_0, t_1, \dots, t_{n-2}$ .

Это свойство характерно для всех видов марковских процессов. Именно поэтому марковские процессы также называют процессами без последействия или без памяти.

Таким образом, последовательность состояний или совокупность дискретных случайных величин образует простую цепь Маркова, если для всех  $n = 1, 2, 3, \dots$  и всех возможных значений случайных величин  $\theta_n$  имеет место соотношение

$$\pi_n(\theta_n | \theta_0, \theta_1, \dots, \theta_{n-1}) = \pi_n(\theta_n | \theta_{n-1}). \quad (4)$$

Условные вероятности  $\pi_n(\theta_n | \theta_{n-1})$  (5) принято называть вероятностями перехода. Нетрудно убедиться, что для простой марковской цепи формула (2) принимает вид

$$P(\theta_0, \theta_1, \dots, \theta_n) = P_0(\theta_0) \prod_{\mu=1}^n \pi_\mu(\theta_\mu | \theta_{\mu-1}). \quad (6)$$

Полное вероятностное описание простой цепи Маркова достигается заданием вероятностей начального состояния и вероятностей перехода

$$\pi_\mu(\theta_\mu | \theta_{\mu-1}) = P\{\theta_\mu = \vartheta_{k_\mu} | \theta_{\mu-1} = \vartheta_{k_{\mu-1}}\}.$$

Пусть процесс  $\theta(t)$  описывает функционирование сети. Введем следующие обозначения для безусловных (абсолютных) и условных вероятностей

$$p_k(n) = P\{\theta_n = \vartheta_k\}, \pi_{jk}(\mu, n) = P\{\theta_n = \vartheta_k | \theta_\mu = \vartheta_j\}, j, k = 1, 2, \dots, K; 0 \leq \mu \leq n; n = \overline{0, N}.$$

Смысл введенных вероятностей следующий. Величина  $p_k(n)$  есть безусловная (абсолютная) вероятность нахождения системы в состоянии  $\vartheta_k$  на  $n$ -м шаге (т.е. в момент времени  $t = t_n$ ), а  $p_k(t_0) = p_k^0$  – начальная вероятность состояния  $\vartheta_k$ . Условная вероятность  $\pi_{jk}(\mu, n)$  определяет вероятность состояния  $\vartheta_k$  при  $t_n$ , если в более ранний момент времени  $t_\mu < t_n$  система находилась в состоянии  $\vartheta_j$ . Эти условные вероятности обычно называют вероятностями перехода из состояния  $\vartheta_j$ , в состояние  $\vartheta_k$  за  $(n - \mu)$  шагов.

Среди простых цепей Маркова различают однородные и неоднородные. Однородная цепь характеризуется тем, что вероятности  $\pi_{jk}(\mu, n)$  зависят только от разности аргументов, т.е.  $\pi_{jk}(\mu, n) = \pi_{jk}(n - \mu)$ ,  $n > \mu \geq 0$ . Согласно теории цепей Маркова, данное предположение в совокупности с предположением о том, что вероятности нахождения системы в том или ином состоянии также не зависят от времени,  $\mathbf{P}^T(n) = \mathbf{P}^T(1) = \mathbf{P}^T = \{p_k\}$ , сводится к рассмотрению  $\theta(t)$  в качестве простой стационарной цепи Маркова. Учитывая вышесказанное, общее выражение (6) для определения совместной вероятности  $P(\theta_0, \theta_1, \dots, \theta_n)$  переписывается в виде

$$P(\theta_0, \theta_1, \dots, \theta_n) = P(\theta_0) \prod_{\mu=1}^n \pi(\theta_\mu | \theta_{\mu-1}),$$

где  $\theta_i$  может принимать любые значения из множества возможных состояний  $\vartheta_1, \vartheta_2, \dots, \vartheta_k, \dots, \vartheta_K$ .

Данное предположение имеет очевидные достоинства, а именно, относительную математическую и вычислительную простоту нахождения вероятности состояния системы в определенный момент времени в будущем, отсутствие сложностей, связанных с временной уникальностью матриц вероятностей одношаговых переходов, в частности, с возведением матриц в степень  $n$  или перемножением  $n$  различных матриц.

Характер сетевых взаимодействий в целом определяется зависимостями более высокого порядка, чем простая цепь Маркова. Однако моделирование системы с помощью случайных процессов высокого порядка является практически неосуществимой задачей ввиду сложности построения формального математического описания и, как следствие, чрезмерно больших вычислительных затрат, особенно при работе с большими объемами данных в режиме реального времени. Более того, часто не очевидно, какой порядок модели следует выбрать для полного описания конкретного процесса. Таким образом, при построении модели процесса сделанные выше предположения об отсутствии эффекта памяти, однородности и стационарности вносят существенные упро-

щения в аналитическую модель и вычислительный алгоритм, что наиболее предпочтительно, особенно на начальных этапах исследований.

Принятую выше модель нормального функционирования сети можно в дальнейшем применять при исследовании реального состояния сети, определяя тем или иным образом различия между предсказанным значением по "шаблону" нормального функционирования и проведенными измерениями. Для реализации такого подхода необходимо "обучить" или настроить алгоритм для работы с сетью. Значения вероятностей начального состояния системы  $\mathbf{P}^T(0) = \{p_k^0\}$ , а также матрицу одношаговых переходов можно вычислить на основе экспериментальных данных о нормальном функционировании системы.

Рассмотрим последовательность  $N$  состояний системы  $\theta_0, \theta_1, \dots, \theta_N$  в моменты времени  $t_0 < t_1 < t_2 < \dots < t_N$ . Тогда матрица-столбец вероятностей начального состояния и стохастическая матрица переходов системы определяются как

$$\mathbf{P}^T(0) = \{p_k^0\} = \left\{ \frac{N_k}{N} \right\},$$

где  $N_k$  – количество значений  $\theta_t$  в состоянии  $\mathcal{S}_k$ ,  $N$  – общее количество наблюдений  $\theta_t$ ,

$$\boldsymbol{\pi} = \{\pi_{ij}\} = \left\{ \frac{N_{ij}}{N_i} \right\},$$

где  $N_{ij}$  – количество пар  $(\theta_{t-1}, \theta_t)$ , в которых  $\theta_{t-1}$  находится в состоянии  $\mathcal{S}_i$ , а  $\theta_t$  – в состоянии  $\mathcal{S}_j$ ,  $N_i$  – количество пар  $(\theta_{t-1}, \theta_t)$ , в которых  $\theta_{t-1}$  находится в состоянии  $\mathcal{S}_i$ , а  $\theta_t$  – в любом другом состоянии  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_k, \dots, \mathcal{S}_K$ . При этом на основании статистических данных об эволюции реальной системы можно получить требуемые характеристики модели, а именно, столбец вероятностей начального состояния и стохастическую матрицу марковской цепи.

### Практическая реализация

Модель, полученную для режима нормального функционирования сети, необходимо сравнивать с текущим состоянием сети. В определении нормального функционирования изначально заложена большая гибкость по отношению к критериям оценки работы сети, а именно, совокупность характеристик определяется в зависимости от целей исследования.

Примем, что нормальное функционирование сети характеризуется интенсивностью входящего брутто-трафика. Для применения обученной модели рассматривается так называемое временное "окно" в прошлом или интервал наблюдения размером  $N$  временных отсчетов. Накладывая окно выбранного размера на непрерывный поток трафика через маршрутизатор, получим отображение работы сети в интервале времени  $[t - N; t]$ , где  $t$  обозначает текущий момент времени.

Предположим, что система последовательно случайным образом изменяла свое состояние в течение фиксированного промежутка времени и принимала состояния  $\theta_{t-N}, \theta_{t-(N-1)}, \dots, \theta_t$  соответственно в моменты времени  $t - N < t - (N - 1) < t - (N - 2) < \dots < t$ . Тогда основным критерием оценки работы системы является вероятность того, что последовательность состояний  $\theta_{t-N}, \theta_{t-(N-1)}, \dots, \theta_t$  является отображением нормальной функционирования системы или, иными словами, вероятность того, что марковская модель нормального функционирования системы поддерживает данную последова-

тельность состояний  $\theta_{t-N}, \theta_{t-(N-1)}, \dots, \theta_t$ . Указанная вероятность определяется, исходя из математической модели, принятой выше, а именно

$$P(\theta_{t-N}, \dots, \theta_t) = P(\theta_{t-N}) \prod_{\mu=t-(N-1)}^t \pi_{k_{\mu} k_{\mu-1}}.$$

Очевидно, что чем выше значение вероятности, вычисленное с помощью данной формулы, тем вероятнее, что последовательность  $\theta_{t-N}, \theta_{t-(N-1)}, \dots, \theta_t$  является отображением нормальной работы сети. В свою очередь, аномалии в функционировании сети приводят к уменьшению получаемых значений, тем самым значительно снижая вероятность поддержки моделью данной последовательности.

Для того, чтобы упростить процесс проверки соответствия модели системы, функционирующей в реальном времени, для эксперимента была выбрана тестовая последовательность отсчетов случайной интенсивности трафика, в процессе исследования которой был сделан вывод о применимости модели в условиях, приближенных к реальным. При этом для системы с искусственным трафиком было показано, что даже при отклонениях в случайной интенсивности трафика от нормального режима функционирования, практически не фиксируемых с помощью субъективных оценок, таких, как, например, наблюдение за изменениями интенсивности брутто-трафика на узловом маршрутизаторе, модель, построенная на основе теории марковских процессов, дает положительные результаты.

Отметим, что для "обучения" системы использовался трафик реально функционирующей сети за период времени две недели, что составило примерно 2000 Гбайт трафика и 3500 отсчетов случайной интенсивности процесса. В результате многочисленных экспериментов и исследований модели был сделан вывод о том, что вероятности поддержки моделью нормального режима работы сети намного выше вероятностей поддержки последовательности, вызванной сбоями при функционировании сети. Кроме этого, экспериментально подтверждена возможность выбора значения пороговой вероятности, достижение которой идентифицирует наличие сбоев в работе сети.

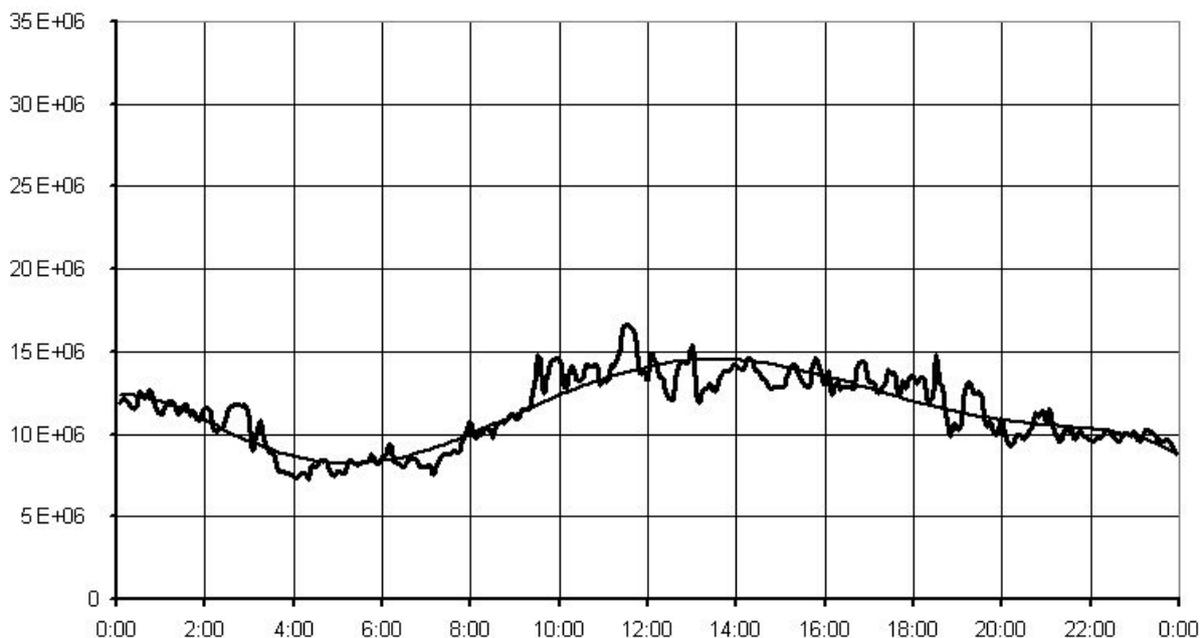


Рис. 1. Последовательность нормальных отсчетов входящего трафика (бит/сек)

На рис. 1, 2 показаны последовательности интенсивности входящего брутто-трафика для режима нормального функционирования (рис. 1) и аномального функционирования (рис. 2).

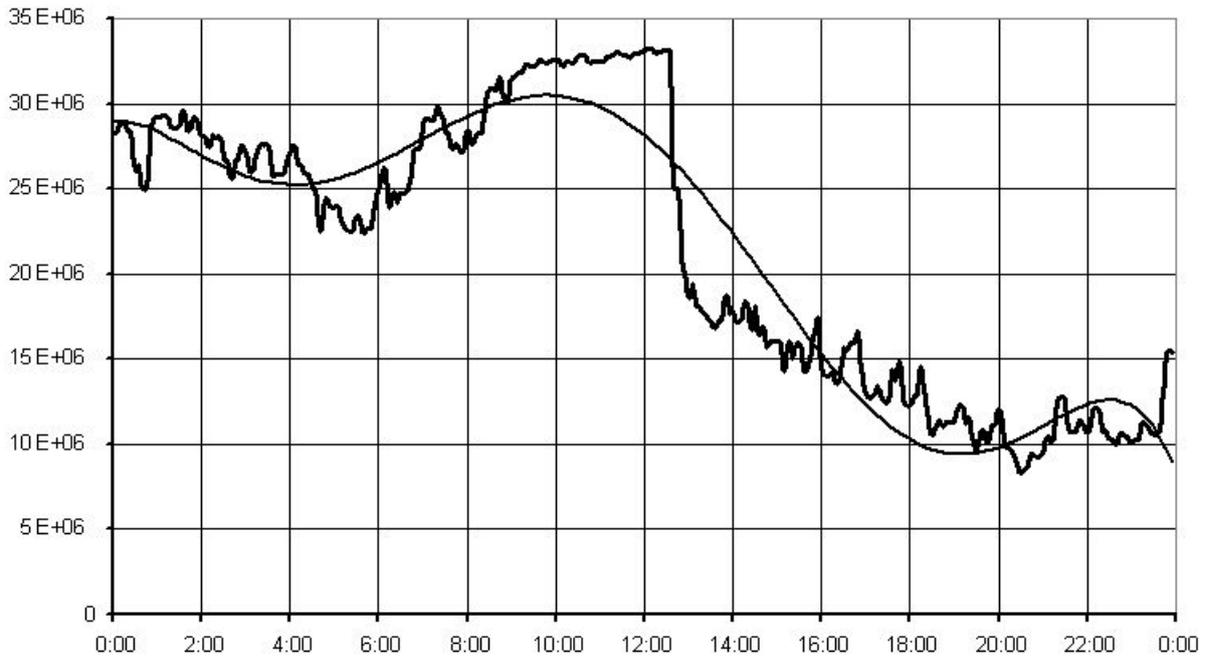


Рис. 2. Последовательность аномальных отсчетов входящего трафика (бит/сек)

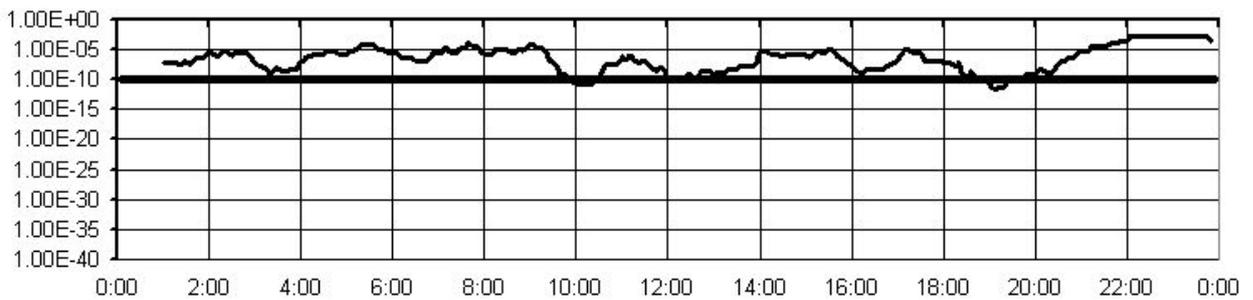


Рис. 3. Вероятность поддержки моделью нормальной последовательности.

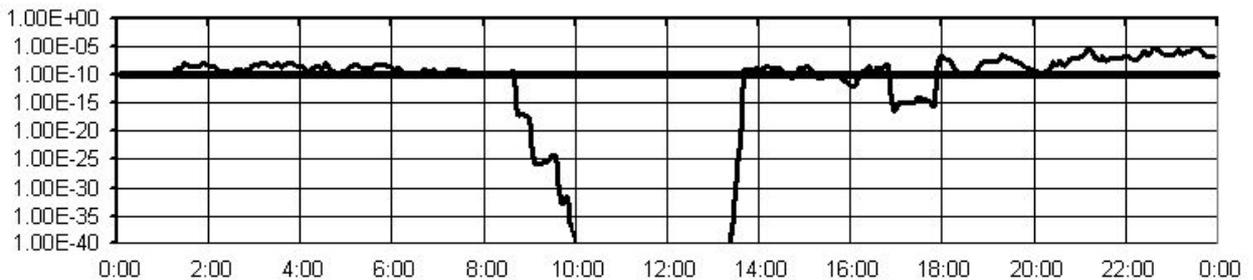


Рис. 4. Вероятность поддержки моделью аномальной последовательности

На рис. 3, 4 приведены сравнительные данные о вероятности поддержки моделью нормального (рис. 3) и аномального (рис. 4) входящего трафика.

### Заключение

Результатом проведенных исследований явилось построение теоретико-технической модели системы, позволяющей определять состояние реально функционирующей сети. Основными входными параметрами, определяющими состояние сети, явились данные о работе магистрального соединения федеральной университетской компьютерной сети RUNNet, составляющие несколько Тбайт входящего IP трафика исследуемой сети, что положительным образом отразилось на результатах работы, а так-

же вызвало необходимость использования в качестве инструментария современных технологий коммутации потоков трафика [11] в комбинации с другими эффективными механизмами получения статистики работы сети.

Принятый подход обеспечивает существенный выигрыш в производительности программно-аппаратного обеспечения и позволяет проводить исследования функционирования глобального объединения сетей в режиме реального времени. Это накладывает определенные ограничения как на сложность практической реализации построенной модели, так и выбранного математического аппарата. Предложенный метод полностью оправдал себя, поскольку использование теории марковских процессов, оставаясь общепризнанным при исследовании свойств сетевого трафика, обеспечило возможность наиболее просто и эффективно реализовать модель системы и тем самым на практике достичь желаемого результата.

Результаты проведенных исследований определяют круг проблем, решение которых позволит дополнительно повысить эффективность использования разработанного метода. В частности, исследование, классификация и математическая формализация сетевых аномалий позволит более эффективно идентифицировать соответствующие изменения в характере сетевого трафика. Уточнение и усложнение модели, построенной на основе теории марковских процессов, обеспечивает возможность детального описания локальных изменений характеристик работы сети [12–14].

Предложенный в работе подход к диагностике сетевого трафика позволяет построить модель суперпозиции различных составляющих трафика и создать автоматизированную систему управления, позволяющую идентифицировать состояние сети, выявлять возможный источник аномалий и предпринимать оперативные действия, необходимые для обеспечения нормального функционирования сети.

## Литература

1. Lizcano P. J. et al. MEHARI: a system for analyzing the use of the Internet services // *Computer Networks*. 1999. V. 31.
2. Uhlig S., Bonaventure O. Implications of Interdomain Traffic Characteristics on Traffic Engineering / *Infonet group, University of Namur, Belgium*. 2001.
3. Uhlig S., Bonaventure O. Analysis of Interdomain Traffic / *Infonet group, University of Namur, Belgium*. 2001.
4. Bradford P., Plonka D. Characteristics of Network Traffic Flow Anomalies / *Proceedings of ACM SIGCOMM Internet Measurement Workshop*. 2001.
5. Cooperative Association for Internet Data Analysis (CAIDA). Cflowd: traffic flow analysis tool / *Technical documentation*. 1998.
6. Plonka D. Flowscan: A network traffic flow reporting and visualization tool / *Proceedings of the USENIX Fourteenth System Administration Conference LISA XIV*. 2000.
7. Hlavac H., Kotsis G., Steinkellner C. Traffic Source Modeling / *Institute of Applied Computer Science and Information Systems, University of Vienna*. 1999.
8. Salvador P., Nogueira A., Valadas R. Modeling local area network traffic with Markovian traffic models / *Institute of Telecommunications, University of Aveiro, Portugal*. 2001.
9. Тихонов В. И. Статистическая радиотехника. М.: Радио и связь, 1982.
10. Тихонов В. И., Миронов М. А. Марковские процессы. М.: Сов. Радио, 1977.
11. NetFlow services and applications / *Technical documentation*. Cisco Systems. 1999.
12. Jagerman D. L., Melamed B., Willinger W. Stochastic modeling of traffic processes // *Frontiers in Queueing: Models, Methods and Problems*. 1996.
13. Kulkarni L. A., Li S. Q. Measurement-Based Traffic Modeling: Capturing important statistics // *Journal of Stochastic Modeling*. 1998. V. 14.
14. Sang A., Li S. Q. A Predictability Analysis of Network Traffic // *Proceedings of IEEE INFOCOM*. 2000.

# РАЗРАБОТКА СИСТЕМЫ ОРГАНИЗАЦИИ И УПРАВЛЕНИЯ САЙТА КАФЕДРЫ УЧЕБНОГО ЗАВЕДЕНИЯ

А. А. Малинин

Научный руководитель – к.т.н., доцент Н.Ю. Иванова

Предлагается оригинальная методика разработки математической модели и программного обеспечения системы создания сайтов высших учебных заведений.

## Введение

Развитие информационных технологий оказывает колоссальное влияние на всю человеческую деятельность, в том числе и на сферу образования. Внедряются системы дистанционного обучения, все чаще компьютерное оборудование используется в лабораторных работах и на практических занятиях, создаются системы дистанционного обучения. Современные технологии позволяют налаживать прямые контакты ученых, преподавателей и студентов высших учебных заведений всех стран мира, проводить совместные конференции, на которых происходит обмен научными идеями и обсуждение возникающих проблем. Преподаватели и студенты вынуждены работать с электронной информацией, постоянно растет потребность в системах внутреннего документооборота в образовательных учреждениях, а также системах, позволяющих получать актуальную информацию от факультетов, кафедр и других подразделений учебных заведений.

В то же время необходимо учитывать тот факт, что образовательные учреждения обладают неплохо развитой коммуникационной инфраструктурой, к тому же им предоставляются льготы при использовании Интернета и при доступе ко многим информационным ресурсам. В результате, рано или поздно, принимается решение о создании собственного сайта в Интернете.

## Выбор методики разработки системы

Можно выделить три характерных метода действий в данном случае:

- разработка статичного сайта и назначение ответственного за его обновление – человека, владеющего навыками работы с HTML, CSS, JavaScript;
- выбор существующей платной или бесплатной системы управления сайтом (CMS) и ее настройка;
- разработка собственной системы организации и управления сайта кафедры учебного заведения.

Плюсом первого варианта является простота разработки, минусом – процесс обновления, который является длительным и достаточно нудным (загрузка файла из Интернета, обновление странички, выгрузка на ФТП). Так, при необходимости поменять шапку всех страниц или меню придется править все файлы сайта.

Второй вариант выглядит очень привлекательным: как правило, используются Open Source CMS products (системы управления содержанием сайта с открытым кодом). Они имеют отличные функциональные возможности, модульную систему, но имеют характерные недостатки: часто в результате анализа кода сторонними лицами обнаруживаются уязвимости данных продуктов, а обновление происходит не так оперативно, как в коммерческих продуктах. К тому же в бесплатных системах иногда находят "backdoors" (буквально – задние двери) – скрытые механизмы для передачи данных, используемых сайтом.

В данной статье рассмотрен третий вариант – создание собственной системы организации и управления сайта кафедры учебного заведения.

## Разработка системы

Предлагается разбить процесс разработки электронного представительства кафедры на следующие этапы:

1. постановка задачи;
2. определение требований к разработке;
3. обзор систем программирования в среде Web;
4. выбор языков программирования и базы данных;
5. разработка алгоритма работы системы;
6. разработка Web-интерфейса;
7. тестирование и отладка системы;
8. разработка документации;
9. наполнение сайта информацией.

Посетителями станут абитуриенты, студенты, преподаватели и те люди, которых интересует информация, представленная на сайте (они попадают на сайт чаще всего через поисковые системы). Желательно предоставить преподавателям возможность загружать файлы на веб-сервер и добавлять новости.

Среди общих задач, характерных для динамических сайтов, можно выделить следующие:

- простота установки сайта и его администрирование;
- максимальная гибкость в выборе операционной системы сервера;
- простота "переезда" сайта на другой хостинг;
- разграничение доступа к информации сайта на уровне пользователей;
- интуитивно понятный, максимально простой и удобный инструмент для редактирования или добавления информации в режиме он-лайн пользователем;
- защищенность скрытой информации от незарегистрированных пользователей,
- защищенность доступа к изменению содержимого сайта, паролей и служебной информации;
- возможность дальнейшей доработки сайта.

Ввиду перечисленных задач необходимо разработать такую систему, которая будет предоставлять необходимый функционал и заданную защищенность доступа к информации и средствам ее модифицирования.

Также необходимо так построить систему и описать ее в документации, чтобы было возможно дополнить систему модулями или доработать определенные функциональные части.

Для разработки системы, удовлетворяющей указанным требованиям, необходимо использовать языки программирования на стороне сервера. Наиболее популярными на данный момент являются ASP, Perl, PHP. ASP поддерживается только веб-серверами семейства Microsoft, так что для создания универсальной системы необходимо использовать Perl или PHP (рис.1).

Хранить ли информацию в отдельных файлах или в базе данных – на этот вопрос нет однозначно верного ответа. Если речь идет не о больших объемах данных, если выборки осуществляются 1–2 раза во время формирования страницы, то стоит задуматься об использовании файлов в качестве хранилища информации, в противном случае следует использовать базы данных. Из коммерческих проектов наиболее распространенными являются MS SQL и Oracle, из продуктов с лицензией GPL – mSQL, MySQL, PostgreSQL. Если до разработки сайта никакое программное обеспечение на кафедре или факультете, взаимодействующее с БД, не использовалось, то следует узнать у будущего провайдера о поддерживаемых хостингом базах данных и затем решать, что выбрать. Автор использовал ПО MySQL (версии 3.23.49) при разработке сайта кафедры ПКС УИТМО. Нареканий к быстродействию и надежности не возникло.

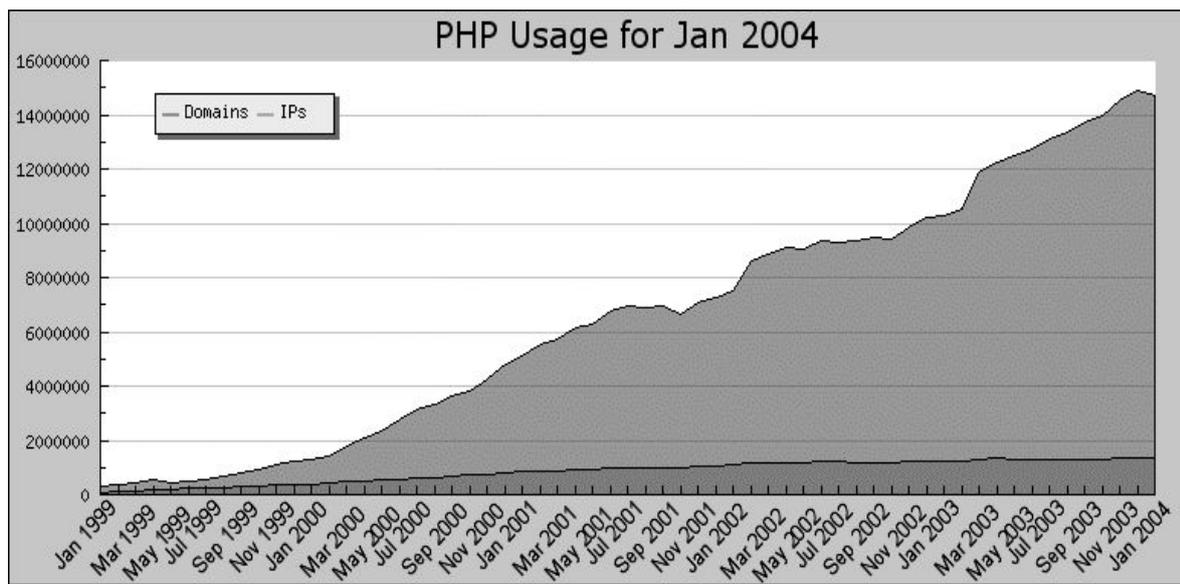


Рис. 1. Использование PHP

Не секрет, что количество атак на сайты в Интернете постоянно возрастает. Часто знания, полученные на занятиях по безопасности информации, используются на примере сайта своего учебного заведения. Необходимо понимать, что, чем больше возможностей предоставляет система, тем больше количество "узких мест" и потенциальных уязвимостей. Поэтому полученный программный код нужно тестировать на реакцию на нестандартные ситуации, внедрение SQL-кода, вывод ошибок в текст страниц. Желательно постоянно следить за активностью посетителей и периодически архивировать содержимое сайта.

Принцип организации сайта отражает реальную организационную структуру кафедры. На рис.2 приведена структурная схема сайта. Поиск информации и навигация должны быть интуитивно понятны любому посетителю, будь то абитуриент или ректор университета.

Как известно, к числу основных составляющих деятельности кафедры относятся учебный процесс, научная и организационная работа. В учебном процессе принимают непосредственное участие преподаватели и студенты, на сайте должны быть представлены расписания, учебные планы и т.д. Каждый из преподавателей может иметь свою страничку.

Научная работа подразумевает подготовку аспирантов и докторантов, публикацию научных трудов и статей, участие в конференциях и защиту диссертаций. Характерно то, что каждое направление деятельности кафедры имеет свою учебно-методическую документацию. Для удобства использования вся литература находится в соответствующем разделе. В случае необходимости в любом разделе можно разместить ссылки на требуемую и сопутствующую литературу.

В целях формирования позитивных впечатлений о сайте и кафедре у новых посетителей желательно не перегружать содержимое графическими и анимированными элементами, необходимо предусмотреть наличие элементов, ставших стандартными де-факто: карта сайта, поиск, быстрый возврат на первую страницу.

Помимо информации официального характера, неплохо размещать на сайте информацию о жизни студентов и преподавателей вне учебы. При этом сайт приобретает эмоциональную окраску, уходя от сухости канцелярского языка.

Часто сайты способны предоставлять разную информацию и набор инструментов своим посетителям. Механизм авторизации позволяет использовать разграничение дос-

тупа к данным на уровне пользователя. На рис. 3 показан алгоритм авторизации реализованный на PHP.

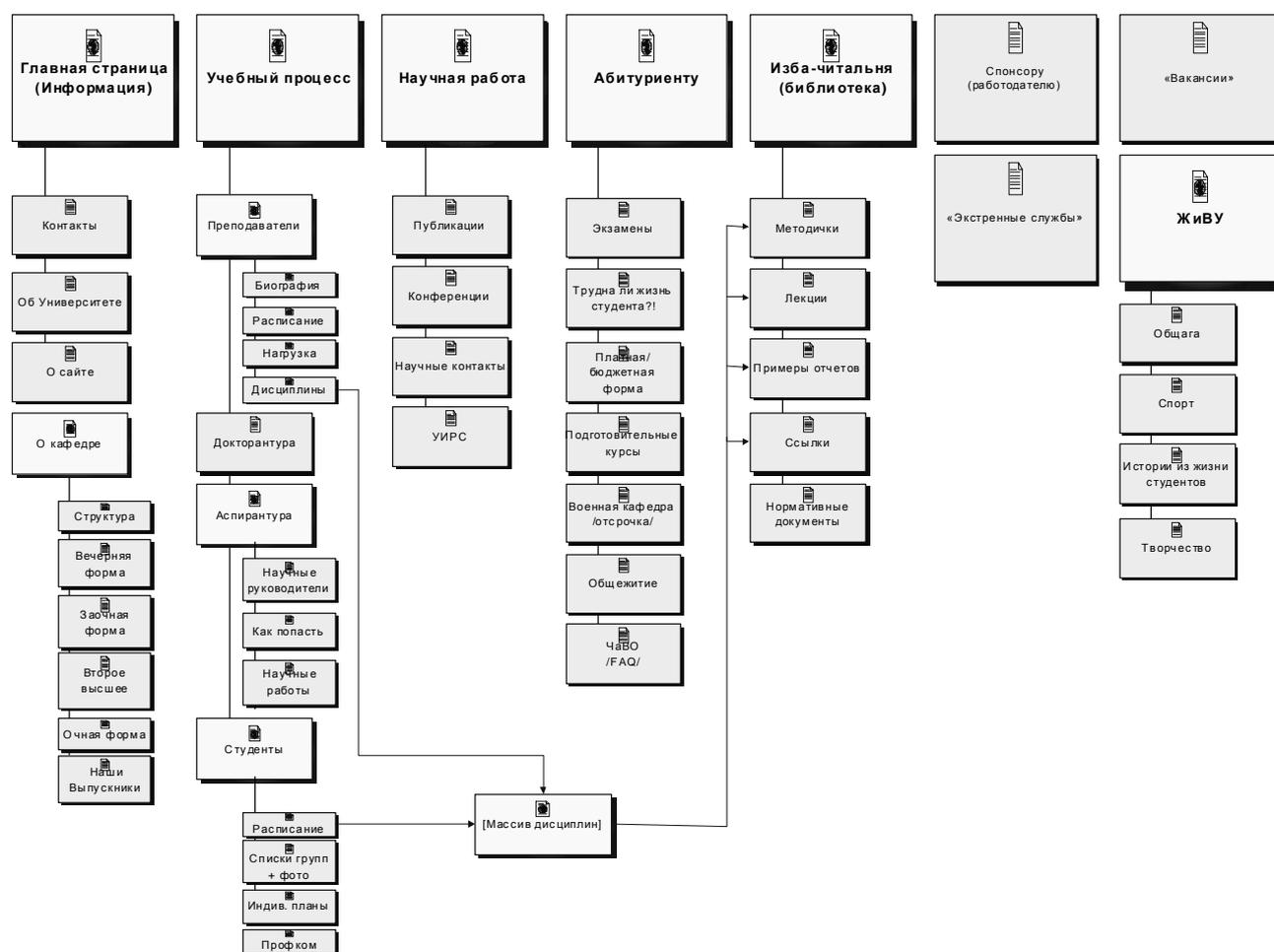


Рис. 2. Структура сайта

В числе зарегистрированных посетителей сайта выделяют особую группу людей – администраторов, которые имеют максимум прав для изменения информации, а иногда и структуры сайта.

Как только разработанная система "узнает" зарегистрированного пользователя, она выводит дополнительную информацию, скрытую от гостей. Это могут быть расширенные личные данные о преподавателях (адрес, телефон и т.д.), новости и тексты, адресованные только данному пользователю и т.п. После заполнения имени пользователя и пароля эти данные плюс идентификатор сессии передаются браузером на сервер файлу-обработчику методом POST. Программа на сервере проверяет, все ли поля заполнены и возможно ли подключение к базе данных, в случае какой-либо ошибки посетителю выводится сообщение с соответствующим содержанием, например "Ошибка авторизации: необходимо заполнить все поля!" Далее выбирает из таблицы "users" все строки, где имя пользователя совпадает с заданным. Затем выбирается та запись, пароль в которой совпадает с введенным.

При возникновении ошибки на этом этапе также выводится соответствующее сообщение типа "Ошибка авторизации: пароль не соответствует введенному имени пользователя!" В случае успеха авторизации в выбранной строке обновляются два поля: идентификатор сессии и время истечения срока действия сессии. При достижении тайм-аута посетителю выдается сообщение о выходе из системы.



Рис. 3. Алгоритм авторизации

В данном случае используется механизм сессий PHP. При отправке браузером любых данных, будь то переменная, данные форм или куки, к содержимому добавляется идентификатор сессии. По умолчанию он добавляется в куки, но при их отключении идентификатор добавляется скрытым полем ко всем формам. Благодаря этому механизму, сервер всегда может однозначно определить текущего посетителя, так как в противном случае все запросы браузеров равнозначны.

Стандартом де-факто стало наличие 'бэк-офиса' на сайте. Back-office – это набор инструментов, позволяющих редактировать содержимое сайта, не покидая его. Инструментарий предоставляет возможность редактировать содержание, добавлять и уничтожать разделы, новости, редактирование и ввод новых пользователей, их прав, паролей. Как было сказано ранее, все посетители разделены на три группы:

- гости;
- зарегистрированные пользователи;
- администраторы.

Соответственно, набор инструментов для каждой группы будет различным. Зарегистрированные пользователи могут редактировать раздел, соответствующий им, а также добавлять новости.

Администраторы способны добавлять, удалять и редактировать все разделы, всех пользователей и их права.

Хорошим тоном является наличие подробной документации по разработке и использованию сайта. Пользователями данной системы являются несколько категорий посетителей, некоторые из них не владеют навыками обращения с языком гиперразметки текста (HTML). Соответственно, документация разделена на несколько частей:

- по установке системы;
- по администрированию;
- для конечных пользователей системы.

Предполагается, что человек, производящий установку системы, знаком с особенностями и принципом функционирования операционной системы и веб-сервера, а также обладает необходимыми правами.

Предполагается, что администратор системы обладает навыками работы с HTML и знанием принципов функционирования системы. Документация по администрированию сайта должна содержать описание инструментария редактирования текущих разделов, создание новых и использование соответствующих возможностей по работе с пользователями. Также должны присутствовать рекомендации по работе с базой данных и управлением загруженных на сервер файлов. Следует обратить внимание на варианты политик безопасности имен пользователей и паролей.

Конечный пользователь может не обладать специализированными навыками работы с HTML. В панель управления системой, используемой авторизованным пользователем, внедрена так называемая контекстно-зависимая помощь. В документации, адресованной данной группе посетителей, должны быть указаны тэги форматирования текста и использования списков, заголовков и таблиц.

Обязательно должна присутствовать информация о требованиях для соответствия стандарту XHTML. Также необходимо указать основные правила безопасности при размещении контактной информации в Интернете.

Все инструкции для наглядности снабжены необходимыми примерами и графическим пояснительным материалом.

Разработанная документация доступна в текстовом формате и представлена в формате презентации Macromedia Flash, это необходимо для повышения удобства пользователей и возможности в режиме офф-лайн ознакомиться с предоставленными материалами.

Хотелось бы перечислить наиболее часто встречающиеся ошибки на кафедральных сайтах: нередко посетители попадают к вам через поисковые машины, поэтому необходимо указывать краткую информацию об институте или университете, подразделением которого является ваша кафедра. Часто с первой же страницы сайта посетитель не видит знакомых русских слов среди научных и каких-либо других терминов и теряется. Думаю, после завершения наполнения сайта информацией, необходимо провести тестирование на предмет логической целостности, наличия грамматических ошибок и уязвимостей системы.

Во время тестирования системы необходимо проследить за реагированием на типичные виды атак на динамичные сайты, работающие с базами данных:

- внедрение SQL-кода в запросы;
- отправка NULL-данных в запросах и переменных;
- форсирование переполнения буфера;
- подмена метода передачи данных форм.

Также следует обратить внимание, как система реагирует на нетипичные ситуации, какие ошибки при этом выводятся в браузере. Желательно, чтобы ни при каких действиях посетитель не мог увидеть сообщения сервера `mysql` с полным текстом выполняемого запроса, с подробным описанием ошибки, – это отличный помощник для людей, желающих получить несанкционированный доступ к данным сайта.

## **Заключение**

В данной статье представлена оригинальная система организации и управления сайта кафедры высшего учебного заведения. Описаны основные этапы разработки системы, перечислены возможные программные средства разработки, указаны типичные ошибки, встречающиеся на существующих подобных сайтах. Идеи этой разработки внедрены на кафедре ПКС факультета КТ и У СПб УИТМО. Наш сайт доступен по адресу: <http://www.ifmo.ru/csd/>

## **Литература**

1. Гласман А.. Маркетинговые принципы построения виртуальных страниц Internet. СПб, 1998.
2. Дидре Хейз. HTML и XHTML. Москва, 2003.
3. Дмитриева М.. JavaScript. СПб, 2002.
4. Кастаньето Д., Х. Рават, С. Шуман, К. Сколло, Д. Велиаф. Профессиональное PHP программирование. СПб, 2001.
5. <http://tidy.sourceforge.prg/>
6. <http://www.mysql.com/>
7. <http://www.netcraft.com/>
8. <http://www.php.net/>
9. <http://www.securityspace.com/>
10. <http://www.w3.org/>

# СОЗДАНИЕ ЭЛЕКТРОННОГО АРХИВА ПРЕДПРИЯТИЯ

О.В. Михайличенко

Научный руководитель – д.т.н., проф. А.Г. Коробейников

Системы электронного технического документооборота – один из важнейших компонентов автоматизации процесса проектирования и конструирования. В статье рассматривается концепция построения электронного архива предприятия и представлено общее описание системы электронного архива на примере системы "ПроДок".

## Введение

Современный документооборот предприятия, когда большое количество документов разрабатывается и хранится в виде компьютерных файлов, требует специальных программных средств, помогающих реализации процедур хранения, разработки и распределения документации.

Традиционные методы управления бумажными документами, разрабатываемые в течение многих лет и регламентированные нормативными актами, не позволяют в должной мере обеспечить прохождение, согласование и конфиденциальность электронных информационных потоков предприятия. В частности, это связано с тем, что при переходе к компьютерной технологии обработки информации предприятия сталкиваются с проблемами, которые не были и не могли быть учтены существующими инструкциями и положениями. Использование компьютерных технологий существенно увеличивает скорость разработки документации, что во многом связано с возможностью коллективной работы над документами. С другой стороны, многократно возрастает риск несанкционированного или одновременного доступа и, следовательно, порчи, несогласованности или уничтожения информации. Эта опасность особенно серьезна при наличии локальной сети, дающей возможность распределенного хранения и доступов к документам. Вторым средством повышения производительности – особенно при внедрении сканерных технологий получения электронных документов – является все увеличивающийся поток файлов с информацией, которые необходимо учитывать, индексировать, заносить в каталоги. Преодолеть эти трудности можно, используя программные средства управления электронным документооборотом [2].

## 1. Общая концепция построения электронного архива

Структура электронного архива (ЭА) изображена на рис. 1 и состоит из следующих подсистем:

- подсистема сканирования – комплекса программно-аппаратных средств для преобразования бумажной документации в электронный вид;
- подсистемы оперативного архива – для хранения информации оперативного использования и накопления информации перед записью его на долговременное хранение;
- подсистемы долгосрочного хранения – для долговременного хранения больших и сверхбольших объемов данных, обращение к которым носит не регулярный характер;
- система пользовательских приложений. Как правило, на этом уровне организуется доступ к информации архива либо в файл серверном режиме, либо посредством использования систем автоматизации, СУБД;
- подсистема тиражирования;
- подсистема технологических решений.

Процесс создания ЭА следует начать с определения объемов документации, которые необходимо "положить" в ЭА. Очень важно определить, какие форматы файлов должны иметь электронные документы [3].

Устройством оперативного накопления может являться жесткий диск или RAID-массив, объем которого определяется общей технологией создания архива.

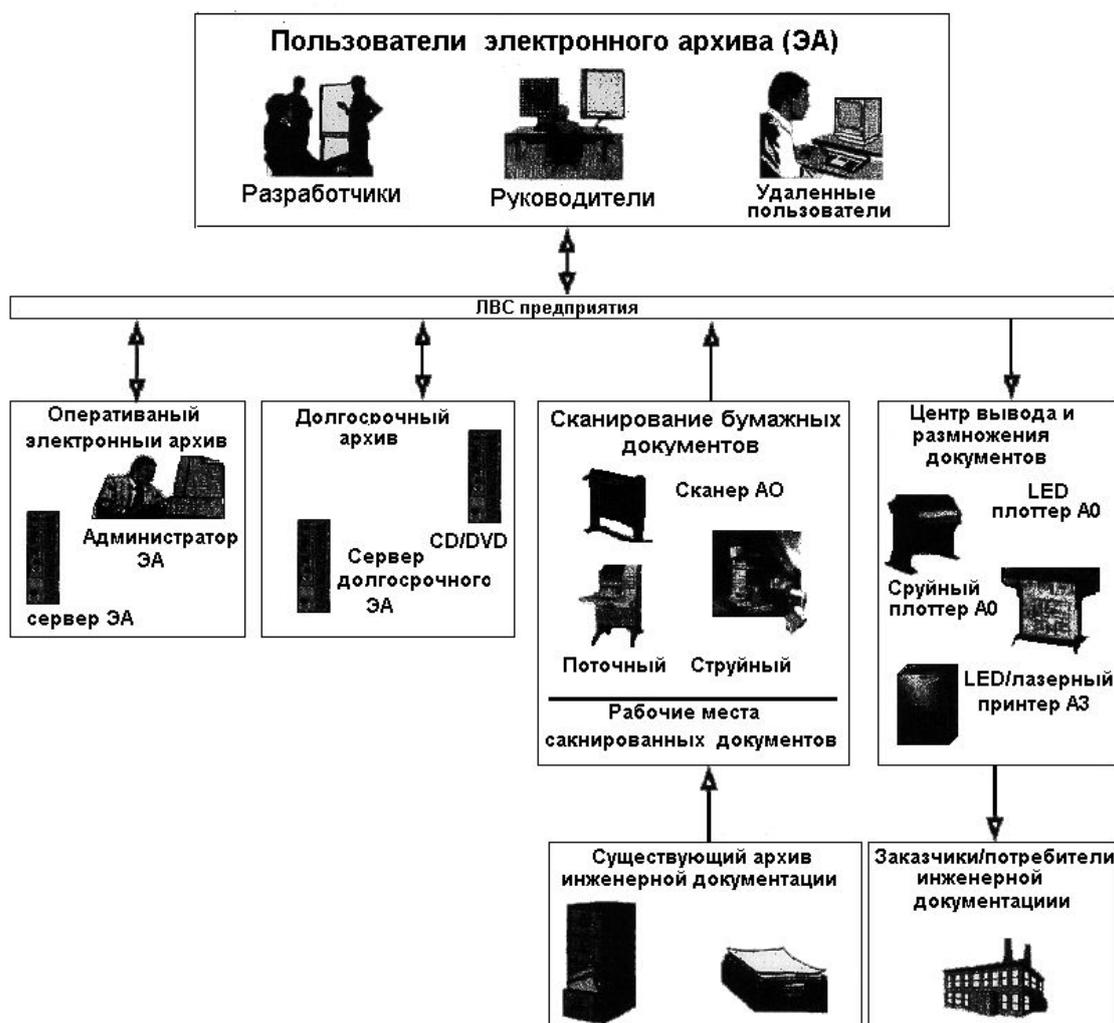


Рис.1. Структура электронного архива (ЭА)

При расчете объема учитывается общий объем информации, подлежащей оперативному накоплению до записи в устройство долговременного хранения и оперативному использованию.

В качестве устройств долговременного хранения могут быть использованы роботизированные CD-DVD-библиотеки под управлением специализированного программного обеспечения.

Система долгосрочного хранения выполняет следующие задачи:

1. организация больших и сверхбольших объемов информации;
2. организация доступа к большим и сверхбольшим объемам информации;
3. организация пополнения хранящейся информации по мере ее накопления на предприятии.

При создании подсистемы тиражирования, как и при создании каждой системы ЭА, прежде всего изучаются конкретные задачи на предприятии. Необходимо уточнить следующие пункты:

1. необходимая производительность подсистемы тиражирования по каждому из форматов;
2. вид печати по каждому из форматов;
3. параметры печати по каждому из форматов;
4. перспективы увеличения объемов по каждому из форматов, виду печати.

Рассмотрим систему пользовательских приложений. Условно, все пользовательские приложения, применяемые в комплексном решении – электронном архиве, можно разделить на следующие группы:

- система управления процессами сканирования;
- средства обработки, векторизации, распознавания изображений (Fine Reader);
- средства разработки – приложения, в которых производится разработка электронных документов;
- средства просмотра и редактирования документов, находящихся в электронном виде;
- средства управления хранилищами данных;
- система управления архивом – специализированные программные пакеты.

Последние условно можно разделить на следующие группы.

Картотеки – основная база данных системы с возможностью учета всех электронных и бумажных документов предприятия по заводимым атрибутам и возможностью быстрого поиска информации о документе. Основными пользователями являются Руководитель, Начальник ПО, Архивариус. Модуль предназначен для хранения, отображения, поиска информации об объектах проектирования, а также для контроля сроков проектирования. При поиске в электронной картотеке используется вся мощь SQL-запросов к серверу базы, позволяющему быстро произвести выборку по одному или нескольким условиям. Пользователь получает информацию о документе, не имея доступа непосредственно к нему для просмотра, редактирования, тиражирования.

Полные архивные системы – в отличие от электронных картотек, они дают доступ не только к карточке документа, но непосредственно к его электронному образу для просмотра, редактирования, тиражирования (в зависимости от прав пользователя). Такие системы позволяют регистрировать документ, содержащийся не только в одном файле, созданном в любых приложениях, но и виде набора файлов, определяющих сложное изделие, имеющие свою структуру [4].

## **2. Полные системы архива и документооборота**

В общем виде комплексная система автоматизации обработки технических документов (в западной терминологии TDM – Technical Data Management) состоит из следующих основных частей:

- подсистемы управления документами (электронного архива);
- подсистемы управления данными о проектах /изделиях (в западной терминологии PDM Product Data Management);
- подсистемы маршрутизации документов и работ (Workflow).

## **3. Общее описание системы электронного архива на примере системы "ПроДок"**

Ядро любого электронного документооборота – подсистема управления документами. В "ПроДок" электронный архив представляет собой базу данных, в которой система хранит сами документы и картотеку с информацией, необходимой для поиска документов, – обозначение, наименование, формат и т.п. Файлы документов хранятся на серверах, что позволяет получать к ним авторизованный доступ с любого компьютера.

Система предназначена для автоматизирования повседневной деятельности сотрудников (в основном сотрудников проектных организаций и научно-исследовательских институтов):

- просмотр, создание, регистрация, хранение, изменение и поиск проектных документов;
- создание, выдача и получение заданий на проектирование;

- согласование заданий на проектирование и проектных документов;
- работа с архивом документов - просмотр, помещение, изменение, удаление;
- контроль исполнения.

Цели системы:

- увеличение производительности труда сотрудников проектного отдела за счет централизованного хранения документов и быстрого получения документов и данных, необходимых для работы;
- сокращение циклов разработки документов, согласования документов и выдачи заданий на проектирование за счет маршрутизации, которая позволяет задавать пути прохождения документов внутри организации и контролировать эти процессы по срокам и исполнителям;
- обеспечение контроля доступа к информации.

Система состоит из четырех видов модулей (рис 2): Справочник сотрудников, Шаблоны документов, Картотека проектов, Объект (множественный модуль, для каждого проекта создается свой собственный).

Все модули, за исключением модуля "Объект", присутствуют в системе в единственном экземпляре. Количество же модулей типа "Объект" должно соответствовать количеству объектов проектирования. Это означает, что работа по проектированию нового объекта начинается с автоматического создания нового модуля "Объект".

Модуль "Картотека проектов" – основная база данных системы. Основными пользователями этого модуля являются Руководитель, Начальник ПО, Архивариус. Модуль предназначен для хранения, отображения, поиска информации об объектах проектирования, а также для контроля сроков проектирования. С помощью Картотеки проектов можно динамически наблюдать за ходом работ над всеми проектами в организации. У каждого проекта есть свой статус, система отслеживает все изменения происходящие внутри проекта.

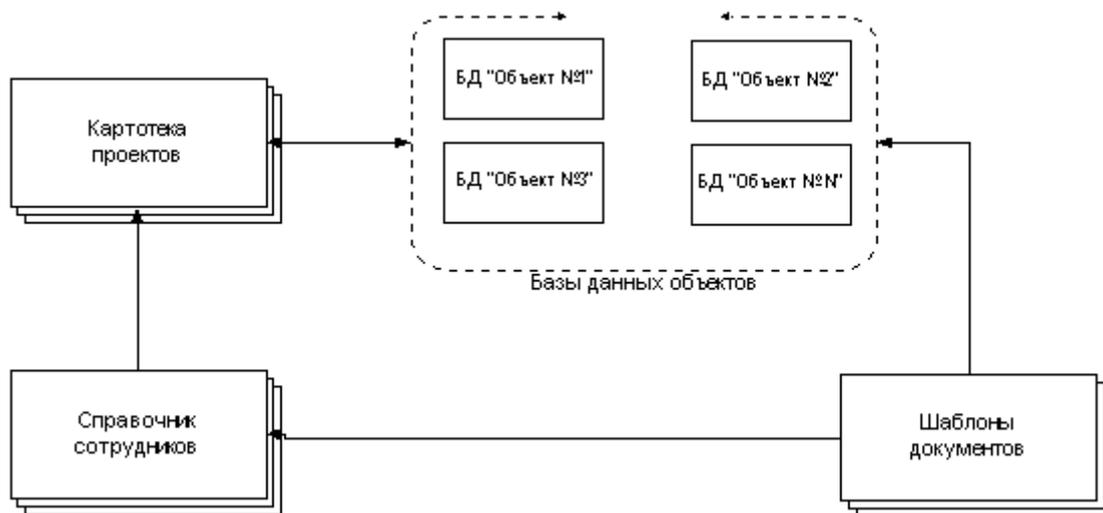


Рис. 2. Структура рабочих модулей системы

Модуль "Объект" – это основной рабочий модуль. Для каждого проекта создается отдельный модуль, в котором в дальнейшем и происходят все действия, касающиеся документов, планов и заданий. Это основная рабочая среда для разработки проектной документации. Основными пользователями модуля являются Ответственный по проекту и исполнители. Модуль предназначен для создания и хранения документов проекта, документов по планированию и выполнению, согласований между разделами, а также для осуществления контроля над ходом проектирования, по разделам проекта.

Модуль "Справочник сотрудников" – справочная база данных. Основным пользователем модуля является Администратор. Модуль предназначен для создания, хране-

ния и удаления информации о сотрудниках организации, а также для хранения общей информации о настройках системы. В дальнейшем остальные модули системы, такие как "Картотека проектов", "Объект", постоянно работают с модулем "Справочник сотрудников", читая оттуда все необходимые для работы данные настроек и данные о сотрудниках. В основном модуль предназначен для работы с ним администратора системы и работников, отвечающих за ведение данных о сотрудниках организации. Также модулем могут пользоваться и остальные сотрудники, но в зависимости от настройки прав доступа могут видеть определенный набор информации.

Модуль "Шаблоны документов" предназначен для создания и хранения шаблонов документов, наиболее часто используемых в организации. Использование этого модуля способствует стандартизации всех создаваемых в организации электронных документов, и облегчает работу сотрудников по созданию новых документов [5].

## **Заключение**

При помощи системы электронного документооборота предприятие получает:

- надежное хранение документов – значительно снижается вероятность потери документа или доступа к нему лиц, не имеющих на это права;
- быстрый поиск документов;
- повышение уровня обслуживания заказчика за счет снижения времени выполнения работы и уменьшения количества ошибок;
- автоматизированный контроль создания и сохранности новых версий документов, исключающий ошибки в их идентификации;
- приближение к стандартам ISO 9000 за счет возможности внедрения стандартизированных и стандартизированных процедур управления документами;
- возможность получать актуальную информацию о статусе проекта(состоянии разработки, загрузки и активности сотрудников);
- планирование загрузки сотрудников работой;
- автоматизации типовых процедур (параллельной разработки изделия, проведения изменений, построения различных отчетов по проекту и т.п.).

## **Литература**

1. Журнал "CADmaster" (журнал для профессионалов в области САПР). 2002. №3.
2. Журнал "CADmaster" (журнал для профессионалов в области САПР). 2002. №5.
3. Компьютер-информ. 2002. №10.
4. Компьютер-информ. 2002. №11.
5. "САПР и ГИС. Электронный документооборот 2003". ОАО "Петростройсистема" – центр информационных технологий.

# РАСПРЕДЕЛЕННЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ КАК НЕПРЕРЫВНОЕ РАЗВИТИЕ ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМ И КОМПЬЮТЕРНЫХ СЕТЕЙ

Н.А. Рубина

Научный руководитель – к.т.н., доц. Ю.Г. Кирчин

В статье обсуждается вопрос применимости метода многослойных логических моделей для анализа и синтеза информационных систем. Рассматривается проблема принадлежности метода, основанного на монотонном развитии основной модели телекоммуникационных компьютерных сетей к специальности "Телекоммуникационные системы и компьютерные сети".

Применение многослойных логических моделей, полученных из стандартных OSI-подобных моделей сетевого взаимодействия [1, 2, 3] посредством их дополнения верхними логическими слоями [4], представляется весьма мощным методом для анализа и синтеза информационных систем [5]. Его особенностью является реальная возможность получения решений уровня программного продукта, сетевого транспорта и аппаратных средств как функции бизнес-задач поддерживаемой предметной области [6]. Вместе с тем публичное обсуждение этого метода, основанного на монотонном развитии основной модели телекоммуникационных компьютерных сетей обычно требует доказательства его принадлежности к области знаний, формально специфицируемой как телекоммуникационные системы и компьютерные сети [7]. Настоящая статья делает попытку специального обращения к названной проблеме.

Относительно методологий анализа и разработки информационных систем существует несколько школ, во многом отражающих историю профессионального становления их приверженцев.

- Информационная система как программа [8]. Сторонники этой концепции исходят из полной априорной ясности информационной задачи, возможности составления непротиворечивого и исполнимого технического задания, предполагают требования к системе неизменными на всем протяжении ее жизненного цикла. Абсолютизируется проблема практического технического исполнения системы в форме создания программ, на практике ставится знак равенства между понятиями "информационная система" и "компьютерная программа".

- Информационная система как база данных [9]. Сторонники этого подхода полагают, что информационная система может рассматриваться в виде комбинации данных, представляющих смысловую информацию, и методов работы с ними, отражающих бизнес-правила смысловой задачи. При этом оба компонента прямо отображаются на, соответственно, таблицы и представления базы данных, и на хранимые процедуры, триггеры и ограничения той же базы данных. Дополнительным объектом является клиентское программное обеспечение, реализующее пользовательский доступ к системе.

- Информационная система как совокупность информации и правил пользования ей [10]. При таком подходе практическая компьютерная реализация может вовсе выводиться за рамки обсуждения как маловажная техническая подробность на фоне глубокой проблематики анализа информации и правил манипулирования ей. При таком подходе особое внимание уделяется сложности формализации реальных информационных ситуаций, неполноте знания о них, изменчивости требований в процессе жизни информационной системы.

- Информационная система как сеть [11]. Сторонники этого направления исходят из совокупности понятий "источник информации", "приемник информации" и "путь передачи и/или преобразования информации", настаивают на том, что при совпадении источника и приемника и при отсутствии пути между ними сама необходимость ин-

формационной системы становится сомнительной. На этом подходе мы остановимся подробнее.

История термина "информация" достаточно сложна и неоднозначна. Основу понятий "информация" и "данные" заложил в 1924 году Г.Найквист, а сам термин "информация" был введен в 1928 году Р.Хартли для количественной характеристики объемов передачи по каналам связи, стоимость которых высока, а производительность ограничена. Таким образом, этот термин был введен в предположении, с одной стороны, необходимости транспортировки сведений, с другой стороны, - баланса их собственной стоимости и стоимости их доставки. Необходимость обеспечить возможно более полную загрузку коммерческих каналов связи оплаченными сообщениями вызвало к жизни статистическую теорию информации, сформулированную в 1945 году К.Шенноном. С этих пор теоретическое понимание термина "информация" неразрывно связано с логарифмической функцией вероятности, с понятиями шумов, фильтрации, различения, опознания и т.п.

Основоположники теории информации всячески подчеркивали, что их как инженеров-связистов не интересует значение передаваемых сведений, их семантика; все внимание сосредоточивалось на количестве, качестве технической передачи, способности восстановить исходное сообщение после воздействия помех. Вероятно, примерно в это же время термин "информация" получил иное значение, более принятое в областях, не связанных непосредственно с проблемами связи: информация – сведения о лицах, предметах, фактах, событиях, явлениях и процессах независимо от формы их представления [12].

Понятие "система" также трактуется достаточно свободно в зависимости от контекста. Обычно представители всех школ соглашаются с мнением, что под системой следует понимать совокупность объектов, совокупные свойства которых превосходят совокупность их собственных свойств. Такое понимание однозначно указывает на отличие системы от множества: система характеризуется наличием и важностью межкомпонентного взаимодействия, а множество – лишь своим элементным составом.

Наиболее сложно обстоит дело с формализацией понятия "сеть". Будучи очень образным и интуитивно понятным, этот термин используется обычно без определения. При этом предполагается очевидное сходство обсуждаемого объекта или его модели с бытовым понятием сети как совокупности связанных между собой узлов, подобия разреженной ткани, покрывающей некоторую поверхность. Поэтому, вероятно, можно согласиться, что, используя понятие "сеть", мы подразумеваем пространственно или логически распределенную совокупность объектов и связей между ними; мы можем говорить о системной совокупности взаимодействующих узлов, системным свойством которой является способность покрывать значительные с точки зрения текущего обсуждения пространственные или логические области.

Появившись и получив быстрое развитие в конце XX века, понятие "компьютерная сеть" семантически очень насыщено и вряд ли допускает какую-либо иную интерпретацию, чем совокупность компьютеров, объединенных для обмена данными и взаимодействия на расстоянии. Уточняющие понятия "локальная сеть" (LAN) и несколько хуже звучащее по-русски "глобальная сеть" (WAN) еще более конкретизируют это понимание. Важно отметить также миграцию с течением времени удельного веса важности сосредоточенной компьютерной работы и коммуникационной функции связанных в сеть компьютеров. Если на заре компьютерных сетей удаленное взаимодействие компьютеров сводилось, скорее, к эпизодическим обменам данными для независимой работы, то в настоящее время, вероятно, большая часть объединенного Интернет компьютерного парка выполняет роль либо сетевых терминалов, используемых исключительно для связи, либо настолько тесно взаимодействует, что прерывание связи между ними делает их функционирование невозможным, а существование бессмысленным.

Динамика наименования специальности 05.13.13 отражает описанную тенденцию [13]. Звучавшее в прежней редакции как "Вычислительные машины, комплексы, системы и сети", это наименование отражало порядком перечисления и важность, и сложность понятий: основным объектом исследования полагались вычислительные машины, которые могли объединяться в более сложные композитные объекты – комплексы и системы, и уже в частном случае – в сети. В настоящей редакции наименование специальности звучит "Телекоммуникационные системы и компьютерные сети", отражая, скорее всего, доминанту распределенных коммуникаций, выполняемых системным образом и на основе компьютерных сетей.

Следует отметить, что лингвистически старая и новая формулировки имеют различную структуру. В первом случае мы имеем последовательность четырех однородных членов предложения, выстроенные по порядку монотонного нарастания некоторого важного свойства (в данном случае сложности, системности) рассматриваемого объекта. Такой тип построения именуемой фразы очень логичен и характерен для формального русского языка; он авторитарен, четок, "инженерен", имеет монотонную динамику – стремление. Во втором случае мы имеем симметричную конструкцию из двух устойчивых словосочетаний. Специфика союза "и" в такой конструкции в формально-декларативном стиле речи состоит в том, что он устанавливает некоторую слабо определенную бинарную симметричную операцию между соединяемыми компонентами: нельзя однозначно сказать, соответствует ли он логическому И или ИЛИ. Обычно при необходимости более строгой логической формулы вводятся дополнительные уточняющие конструкции ("включая", "за исключением", "кроме" и т.п.). Такая же конструкция определяющей фразы воспринимается как инструкция соответствия исследуемого объекта либо первой части фразы, либо второй, либо обеим в произвольном весовом наполнении; эта конструкция симметрична, сбалансирована, демократична. Новая формулировка допускает трактовки: телекоммуникационные системы на основе компьютерных сетей; компьютерные сети на основе телекоммуникационных систем; системы, основанные на равном использовании телекоммуникаций, компьютеров и сетей и т.п.

Широта взгляда и плюралистичность новой формулировки адекватно отражает реальную ситуацию в рассматриваемой области знаний. Примером тому может послужить IP телефония. Действительно, мы вполне можем установить IP соединение между удаленными компьютерами по обычной телефонной сети коллективного пользования – при этом у нас реализуется схема "IP по голосовой линии" (IP-over-voice); затем мы установим на связанных компьютерах программное обеспечение, предоставляющее возможность голосовой связи телефонного стандарта – у нас образуется схема "голос по IP" (voice-over-IP); в результате же мы имеем схему "голос поверх IP, который сам поверх голоса". Если при этом учесть, что современные общественные телефонные компании все больше используют для передачи голоса компьютерные технологии и цифровую связь, то мы получим логический слоеный пирог вида (голос поверх данных)<sup>n</sup>. Прочитав примечание в конце паспорта специальности 05.13.13, уточняющее, что "Специальность <Телекоммуникационные системы и компьютерные сети> не включает исследование и разработки в области систем, сетей и устройств телекоммуникаций. Эту область ... включает специальность 05.12.13" [13], впору воскликнуть, как профессор Преображенский: "Простите, кто на ком стоял?!"

Строго ответить на этот вопрос позволяют слоистые логические модели, используемые для анализа сетей, простейшей и одновременно наиболее используемой из которых является модель ТСП/IP. Модель ТСП/IP состоит из четырех логических слоев или уровней, каждый из которых обслуживает верхний по отношению к нему и, соответственно, обслуживается нижним слоем:

APPL	Приложения	Смысловые программы, с которыми работает пользователь; то, что видно на экране
TCP	Контроль доставки	Виртуальное соединение между отправителем и получателем сообщения
IP	Адресация и маршрутизация	Реальный путь частей сообщения по сети; механизмы определения этого пути
RNY	Аппаратные средства	Компьютерное и связанное оборудование, включая линейные сооружения

Таблица 1. Модель TCP/IP

Классическая модель TCP/IP, равно как и более общая и несколько более теоретическая семиуровневая модель ISO/OSI были созданы для внутреннего анализа сетевого взаимодействия; они предполагают взаимодействующие приложения (уровень APPL) конечными точками логического пути; причины происходящего взаимодействия этих приложений не рассматриваются. Поэтому понятие "сеть" применяется обычно лишь на нижних трех уровнях: сети электросвязи, IP сети, TCP/IP сети. Сеть уровня приложений как правило в общей практике анализа Интернет не используется, хотя группа протоколов, действующих над уровнем TCP, весьма обширна: telnet, ftp, http, SMTP, POP3, IMAP4 и т.п. Как правило, рассматривается лишь парное или клиент-серверное (по сути также парное, хотя и асимметричное) взаимодействие приложений.

Между тем легко привести пример одноранговой сети уровня приложений, действующей строго над уровнем TCP. В технологии Oracle важное место занимает компонент, именованный SQL\*Net в версии 7 сервера и, соответственно, Net8 в версии 8 сервера Oracle. С его помощью формируется сеть взаимодействующих именованных серверов баз данных, а также клиентских программных продуктов. Эта сеть, с одной стороны, обладает таким фундаментальным понятием, роднящим ее с IP сетями, как пространство имен (включая систему разрешения имен, напоминающую DNS), с другой стороны, может строиться на разных протоколах транспортного уровня, включая TCP/IP, но не ограничиваясь им. Это значит, что такая сеть может непрерывно пересекать границы разнопротокольных транспортных сетей подобно тому, как сами транспортные сети могут сохранять непрерывность над канальными сетями различных технологий (кабель, волокно, радиоканал).

На рис. 1 показано три транспортные сети, используемые авторами в своих работах: Private Net 1, Private Net 2 и Public Net. Первые две сети технологически представляют собой локальные TCP/IP сети, не взаимодействующие между собой непосредственно, но каждая из них имеет возможность обмена данными по протоколу TCP/IP с мобильным компьютером, периодически перемещаемым между рабочими площадками. Взаимодействие с удаленным сервером, расположенным в третьей сети (Public Net) происходит по тому же транспортному протоколу TCP/IP. Private Net 1 обслуживает проект 1, Private Net 2 обслуживает разработку, материалы которой периодически переносятся на мобильном компьютере на компьютеры проекта 1. Структуры баз данных непосредственно взаимодействуют по протоколу и в рамках стабильной сети SQL\*Net.

Без этих комментариев, вероятно, верхнюю часть рисунка можно было бы принять за схему традиционной транспортной сети. Также, вероятно, определенных усилий потребует расположение по логическим уровням OSI-подобной модели следующие команды:

```
select * from scott.emp@my_dblink
http://www.ifmo.ru/nayka/index.htm
telnet spb.runnet.ru 25
ping www.runnet.ru
```

которые, по сути, все являются сетевыми вызовами.

По-видимому, не так трудно согласиться, что вполне "техническое" взаимодействие серверов баз данных по dblink'ам образует компьютерную сеть: сетевой характер такого взаимодействия достаточно наглядно прослеживается даже на такой простой схеме, а без применения компьютеров технология Oracle очевидно существовать не может. Более далекий логический шаг приходится делать при анализе, например, навигационных сетей Web сайтов. На рис. 2 представлена топология взаимных ссылок виртуальных университетов Информационно-Образовательной Среды Открытого Образования [14] с проекцией на связывающие их транспортные сети:

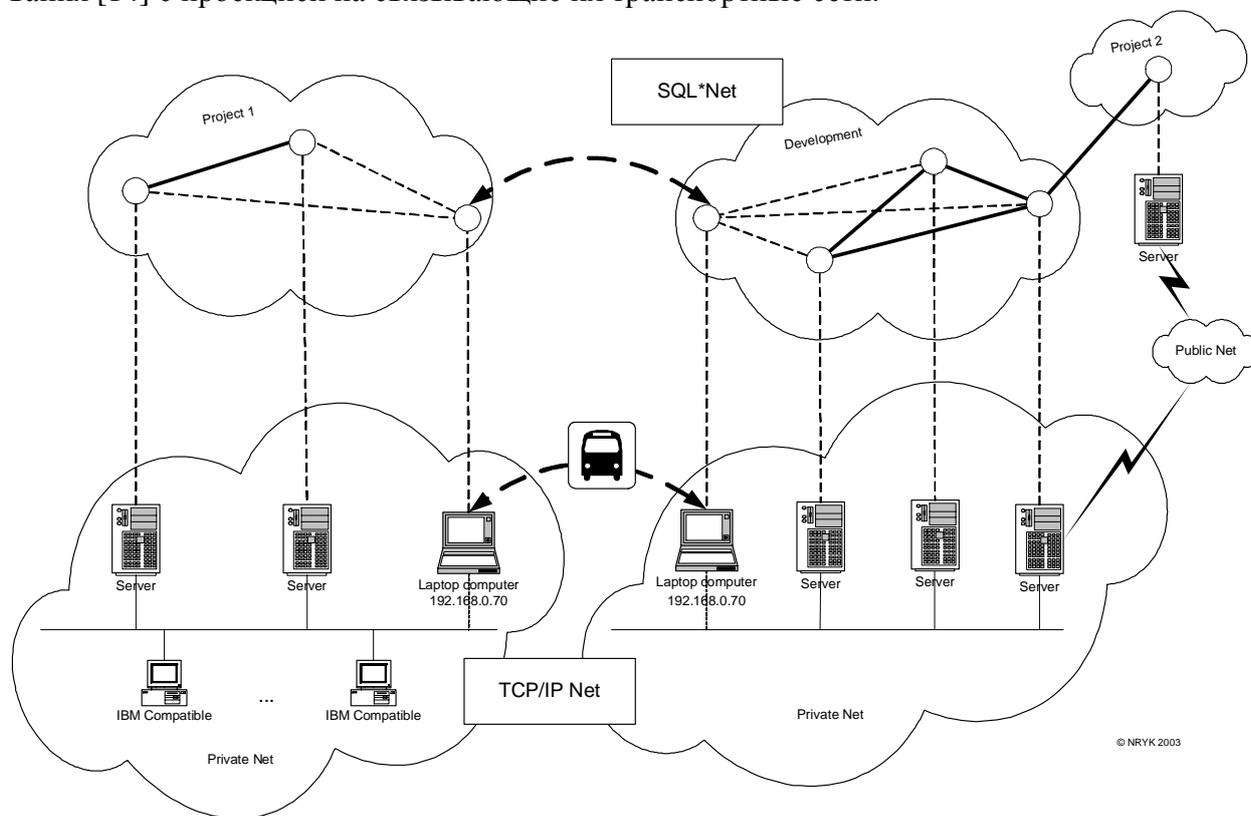


Рис. 1. Сеть уровня приложений SQL\*Net над транспортной сетью TCP/IP

Сеть взаимных ссылок объединенных виртуальных университетов достаточно наглядна; гораздо большее удивление вызывает анализ топологии связывающей их транспортной сети.

Проведя некоторые рассуждения и проанализировав простые примеры, взятые из каждодневной практики, мы приходим к выводу о том, что сети над традиционными транспортными сетевыми уровнями телекоммуникационных систем встречаются повсеместно, они многообразны и сложны. Вместе с тем, они – в полном соответствии со стандартными многоуровневыми логическими моделями – не могут существовать без подстилающих традиционных транспортных сетей, являются их историческим и логическим развитием, неразрывно с ними связаны. С другой стороны, все более приближаясь к задачам предметных областей реального мира, эти сети во все большей мере играют причинную роль в отношении транспортных сетей, оправдывают их существование и определяют требования к ним. Мысля рекурсивно, можно предположить, что дополнение стандартных многоуровневых моделей новыми верхними уровнями носит объективный характер, а сам этот процесс направлен на более точное следование требованиям реального мира. Непрерывность использования строгих методов технических наук, заложенных в стандартных сетевых моделях, при их распространении на новые

слои понятий определяет высокую научную ценность нового и одновременно преемственного подхода в рамках традиционной области исследований.

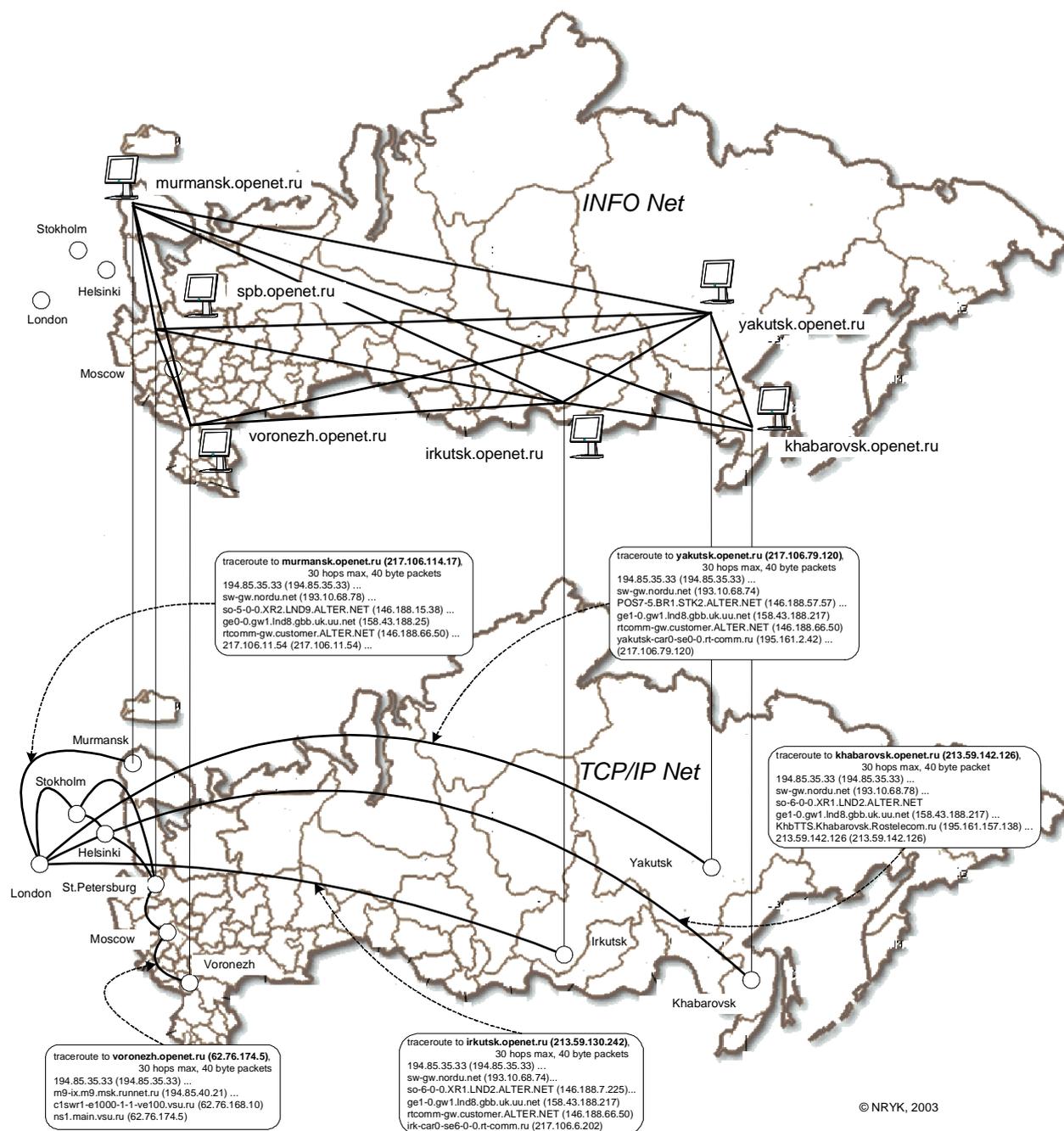


Рис. 2. Сеть взаимных http-ссылок над транспортной сетью TCP/IP

Web версия статьи находится по адресу:  
[http://risbank.spb.ru/risbank2/Nina\\_PhD\\_RUS/2003/ITMO\\_Students\\_Conference/](http://risbank.spb.ru/risbank2/Nina_PhD_RUS/2003/ITMO_Students_Conference/)

## Литература

1. Information Processing Systems - OSI Reference Model - The Basic Model. ISO/IEC 7498-1: 1994(E), ITU-T Rec. X.200 (1994 E), [http://www.acm.org/sigcomm/standards/iso\\_std/OSI\\_MODEL/ISO\\_IEC\\_7498-1.TXT](http://www.acm.org/sigcomm/standards/iso_std/OSI_MODEL/ISO_IEC_7498-1.TXT)
2. D.Wetteroth. OSI Reference Model for Telecommunications. McGraw-Hill Professional, 2001, ISBN 0-07138-0418
3. M.Miller. Troubleshooting TCP/IP. 3rd Edition. John Wiley & Sons Inc, 1999, ISBN 0-76457-0129

- 
4. Кирчин Ю.Г., Рубина Н.А., Танаев А.В. Информационный оборот предприятия. Web статья, декабрь 1997, <http://risbank.spb.ru/risbank2/papers/InfoSupport/Infoexchange.htm>
  5. Y.Kirchin, N.Rubina. Information Spaces. Proc. of Telematics-2001 International Conference on Telematics and Web-Based Education, 18-21 June, 2001, SPb, Int.Vol.,  
[http://risbank.spb.ru/risbank2/Telematics\\_2001/Web\\_Papers/348.htm](http://risbank.spb.ru/risbank2/Telematics_2001/Web_Papers/348.htm)
  6. N.Rubina. The Application of Inormation Technology to Support Real-Time Business Services. A Study. De Montfort Univercity, Leicester, UK, 2001,  
[http://risbank.spb.ru/risbank2/Nina\\_PhD/Transfer\\_Report/Transfer\\_Report\\_last\\_version.pdf](http://risbank.spb.ru/risbank2/Nina_PhD/Transfer_Report/Transfer_Report_last_version.pdf)
  7. Web сайт <http://www.aspirantura.spb.ru/>, раздел «Паспорта специальностей научных работников»,  
05.13.13 Телекоммуникационные системы и компьютерные сети  
[http://www.aspirantura.spb.ru/pasp/5\\_13\\_13.html](http://www.aspirantura.spb.ru/pasp/5_13_13.html)
  8. Сайт кафедры Информационные Системы СПбИТМО, <http://is.ifmo.ru/>
  9. Wlliam G.Page, Jr. Using Oracle8/8i. Special Edition, Que, 1999, ISBN 0-7897-1975-4
  10. T.Lucey. Management Information Systems. 8<sup>th</sup> Edition, Letts Educational, 1997, ISBN 1-85805-303-X
  11. Баженов А.Г. Построение систем электронного документооборота на уровне информационной сети предприятия. Научно-технический вестник СПбГИТМО(ТУ), Санкт-Петербург, 2002, стр.46-51.
  12. Федеральный закон № 24-ФЗ «Об информации, информатизации и защите информации», принят Государственной Думой 25 января 1995 года
  13. Web сайт <http://www.ifmo.ru/наука/>, раздел «Аспирантура и докторантура»
  14. Web сайт Российского Портала Открытого Образования Консалтингового Центра Информационно-Образовательной Среды Открытого Образования <http://www.openet.ru/>

<b>Предисловие</b> .....	3
<b>Секция 1. Оптические технологии</b> .....	5
<b>Макаров Н.С.</b> (научный руководитель – д.ф.-м.н., проф. В.Г. Беспалов). Генерация антистоксового излучения при вынужденном комбинационном рассеянии в условиях фазового квазисинхронизма.....	5
<b>Штумпф С.А.</b> (научные руководители – д.ф.-м.н., проф. С.А. Козлов, к.ф.-м.н., доц. А.А. Королёв). Уширение "фиолетового" крыла фемтосекундного спектрального суперконтинуума из-за дисперсии нелинейного показателя преломления среды.....	14
<b>Серебрякова М.К.</b> (научный руководитель – к.ф.-м.н., ст.н.с. В.Н. Михайлов). Методика обработки отражательных голограмм, записанных на фотопластинах ВРП при импульсном экспонировании .....	24
<b>Вознесенская Н.Н.</b> (научный руководитель – д.т.н., проф. В.П. Вейко). Компьютерное моделирование и анализ эффективности соединения типа канальный волновод – оптическое стекло.....	29
<b>Ерофеева М.С.</b> (научный руководитель – д.т.н., проф. А.Л. Дмитриев). Эффект Фарадея в волоконных световодах .....	36
<b>Вознесенская А.О.</b> (научный руководитель – д.т.н., проф. И.К. Мешковский). Расчёт функции пропускания и чувствительности волоконнооптического датчика с модулирующим отражательным элементом.....	42
<b>Фролков В.Н., Рядчин Ю.И.</b> (научный руководитель – д.т.н., проф. Ю.А. Гатчин). Основные методы изготовления преформ для вытяжки оптического волокна.....	48
<b>Фролков В.Н., Рядчин Ю.И.</b> (научный руководитель – д.т.н., проф. Ю.А. Гатчин). Вытягивание волоконных световодов .....	57
<b>Фролков В.Н.</b> (научный руководитель – д.т.н., проф. Ю.А. Гатчин). Контроль качества оптического волокна методом газоразрядной визуализации .....	66
<b>Ильин А.В.</b> (научный руководитель – к.т.н. Д.А. Гаврилин). Расшифровка интерферограмм фазового сдвига.....	69
<b>Беляев Ю.В.</b> (научные руководители – д.т.н., проф. В.К. Кирилловский, к.т.н., доц. В.М. Домненко). Разработка программного обеспечения для компьютерной изофотометрии ФРЛ и ЧКХ оптической системы .....	74
<b>Романова Г.Э.</b> (научный руководитель – д.т.н. проф. В.А. Зверев). Композиция зеркально-линзового объектива, основанная на применении коррекционных свойств плоскопараллельной пластинки .....	80
<b>Точилина Т.В.</b> (научный руководитель – д.т.н. проф. В.А. Зверев). Взаимосвязь параметров перемещения элементов принципиальной схемы оптической системы переменного увеличения .....	91
<b>Виноградова О.А., Фролов Д.Н.</b> (научный руководитель – д.т.н. проф. В.А. Зверев) Влияние параметров элементов оптической системы микроскопа на его разрешающую способность .....	94
<b>Ежова К.В.</b> (научный руководитель – к.т.н., доц. Н.Д. Толстоба). Анализ децентрировок поверхностей в оптической системе на основе методов гауссовой оптики .....	98

<b>Гаврилов Е.В.</b> (научные руководители – д.т.н., проф. Н.Б. Вознесенский, д.т.н., проф. В.К. Кирилловский). Методы проектирования и контроля зеркального объектива для рентгеновской фотолитографии .....	102
<b>Иванов К.А.</b> (научный руководитель – д.т.н., проф. С.А. Сухопаров). Компьютерный контроль и юстировка сканирующих систем .....	109
<b>Гантворг М.Л.</b> (научный руководитель – к.т.н., доц. Т.В. Иванова). Разработка алгоритмов анализа аберраций оптических систем с использованием дифференциалов реального луча .....	114
<b>Федорова И.Ю., Фролова С.Г., Златина С.В., Волкова М.А.</b> (научный руководитель – к.т.н., доц. <b>А.А. Цуканов</b> ). Моделирование ложного двойного лучепреломления .....	120
<b>Лаптева Н.А.</b> (научный руководитель – д.т.н., проф. Л.Н. Андреев). Объективы для оптической записи и считывания информации .....	124
<b>Дубовиков А.Л., Волкова М.А., Златина С.В.</b> (научные руководители – к.т.н., доц. <b>А.А. Цуканов</b> , д.т.н., проф. Г.Н. Грязин). Светодиоды в микроскопии .....	128
<b>Карасев Н.Н., Шакин А.О.</b> (научный руководитель – д.т.н., проф. Э.С. Путилин). Исследование структуры тонких диэлектрических слоев оптическими методами .....	133
<b>Секция 2. Информационные технологии .....</b>	<b>139</b>
<b>Дорожкин С.К.</b> (научный руководитель – к.т.н., доц. А.А. Приблуда). Балансировка нагрузки в распределённых вычислительных системах .....	139
<b>Елькин И.В.</b> (научный руководитель – к.т.н., доц. А.Е. Платунов). Проектирование управляющих систем на основе высокоуровневых описаний .....	143
<b>Яковлев Ю.А.</b> (научный руководитель – к.т.н., доц. Т.А. Павловская). Критерии полноты тестирования программ .....	149
<b>Оскотский С.З., Григорьев И.В.</b> (научный руководитель – д.т.н., проф. И.П. Гуров). Система управления качеством сервиса на основе политик – разработка сервера конфигурирования сетевых устройств .....	154
<b>Харитонов А.Е.</b> (научный руководитель – к.т.н., доц. А.А. Приблуда). Семантический анализ технической документации в CASE-технологиях .....	162
<b>Казаков М.А.</b> (научный руководитель – д.т.н., проф. А.А. Шалыто). Использование автоматного подхода для построения визуализаторов .....	166
<b>Казаков М.А.</b> (научный руководитель – д.т.н., проф. А.А. Шалыто). Система автоматического тестирования программных решений и проведения соревнований в режиме on-line .....	181
<b>Степанюк С.В.</b> (научный руководитель – к.ф.-м.н., доц. А.В. Норин). Автоматический генератор тестовых заданий .....	190
<b>Койнов Р.В.</b> (научный руководитель – д.т.н., проф. С.К. Стафеев). Подходы к созданию и переносу электронных учебных пособий на карманные персональные компьютеры .....	196
<b>Заколдаев Д.А.</b> (научный руководитель – к.т.н., проф. Н.Д. Фролов). Процессный подход к управлению информационными потоками при внедрении корпоративной информационной системы .....	203
<b>Шилов К.Н.</b> (научный руководитель – д.т.н., проф. А.Ю. Щеглов). Проблема и выбор способа противодействия вредоносным программам, использующим интерпретаторы .....	208

<b>Дорожкин А.К.</b> (научный руководитель – к.т.н., проф. В.В. Кириллов). Объединение реляционного и многомерного представления данных на примере Oracle 9i.....	212
<b>Антонов С.В., Соловьёв Д.В.</b> (научный руководитель – к.т.н., доц. М.Ю. Юдин). Управление памятью как основополагающий фактор быстродействия приложений .....	216
<b>Антонов С.В., Соловьёв Д.В.</b> (научный руководитель – к.т.н., доц. М.Ю. Юдин). Новый протокол IPv6: преимущества, модели перехода.....	219
<b>Антонов С.В.</b> (научный руководитель – к.т.н., доц. М.Ю. Юдин). Распределение виртуальной памяти в среде Win32. Управление малыми блоками.....	222
<b>Богданов К.В.</b> (научный руководитель – д.т.н., проф. А.Г. Коробейников). Проектирование математического обеспечения САПР систем автоматизации производства оптических материалов на основе нейросетевых компонентов .....	225
<b>Богданов К.В.</b> (научный руководитель – д.т.н., проф. А.Г. Коробейников). Оптимизация проектных решений в САПР на основе нейросетевых алгоритмов .....	229
<b>Семенюк В.В.</b> (научный руководитель – д.т.н., проф. В.Н. Васильев). Метод эффективного моделирования выборки источников двоичной информации.....	233
<b>Семенюк В.В.</b> (научный руководитель – д.т.н., проф. В.Н. Васильев). Эффективный алгоритм экономного кодирования коэффициентов дискретного косинусного преобразования.....	238
<b>Семенюк В.В.</b> (научный руководитель – д.т.н., проф. В.Н. Васильев). Эффективный алгоритм экономного кодирования изображений на основе дискретного вейвлет- преобразования.....	243
<b>Таратин М.А., Захаров А.С.</b> (научный руководитель – д.т.н., проф. И.П. Гуров). Анализ вычислительного процесса нелинейной дискретной фильтрации Калмана с использованием процессора Intel Pentium 4. ....	249
<b>Гришин М.В.</b> (научные руководители – д.т.н., проф. А.А. Ожиганов, д.т.н., проф. А.Ю. Тропченко). Использование дискретных вейвлет-преобразований для сжатия полутонных изображений .....	254
<b>Шалаев М.П.</b> (научный руководитель – д.т.н., проф. И.П. Гуров). Моделирование трафика в компьютерных сетях на основе марковских процессов .....	262
<b>Малинин А.А.</b> (научный руководитель – к.т.н., доц. Н.Ю. Иванова). Разработка системы организации и управления сайта кафедры учебного заведения .....	269
<b>Михайличенко О.В.</b> (научный руководитель – д.т.н., проф. А.Г. Коробейников). Создание электронного архива предприятия .....	276
<b>Рубина Н.А.</b> (научный руководитель – к.т.н., доцент Ю.Г. Кирчин). Распределенные информационные системы как непрерывное развитие телекоммуникационных систем и компьютерных сетей .....	281

**Вестник конференции молодых ученых СПбГУ ИТМО. Сборник научных трудов / Под ред. В.Л. Ткалич. Том 1. – СПб: СПбГУ ИТМО, 2004. 291 с.**

**ВЕСТНИК КОНФЕРЕНЦИИ МОЛОДЫХ УЧЕНЫХ  
СПБГУ ИТМО**

**Сборник научных трудов**

**Том 1**

Научный редактор – д.т.н., профессор В.Л. Ткалич

Дизайн обложки Т.В. Точилина

Редакционно-издательский отдел СПбГУ ИТМО

Зав. РИО Н.Ф. Гусарова

Лицензия ИД № 00408 от 05.11.99.

Подписано в печать 22.06.04.

Заказ 780. Тираж 100 экз.