

**ПРИОРИТЕТНЫЙ НАЦИОНАЛЬНЫЙ ПРОЕКТ «ОБРАЗОВАНИЕ»  
РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ**

---

**К.П. ЛОВЕЦКИЙ, Л.А. СЕВАСТЬЯНОВ,  
О.Н. БИКЕЕВ, М.В. ПАУКШТО**

**МАТЕМАТИЧЕСКИЙ СИНТЕЗ  
ОПТИЧЕСКИХ НАНОСТРУКТУР**

**Учебное пособие**

**Москва**

**2008**

*Инновационная образовательная программа  
Российского университета дружбы народов*

**«Создание комплекса инновационных образовательных программ  
и формирование инновационной образовательной среды,  
позволяющих эффективно реализовывать государственные интересы РФ  
через систему экспорта образовательных услуг»**

Экспертное заключение –

доктор физико-математических наук, профессор *А.В. Крянев*

**Ловецкий К.П., Севастьянов Л.А., Бикеев О.Н., Паукшто М.В.**

Математический синтез оптических наноструктур: Учеб. пособие. – М.: РУДН, 2008. – 145 с.

Данное пособие является завершающим в цикле учебно-методических пособий для магистерской программы «Оптика наноструктур». В нем на основе методов и алгоритмов, развитых в предыдущих курсах, предлагается общая схема решения задачи проектирования сложных периодических многослойных оптических покрытий для современных технологий солнечной энергетики, «умных» материалов, дисплейных покрытий с нетривиальными электромеханическими свойствами и т.п. Приводятся примеры решения актуальных научно-исследовательских и конструкторских разработок, пользующихся спросом у зарубежных производителей.

Для магистров и аспирантов, обучающихся по направлению «Прикладная математика и информатика».

*Учебное пособие выполнено в рамках инновационной образовательной программы Российского университета дружбы народов, направление «Комплекс экспортноориентированных инновационных образовательных программ по приоритетным направлениям науки и технологий», и входит в состав учебно-методического комплекса, включающего описание курса, программу и электронный учебник.*

© Ловецкий К.П., Севастьянов Л.А., Бикеев О.Н., Паукшто М.В., 2008

## Содержание

Общее описание курса.....	5
Иновационность курса .....	7
Тема 1. Методы оптимизации математического синтеза оптических покрытий.....	10
1.1. Симплекс-метод Нелдера — Мида .....	11
1.1.1. Алгоритм Нелдера—Мида .....	13
1.1.2. Описание алгоритма .....	14
1.1.3. Свойства метода Нелдера—Мида.....	18
1.2. Оптимизация методом наименьших квадратов.....	20
1.2.1. Метод Ньютона—Гаусса.....	22
1.2.2. Метод Левенберга—Марквардта .....	23
1.2.3. Реализация метода Ньютона—Гаусса .....	24
1.3. Алгоритм Левенберга—Марквардта .....	25
1.3.1. Постановка задачи .....	25
1.3.2. LMA как комбинация простейшего градиентного метода и метода Ньютона— Гаусса.....	27
1.3.3. LMA как метод доверительных интервалов.....	30
Тема 2. Брэгговское зеркало .....	33
2.1. Постановка задачи .....	34
2.2. Анализ дисперсионного уравнения. Обсуждение физического смысла решений.....	41
2.3. Амплитуды пространственных гармоник.....	51
2.4. Диэлектрическое зеркало, основанное на эффекте брэгговского отражения .....	56
Тема 3. Метод Тихонравова синтеза многослойного зеркала .....	65
3.1. Физическое содержание рассматриваемой задачи синтеза .....	65
3.2. Математическая постановка задачи синтеза .....	68
3.3. Дискретная формулировка задачи синтеза.....	70
3.4. Алгоритм минимизации функционала и некоторые результаты расчетов .....	73
Тема 4. Теоретические аспекты задачи синтеза оптических покрытий.....	81
4.1. Качественные закономерности структуры оптических покрытий....	81
4.2. Наклонное падение волны на слоистую структуру .....	89
4.2.1. Наклонное падение ТЕ-волны на систему непоглощающих магнитоэлектрических слоев .....	89
4.2.2. Наклонное падение ТМ-волны на систему непоглощающих магнитодиэлектрических слоев .....	95
4.2.3. Наклонное падение монохроматической электромагнитной волны с вертикальной поляризацией (ТМ-волны) на систему непоглощающих магнитодиэлектрических слоев .....	96

Тема 5. Фотонные кристаллы .....	98
5.1. Аналитические и численные методы теоретического исследования фотонных кристаллов .....	100
5.2. Распространение света в фотонных кристаллах .....	102
5.2.1. Спектр отражения и пропускания одномерных фотонных кристаллов .....	102
5.2.2. Дифракция света в двумерных фотонных кристаллах .....	106
Тема 6. Проектирование зеркал: многослойное покрытие, дифракционная однослойная решетка .....	113
Литература .....	121
Описание курса и программа.....	124

## **Общее описание курса**

Данное пособие является завершающим в цикле учебных пособий для магистерской программы «Оптика наноструктур». В нем на основе методов и алгоритмов, развитых в предыдущих курсах, предлагается общая схема решения задачи проектирования покрытий из нанотрубок и нанопроволок для современных технологий солнечной энергетики, умных материалов, дисплейных покрытий с нетривиальными электромеханическими свойствами и т.п. Приводятся примеры решения актуальных научно-исследовательских и конструкторских разработок, пользующихся спросом у зарубежных производителей.

Курс «Математический синтез оптических наноструктур» является составной частью магистерской программы «Оптика наноструктур». Магистерская программа «Оптика наноструктур» реализуется в рамках направления «Прикладная математика и информатика» и направления «Прикладная математика и физика», а возможно, и других направлений. В составе магистерской программы «Оптика наноструктур» курс «Математический синтез оптических наноструктур» является обязательным. Для других магистерских программ этот курс предлагается по выбору без привязки к семестру или факультативно на усмотрение методической комиссии программы. Курс носит теоретический и практический характер.

Цель курса - подробное ознакомление студентов с устойчивыми современными методами численного решения математических задач, возникающих при проектировании дифракционных нанометровых оптических элементов и устройств, разработке тонкопленочных покрытий с характерными толщинами порядка длины волны излучения. Задачи

проектирования стали особенно актуальны в последние годы в связи с широким применением наноэлементов и тонких (менее одного микрометра толщиной) пленок в производстве жидкокристаллических дисплеев, солнечных батарей на основе диэлектриков, фотоэмиссионных диодов, просветляющих покрытий, поляризаторов, миниатюрных лазеров, управляемых оптических элементов. Задачи оптики наноструктур практически не поддаются аналитическому решению, поэтому важным является изучение эффективных численных методов, используемых при решении задач математического синтеза, приобретение навыков создания программного обеспечения для компьютерного дизайна оптических наноструктур.

Задачей курса «Математический синтез оптических наноструктур» является формирование у студентов навыков работы на современной измерительной аппаратуре, обучение использованию строгих методов связанных волн при решении задач моделирования современных оптических устройств на основе тонкопленочных покрытий и дифракционных оптических элементов. Это позволит при необходимости разрабатывать новое программное обеспечение. Безусловной задачей курса является также освоение существующего программного обеспечения, ориентированного на расчет и проектирование оптических покрытий. В результате обучения студенты получают умение и навыки правильно оценивать сложность научно-исследовательских и конструкторских заданий на разработку дифракционных оптических элементов и устройств, аргументированно выбирать метод решения конструкторской задачи, а затем экономично и эффективно выполнять компьютерный дизайн требуемого дифракционного оптического покрытия или устройства.

## **Инновационность курса**

Курс является инновационным по содержанию и по литературе, он включает в себя последние научные достижения в области решения задач дифракционной оптики, когда характерные размеры исследуемых объектов не превышают либо сравнимы с длиной волны оптического излучения. Эта область знаний интенсивно развивалась в последнее время, но лишь недавно были созданы устойчивые алгоритмы и разработаны численные методы решения задач для многослойных решеток. Следует отметить, что для оптических однослойных и многослойных дифракционных решеток с характерными размерами больше длины волны оптического излучения устойчивые методы решения известны с середины прошлого века. Сейчас алгоритмы решения оптических задач в субволновой области распространяются на объекты со сложной геометрией, такие как двумерные решетки с произвольным профилем, трехмерные решетки (фотонные кристаллы), и на анизотропные материалы. Они востребованы, поскольку позволяют создавать математические модели взаимодействия излучения с веществом в наномасштабах, а затем с их помощью проектировать новые эффективные устройства в высокотехнологичных областях медицины, энергетики, инфокоммуникаций и приборостроения.

В ходе проведения занятий по этому курсу разработчики предполагают использование традиционных методик преподавания, принятых в странах болонской системы образования, т. е. с использованием кредитной системы оценки знаний.

Наряду с традиционными элементами преподавания математических методов решения прикладных задач, разработчики курса предполагают воспользоваться хорошо зарекомендовавшим себя опытом МФТИ и подобных вузов. Для этого в рамках подпрограммы «Оптика

наноструктур» осуществляется закупка уникального измерительного и аналитического оборудования для выполнения измерений разнообразных характеристик оптических наноустройств с целью использования этого оборудования в учебном процессе и для проведения научно-исследовательских работ преподавателями, аспирантами и студентами.

По окончании магистратуры по направлению «Оптика наноструктур» выпускники Российского университета дружбы народов станут конкурентоспособными специалистами в области проектирования современных оптических устройств, и не будут испытывать затруднений при последующем трудоустройстве.

Данное направление научно-практических разработок сформировалось лишь в последние 10 – 15 лет. Поэтому наблюдается сильный дефицит учебно-методической литературы не только в России, но и во всем мире. Разрабатываемые в рамках инновационной программы «Оптика наноструктур» учебные пособия восполнят в некоторой степени этот пробел и составят основной список литературы для слушателей курсов. Вместе с ними следует использовать несколько учебников и монографий, вышедших в свет к настоящему времени и перечисленных в списке литературы. Курс базируется на публикациях научных статей мировых лидеров исследований в данной области в научной периодике, диссертационных работах их учеников, включающих работы по непосредственному моделированию, дизайну и последующему изготовлению лабораторных образцов оптических элементов и устройств. В список дополнительной и рекомендуемой литературы включены научно-исследовательские публикации, положенные в основу предлагаемого курса.

В качестве практических заданий, курсовых работ и тем рефератов слушателям магистерской программы будут предложены актуальные



проблемы и задачи, решение которых востребовано современным уровнем развития высокотехнологичных отраслей промышленности и научно-исследовательских лабораторий.

## **Тема 1. Методы оптимизации математического синтеза оптических покрытий**

Задачи математического синтеза многослойных оптических покрытий являются нетривиальной проблемой в силу того, что сложны сами математические модели рассеяния электромагнитного излучения на дифракционных оптических системах. Поэтому в большинстве случаев используется следующая идеология. Осуществляется итеративный поиск оптимальной конструкции, на каждом шаге которого решается прямая задача дифракции света на многослойной системе. Переход от шага к шагу осуществляется методом улучшения целевой функции, характеризующей отклонения рассчитанных характеристик текущей системы от технического задания, т. е. решается задача оптимизации.

Каждая задача оптимизации начинается с задания начальной конфигурации системы, поэтапного сравнения характеристик нового варианта с предыдущим и прекращения процесса оптимизации по критерию качества.

Методы оптимизации традиционно подразделяются на три основных класса: методы нулевого порядка, методы первого и второго порядков. Методы нулевого порядка требуют вычисления лишь значений функции качества. Методы первого порядка используют в своей реализации производные функции качества по параметрам проектируемой системы. Методы второго порядка дополнительно требуют знания вторых производных. Теоретически методы второго порядка сходятся за меньшее количество итераций, чем методы первого, и тем более нулевого порядков. Однако количество вычислений на каждой итерации существенно возрастает с повышением порядка метода. Поэтому в каждом конкретном случае необходимо выбирать компромиссное решение, минимизирующее

совокупные затраты вычислительного процесса.

Опишем кратко алгоритмы, наиболее часто используемые при проектировании дифракционных оптических систем.

### **1.1. Симплекс-метод Нелдера — Мида**

Симплекс-метод Нелдера—Мида, впервые опубликованный в 1965 г., приобрел с тех пор необыкновенную популярность при решении задач многомерной безусловной оптимизации. Однако, несмотря на такую популярность, простого и подробного доказательства сходимости метода для многих классов функций все еще нет. Приведем [1] некоторые теоретические результаты сходимости метода Нелдера—Мида для строго выпуклых функций одной и двух переменных. Здесь будут даны теоремы сходимости для функций размерности 1 и некоторые (ограниченные) результаты о сходимости в двумерном пространстве. Для более общего представления о свойствах этого метода рассмотрим также контрпример МакКиннона, в котором при выборе специального начального приближения минимизация строго выпуклой функции не приводит к точке минимума.

Симплекс метод Нелдера—Мида [2] стал в последние годы наиболее распространенным методом для решения задач оптимизации без ограничений. Не следует путать его с (возможно?) гораздо более известным методом — симплекс-алгоритмом Данцига для решения задач линейного программирования. Оба алгоритма основаны на последовательном построении и использовании симплексов в многомерных пространствах, но во всем остальном — это совершенно различные алгоритмы, ничем между собой не связанные.

Особенно популярен алгоритм Нелдера—Мида при решении

прикладных задач, возникающих в физике, химии, медицине. Это, в первую очередь, объясняется его простотой как в программной реализации, так и в понимании. Скорость сходимости метода в практических приложениях сравнима, а часто и превосходит градиентные алгоритмы первого порядка (требующие вычисления первых производных функции по всем аргументам). К тому же по вычислительным затратам он гораздо экономичнее.

Метод Нелдера—Мида предназначен для минимизации функции  $n$  действительных переменных с использованием лишь вычисляемых на каждом шаге значений минимизируемой функции (метод нулевого порядка). Для него не требуется и в нем не используется (явно или неявно) информация о производных минимизируемой функции. Таким образом метод Нелдера—Мида относится к общему классу *прямых методов поиска* минимума функций. Как и многие другие методы, относящиеся к подмножеству прямых методов, метод Нелдера—Мида на каждом шаге итеративного процесса хранит невырожденный *симплекс* – геометрический объект в  $n$ -мерном пространстве ненулевого объема, являющийся выпуклой оболочкой, натянутой на  $n + 1$  вершину.

Каждая итерация прямого симплекс-метода поиска минимума начинается с построения симплекса, который задается своими  $n + 1$ -ой вершинами и вычисляемыми в этих вершинах значениями функции. Затем многогранник дополняется одной либо несколькими точками вместе со значениями функции в них. Одна или несколько вершин после этого отбраковывается. Итерационный процесс завершается тогда, когда вершины симплекса и вычисленные в них значения функции при сравнении с предыдущей итерацией удовлетворяют некоторым условиям сходимости.

Среди алгоритмов такого типа метод Нелдера—Мида, пожалуй,

самый экономичный – число вычислений значений минимизируемой функции минимально. На практике обычно требуется одно-два вычисления функции для построения нового симплекса.

Метод Нелдера—Мида весьма широко используется для решения прикладных задач.

Приведем известные [1] результаты о сходимости оригинального метода (сформулированного авторами).

- Для функций одной переменной алгоритм сходится к точке минимума. Скорость сходимости  $M$ -линейна в том случае, когда коэффициент отражения  $\rho = 1$ . (Под  $M$ -линейной скоростью сходимости подразумевается, что существует такое число  $M$ , не зависящее от минимизируемой функции, что диаметр симплекса уменьшается не менее чем в два раза после  $M$  итераций.)
- В двумерном пространстве значения минимизируемой функции во всех вершинах симплекса сходятся к одному и тому же значению (в случае стандартного симплекс-метода).
- В двумерном пространстве в случае стандартного метода Нелдера—Мида диаметр симплексов стремится к нулю при стремлении числа итераций к  $\infty$ .

Интересно отметить, что последнее утверждение не предполагает, что симплексы сходятся в одну точку. До сих пор не известны случаи, когда бы алгоритм не приводил к окончательному решению в виде одной точки, однако доказательства этого факта пока нет.

### 1.1.1. Алгоритм Нелдера—Мида

Алгоритм Нелдера—Мида был предложен [2] в качестве метода

минимизации действительной функции  $f(\mathbf{x})$  для  $\mathbf{x} \in R^n$ . Для полного задания симплекс-метода необходимо определить четыре константы: коэффициент отражения (reflection)  $\rho$ , коэффициент растяжения (expansion)  $\chi$ , коэффициент усечения множества (contraction)  $\gamma$  и коэффициент сжатия (shrinkage)  $\sigma$ . Согласно оригинальному алгоритму, эти константы должны удовлетворять условиям:

$$\rho > 0, \quad \chi > 1, \quad \chi > \rho, \quad 0 < \gamma < 1, \quad 0 < \sigma < 1. \quad (1.1)$$

Неравенство  $\chi > \rho$ , хотя и не упоминается в оригинальной формулировке, однако в описании реализации алгоритма присутствует. Практически во всех стандартных реализациях метода коэффициенты выбираются равными:

$$\rho = 1, \quad \chi = 2, \quad \gamma = 1/2, \quad \sigma = 1/2. \quad (1.2)$$

При рассмотрении одномерного случая воспользуемся общими условиями (1.1), а при доказательстве сходимости для двумерного случая ограничимся условиями (1.2).

### 1.1.2. Описание алгоритма

Рассмотрим  $k$ -ую итерацию алгоритма. Для каждого  $k \geq 0$  дан невырожденный симплекс  $\Delta_k$ , состоящий из  $(n + 1)$ -ой вершины, каждая из которых является точкой в  $R^n$ . Будем далее предполагать, что каждая итерация начинается с упорядочения вершин  $x_1^{(k)}, \dots, x_{n+1}^{(k)}$  таким образом, что

$$f_1^{(k)} \leq f_2^{(k)} \leq \dots \leq f_{n+1}^{(k)}, \quad (1.3)$$

где через  $f_i^{(k)}$  обозначено значение целевой функции  $f(x_i^{(k)})$  в точке  $x_i^{(k)}$ . На  $k$ -ой итерации строится новый симплекс для  $k + 1$ -ой итерации, такой, что  $\Delta_{k+1} \neq \Delta_k$ . Поскольку мы ищем минимум функции  $f$ , то считаем  $x_1^{(k)}$

лучшей вершиной, а  $x_{n+1}^{(k)}$  - *наихудшей* вершиной. Аналогично, значение  $f(x_{n+1}^{(k)})$  будем считать *наихудшим* значением функции.

Опишем одну итерацию симплекс-метода Нелдера—Мида. Не ограничивая общности, будем опускать индекс  $k$  в дальнейших выкладках. Результатом каждой итерации будет:

- единственная новая вершина симплекса, замещающая  $x_{n+1}^{(k)}$  во множестве вершин при переходе к следующей итерации;
- при проведении сжатия симплекса вычисляется  $n$  новых точек, которые вместе с  $x_1$  образуют новый симплекс на следующей итерации.

Опишем одну итерацию алгоритма Нелдера—Мида.

- **1. Сортировка.** Отсортировать вершины, используя правило отсечения элементов (смотри ниже) так, чтобы они удовлетворяли условию

$$f(x_1) \leq f(x_2) \leq \dots f(x_{n+1}).$$

- **2. Отражение.** Вычислить точку отражения  $x_r$  из условия

$$x_r = \bar{x} + (\bar{x} - x_{n+1}) = (1 + \rho)\bar{x} - \rho x_{n+1}, \quad (1.4)$$

где  $\bar{x} = \sum_{i=1}^n x_i / n$  - центр тяжести  $n$  лучших точек (т. е. всех вершин симплекса, за исключением  $x_{n+1}$ ). Вычислить  $f_r = f(x_r)$ .

- **3. Растяжение.** Если  $f_r < f_1$ , т. е.  $\bar{x}_r$  оказывается новой лучшей точкой, то направление отражения признается удачным и делается попытка растянуть многогранник в этом направлении. Для этого вычисляется *точка растяжения*  $\bar{x}_e$  (expansion point)

$$x_e = \bar{x} + \chi(x_r - \bar{x}) = \bar{x} + \rho\chi(\bar{x} - x_{n+1}) = (1 + \rho\chi)\bar{x} - \rho\chi x_{n+1} \quad (1.5)$$

и вычисляется  $f_e = f(x_e)$ . Если  $f_e < f_r$ , то растяжение оказалось успешным, точку  $x_e$  добавляется к множеству точек симплекса; иначе

(если  $f_e \geq f_r$ ), точка  $x_r$  добавляется к множеству точек симплекса. На этом итерация заканчивается

- **4. Сжатие.** Если  $f_r \geq f_n$ , то считаем, что многогранник слишком велик и его надо сжать. Сжатие (стягивание, усадка) проводится между  $\bar{x}$  и лучшей из точек  $x_{n+1}$  и  $x_r$ .

- **а) Внешнее сжатие.** Если  $f_n \leq f_r < f_{n+1}$  (т. е.  $x_r$  строго лучше, чем  $x_{n+1}$ ) надо провести внешнее сжатие: вычислить

$$x_c = \bar{x} + \gamma(x_r - \bar{x}) = \bar{x} + \gamma\rho(\bar{x} - x_{n+1}) = (1 + \rho\gamma)\bar{x} - \rho\gamma x_{n+1} \quad (1.6)$$

и  $f_c = f(x_c)$ . Если  $f_c < f_r$ , добавляем точку  $\bar{x}_c$  к множеству точек симплекса и заканчиваем итерацию, иначе переходим к пункту 5 и проводим операцию усечения.

- **б) Внутреннее сжатие.** Если  $f_r > f_{n+1}$  то следует провести внутреннее сжатие: вычислить

$$x_{cc} = \bar{x} - \gamma(\bar{x} - x_{n+1}) = (1 - \gamma)\bar{x} + \gamma x_{n+1} \quad (1.7)$$

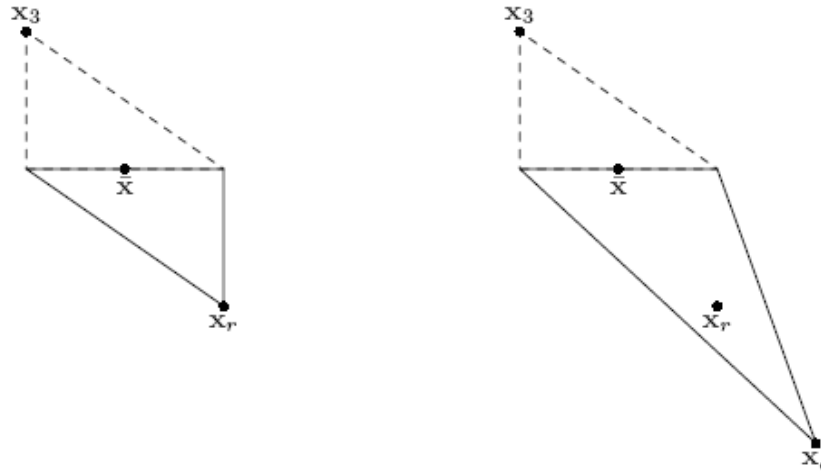
и  $f_{cc} = f(x_{cc})$ . Если  $f_{cc} < f_{n+1}$ , то добавляем точку  $x_{cc}$  к множеству точек симплекса и заканчиваем итерацию, иначе переходим к пункту 5 и проводим операцию усечения.

- **5. Усечение.** Вычисляем значения функции  $f$  в  $n$  точках

$v_i = x_1 + \sigma(x_i - x_1)$ ,  $i = 2, \dots, n+1$ . Неупорядоченные вершины симплекса на следующей итерации состоят из точек  $x_1, v_2, \dots, v_{n+1}$ .

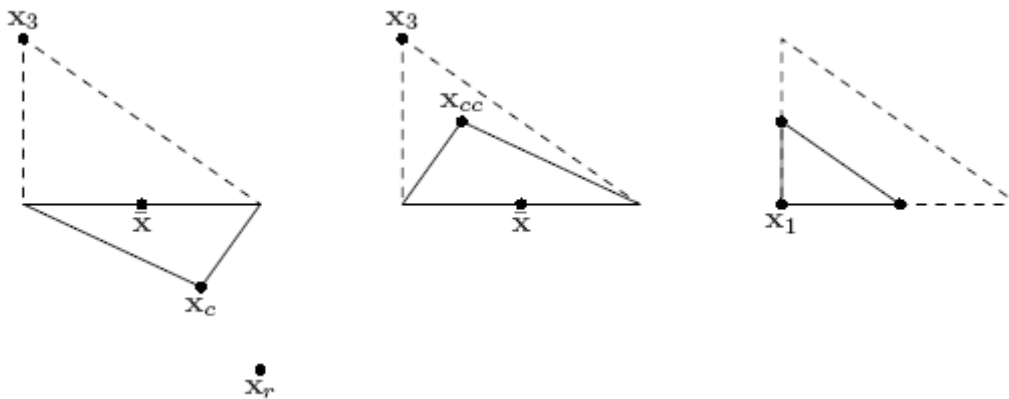
На рис. 1 и 2 продемонстрированы эффекты отражения, растяжения, сжатия и усечения для симплекса в двумерном случае. Исходные симплексы обозначены пунктирной линией.





**Рис. 1. Отраженная и продвинутая точки симплекс-метода в двумерном случае**

Для решения демонстрационных примеров использовался стандартный набор коэффициентов:  $\rho = 1$ ,  $\chi = 2$ ,  $\gamma = 1/2$ , и  $\sigma = 1/2$ .



**Рис. 2. Симплексы метода Нелдера—Мида после внешнего и внутреннего сжатия и после усечения**

Отметим, что во всех случаях, кроме усечения, новая вершина всегда лежит на прямой, проходящей через точки  $\bar{x}$  и  $x_{n+1}$ .

В случае, когда в вершинах симплекса целевая функция принимает одинаковые значения, мы будем упорядочивать вершины (при добавлении новой точки) в соответствии с соотношением

$$f(x_1^{(k+1)}) \leq f(x_2^{(k+1)}) \leq \dots \leq f(x_{n+1}^{(k+1)}).$$

**Упорядочивание точек без усечения.** В том случае, когда итерация заканчивается пунктами 2 или 3 (без поведения усечения), точка  $x_{n+1}^{(k)}$  исключается из множества вершин на следующей итерации. Введенная же в это множество на  $k$ -ой итерации точка, обозначаемая  $v^{(k)}$ , становится новой вершиной и занимает позицию с номером  $j+1$  во множестве вершин  $\Delta_{k+1}$ , где

$$j = \max_{0 \leq l \leq n} \{l \mid f(v^{(k)}) < f(x_{l+1}^{(k)})\};$$

все же остальные вершины сохраняют свой относительный порядок и на новой итерации.

**Упорядочивание точек после операции усечения.** В случае, когда итерация заканчивается усечением многогранника, всего лишь одна точка из старого набора  $\Delta_k$  переходит в новое множество вершин  $\Delta_{k+1}$ , это точка  $x_1^{(k)}$ . Здесь задействовано лишь одно правило: если во вновь вычисленном наборе вершин имеется несколько точек, значения целевой функции в которых совпадают, то в новый набор переносится точка  $x_1^{(k)}$ . То есть если

$$\min \{f(v_2^{(k)}) \leq \dots \leq f(v_{n+1}^{(k)})\} = f(x_1^{(k)}),$$

тогда  $x_1^{(k+1)} = x_1^{(k)}$ . Кроме того, после усечения обязательно проводится упорядочивание вершин в соответствии с применяемым правилом сортировки.

### 1.1.3. Свойства метода Нелдера—Мида

Опишем основные свойства алгоритма оптимизации Нелдера—Мида, следующие непосредственно из определения итерационного процесса.

- 1. На каждой итерации метода Нелдера—Мида необходимо одно вычисление функции в новой вершине, если итерация заканчивается шагом 2, два вычисления целевой функции, если итерация

заканчивается шагами 3 или 4, и  $n + 2$  вычисления целевой функции, если итерация заканчивается шагом 5.

- 2. Шаг «отражение» назван так по той причине, что отраженная точка  $x_r$  (1.4) (отмасштабированное) — отражение худшей точки  $x_{n+1}$  относительно точки  $\bar{x}$  (центра тяжести всех вершин, кроме лучшей) вдоль прямой, проходящей через  $x_{n+1}$  и  $\bar{x}$ . Это чистое отражение в случае, когда  $\rho = 1$ .
- 3. Для произвольных функций усечение может, по-видимому, приводить к увеличению значений в каждой точке по сравнению с предыдущим шагом, за исключением значения  $f_1$  функции в точке  $x_1$ , т. е. возможно, что  $f_i^{(k+1)} > f_i^{(k)}$  для  $2 \leq i \leq n + 1$ . Дополнительно можно отметить, что в случае 4а (внешнего сжатия) по алгоритму требуется провести усечение многогранника, если  $f(x_c) > f(x_r)$ , хотя вычисленное значение функции в точке  $x_r$  значительно улучшает ситуацию, поскольку  $f(x_r) < f(x_{n+1})$ .
- 4. В оригинальном методе Нелдера—Мида при растяжении в случае  $f(x_c) < f(x_1)$  выбирается точка  $x_c$ , в противном случае — точка  $x_r$ . Ныне обычно выбирается лучшая из точек  $x_c$  и  $x_r$ , если обе дают улучшение ситуации по сравнению с точкой  $x_1$ .

Удивительно, но строгого доказательства сходимости в  $R^n$  для столь простого с первого взгляда метода до сих пор нет. Имеются довольно сильные результаты о сходимости в одномерном пространстве для строго выпуклых функций и гораздо более слабые — о сходимости процесса для функций двух переменных [1].

Для функций одной переменной удается показать, что метод всегда сходится к точке минимума, когда  $\rho\chi \geq 1$ ; при  $\rho = 1$  сходимость М-линейна, а при  $\rho\chi < 1$  алгоритм не всегда сходится к точке минимума.

На настоящий момент главной интригой для исследователей метода Нелдера—Мида являются не возможные доказательства его сходимости для наиболее широкого класса функций, а удивительная способность алгоритма при решении практических задач строить такие последовательности вершин, на которых значения целевой функции убывают весьма быстро.

## 1.2. Оптимизация методом наименьших квадратов

Используемые в сочетании с квазиньютоновским методом процедуры линейного поиска получили широкое применение. Эти методы также используются и в подпрограммах нелинейной оптимизации методом наименьших квадратов. В задачах на метод наименьших квадратов подлежащая минимизации функция  $f(x)$  представляет собой сумму квадратов:

$$\min_{x \in R^n} f(x) = \frac{1}{2} \|F(x)\|_2^2 = \frac{1}{2} \sum_i F_i(x)^2. \quad (1.8)$$

Подобного типа задачи широко распространены и имеют ряд практических применений, особенно при подборе модельной функции для некоего набора данных, т.е. определение нелинейных параметров модели. Эти задачи также широко распространены в теории управления, где в конечном итоге необходимо получить некую  $y(x, t)$ , соответствующую некой непрерывной модельной траектории  $\varphi(t)$  для вектора  $x$  и скаляра  $t$ . Данная задача может быть сформулирована как:

$$\min_{x \in R^n} \int_{t_2}^{t_1} (y(x, t) - \varphi(t))^2 dt, \quad (1.9)$$

где  $y(x, t)$  и  $\varphi(t)$  есть некие скалярные функции.

При дискретизации интеграла посредством подходящих квадратурных формул уравнение (1.9) может быть сформулировано как

задача на метод наименьших квадратов:

$$\min_{x \in R^n} f(x) = \sum_i (\bar{y}(x, t_i) - \varphi(t_i))^2, \quad (1.10)$$

где -  $\bar{y}$  и  $\bar{\varphi}$  включают в себя веса квадратичной схемы. Отметим, что в данной задаче под вектором  $F(x)$  понимается:

$$F(x) = \begin{bmatrix} \bar{y}(x, t_1) - \varphi(t_1) \\ \bar{y}(x, t_2) - \varphi(t_2) \\ \dots \\ \bar{y}(x, t_m) - \varphi(t_m) \end{bmatrix}.$$

В задачах данного типа невязка  $\|F(x)\|$ , по-видимому, должна быть наименьшей в точке оптимума, поскольку согласно общепринятой практике необходимо провести как можно ближе к реальной траектории. Хотя приведенная функция для метода наименьших квадратов (уравнение (1.9)) может быть минимизирована с помощью общего метода оптимизации без наличия ограничений, определенные характеристики данной задачи часто могут быть использованы для улучшения итеративной эффективности данной методики решения. Градиент и матрица Гессе для задачи метода наименьших квадратов (1.9) имеют особую структуру.

После обозначения матрицы Якобиана для  $F(x)$  размерностью  $m \times n$  через  $J(x)$ , вектора градиента функции  $f(x)$  через  $G(x)$ , матрицы Гессе для  $f(x)$  через  $H(x)$  и матрицы Гессе для каждой  $F_i(x)$  через  $H_i(x)$  получим

$$G(x) = 2J(x)^T F(x), \quad (1.11)$$

где

$$Q(x) = \sum_{i=1}^m F_i(x)H_i(x).$$

Матрица  $Q(x)$  обладает тем свойством, что когда невязка  $\|F(x)\|$  стремится к нулю при стремлении  $x_k$  к точке решения, то и сама

матрица стремится к нулю. Таким образом, при небольших значениях  $\|F(x)\|$  в точке решения одним из наиболее эффективных методов является использование направления Ньютона—Гаусса в качестве основы для процедуры и оптимизации.

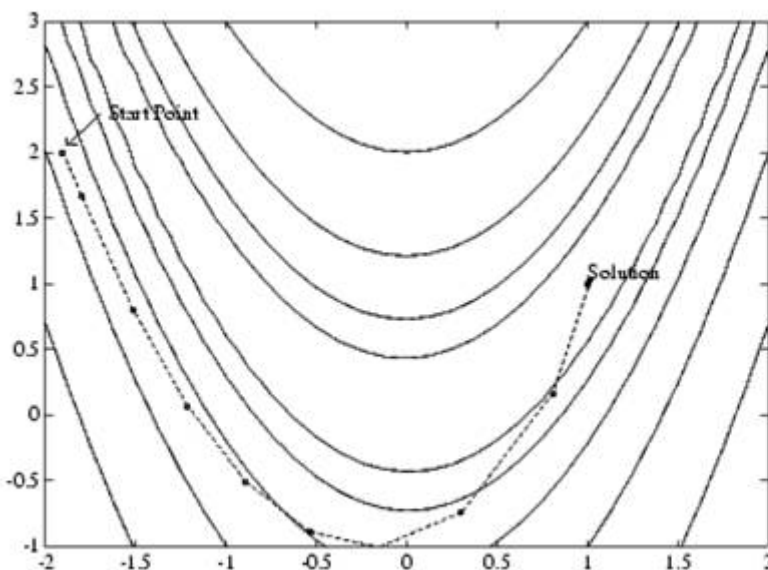
### 1.2.1. Метод Ньютона—Гаусса

Согласно методу Ньютона—Гаусса направление поиска  $d_k$  находится на каждой итерации  $k$  решением линейной задачи минимизации по методу наименьших квадратов:

$$\min_{x \in R^n} \|J(x_k)d_k - F(x_k)\|_2^2. \quad (1.12)$$

Полученное согласно данному методу направление является эквивалентом направления Ньютона при пренебрежении члена  $Q(x)$ . Направление поиска  $d_k$  может быть использовано в качестве составляющей стратегии линейного поиска, что обеспечивает условие: на каждой итерации идет уменьшение функции  $f(x)$ .

Рассмотрим возможные преимущества метода Ньютона—Гаусса. На рисунке 3 представлена траектория поиска минимума для функции Розенброка при использовании постановки задачи как метода наименьших квадратов. Метод Ньютона—Гаусса дает сходимость после 48 обращений к расчету функции при конечно-разностном расчете градиента по сравнению с 140 итерациями для BFGS метода без наличия ограничений.



**Рис. 3. Применение метода Ньютона—Гаусса для функции Розенброка**

В методе Ньютона—Гаусса часто встречается ряд проблем в ситуации, когда член второго порядка  $Q(x)$  в уравнении (1.11) достаточно значителен по своей величине. Методом, который преодолевает такие трудности, является метод Левенберга—Марквардта.

### 1.2.2. Метод Левенберга—Марквардта

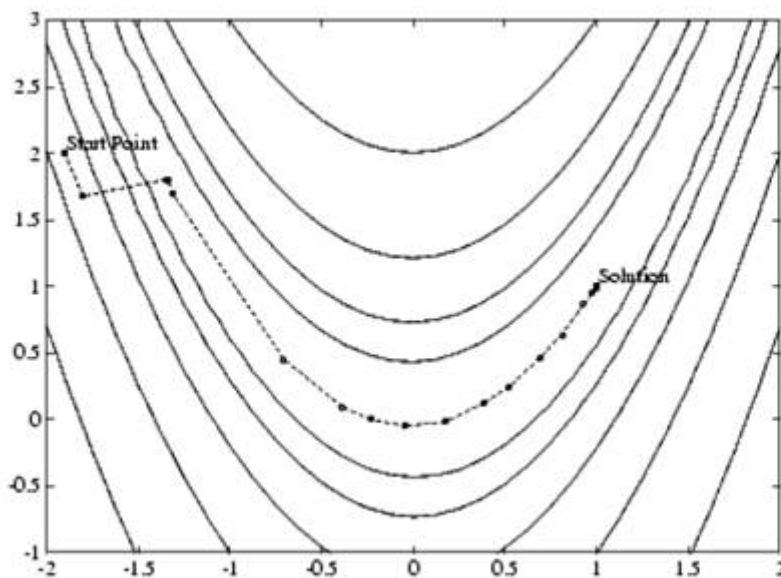
В основу метода Левенберга-Марквардта [3-6] положено направление поиска, которое находится при решении системы линейных уравнений:

$$(J(x_k)^T J(x_k) + \lambda_k I) d_k = -J(x_k) F(x_k), \quad (1.13)$$

где скаляр  $\lambda_k$  задает как величину, так и направление параметра  $d_k$ . Когда  $\lambda_k$  равен нулю, то направление  $d_k$  будет идентично этому же параметру из метода Ньютона—Гаусса. По мере того как  $\lambda_k$  стремится к бесконечности, то  $d_k$  стремится к вектору с нулевыми компонентами и направлению наискорейшего спуска. В данном случае предполагается, что для достаточно больших значений  $\lambda_k (F(x_k) + d_k) < F(x_k)$  остается справедливым. Следовательно, член  $\lambda_k$  может быть контролируемым с целью обеспечения спуска в случае необходимости учета членов второго

порядка, которые, в свою очередь, заметно ограничивают эффективность метода Ньютона—Гаусса.

Отсюда следует, что метод Левенберга—Марквардта основан на направлении поиска, являющегося сочетанием направления Ньютона—Гаусса и наискорейшего спуска (рис. 4). Решение для функции Розенброка сходится после 90 обращений к расчету функции по сравнению с 48 для метода Ньютона—Гаусса. Такая низкая эффективность отчасти объясняется тем, что метод Ньютона—Гаусса обычно более эффективен в случае, когда в решении невязка равна нулю. Однако такая информация не всегда является заранее доступной, и повышенная устойчивость метода Левенберга—Марквардта компенсирует его иногда имеющую место слабую эффективность.



**Рис. 4. Метод Левенберга—Марквардта для функции Розенброка**

### **1.2.3. Реализация метода Ньютона—Гаусса**

Метод Ньютона—Гаусса реализуется с помощью стратегии полиномиального линейного поиска, аналогичного тому, что было рассмотрено в разделе применительно к оптимизации без ограничений.



При решении линейной задачи методом наименьших квадратов (уравнение (1.8)) путем использования разложения QR для  $J(x_k)$  и подобного разложения для  $F(x_k)$  можно избежать дополнительных возмущений. Такой подход отличается от применения инверсии явной матрицы  $J(x_k)^E J(x_k)$ , где возможно проявление неожиданных ошибок.

Для обеспечения устойчивости данного метода возможны дополнительные меры. Эти меры включают в себя корректировку алгоритма метода Левенберга—Марквардта в случае, когда длина шага становится ниже некоего порогового значения ( $10^{-15}$ ) или число обусловленности  $J(x_k)$  будет меньше  $10^{-10}$ . Под числом обусловленности в данном случае понимается отношение наибольшего сингулярного значения к наименьшему.

### **1.3. Алгоритм Левенберга—Марквардта**

Алгоритм Левенберга—Марквардта (Levenberg-Marquardt Algorithm, LMA) является наиболее распространенным алгоритмом оптимизации. Он превосходит по производительности метод наискорейшего спуска и другие методы сопряженных градиентов в различных задачах. Изначально считалось, что LMA – это комбинация простейшего градиентного метода и метода Гаусса-Ньютона, однако впоследствии выяснилось, что данный алгоритм можно также рассматривать как метод доверительных интервалов. В работе [3] дается интуитивно понятное объяснение LMA.

#### **1.3.1. Постановка задачи**

LMA решает задачу нелинейной минимизации методом наименьших квадратов. Это означает, что функция, которую необходимо минимизировать, выглядит следующим образом:

$$f(x) = \frac{1}{2} \sum_{j=1}^m r_j^2(x),$$

где  $x = (x_1, x_2, \dots, x_n)$  – вектор, а  $r_j$  – функция отображения из  $R^n$  в  $R$ .

Функцию  $r_j$  называют невязкой в предположении, что  $m \geq n$ .

Для простоты функция  $f$  представляется вектором невязки вида:

$$r(x) = (r_1(x), r_2(x), \dots, r_m(x)).$$

Теперь  $f$  можно переписать как  $f(x) = \frac{1}{2} \|r(x)\|^2$ , а ее производные

представить с помощью матрицы Якоби  $J(x) = \frac{\partial r_j}{\partial x_i}$ ,  $1 \leq j \leq m$ ,  $1 \leq i \leq n$ .

Рассмотрим линейный случай, когда каждая функция  $r_j$  линейна.

Здесь якобиан равен константе,  $r$  можно представить как гиперплоскость в

пространстве, а  $f(x) = \frac{1}{2} \|Jx + r(0)\|^2$ . Тогда градиент функции

$\nabla f(x) = J^T (Jx + r)$  и  $\nabla^2 f(x) = J^T J$ . Решая задачу минимума  $\nabla f(x) = 0$ ,

получим  $x_{\min} = -(J^T J)^{-1} J^T r$ , т.е. решение системы нормальных уравнений

$$(J^T J)x = -J^T r.$$

Возвращаясь к общему (нелинейному) случаю, получим:

$$\nabla f(x) = \sum_{j=1}^m r_j(x) \nabla r_j(x) = J(x)^T r(x), \quad (1.14)$$

$$\nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x). \quad (1.15)$$

Отличительной особенностью метода наименьших квадратов является то, что, имея матрицу Якоби  $J$ , легко получить гессиан  $\nabla^2 f(x)$ , если функции  $r_j$  можно аппроксимировать линейными приближениями

(т.е.  $\nabla^2 r_j(x)$  малы) или если  $r_j(x)$  малы сами по себе. Тогда гессиан, как и в линейном случае, будет равен:

$$\nabla^2 f(x) = J(x)^T J(x). \quad (1.16)$$

Важно отметить, что уравнение (1.16) верно только для малых невязок. Проблемы больших невязок не могут быть решены с помощью квадратичной аппроксимации, и, следовательно, производительность алгоритма, представленного в этом разделе, в таких случаях невелика.

### **1.3.2. LMA как комбинация простейшего градиентного метода и метода Ньютона— Гаусса**

Простейший градиентный метод – это наиболее интуитивно понятный способ нахождения минимума функции. Вычисление параметра на очередном шаге выполняется путем вычитания градиента функции, умноженного на заданный положительный коэффициент:

$$x_{i+1} = x_i - \lambda \nabla f. \quad (1.17)$$

Однако при таком подходе имеют место различные проблемы сходимости. Логично предположить, что желательно было бы осуществлять большие шаги по направлению градиента там, где градиент мал (т.е. наклон пологий), и, наоборот, маленькие шаги там, где градиент большой, чтобы не пропустить минимум. Вместе с тем, в формуле (1.17) выполняются прямо противоположные действия. Другая проблема заключается в том, что кривизна поверхности невязки может быть не одинаковой по всем направлениям. К примеру, если есть длинная и узкая впадина на поверхности невязки, компонент градиента в направлении, указывающем вдоль основания впадины, очень мал, а компонент градиента вдоль стенок впадины, наоборот, велик. Это приводит к движению по направлению к стенкам впадины, тогда как необходимо перемещаться на большие расстояния вдоль основания впадины и на малые – вдоль ее

стенки.

Ситуацию можно улучшить, если учитывать информацию о кривизне и градиенте, т. е. вторые производные. Один из способов сделать это – использовать метод Ньютона для решения уравнения  $\nabla f(x) = 0$ . Раскладывая градиент  $f$  в ряд Тейлора вокруг текущего состояния  $x_0$ , получим

$$\nabla f(x) = \nabla f(x_0) + (x - x_0)^T \nabla^2 f(x_0) + \dots \quad (1.18)$$

Пренебрегая членами более высокого порядка (считая  $f$  квадратичной вблизи  $x_0$ ) и решая задачу минимума, приравняв левую часть уравнения (1.18) к нулю, получим правило вычисления параметра на очередном шаге по методу Ньютона:

$$x_{i+1} = x_i - (\nabla^2 f(x_i))^{-1} \nabla f(x_i). \quad (1.19)$$

Поскольку метод Ньютона напрямую использует предположение о квадратичности (пренебрегая членами более высоких порядков при разложении в ряд Тейлора), нет необходимости точно вычислять гессиан, а достаточно использовать его аппроксимацию (1.16). Главное достоинство такого подхода – быстрая сходимость. Однако скорость сходимости зависит от начального положения (если быть более точным – от линейности вокруг начального положения).

Легко заметить, что простейший градиентный метод и метод Ньютона— Гаусса дополняют друг друга с точки зрения предоставляемых преимуществ. Основываясь на этом наблюдении, Левенберг предложил алгоритм, в котором правило вычисления параметра есть комбинация правил (1.17) и (1.19):

$$x_{i+1} = x_i - (H + \lambda I)^{-1} \nabla f(x_i), \quad (1.20)$$

где  $H$  – матрица Гессе, вычисленная в точке  $x_i$ . Данное правило

используется следующим образом: если на очередной итерации невязка сокращается, это значит, что предположение о квадратичности  $f(x)$  работает, и мы уменьшаем  $\lambda$  (обычно в 10 раз), чтобы понизить влияние градиентного спуска. С другой стороны, если невязка увеличивается, необходимо следовать направлению градиента, и мы увеличиваем  $\lambda$  (во столько же раз).

Таким образом, алгоритм Левенберга представляется в виде последовательности действий:

- 1. Вычислить параметр на очередной итерации по правилу(1.20).
- 2. Оценить невязку в новом векторе параметров.
- 3. Если в результате вычисления параметра невязка увеличилась, вернуться на шаг назад (т.е. восстановить прежние значения весов) и увеличить  $\lambda$  в 10 раз. Затем повторить выполнение, начиная с шага 1.
- 4. Если в результате вычисления параметра невязка уменьшилась, принять текущий шаг (т.е. оставить новые значения весов) и уменьшить  $\lambda$  в 10 раз.

Недостатком данного алгоритма является то, что если значение  $\lambda$  велико, вычисленная матрица Гессе никак не используется. Однако можно извлечь некоторую выгоду из второй производной даже в этом случае, масштабируя каждый компонент градиента согласно кривизне. Это должно привести к увеличению шага вдоль направлений, где градиент мал, так что классическая проблема впадины больше не возникнет. Этот ключевой момент был замечен Марквардтом. Он заменил единичную матрицу в формуле (1.20) на диагональ гессиана, получив таким образом следующее правило:

$$x_{i+1} = x_i - (H + \lambda \text{diag}[H])^{-1} \nabla f(x_i). \quad (1.21)$$

Поскольку гессиан пропорционален кривизне  $f$ , правило (1.21)

приведет к большим шагам при малой кривизне (т.е. для почти плоской поверхности) и к малым шагам при большой кривизне (т.е. для крутого наклона).

Стоит отметить, что хотя LMA является не оптимальным, а лишь эвристическим методом, он очень хорошо работает на практике. Единственный его недостаток заключается в необходимости обращения матрицы на каждом шаге. Даже несмотря на то, что нахождение обратной матрицы обычно выполняется с использованием быстрых методов псевдообращения (таких, как разложение по сингулярным числам матрицы), время одной итерации становится неприемлемым для нескольких тысяч параметров. Для моделей же средних размеров (с несколькими сотнями параметров) LMA работает даже быстрее, чем простейший градиентный метод.

### 1.3.3. LMA как метод доверительных интервалов

Изначально LMA был представлен Марквардтом так, как показано в предыдущем разделе, где нахождение минимума производится за счет изменения параметра  $\lambda$ . Позже было предложено рассматривать LMA как метод доверительных интервалов.

Алгоритмы, основанные на доверительных интервалах, принципиально отличаются от алгоритмов, представленных нами ранее и основанных на так называемом методе линейного поиска. В методе линейного поиска выбирается направление спуска и  $\alpha_k$  – размер шага, затем на новом шаге вычисляется  $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$ , а размер очередного шага является решением подзадачи:

$$\min f(x^{(k)} + \alpha_k p^{(k)}), \quad \forall \alpha_k > 0.$$

В алгоритме, основанном на доверительных интервалах, сначала строится модель  $m^{(k)}$ , которая аппроксимирует функцию  $f$  на конечном

интервале  $\Delta$  вокруг  $x^{(k)}$ . Этот интервал называется доверительным интервалом. На каждой итерации с помощью некоторых эвристик интервал  $\Delta$  изменяют. Модель  $m^{(k)}$  чаще всего представляет собой квадратичную функцию, полученную разложением функции  $f$  в ряд Тейлора вблизи  $x^{(k)}$ :

$$m^{(k)} = f(x^{(k)}) + \nabla f(x^{(k)}) \cdot p + \frac{1}{2} p^T H p, \quad (1.22)$$

где  $H$  – гессиан или приближение гессиана. Подзадача для нахождения размера очередного шага выглядит следующим образом:

$$\min_{\|p\| \leq \Delta} \left( f(x^{(k)}) + \nabla f(x^{(k)}) \cdot p + \frac{1}{2} p^T H p \right), \quad (1.23)$$

а сам шаг –  $x^{(k+1)} = x^{(k)} + p$ . Таким образом, алгоритм, основанный на методе доверительных интервалов, может быть представлен как последовательность итераций, на каждой из которых функция  $f$  моделируется квадратичным приближением, а затем происходит скачок на его минимум.

Решение задачи (1.23) удовлетворяет следующей теореме

**Теорема.**  $p^*$  – общее решение задачи

$$\min_{\|p\| \leq \Delta} \left( f(x^{(k)}) + \nabla f(x^{(k)}) \cdot p + \frac{1}{2} p^T H p \right),$$

если

$\|p^*\| \leq \Delta$  и существует скаляр  $\lambda$  такой, что:

$$(H + \lambda I)p^* = -g \quad (1.24)$$

$$\lambda(\Delta - \|p^*\|) = 0 \quad (1.25)$$

и  $(H + \lambda I)$  – положительная полуопределенная матрица,  $g, f \in R^n$ .

Легко заметить, что уравнение (1.24) есть ни что иное, как уравнение (1.20). Равенство (1.25) устанавливает, что если  $\|p^*\| \leq \Delta$ , то  $\lambda = 0$ , но не наоборот. Следовательно, мы приходим к тому же правилу вычисления параметра, что и в предыдущем разделе.

Эвристика для вычисления размера доверительного интервала обычно зависит от отношения полученного и предсказываемого изменения  $f$ , т.е.

$$\rho_k = \frac{f(w^{(k)}) - f(w^{(k)} + p^*)}{f(w^{(k)}) - m^{(k)}(p^*)}. \quad (1.26)$$

Если разница между полученным и предсказываемым значением невелика ( $\rho_k \approx 1$ ), то интервал  $\Delta$  увеличивают, если же разница существенная ( $\rho_k$  мало), то  $\Delta$  уменьшают. Если  $\rho_k$  меньше порогового значения (обычно  $\sim 10^{-4}$ ), итерация отклоняется, и значение  $w^{(k)}$  сохраняется, но интервал  $\Delta$  уменьшают, как и ранее. Таким образом, данный алгоритм схож с алгоритмом, представленным в предыдущем пункте, однако вместо  $\lambda$  на каждом шаге изменяется доверительный интервал  $\Delta$ .



## Тема 2. Брэгговское зеркало

Распространение электромагнитных волн в средах с периодически изменяющимися свойствами сопровождается появлением новых качественных особенностей, наиболее заметных в тех случаях, когда длина волны становится сравнима с характерным пространственным периодом изменения свойств структуры. В этом разделе мы рассмотрим наиболее простую задачу, которая, однако, позволит сделать достаточно общие выводы и продемонстрирует используемые математические приемы расчета. Речь идет о так называемом брэгговском зеркале, представляющим собой некоторое ограниченное количество диэлектрических слоев с отличающимися параметрами, расположенными на диэлектрической же подложке. Как отмечалось, подобные структуры обладают способностью отражать электромагнитную волну почти со стопроцентной эффективностью, именно поэтому они находят широкое применение в устройствах лазерной техники, спектроскопии и т.д.

Указанную задачу будем ставить как задачу *одномерную*, т. е. все изменения параметров диэлектрических слоев будут происходить только лишь вдоль одной пространственной координаты, вдоль этой же координаты будет распространяться и электромагнитная волна. Как правило, необходимый для решения математический аппарат именно в одномерном случае оказывается наиболее простым. Благодаря этому часто удается получить точное решение задачи, которое в дальнейшем может служить отправным пунктом и тестом для приближенных методов, описывающих более сложные физические ситуации.

Проблема распространения волн в одномерных периодических средах в своем развитии прошла два ярко выраженных пика. Первый связан с исследованием структуры энергетических зон одномерных кристаллов. Будучи первоначально лишь удобной физической моделью, электронная

теория одномерного кристалла обрела впоследствии реальную экспериментальную базу в виде полупроводниковых сверхрешеток. Другой пик интереса к изучению распространения волн в одномерных периодических структурах был стимулирован интенсивным развитием в середине 60-х годов и начале 70-х годов таких областей прикладной физики, как оптическая голография, акустооптика, интегральная и оптическая электроника и др. Обращение вновь к одномерным задачам было в этом случае продиктовано стремлением на новом уровне (определяемом возникшими задачами) и с новыми возможностями (в частности, вычислительной техники) изучить особенности распространения и рассеяния волн в одномерных структурах с целью практического их использования. И хотя к реальным объектам модель одномерной периодической среды чаще всего неприменима, результаты, полученные на ее основе, полезны для качественного анализа и приближенного описания физики протекающих процессов. Таким образом, одномерная задача приобрела характер ключевой или базовой модели, предназначенной, с одной стороны, для обогащения нашего понимания происходящих в периодической структуре процессов и, с другой стороны, для разработки и тестирования количественных методов их анализа.

## ***2.1. Постановка задачи***

В самом общем виде задача об отражении плоской электромагнитной волны от ограниченной среды с периодически изменяющимися параметрами ставится следующим образом [7]. Между двумя однородными средами с заданными параметрами расположена ограниченная по одной из декартовых координат среда, параметры которой периодически меняются вдоль этой пространственной координаты. Требуется определить коэффициент отражения (пропускания)

по мощности для электромагнитной волны, падающей на такую структуру, причем направление распространения волны совпадает с направлением стратификации периодически модулированной среды.

Для решения поставленной задачи ее необходимо разбить на две части; в первой нужно определить закон дисперсии для электромагнитной волны в неограниченной периодически модулированной среде, а во второй методом «сшивания» полей на границах этой периодически модулированной среды непосредственно вычислить искомый коэффициент отражения. При этом предполагается, что закон дисперсии для электромагнитной волны в однородных полубесконечных средах задан диэлектрическими проницаемостями этих сред.

Итак, в первую очередь займемся определением закона дисперсии для волны в неограниченной среде с периодической модуляцией. Поскольку в большинстве случаев в оптическом диапазоне длин волн мы имеем дело с немагнитными и непоглощающими средами, то естественно предположить, что под периодически меняющимися параметрами среды мы будем понимать изменения действительной части диэлектрической проницаемости такой среды, полагая мнимую часть равной нулю. Возможны два типа модуляции или два типа одномерных периодических структур: слоистые среды со ступенчатым (кусочно-постоянным) законом изменения диэлектрической проницаемости и среды с гармонической модуляцией. Для слоистых сред разработаны эффективные матричные методы анализа, которые будут обсуждаться более подробно в дальнейшем. В случае же гармонической модуляции дисперсионное уравнение может быть получено, как это отмечалось выше, с помощью уравнения Матье. Остановимся в нашем рассмотрении именно на этом случае.

Монохроматическое поле электромагнитной волны с частотой  $\omega$  в среде с неоднородной диэлектрической проницаемостью описывается следующим однородным уравнением:

$$\nabla^2 \vec{E} + \frac{\omega^2}{c^2} \varepsilon \vec{E} + \text{grad}\left(\frac{\text{grad} \varepsilon}{\varepsilon}\right) = 0. \quad (2.1)$$

Поскольку среда, как мы договорились, неоднородна только в одном направлении, например, в направлении оси  $Z$ , то общее решение векторного уравнения (2.1) всегда может быть представлено спектром частных решений, каждое из которых будет однородным в каком-либо направлении в плоскости, нормальной оси  $Z$ . Если для каждого такого решения выбрать систему координат таким образом, что направление, в котором решение однородно, будет совпадать, например, с осью  $X$ , то векторное уравнение (2.1) сведется к двум линейно-независимым скалярным уравнениям относительно проекций электрического и магнитного полей волны на ось  $X$ :

$$\begin{cases} \nabla_{YZ}^2 E_X + \frac{\omega^2}{c^2} \varepsilon E_X = 0 \\ \nabla_{YZ}^2 H_X + \frac{\omega^2}{c^2} \varepsilon H_X + \frac{1}{\varepsilon} \frac{\partial \varepsilon}{\partial z} H_X = 0. \end{cases} \quad (2.2)$$

$$\text{Здесь } \nabla_{YZ}^2 = \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Заметим, что второе уравнение системы (2.2) подстановкой  $U = H_X \sqrt{\varepsilon}$  сводится к уравнению, аналогичному первому:

$$\nabla_{YZ}^2 U + \frac{\omega^2}{c^2} \varepsilon_{\text{экв}} U = 0, \quad (2.3)$$

$$\text{где } \varepsilon_{\text{экв}} = \varepsilon \left[ 1 - \frac{1}{2} \frac{c^2}{\omega^2 \varepsilon^2} \frac{\partial^2 \varepsilon}{\partial z^2} + \frac{1}{4} \frac{c^2}{\omega^2 \varepsilon^2} \left( \frac{\partial \varepsilon}{\partial z} \right)^2 \right].$$

Как видно из уравнений (2.2) и (2.3), задача о собственных решениях для электромагнитной волны (задача о собственных волнах) в среде с

одномерной модуляцией  $\varepsilon$  сводится к двум скалярным уравнениям. Решениями этих скалярных уравнений являются электромагнитные поля, поляризованные по оси  $X$  и по нормали к оси  $X$ . Поскольку уравнения для полей указанных поляризаций одинаковы, то достаточно рассмотреть решения лишь одного из них.

Ограничим дальнейшее рассмотрение анализом уравнения:

$$\nabla_{YZ}^2 E_X + \frac{\omega^2}{c^2} \varepsilon E_X = 0 \quad (2.4)$$

при гармонической модуляции диэлектрической проницаемости среды

$$\varepsilon = \varepsilon_1(1 + m \cos Kz).$$

Здесь  $\varepsilon_1$  — среднее значение диэлектрической проницаемости;  $m$  — глубина модуляции;  $K$  — пространственная частота модуляции ( $K = \frac{2\pi}{\Lambda}$ );  $\Lambda$  — пространственный период модуляции (рис. 5).

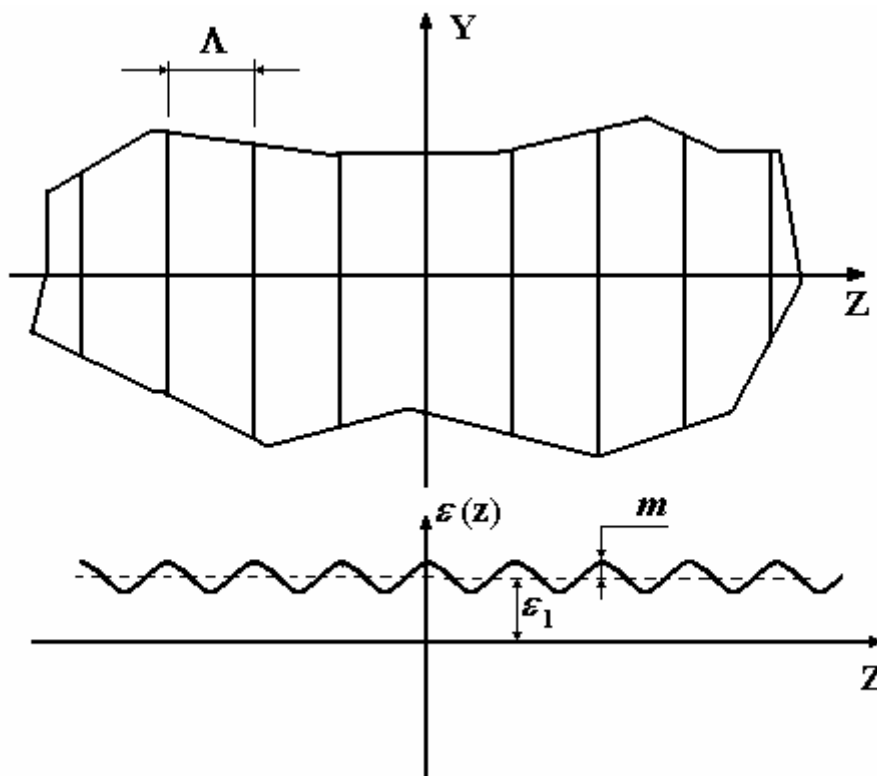


Рис. 5. Среда с одномерной гармонической модуляцией

Общее решение уравнения (2.4) будем разыскивать в виде спектра решений с разделяющимися переменными:

$$E_x = \int_{-\infty}^{+\infty} V(\beta, z) \exp(-i\beta y) d\beta. \quad (2.5)$$

При таком выборе решения зависимость от  $y$  в любом сечении  $z = \text{const.}$  представляется интегралом по спектру Фурье.

Подставляя (2.5) в (2.4) и учитывая ортогональность функций  $\exp(i\beta y)$ , получим

$$\frac{d^2V}{dz^2} + \left[ (k^2 \varepsilon_1 - \beta^2) + k^2 \varepsilon m \cdot \cos Kz \right] \cdot V = 0, \quad (2.6)$$

где  $k = \frac{\omega}{c}$ .

Заменой переменной  $z = \frac{2\tau}{K}$  и введением обозначений

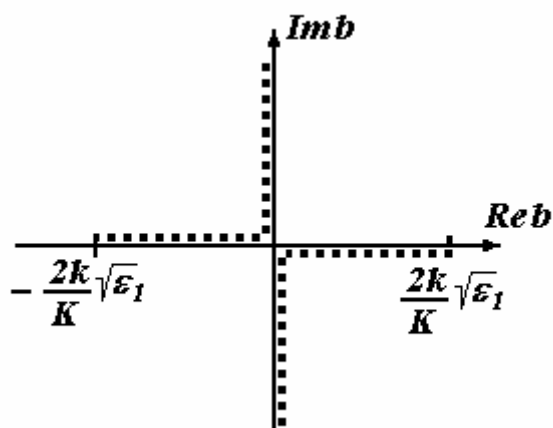
$b^2 = \frac{4}{K^2} (k^2 \varepsilon_1 - \beta^2)$ ,  $4q = \frac{k^2 \varepsilon_1}{K^2} m$  последнее уравнение приводится к привычной форме уравнения Матъе:

$$\frac{d^2V}{d\tau^2} + (b^2 + 4q \cos Kz) \cdot V = 0. \quad (2.7)$$

Поскольку  $\beta$  в (2.5) может принимать любые действительные значения на интервале от  $+\infty$  до  $-\infty$ , то величина  $b$  в уравнении (2.7) может быть либо действительной на интервале от 0 до  $\frac{2k\sqrt{\varepsilon_1}}{K}$ , либо чисто мнимой на интервале от  $-i\infty$  до  $+i\infty$ . На рис. 6 представлены две ветви значений  $b$  на комплексной плоскости для действительных значений  $\beta$  в интервале от  $-\infty$  до  $+\infty$ .

Как видно из обозначений, действительные значения  $b$  пропорциональны постоянным распространения плоских волн — прямой

( $+b$ ) и обратной ( $-b$ ) — в однородной среде с диэлектрической проницаемостью  $\varepsilon_1$  при  $q \rightarrow 0$ .



**Рис. 6.** Две ветви значений  $b$  для действительных значений  $\beta$ .

Мнимые значения при этом соответствуют неоднородным плоским волнам, т.е. волнам прямым, с экспоненциально уменьшающейся амплитудой и волнам обратным, с экспоненциально нарастающей амплитудой. Следовательно, ветвь  $b$  в четвертом квадранте на рис. 6 соответствует прямым плоским волнам (однородным и неоднородным), а ветвь  $b$  во втором квадранте — обратным волнам.

В соответствии с теоремой Флоке [8] решение уравнения (2.7) будем разыскивать в виде произведения экспоненты на периодическую функцию с периодом  $\pi$ . Последняя может быть представлена рядом Фурье:

$$V(\beta, z) = \exp(-i\mu\tau)F(\beta, \tau) = \exp(-i\mu\tau) \sum_{n=-\infty}^{+\infty} a_n \exp(-i2n\pi). \quad (2.8)$$

Решение (2.8) можно трактовать как волну, состоящую из суммы пространственных гармоник с постоянными распространения ( $\mu + 2n$ ) и амплитудами  $a_n$ . Подставляя решение (2.8) в уравнение (2.7), получим однородную бесконечную систему уравнений относительно амплитуд  $a_n$ :

$$p_n a_{n-1} + a_n + p_n a_{n+1} = 0, \quad (2.9)$$

где

$$p_n = -\frac{2q}{(2n + \mu)^2 + b^2}.$$

Условием совместности уравнений системы (2.9) является равенство нулю ее определителя:

$$\Delta(\mu) = \|p_n(\mu)\| = 0. \quad (2.10)$$

Уравнение (2.10) является искомым дисперсионным уравнением, определяющим связь между  $\mu$  и  $b$ . Его можно преобразовать к следующему виду [8]:

$$\sin^2\left(\frac{1}{2}\pi\mu\right) = \Delta(0)\sin^2\left(\frac{1}{2}\pi b\right), \quad (2.11)$$

где  $\Delta(0)$  - значение определителя при  $\mu = 0$ .

Уравнение (2.11) имеет бесконечное множество корней  $\mu_k$ , отличающихся на целое четное число. Каждому корню  $\mu_k$  соответствует свое, линейно-независимое от других,  $k$ -е решение уравнения (2.7). Общее решение уравнения (2.7) при этом можно записать как сумму линейно-независимых решений с произвольными коэффициентами

$$\begin{aligned} V(b, \tau) &= \sum_{-\infty}^{+\infty} A_k \exp[-i(\mu_0 + 2k)\tau] F_k(b, \tau) = \\ &= \sum_{-\infty}^{+\infty} A_k \exp[-i(\mu_0 + 2k)\tau] \sum_{-\infty}^{+\infty} a_{kn}(b) \exp(-i2n\tau), \end{aligned} \quad (2.12)$$

где  $a_{kn}$  - коэффициенты ряда Фурье периодической функции  $F_k$   $k$ -го частного решения.

Поскольку уравнение (2.11) четно относительно  $\mu$ , то, очевидно, что  $-\mu_k$  также будут решениями уравнения (2.11). Заметим, что изменение знака при  $\mu$  эквивалентно изменению знака при  $\tau$ . Таким образом, к решению для  $V(\tau)$  необходимо добавить еще такое же решение со знаком «минус» при  $\tau$ . Решение  $V(-\tau)$  в дальнейшем будем называть обратным, в отличие от прямого решения  $V(+\tau)$ .



Главное, или нулевое, решение  $\mu_0$ , определяемое дисперсионным уравнением (2.11) с точностью до слагаемого, равного целому четному числу, доопределим таким образом, чтобы разность  $|\mu_0 - b|$  была минимальной. При таком выборе главного значения  $\mu_0$  при  $q \rightarrow 0$  одно из решений уравнения (2.6), а именно решение при  $k = 0$  переходит в чисто гармоническое решение с пространственной частотой  $\sqrt{k^2 \varepsilon_1 - \beta^2}$ , что соответствует решению уравнения (2.4) в виде плоской волны с постоянными распространения  $\beta$  и  $\sqrt{k^2 \varepsilon_1 - \beta^2}$  вдоль осей  $Y$  и  $Z$ . Аналогично, обратное решение при  $k = 0$  переходит в волну с постоянными  $\beta$  и  $-\sqrt{k^2 \varepsilon_1 - \beta^2}$ , т.е. в волну, распространяющуюся в противоположном направлении по оси  $Z$ .

## 2.2. Анализ дисперсионного уравнения. Обсуждение физического смысла решений

Наибольшая сложность в анализе дисперсионного уравнения заключается в вычислении значения  $\Delta(0)$ . Рассмотрим один из вариантов его вычисления, позволяющий представить  $\Delta(0)$  в виде ряда по четным степеням  $q$ .

Ограничим бесконечный определитель  $(2N+1)$  столбцами и строками

$$\Delta_N(0) = \begin{vmatrix} 1 & p_{-N} & 0 & 0 & \cdot & \cdot & \cdot \\ p_{-N+1} & 1 & p_{-N+1} & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & p_0 & 1 & p_0 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & p_{N-1} & 1 & p_{N-1} \\ \cdot & \cdot & \cdot & 0 & 0 & p_N & 1 \end{vmatrix},$$

где  $p_n(\mu = 0) = -\frac{2q}{4n^2 - b^2}$ .

Раскрывая определитель, начиная с верхней строки, можно получить следующий результат:

$$\Delta_N(0) = 1 - \sum_{-N}^{N-1} p_n p_{n+1} \left( 1 - \sum_{n+2}^{N-3} p_m p_{m+1} \left( 1 - \sum_{m+2}^{N-5} p_l p_{l+1} \left( 1 - \dots \right. \right. \right.$$

Раскрывая скобки и переходя к пределу при  $N \rightarrow \infty$ , имеем

$$\begin{aligned} \Delta(0) = 1 - \sum_{-\infty}^{+\infty} p_n p_{n+1} + \sum_{-\infty}^{+\infty} p_n p_{n+1} \cdot \sum_{n+2}^{+\infty} p_m p_{m+1} - \\ - \sum_{-\infty}^{+\infty} p_n p_{n+1} \cdot \sum_{n+2}^{+\infty} p_m p_{m+1} \cdot \sum_{m+2}^{+\infty} p_l p_{l+1} + \dots \end{aligned}$$

Учитывая, что  $p_n$  пропорциональны  $2q$ , последнее выражение можно рассматривать как разложение  $\Delta(0)$  по степеням  $(2q)^2$ . Если  $q \ll 1$ , то можно ограничиться учетом только  $(2q)^2$ . В этом случае

$$\Delta(0) = 1 - 4q^2 \sum_{-\infty}^{+\infty} \frac{1}{(4n^2 - b^2)[4(n+1)^2 - b^2]}.$$

Вычисление суммы в последнем выражении дает следующий результат:

$$\sum_{-\infty}^{+\infty} \frac{1}{(4n^2 - b^2)[4(n+1)^2 - b^2]} = -\pi \frac{\operatorname{ctg}\left(\frac{1}{2}\pi b\right)}{4b(1-b^2)}.$$

Следовательно,  $\Delta(0)$  с учетом первого порядка  $(2q)^2$  имеет следующее значение:

$$\Delta(0) = 1 + q^2 \frac{\pi \operatorname{ctg}\left(\frac{1}{2}\pi b\right)}{b(1-b^2)}. \quad (2.13)$$

Учет второго порядка  $(2q)^2$  приводит к следующему результату:

$$\Delta(0) = 1 + q^2 \frac{\pi \operatorname{ctg}\left(\frac{1}{2}\pi b\right)}{b(1-b^2)} + q^4 \frac{\pi \operatorname{ctg}\left(\frac{1}{2}\pi b\right)}{2b^2(1-b^2)} \left[ \pi \operatorname{ctg}(\pi b) - \frac{2\pi \frac{35}{4}b^2 + \frac{15}{4}b^4}{16b(1-b^2)(4-b^2)} \right].$$

В дальнейшем ограничимся анализом дисперсионного уравнения в первом приближении, учитывающим только первый порядок малой величины  $(2q)^2$ .

Подставляя  $\Delta(0)$  в первом приближении из (2.13) в дисперсионное уравнение (2.11), получим

$$\cos(\pi\mu) = \cos(\pi b) - q^2 \frac{\pi \sin(\pi b)}{b(1-b^2)}. \quad (2.14)$$

Из уравнения (2.14), учитывая малость  $q$ , для всех  $b$ , исключая окрестности  $b = 0$  и  $|b|=1$ , можно получить следующее приближенное выражение для  $\mu$ :

$$\mu_k = \pm \left[ b + \frac{q^2}{b(1-b^2)} + 2k \right].$$

Главное значение  $\mu_0$  при  $b \neq 0$  и  $|b| \neq 1$  будет

$$\mu_0 = \pm \left[ b + \frac{q^2}{b(1-b^2)} \right]. \quad (2.15)$$

В окрестностях точек  $b = 0$  и  $|b|=1$ , раскрывая неопределенности  $\frac{\sin(\pi b)}{b}$  и  $\frac{\sin(\pi b)}{1-b^2}$ , имеем

$$\mu_0 = \left[ 1 \mp iq \sqrt{1 - \frac{(b-1)^2}{q^2}} \right] \text{ в окрестности } b=1, \quad (2.16)$$

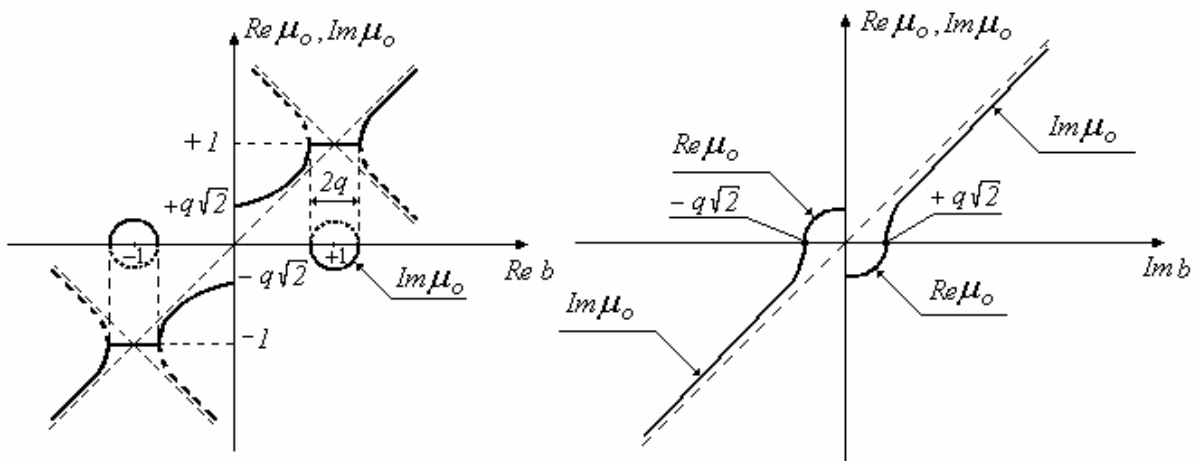
$$\mu_0 = \pm \sqrt{b^2 + q^2} \text{ в окрестности } b=0, \quad (2.17)$$

$$\mu_0 = - \left[ 1 \mp iq \sqrt{1 - \frac{(b-1)^2}{q^2}} \right] \text{ в окрестности } b = -1. \quad (2.18)$$

Знаки перед квадратными корнями в выражениях (2.16) – (2.18) следует выбирать таким образом, чтобы разность  $|\operatorname{Re}\mu_0 - b|$  была минимальной. В окрестности  $b = 1$  из этого условия следует, что при  $|b-1| > q$  знак в (2.16) должен быть «минус», если  $b > 1$ , и «плюс», если  $b < 1$ . Можно показать, что противоположные знаки соответствуют решениям дисперсионного уравнения  $-\mu_{-1} = -(\mu_0 - 2)$ .

Представляет интерес поведение  $\mu_0$  в окрестности  $b = 0$ . При  $q \rightarrow 0$  точка  $b = 0$  соответствует границе между гармоническими и экспоненциально убывающими решениями. При  $q \neq 0$ , как это видно из (2.17), границей между гармоническими и экспоненциально убывающими решениями будет уже не  $b = 0$ , а  $b = \pm iq\sqrt{2}$ .

Зависимости  $\mu_0$  от  $b = \frac{2}{K} \sqrt{k^2 \varepsilon_1 - \beta^2}$  для действительных значений  $\beta$ , рассчитанные по формулам (2.15) – (2.18), представлены на рис. 7.



**Рис. 7. Зависимость  $\operatorname{Re}\mu_0$  и  $\operatorname{Im}\mu_0$  от действительных и мнимых значений  $b$**

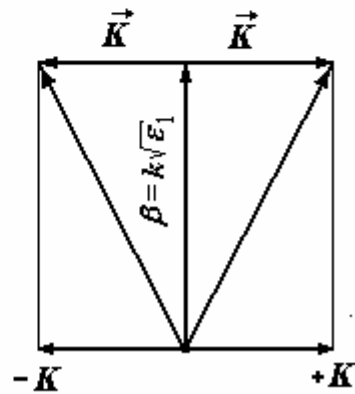
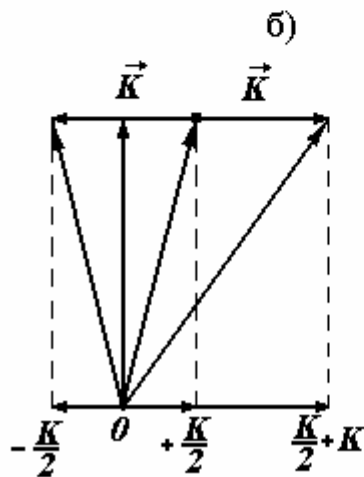
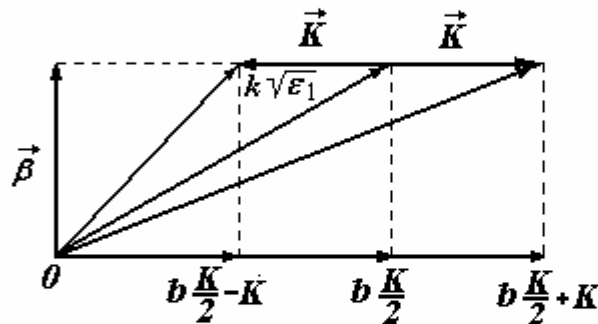
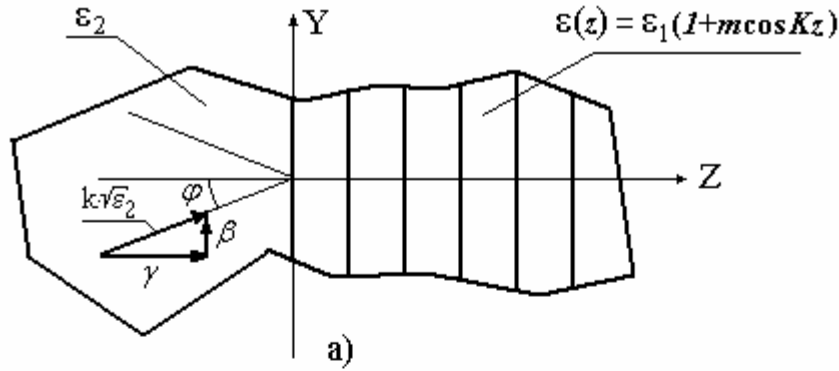
Остановимся на физическом смысле полученных решений. Для определенности допустим, что поле внутри модулированной среды возбуждается на плоской границе модулированного и немодулированного диэлектриков плоской волной, падающей из немодулированного диэлектрика под углом  $\phi$ . Допустим, что  $\varepsilon_2 > \varepsilon_1$ . Из условий непрерывности тангенциальных компонент поля следует, что решения вдоль оси  $Y$  справа и слева от  $Z = 0$  должны быть одинаковы. В рассматриваемом случае это будет гармоническое решение вида  $\exp(-i\beta y)$ , где  $\beta = k\sqrt{\varepsilon_2}\sin\phi$ . Учитывая однородность модулированной среды вдоль оси  $Y$ , можно утверждать, что такого вида решение вдоль оси  $Y$  будет сохраняться в любом другом сечении  $Z$ .

Качественно картину распространения волн в модулированной среде рассмотрим методом последовательных приближений. В нулевом приближении положим, что  $q = m \frac{k^2 \varepsilon_1}{4K^2} = 0$ , т.е. в нулевом приближении при  $Z > 0$  среда однородна. Решение для поля в этом случае представляет собой плоскую волну с постоянными распространения  $\beta$  и  $b \frac{K}{2} = \sqrt{k^2 \varepsilon_1 - \beta^2}$  вдоль осей  $Y$  и  $Z$ , соответственно. При  $\beta^2 > k^2 \varepsilon_1$  величина  $b$  является мнимой, что соответствует полному внутреннему отражению на границе  $Z = 0$ . При этом для  $Z > 0$  решение вдоль  $Z$  является экспоненциально убывающим.

В первом приближении будем считать, что рассеяние плоской волны на гармонической неоднородности  $\varepsilon$  является очень слабым, и поэтому можно не учитывать изменение амплитуды рассеиваемой волны, амплитуда которой определяется нулевым приближением. Гармоническую неоднородность  $\varepsilon$  можно рассматривать как две встречных плоских волны  $\cos Kz = \frac{1}{2} [\exp(+iKz) + \exp(-iKz)]$  с волновыми векторами  $\pm K$  и

нулевыми фазовыми скоростями. В результате рассеяния электромагнитной волны на этих двух волнах неоднородности  $\epsilon$  возникают две вынужденные волны с компонентами волновых векторов  $\left(\beta, b\frac{K}{2} + K\right)$  и  $\left(\beta, b\frac{K}{2} - K\right)$ . Амплитуды этих вынужденных волн, очевидно, должны быть пропорциональны  $q$ . Таким образом, полное поле в первом приближении представляет собой плоскую рассеиваемую волну, которую в дальнейшем будем называть нулевой пространственной гармоникой, и две рассеянные плоские волны, которые мы будем называть плюс и минус первыми пространственными гармониками полного поля в модулированной среде (рис. 8).

На рис. 8а приведено схематическое изображение возбуждения модулированной среды плоской волной, падающей из немодулированного диэлектрика. На рис. 8б изображены направления распространения нулевой и плюс первой гармоник при  $b \neq 0$  и  $b \neq 1$ . На рис. 8в представлены направления распространения нулевой и плюс первой гармоник при  $b = 1$ , а на рис 8г - направления распространения нулевой и плюс первой гармоник при  $b = 0$ .



**Рис. 8. Геометрия волновых векторов решетки.**

В следующем, втором приближении необходимо учесть вторичное рассеяние, т.е. рассеяние  $+1$  и  $-1$  гармоник на неоднородности  $\varepsilon$ . При вторичном рассеянии возникают плоские волны с постоянными распространения  $\left(\beta, b\frac{K}{2} + 2K\right)$  и  $\left(\beta, b\frac{K}{2} - 2K\right)$ , т.е. появляются  $+2$  и  $-2$  пространственные гармоники с амплитудами, пропорциональными  $q^2$ , и,

кроме того, нулевая пространственная гармоника  $\left(\beta, b\frac{K}{2}\right)$ . Наличие нулевой пространственной гармоники во втором приближении соответствует учету обратного влияния рассеянного поля на рассеиваемое поле. Это влияние приводит к сдвигу постоянной распространения нулевой гармоники относительно значения, определяемого в нулевом приближении, на величину, пропорциональную  $q^2$ . В нашем решении это имеет место для всех  $b$ , за исключением точек  $b = 0, |b| = 1$  (см. выражение (2.15) для  $\mu_0$ ).

В окрестности точки  $b = 1$  значение волнового вектора -1 гармоники по абсолютной величине оказывается равным постоянной распространения нулевой гармоники, а направление его проекции на ось  $Z$  обратно направлению волнового вектора нулевой гармоники. Эта точка соответствует условию Брэгга, при котором поля, отраженные от каждого периода неоднородностей, складываются синфазно. В этом случае амплитуда -1 гармоники (отраженной волны) оказывается максимально возможной и равной амплитуде рассеиваемой (отражаемой) нулевой гармоники.

Решение для  $\mu_0$  в окрестности этой точки оказывается комплексным (см. (2.16)). Мнимая часть  $\mu_0$  соответствует экспоненциальному убыванию амплитуды поля нулевой пространственной гармоники, которое можно объяснить качественно тем, что при отражении от каждого периода неоднородности часть мощности падающей волны (нулевой гармоники) отражается (рассеивается) в минус первую гармонику. Вторично рассеянное поле на том же периоде оказывается точно в противофазе с полем падающей волны (при  $b = 1$ ). Следовательно, при обратном рассеянии нет возврата энергии из отраженной волны в падающую, как это имеет место при  $b \neq 1$ , и амплитуда падающей волны на каждом периоде



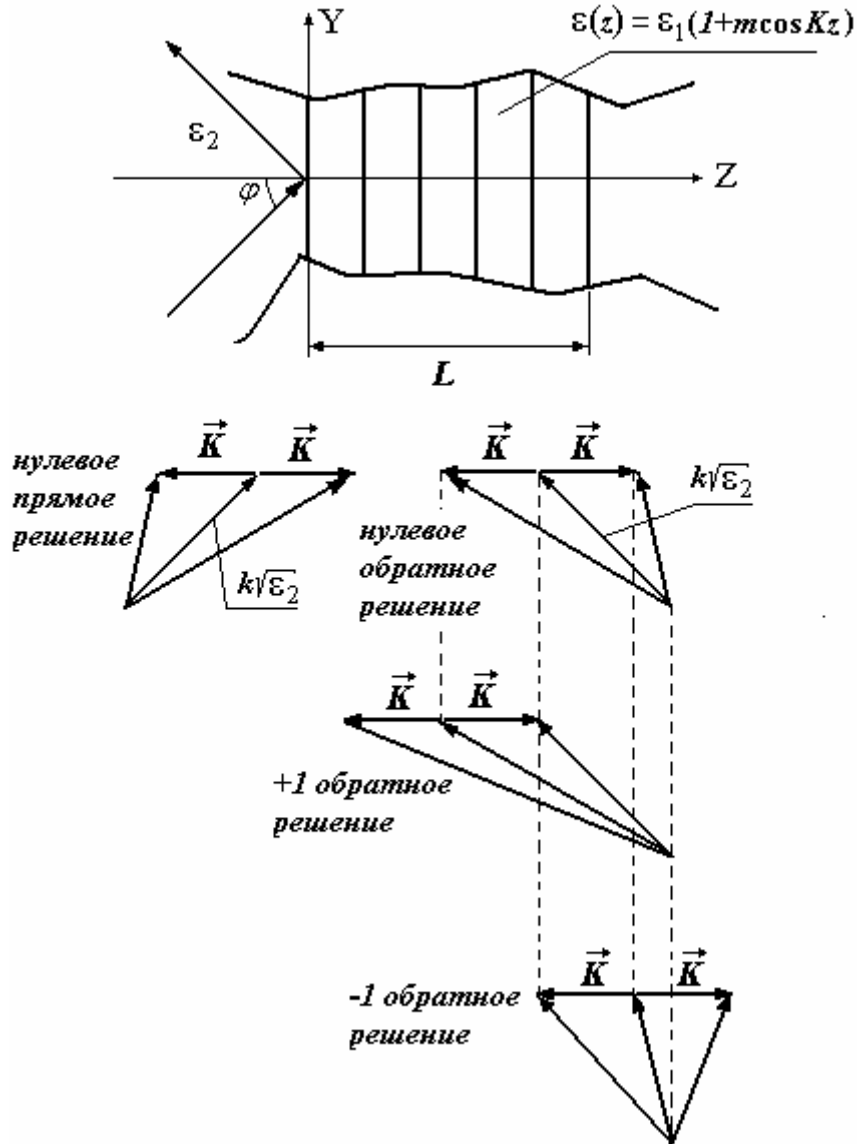
неоднородности уменьшается на величину, пропорциональную  $q$ . При небольшой расстройке от  $b = 1$  экспоненциальная зависимость сохраняется в области, ширина которой равна  $q$  (см. рис. 7). Полоса  $\Delta b$  в окрестности  $b = 1$ , в которой  $\text{Im}\mu_0 \neq 0$ , является *полосой брэгговского отражения* первого порядка. Поскольку  $b\frac{K}{2} = \sqrt{k^2\varepsilon_1 - \beta^2} = \frac{\omega}{c}\sqrt{\varepsilon_2}\cos\phi$  зависит от частоты и угла падения, то, следовательно, существуют *частотная и угловая полосы брэгговского отражения*, пропорциональные глубине модуляции диэлектрической проницаемости. Заметим, что действительная часть постоянной распространения нулевой гармоники внутри полосы брэгговского отражения остается постоянной и равной  $\frac{K}{2}$ .

В среде с гармонической модуляцией диэлектрической проницаемости кроме полосы брэгговского отражения первого порядка при  $b = 1$  существуют полосы отражения второго (в окрестности  $b = 2$ ), третьего (при  $b = 3$ ) и т.д. порядков. Мнимые части постоянных распространения и ширина полосы отражения в окрестностях этих точек пропорциональны  $q^n$ , где  $n$  – порядок брэгговского отражения. Дисперсионное уравнение (2.14), учитывающее только первый порядок малой величины  $(2q)^2$ , описывает брэгговское отражение только первого порядка. Для описания более высокого порядка в дисперсионном уравнении необходимо учесть более высокие порядки  $(2q)^2$  при вычислении  $\Delta(\mu = 0)$ .

Приведенная выше физическая картина объясняет возбуждение в модулированной среде одного из решений волнового уравнения, которое описывается набором пространственных гармоник. Это решение при  $q \rightarrow 0$  переходит в плоскую волну, так как при  $q \rightarrow 0$  амплитуды всех пространственных гармоник, кроме амплитуды нулевой, стремятся к нулю,

а постоянная распространения нулевой гармоники стремится к постоянной распространения плоской волны в однородном диэлектрике. Это решение называется прямым нулевым. Возбуждение нулевого обратного и решений с другими номерами (прямых и обратных) можно пояснить на примере модулированной среды, ограниченной по оси  $Z$  длиной  $L$  (рис. 9).

Пусть слева на границу модулированной среды при  $Z = 0$  под углом  $\phi$  падает плоская волна. Эта волна возбуждает в модулированной среде спектр пространственных гармоник (нулевое решение), каждая из которых распространяется со своей постоянной распространения в положительном направлении оси  $Z$ . На другой границе при  $Z = L$  происходит частичное отражение, при котором компоненты спектра нулевого решения являются источниками обратных решений – нулевого,  $\pm 1$  и т.д. При частичном отражении обратных решений от границы при  $Z = 0$  кроме нулевого решения возбуждаются  $\pm 1$ ,  $\pm 2$  прямые решения и т.д. Амплитуды решений определяются из граничных условий при  $Z = 0$  и  $Z = L$ , а соотношения между амплитудами гармоник в каждом решении — значениями  $b$  и  $q$ , т.е. параметрами модулированной среды.



**Рис.9. Ограниченная модулированная среда и направление волновых векторов пространственных гармоник в решениях с различными номерами**

### **2.3. Амплитуды пространственных гармоник**

Решение задачи о распространении электромагнитных волн в гармонически модулированной среде, как это показано выше, сводится к уравнению Матве, частными решениями которого являются функции:

$$V_k(\beta, +z) = V_k(\beta, +\tau) = \exp[-i(\mu_0 + 2k)\tau] \cdot F_k(\beta, +\tau),$$

$$V_k(\beta, -z) = V_k(\beta, -\tau) = \exp[+i(\mu_0 + 2k)\tau] \cdot F_k(\beta, -\tau),$$

где  $F_k$  - периодические функции с периодом  $\pi$ , которые могут быть представлены рядом Фурье:

$$F_k(\beta, \tau) = \sum a_{kn}(\beta) \cdot \exp(-i2n\tau),$$

где  $a_{kn}(\beta)$  - амплитуды спектра Фурье, в рассматриваемой задаче имеющие смысл амплитуд пространственных гармоник  $k$ -го решения.

Значения  $a_{kn}$  вычисляются из однородной системы (2.9) при  $\mu_k = \mu_0 + 2k$  и  $b$ , определяемом дисперсионным уравнением. Из системы (2.9)  $a_{kn}$  определяются с точностью до постоянного множителя. Если постоянный множитель выбрать так, чтобы  $a_{kk} = 1$ , то  $a_{kn}$  при  $k \neq n$  представляет собой отношение  $\frac{a_{kn}}{a_{kk}}$ . Полагая в (2.9) значения  $a_{kk} = 1$  и разыскивая в виде разложения в ряд по степеням  $q$ , для  $a_{kn}$ , учитывая слагаемые с наименьшей степенью  $q$ , имеем:

$$a_{k, k \pm S} = (-1)^S \prod_{r=\pm 1}^{\pm S} p_{k+r}, \quad (2.19)$$

где  $S = 1, 2, 3, \dots$ . При  $S = 0$   $a_{kk} = 1$ .

Вычислим  $a_{kn}$  для частного случая, когда  $b$  считается заданным. Для этого подставим в (2.19) значения  $b$  и  $\mu_k$  из решения дисперсионного уравнения.

При всех  $b$ , исключая окрестности  $|b|=1$  и  $b=0$ , связь между  $b$  и  $\mu_k$  определяется выражением (2.15). Подставляя (2.15) в (2.19), имеем:

$$a_{k, k \pm S} = (-2q)^S \prod_{r=\pm 1}^{\pm S} p_{k+r} = (-q)^S \prod_{r=\pm 1}^{\pm S} \frac{1}{2r(r+b)}. \quad (2.20)$$

В окрестности  $b = 1$ , учитывая (2.16)

$$a_{k,k\pm S} = (-2q)^S \prod_{r=\pm 1}^{\pm S} \frac{1}{2r(r+1) - (b-1) \mp iq(1+2r) \sqrt{1 - \left(\frac{b-1}{q}\right)^2}}. \quad (2.21)$$

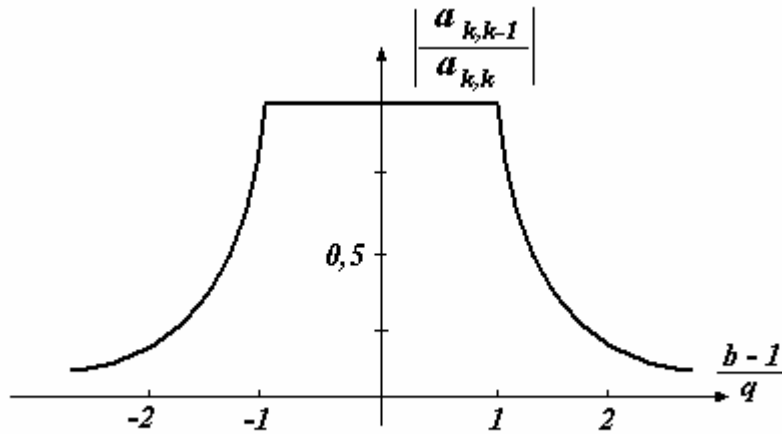
В окрестности  $b = 0$

$$a_{k,k\pm S} = (-2q)^S \prod_{r=\pm 1}^{\pm S} \frac{1}{2r^2}. \quad (2.22)$$

Как видно из выражений (2.20)-(2.22), амплитуды пространственных гармоник  $k$ -го решения убывают с ростом  $S$  – номера гармоники пропорционально  $q^S$ , и при  $q \ll 1$  малы ( $a_{k,k\pm S} \ll a_{kk}$ ). Исключением является лишь окрестность  $b = 1$ , в которой амплитуда  $(k-1)$ -ой гармоники  $k$ -го решения равна амплитуде  $k$ -ой гармоники того же решения. На самом деле, из (2.21) для  $a_{k,k-1}$  можно получить:

$$a_{k,k-1} = \frac{1}{\frac{b-1}{q} \pm \sqrt{1 - \left(\frac{b-1}{q}\right)^2}}. \quad (2.23)$$

Здесь знак плюс перед квадратным корнем соответствует  $b < 1$ , а знак минус -  $b > 1$ . Из этого выражения видно, что в полосе  $|b-1| \leq q$  значение  $|a_{k,k-1}| = 1$ , а при  $|b-1| > q$  быстро убывает с ростом  $|b-1|$  (рис. 10). Полоса  $|b-1| = q$  является полосой брэгговского отражения. В пределах этой полосы амплитуда  $(k-1)$ -ой и  $k$ -ой гармоник  $k$ -го решения равны по абсолютной величине.



**Рис. 10. Зависимость отношения амплитуд  $(k-1)$ -ой и  $k$ -ой гармоник от расстройки относительно брэгговского условия**

Зависимость  $|a_{k,k-1}|$  от расстройки  $(b-1)$  в окрестности брэгговского условия  $b=1$  получена в предположении, что величина  $b$  является заданной. Как уже отмечалось, дисперсионное уравнение определяет лишь связь между  $b$  и  $\mu_0$ , и вопрос о том, какую из этих величин можно считать заданной, может быть решен с учетом условий возбуждения поля в периодически модулированной среде. Одним из возможных вариантов возбуждения поля, при котором заданной можно считать величину  $b$ , является возбуждение периодически модулированной среды на плоской границе, нормальной вектору решетки (см. рис. 8).

Из анализа выражений для амплитуд спектра Фурье любого  $k$ -го решения можно заключить, что при малой глубине модуляции диэлектрической проницаемости ( $q \ll 1$ ) и при достаточно большой расстройке от условия Брэгга  $|b-1| \gg q$ ,  $|\mu_0 - 1| \gg q$  решения для электромагнитного поля в модулированной среде являются почти гармоническими. В окрестности брэгговского условия, где амплитуды  $k$ -ой и  $(k-1)$ -ой гармоник  $k$ -го решения сравнимы по величине, а амплитуды остальных  $\ll 1$ , решения являются бигармоническими.

Если в решении для электромагнитного поля ограничиться учетом только нулевых решений (прямого и обратного), а в них учесть только нулевые и минус первые гармоники, то в результате получим решение, которое совпадает с решением, получаемом методом связанных волн в двухволновом приближении.

В качестве заключения приведем выражения для электромагнитного поля в бигармоническом приближении для окрестности  $b=1$ . Если заданной величиной является  $b$ , то из (2.5) можно записать:

$$E_x(y, z) = [V_0(\beta, +z) + V_0(\beta, -z)] \cdot \exp(-i\beta y),$$

где

$$V_0(\beta, +z) = V(\beta, \tau) = A_0 (1 + a_{0,-1} \cdot \exp(+i2\tau)) \cdot \exp(-i\mu_0 \tau),$$

$$V_0(\beta, -z) = V(\beta, \tau) = B_0 (1 + a_{0,-1} \cdot \exp(-i2\tau)) \cdot \exp(+i\mu_0 \tau),$$

$$b = \frac{2}{K} \sqrt{k^2 \varepsilon_1 - \beta^2}, \quad \tau = \frac{K}{2} z,$$

$$(\mu_0 - 1) = \mp q \sqrt{1 - \left(\frac{b-1}{q}\right)^2}, \quad a_{0,-1} = \frac{1}{\frac{b-1}{q} \mp i \sqrt{1 - \left(\frac{b-1}{q}\right)^2}}.$$

Подставляя  $V(b, \pm\tau)$  в выражение для поля и опуская индекс 0, имеем

$$\begin{aligned} E_x(y, z) = & \\ = & \left\{ A \cdot \exp\left(-q \frac{Kz}{2} \sqrt{1 - \theta^2}\right) \left[ \exp\left(-i \frac{Kz}{2}\right) + \left(\theta + i \sqrt{1 - \theta^2}\right) \cdot \exp\left(+i \frac{Kz}{2}\right) \right] + \right. \\ & \left. + B \cdot \exp\left(+q \frac{Kz}{2} \sqrt{1 - \theta^2}\right) \left[ \exp\left(-i \frac{Kz}{2}\right) + \left(\theta - i \sqrt{1 - \theta^2}\right) \cdot \exp\left(+i \frac{Kz}{2}\right) \right] \right\} e^{-i\beta y}. \end{aligned} \quad (2.24)$$

Здесь

$$\theta = \frac{b-1}{q} = \frac{\sqrt{k^2 \varepsilon_1 - \beta^2} - \frac{K}{2}}{q \frac{K}{2}}, \quad q = m \frac{k^2 \varepsilon_1}{4K^2}.$$

Как видно, решение (2.24) при  $\theta < 1$  представляет собой сумму двух решений в виде стоячих волн, одна из которых убывает, а другая нарастает по амплитуде вдоль оси  $Z$  по экспоненциальному закону с постоянной равной  $q \frac{K}{2} \sqrt{1 - \theta^2}$ . При  $\theta > 1$  полное решение представляет собой биение с пространственной частотой  $q \frac{K}{2} \sqrt{1 - \theta^2}$ .

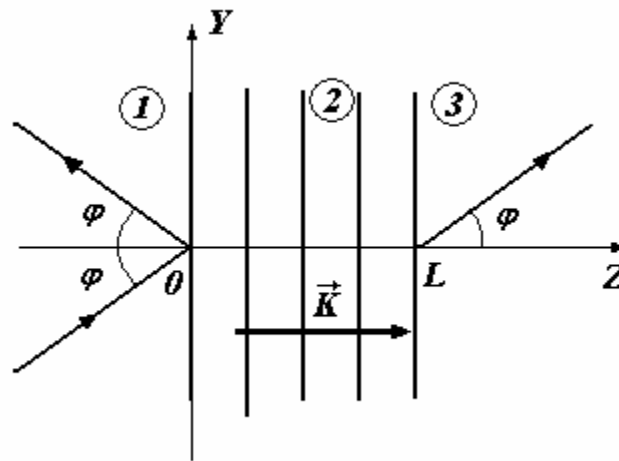
#### **2.4. Диэлектрическое зеркало, основанное на эффекте брэгговского отражения**

В предыдущих параграфах представлены решения для электромагнитного поля в среде с одномерной модуляцией. Показано, что при малой глубине модуляции в окрестности брэгговского условия решение с достаточной точностью может быть представлено двумя пространственными гармониками. Полученные результаты мы будем использовать в этом параграфе для анализа свойств брэгговского отражателя, представляющего собой участок модулированной среды с одномерной модуляцией (рис. 11), ограниченный параллельными плоскостями.

Результаты следующего ниже анализа справедливы не только для среды с одномерной модуляцией диэлектрической проницаемости. Они могут быть распространены на брэгговские отражатели в открытых волноводах для электромагнитных волн, а также на брэгговские отражатели поверхностных акустических волн. Возможность распространения результатов анализа для среды с одномерной модуляцией



на другие устройства основана на том, что в окрестности брэгговской частоты во всех устройствах распределение поля внутри модулированной области определяется приближением двух встречных волн, постоянные распространения которых однозначно задаются значением постоянной решетки и шириной брэгговской полосы.



**Рис. 11. Брэгговское зеркало**

Ниже рассматривается наиболее простая конфигурация брэгговского отражателя, в которой плоскости, ограничивающие модулированную среду, нормальны вектору решетки  $\vec{K}$  (рис. 11). В дальнейшем такой отражатель будем называть брэгговским зеркалом.

Для упрощения анализа предположим, что среднее значение диэлектрической проницаемости в области с модуляцией равно диэлектрической проницаемости в областях без модуляции.

Задачей следующего ниже анализа является отыскание зависимостей коэффициента отражения и пропускания брэгговского зеркала от угла падения и частоты падающей плоской монохроматической волны.

Решение поставленной задачи строится следующим образом. В немодулированной области 1 при  $Z < 0$  поле записывается в виде падающей и отраженной плоских волн, а в области 3 при  $Z > L$  – в виде плоской прошедшей волны. В модулированной области 2 ( $0 < Z < L$ ) поле

записывается в виде прямого и обратного решений, каждое из которых является суммой двух пространственных гармоник (см. (2.24)).  
 Неизвестные амплитуды отраженной и прошедшей волн, а также амплитуды прямого и обратного решений в модулированной области находятся из условий непрерывности компонент поля, тангенциальных границам раздела областей.

Отношение амплитуд отраженной и падающей волн равно коэффициенту отражения, а отношение амплитуд прошедшей волны падающей равно коэффициенту пропускания брэгговского зеркала.

Итак, запишем поле в области 1 (при  $Z < 0$ ):

$$E_{x1} = [A_{nad} \exp(-i\gamma z) + A_{omp} \exp(+i\gamma z)] \cdot \exp(-i\beta y), \quad (2.25)$$

где  $A_{nad}$  и  $A_{omp}$  - амплитуды падающей и отраженной волн, а  $\beta$  и  $\gamma$  - проекции волновых векторов  $\vec{k}\sqrt{\epsilon}$  падающей и отраженной волн на оси  $Y$  и  $Z$  соответственно.

Поле в области 2 (при  $0 < Z < L$ ), пользуясь (24), можно записать следующим образом:

$$E_{x2} = \left\{ A \cdot \exp(-q \frac{Kz}{2} \sqrt{1-\theta^2}) \left[ \exp(-i \frac{Kz}{2}) + (\theta + i\sqrt{1-\theta^2}) \cdot \exp(+i \frac{Kz}{2}) \right] + \right. \\ \left. + B \cdot \exp(+q \frac{Kz}{2} \sqrt{1-\theta^2}) \left[ \exp(-i \frac{Kz}{2}) + (\theta - i\sqrt{1-\theta^2}) \cdot \exp(+i \frac{Kz}{2}) \right] \right\} e^{-i\beta y}, \quad (2.26)$$

где  $A$  и  $B$  - амплитуды прямого и обратного решений, а  $\theta = \frac{\sqrt{k^2 \epsilon_1 - \beta^2} - \frac{K}{2}}{q \frac{K}{2}}$

- относительная расстройка от брэгговского условия.

Поле в области 3 (при  $Z > L$ )

$$E_{x3} = A_{прои} \cdot \exp[-i\gamma(z-L)] \cdot \exp(-i\beta y), \quad (2.27)$$

где  $A_{\text{прои}}$  - амплитуда прошедшей волны в точке  $y = 0, z = L$ .

Удовлетворяя граничным условиям:

$$E_{x1} = E_{x2}, \quad \frac{\partial E_{x1}}{\partial z} = \frac{\partial E_{x2}}{\partial z} \quad \text{при} \quad z = 0$$

$$E_{x2} = E_{x3}, \quad \frac{\partial E_{x2}}{\partial z} = \frac{\partial E_{x3}}{\partial z} \quad \text{при} \quad z = L,$$

получим следующую систему алгебраических уравнений относительно неизвестных амплитуд  $A, B, A_{\text{отр}}$  и  $A_{\text{прои}}$ :

$$\left\{ \begin{array}{l} A_{\text{над}} + A_{\text{отр}} = A \left[ 1 + \left( \theta + i\sqrt{1-\theta^2} \right) \right] + B \left[ 1 + \left( \theta - i\sqrt{1-\theta^2} \right) \right] \\ A_{\text{над}} - A_{\text{отр}} \approx A \left[ 1 - \left( \theta + i\sqrt{1-\theta^2} \right) \right] + B \left[ 1 - \left( \theta - i\sqrt{1-\theta^2} \right) \right] \\ A_{\text{прои}} = A \exp\left(-q \frac{KL}{2}\right) \left[ \exp\left(-i \frac{KL}{2}\right) + \left( \theta + i\sqrt{1-\theta^2} \right) \exp\left(+i \frac{KL}{2}\right) \right] + \\ \quad + B \exp\left(+q \frac{KL}{2}\right) \left[ \exp\left(-i \frac{KL}{2}\right) + \left( \theta - i\sqrt{1-\theta^2} \right) \exp\left(+i \frac{KL}{2}\right) \right] \\ A_{\text{прои}} \approx A \exp\left(-q \frac{KL}{2}\right) \left[ \exp\left(-i \frac{KL}{2}\right) - \left( \theta + i\sqrt{1-\theta^2} \right) \exp\left(+i \frac{KL}{2}\right) \right] + \\ \quad + B \exp\left(+q \frac{KL}{2}\right) \left[ \exp\left(-i \frac{KL}{2}\right) - \left( \theta - i\sqrt{1-\theta^2} \right) \exp\left(+i \frac{KL}{2}\right) \right]. \end{array} \right. \quad (2.28)$$

Во втором и четвертом уравнениях системы (2.28) использовано следующее приближение: множитель при  $A$  и  $B$ , равный  $\frac{K}{2\gamma}$ , полагается равным единице. При небольших расстройках от брэгговского условия  $|2\gamma - K| \ll K$  это приближение вполне оправданно. Физически оно соответствует пренебрежению отражением на границе между модулированной и немодулированной областями.

Вычисление  $\Gamma = \frac{A_{\text{отр}}}{A_{\text{над}}}$  и  $T = \frac{A_{\text{прои}}}{A_{\text{над}}}$  из системы (2.28) приводит к

следующим результатам:

$$\Gamma = i \frac{\operatorname{sh}\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right)}{\sqrt{1-\theta^2} \operatorname{ch}\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right) + i \operatorname{sh}\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right)}, \quad (2.29)$$

$$T = \frac{\sqrt{1-\theta^2} \exp\left(-i \frac{KL}{2}\right)}{\sqrt{1-\theta^2} \operatorname{ch}\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right) + i \operatorname{sh}\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right)}. \quad (2.30)$$

Для амплитуд прямого и обратного решений в модулированной среде из (2.28) получаем следующие выражения:

$$A = A_{nad} \frac{2i\left(\theta - i\sqrt{1-\theta^2}\right) \exp\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right)}{\sqrt{1-\theta^2} \operatorname{ch}\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right) + i \operatorname{sh}\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right)}, \quad (2.31)$$

$$B = -A_{nad} \frac{2i\left(\theta + i\sqrt{1-\theta^2}\right) \exp\left(-q \frac{KL}{2} \sqrt{1-\theta^2}\right)}{\sqrt{1-\theta^2} \operatorname{ch}\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right) + i \operatorname{sh}\left(q \frac{KL}{2} \sqrt{1-\theta^2}\right)}. \quad (2.32)$$

Легко убедиться, что сумма коэффициентов отражения и пропускания по мощности  $\Gamma \cdot \Gamma^* + T \cdot T^* = 1$ .

Отметим основные особенности зависимостей  $\Gamma$  и  $T$  от относительной расстройки  $\theta$ .

В центре брэгговской полосы ( $\theta = 0$ ):

$$\Gamma = i \cdot \operatorname{th}\left(q \frac{KL}{2}\right), \quad T = \frac{\exp\left(-i \frac{KL}{2}\right)}{\operatorname{ch}\left(q \frac{KL}{2}\right)}.$$

На краю брэгговской полосы ( $\theta = \pm 1$ ):

$$\Gamma = i \frac{qKL}{2} \left( 1 + i \frac{qKL}{2} \right)^{-1}, \quad T = \exp\left(-i \frac{KL}{2}\right) \left( 1 + i \frac{qKL}{2} \right)^{-1}.$$

В точках  $q \frac{KL}{2} \sqrt{1 - \theta^2} = i\pi \cdot n$  ( $n = 1, 2, 3, \dots$ ) коэффициент отражения  $\Gamma_n = 0$ , а  $|T_n| = 1$ .

В точках  $q \frac{KL}{2} \sqrt{1 - \theta_m^2} = i \frac{1}{2} (2m + 1) \pi$  ( $m = 1, 2, 3, \dots$ ) коэффициент отражения принимает максимальные значения, а коэффициент пропускания – минимальные.

$$\Gamma_m = \frac{1}{\theta_m}, \quad T_m = \frac{\sqrt{\theta_m^2 - 1}}{\theta_m} \exp\left(-i \frac{KL}{2}\right).$$

Зависимость коэффициента отражения от  $\theta$  при  $\theta \gg 1$  асимптотически стремится к зависимости

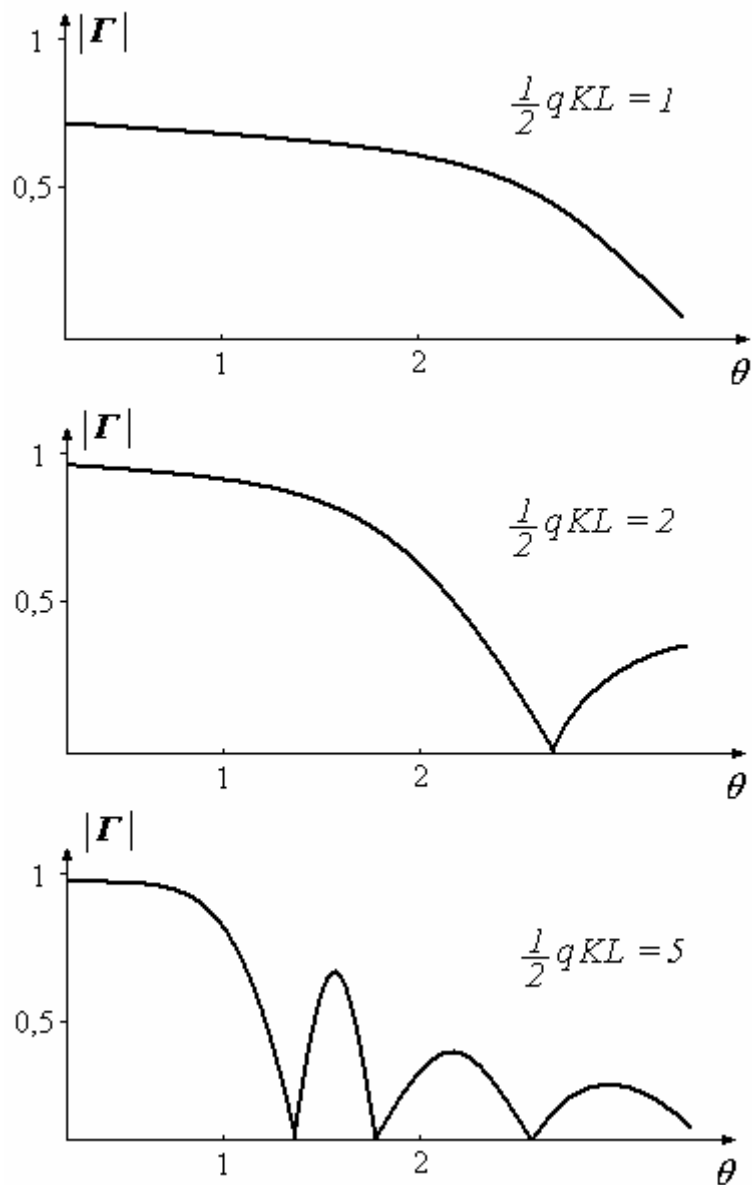
$$\Gamma = i \frac{qKL}{2} \exp\left(iq \frac{KL}{2} \theta\right) \frac{\sin\left(q \frac{KL}{2} \theta\right)}{q \frac{KL}{2} \theta}.$$

Зависимости  $|\Gamma|$  и  $|T|$  от относительной расстройки  $\theta$  для нескольких значений  $\left(\frac{qKL}{2}\right)$  приведены на рис. 12. Из этих графиков и из выражений для  $\Gamma$  и  $T$  видно, что в пределах полосы брэгговского отражения  $|\theta| \leq 1$  модуль коэффициента отражения может быть получен сколь угодно близким к единице за счет увеличения глубины модуляции или посредством увеличения длины зеркала. При этом увеличение коэффициента отражения за счет увеличения глубины модуляции приводит к увеличению полосы брэгговского отражения, в то время как увеличение длины зеркала не изменяет величину полосы отражения.

Используя длинные зеркала с малой глубиной модуляции, возможно получение большого коэффициента отражения в сколь угодно узкой полосе.

Колебания коэффициента отражения за пределами брэгговского отражения ( $|\theta| > 1$ ) можно объяснить следующим образом. Поле в модулированной области при  $|\theta| > 1$ , описываемое суммой двух пространственных гармоник, представляет собой биения с пространственной частотой, равной разности постоянных распространения указанных гармоник -  $qK\sqrt{\theta^2 - 1}$ . При изменении частоты или угла падения период биений изменяется. Если на длине зеркала укладывается нечетное число четвертей биений, то коэффициент отражения принимает максимальное значение. Ситуация здесь аналогична той, которая имеет место при отражении от плоскопараллельной пластинки с толщиной, кратной нечетному числу четвертей длин волн, когда отражения от краев пластинки складываются в фазе. Поскольку амплитуда биений тем больше, чем ближе  $\theta_m$  к единице, то первый максимум коэффициента отражения за пределами полосы оказывается наибольшим. При больших  $\theta$  ( $|\theta_m| \gg 1$ ) значения  $|r_m|$  убывают пропорционально  $\theta^{-1}$ .

Значения  $\theta_n$  соответствуют таким значениям частоты или угла падения падающей волны, при которых на длине зеркала укладывается целое число полупериодов биений. При этом коэффициент отражения оказывается равным нулю.



**Рис. 12. Зависимость коэффициента отражения брэгговского зеркала**

**от расстройки при различных значениях параметра  $\frac{qKL}{2}$**

Следует обратить внимание на распределение поля при расстройках  $\theta_n$  ( $q \frac{KL}{2} \sqrt{1 - \theta_n^2} = i\pi \cdot n$ ). Подставляя  $\theta_n$  в выражения для амплитуды поля в модулированной области (2.31) и (2.32), имеем

$$A = -A_{nad} \frac{\theta_n - i\sqrt{1 - \theta_n^2}}{2i\sqrt{1 - \theta_n^2}}, \quad B = A_{nad} \frac{\theta_n + i\sqrt{1 - \theta_n^2}}{2i\sqrt{1 - \theta_n^2}}.$$

Подставляя эти выражения в выражение (2.26) для поля в модулированной области при расстройках  $\theta = \theta_n$ , имеем

$$E_{x2} = (-1)^n A_{nao} \left\{ iq \frac{KL}{n\pi} \cdot \sin\left(n\pi \frac{z}{L}\right) \cdot \cos\left(\frac{K}{2}z\right) + \left[ \cos\left(n\pi \frac{z}{L}\right) - i \left( q \frac{KL}{2n\pi} - \sqrt{1 - \left( q \frac{KL}{2n\pi} \right)^2} \right) \sin\left(n\pi \frac{z}{L}\right) \right] \cdot \exp\left(-i \frac{Kz}{2}\right) \right\} \quad (2.33)$$

Первое слагаемое в фигурных скобках (2.33) представляет собой стоячую волну, амплитуда которой изменяется вдоль  $z$  по гармоническому закону. При этом амплитуда стоячей волны равна нулю на краях зеркала. Второе слагаемое описывает бегущую волну с амплитудой, медленно меняющейся вдоль  $z$ . Из приведенного выражения видно, что в пучностях стоячей волны амплитуда поля в модулированной области всегда превышает амплитуду падающей волны. Особенно четко это заметно при  $q \frac{KL}{n\pi} \gg 1$ . В этом случае максимум амплитуды бегущей волны приблизительно равен амплитуде падающей волны, а пучности стоячей волны в  $q \frac{KL}{n\pi}$  раз больше. Это соответствует значительной запасенной энергии внутри модулированной среды. При этом запасенная энергия при  $n = 1$  оказывается наибольшей. Значительное накопление энергии в модулированной области при  $\theta = \theta_n$  позволяет рассматривать значения  $\theta_n$  как резонансные, а участок модулированной среды — как проходной резонатор. В точках резонанса  $\theta = \theta_n$  коэффициент пропускания такого резонатора, как и любого другого проходного резонатора без потерь, равен единице.



## **Тема 3. Метод Тихонравова синтеза многослойного зеркала**

В работе [9] на основе идей регуляризации дается корректная постановка задачи синтеза многослойных оптических систем с заданными коэффициентами пропускания. Приводятся примеры синтеза некоторых практически важных систем.

Задачи синтеза различных физических систем привлекают в последнее время все большее внимание исследователей. Новые возможности для изучения и решения этих задач дает принцип регуляризации [10-12].

К этому же классу относится задача о синтезе оптических систем с заданными характеристиками (энергетическими коэффициентами пропускания и отражения). Такие системы могут быть получены путем создания многослойных покрытий, наносимых в виде тонких пленок на поверхность «подложки» (см., например, [13]). Достаточно полно были исследованы системы с небольшим числом слоев равной оптической толщины (произведения показателя преломления на толщину слоя).

Однако задача синтеза многослойных покрытий требует дальнейшего изучения как в смысле постановки, так и разработки устойчивых алгоритмов решения, реализуемых на компьютере.

Работа [9] посвящена именно этим вопросам, рассматриваемым на примере систем однородных, изотропных, не поглощающих диэлектрических пленок при нормальном падении световой волны.

### **3.1. Физическое содержание рассматриваемой задачи синтеза**

Рассмотрим пластинку  $0 \leq z \leq H$  с показателем преломления  $n = n(z)$ , ( $n = \sqrt{\epsilon\mu}$ ), на которую падает нормально монохроматическая световая

волна длины  $\lambda = 2\pi c / \omega$ . Амплитуда электрического поля волны  $E(z)$  определяется условиями следующей краевой задачи:

$$\frac{\partial^2 E}{\partial z^2} + n^2 \omega^2 E = 0, n = \begin{cases} n_0 = 1, z \leq 0, \\ n(z), 0 < z < H, \\ n(H) = \text{const}, z \geq H \end{cases}$$

$$[E]_{z=0} = E_0, \left[ \frac{dE}{dz} \right]_{z=0} = -i \frac{\omega}{c} E_0, E|_{z \rightarrow +\infty} \propto e^{-ikz}, \quad (3.1)$$

$$[E]_{z=H} = 0, \left[ \frac{dE}{dz} \right]_{z=H} = 0, E|_{z \rightarrow +\infty} \propto e^{-ikz},$$

где  $E_0$  - амплитуда падающей волны. Если  $n(z)$  – кусочно-непрерывная функция, то задача дополняется однородными условиями сопряжения, соответствующими третьей строке (3.1), на внутренних границах.

Пропускательная способность системы в рассматриваемом случае дается «коэффициентом пропускания»

$$T = \frac{n(H)}{n_0} \left| \frac{E}{E_0} \right|_{z=H}^2 = T(\lambda), \quad (3.2)$$

однозначно определяемым для заданной системы как функция длины падающей волны. Вместо  $T$  в других аспектах задачи можно рассматривать коэффициент отражения  $R = 1 - T$ .

Рассматриваемая нами задача является обратной по отношению к задаче (3.1) — (3.2) и состоит в том, чтобы по заданной априори функции  $T(\lambda)$  определить  $n(z)$ .

Однако такая постановка задачи не является корректной, что характерно для задач синтеза, как это отмечалось в [11], по следующим причинам. Во-первых, может не существовать системы  $(n(z), H)$ , имеющей требуемый коэффициент пропускания. Во-вторых, данному коэффициенту пропускания (например, близкому в каком-то смысле к  $T(\lambda)$ ) могут отвечать различные системы [12]. Наконец, не всякая система

практически реализуема, что также должно быть принято во внимание при постановке задачи.

Мы будем рассматривать задачу синтеза в классе кусочно-постоянных функций  $n(z)$ , т. е. для слоистых систем, при условиях типов: показатели преломления слоев, во-первых, имеют определенные для каждого слоя значения ( $n_j$  для  $j$ -го слоя,  $j=1,2,\dots,N$ ), во-вторых, варьируются в заданных пределах:  $n_{\min} \leq n \leq n_{\max}$ ; при этом толщины слоев произвольны;  $N$  не является априори заданной величиной.

Постановка задачи, отвечающая предположению о том, что показатели преломления принимают ограниченное число дискретных значений, а толщины могут произвольно варьироваться, в ряде случаев представляется достаточно естественной по соображениям практической реализуемости. Действительно, толщины слоев, наносимых на вакуумных напылительных установках, могут колебаться в достаточно широких пределах и контролироваться с высокой степенью точности. В то же время на практике часто для напыления используются всего два вещества — с высоким и низким показателями преломления.

Имеется ряд экспериментальных работ (например, [14]) по синтезу многослойных систем с показателями преломления, принимающими произвольные значения в некоторых пределах. Поэтому предлагаемая выше более общая постановка задачи, связанная с варьированием показателей преломления, представляется целесообразной. Необходимо также проанализировать, что дает переход к такой более общей постановке.

Ниже мы рассматриваем случай нормального падения световой волны. Однако это не ограничивает общности постановки задачи. В самом деле, в случае наклонного падения волны достаточно заменить  $n_j$  на

$p_j = n_j \sin \theta_j$  и воспользоваться законом Снеллиуса:  $n_j \sin \theta_j = n_0 \sin \theta_0$ .

Аналогично этому могут быть рассмотрены системы с поглощением.

### 3.2. Математическая постановка задачи синтеза

Приводимая ниже постановка задачи синтеза оптических систем базируется на принципах, изложенных в [11].

Систему из  $N$  слоев будем описывать  $2N$ -мерным вектором  $x = \{d_1, d_2, \dots, d_N, n_1, n_2, \dots, n_N\}$ , координатами которого  $d_j, n_j$  являются толщины и показатели преломления слоев. Пусть  $\bar{T}(\lambda)$  - заданная на интервале длин волн  $[\lambda_1, \lambda_2]$  пропускательная способность. Будем считать  $\bar{T}(\lambda)$  функцией весьма общего вида, в частности на  $L_2[\lambda_1, \lambda_2]$ .

Задача (3.1) сопоставляет любому конечномерному вектору  $x$  некоторую функцию  $T(\lambda)$  (см. (3.2)). Тем самым определен нелинейный оператор  $A$ :

$$A(x, \lambda) = T(\lambda).$$

В случае, если все  $d_j > 0, n_{\min} \leq n_j \leq n_{\max}$ , функция  $T(\lambda)$  будет пропускательной способностью системы слоев, описываемой вектором  $x$ .

Пусть  $E_{2N}$  есть  $2N$ -мерное векторное пространство,  $D_{2N}$  - замкнутое выпуклое множество в нем, определяемое так:

$$D_{2N} = \{x \in E_{2N} : d_j \geq 0, n_{\min} \leq n_j \leq n_{\max}, j = 1, 2, \dots, N\}.$$

Примем за меру близости  $T(\lambda)$  и  $\bar{T}(\lambda)$  величину

$$\rho(T, \bar{T}) = \left\| A(x, \lambda) - \bar{T}(\lambda) \right\|_{L_2}, x \in D_{2N}. \quad (3.3)$$

Назовем

$$\delta_N = \inf \left\| A(x, \lambda) - \bar{T}(\lambda) \right\|_{L_2}$$

максимально достижимой точностью приближения на  $N$ -слойных

системах.

Ясно, что максимально достижимые точности удовлетворяют оценкам  $\delta_1 \geq \delta_2 \geq \dots \geq \delta_N \geq \dots \geq 0$ , и в совокупности ограничены снизу величиной  $\delta = \lim_{N \rightarrow \infty} \delta_N$ , которую назовем предельно достижимой точностью.

Основная задача состоит в том, чтобы приблизить в смысле (3.3) требуемую характеристику с некоторой априори заданной точностью  $\delta$  (такая задача имеет смысл, если  $\delta > \delta_0$ ) при дополнительных ограничениях на синтезируемую структуру. Такими дополнительными требованиями являются, в первую очередь, требование минимальности слоев системы  $N$ , а также минимальности общей толщины покрытия  $\sum_{j=1}^N d_j$ . Отметим, что оба требования чрезвычайно важны на практике. Они определяются требованиями устойчивости к внешним воздействиям (при большом числе слоев системы легко разрушаются), возможностями напылительной аппаратуры.

Тогда основная задача сводится к следующей: если

$$F_N^\alpha[x] = \left\| A(x, \lambda) - \bar{F}(\lambda) \right\|_{L_2}^2 + \alpha \|d\|^2, \quad (3.4)$$

где

$$\|d\| = \sum_{j=1}^N d_j, x \in D_{2N}, \quad (3.5)$$

то требуется определить вектор  $x$ , минимизирующий функционал  $F_N^\alpha[x]$  при условиях

$$\left\| A(x, \lambda) - \bar{F}(\lambda) \right\|_{L_2}^2 = \delta^2, N = \min k, k = 1, 2, \dots$$

Задачу будем решать следующим образом: последовательно увеличивая число  $N$  слоев системы, ищем такое  $N$ , при котором удастся достичь заданной точности  $\delta$ . Пусть найдено минимальное  $N = N^*$ , при

котором достигается точность  $\delta$ ; тогда можно потребовать, чтобы при данном числе слоев была минимальной суммарная толщина покрытия.

Последняя задача есть задача минимизации функционала  $F_N^\alpha[x]$  на множестве  $D_{2N}$ . Причем параметр  $\alpha$  выбирается из условия

$$\|A(x_{N^*}^\alpha, \lambda) - \bar{T}(\lambda)\| = \delta.$$

Соответствующий минимизирующий элемент  $x_{N^*}^\alpha$ , очевидно, существует, так как  $D_{2N}$  - замкнутое выпуклое множество конечномерного пространства. Найденное таким образом решение будет минимальным по норме (5) из всех  $x \in D_{2N}$  и удовлетворяющих требованию  $\|A(x, \lambda) - \bar{T}(\lambda)\| \leq \delta$ .

### 3.3. Дискретная формулировка задачи синтеза

В случае системы с произвольным  $n(z)$  для вычисления  $A(x, \lambda)$  и явного (через  $A(x, \lambda)$ ) выражения его производной можно использовать алгоритм, аналогичный изложенному в [15].

В последующих расчетах мы ограничимся рассмотрением характерных для практики систем с кусочно-постоянной  $n(z)$ . В этом случае оказывается удобной матричная форма рекуррентных представлений оператора  $A(x, \lambda)$ , принятая в работах по изучению оптических систем [13]. Соответственно

$$A(x, \lambda) \equiv T(\lambda) = \frac{4n_0 n_{N+1}}{(n_0 m_{11} + n_{N+1} m_{22})^2 + (m_{21} + n_0 n_{N+1} m_{12})^2}, \quad (3.6)$$

где  $n_0, n_{N+1} = n(H)$  - заданные показатели преломления обрамляющих покрытие сред (рис. 13),  $m_{ik}$  - элементы полной характеристической матрицы  $\hat{M}(x)$ , определенные формулой

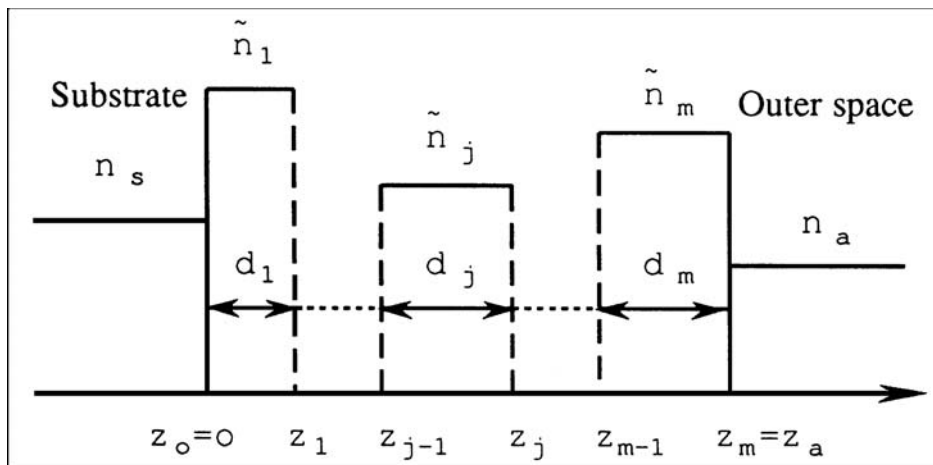
$$\hat{M}(x) = \begin{vmatrix} m_{11} & im_{12} \\ im_{21} & m_{22} \end{vmatrix} = \prod_{j=1}^N \hat{M}(x_j),$$

$$\hat{M}(x_j) = \begin{vmatrix} \cos \frac{2\pi}{\lambda} n_j d_j & -\frac{i}{n_j} \sin \frac{2\pi}{\lambda} n_j d_j \\ -in_j \sin \frac{2\pi}{\lambda} n_j d_j & \cos \frac{2\pi}{\lambda} n_j d_j \end{vmatrix},$$

где  $d_j = z_{j+1} - z_j$  - толщина слоя.

В случае наклонного падения здесь следует заменить  $n_j$  на  $n_j \cos \theta_j$ , где  $\theta_j$  - угол преломления на границе  $j$ -го слоя, легко рассчитываемый (при известных  $n_j$ ) через заданный угол падения  $\theta_0$ .

Операторная производная  $A'(x, \lambda)$ , которая нам потребуется для минимизации функционалов, может быть представлена аналитически.



**Рис. 13. Структура многослойного покрытия**

Положим, не меняя обозначений,  $A(x, \lambda) = T(\hat{M}(x))$ , где  $T$  - оператор, отображающий матрицу  $\hat{M}(x)$  в функцию  $T(\lambda)$ . Тогда [16]

$$A'(x, \lambda) = T'[\hat{M}(x)] \hat{M}'(x).$$

Частная производная матрицы  $\hat{M}(x)$  по  $d_j$  будет иметь вид:

$$\frac{\partial}{\partial d_j} [\mathbb{M}(x)] = -\frac{2\pi}{\lambda} n_j \prod_{k=1}^{j-1} M(x_k) \left\| \begin{array}{cc} \sin \frac{2\pi}{\lambda} n_j d_j & \frac{i}{n_j} \cos \frac{2\pi}{\lambda} n_j d_j \\ i n_j \cos \frac{2\pi}{\lambda} n_j d_j & \sin \frac{2\pi}{\lambda} n_j d_j \end{array} \right\| \times \\ \times \prod_{k=j+1}^N M(x_k) = \mathbb{M}^{j,1}(x),$$

а частная производная по  $n_j$ :

$$\frac{\partial}{\partial n_j} [\mathbb{M}(x)] = -\frac{2\pi}{\lambda} d_j \prod_{k=1}^{j-1} M(x_k) \times \\ \left\| \begin{array}{cc} \sin \frac{2\pi}{\lambda} n_j d_j & \frac{i}{n_j} \cos \frac{2\pi}{\lambda} n_j d_j - \frac{\lambda \sin(2\pi / \lambda) n_j d_j}{n_j^2 2\pi d_j} \\ i n_j \cos \frac{2\pi}{\lambda} n_j d_j - \frac{\lambda \sin(2\pi / \lambda) n_j d_j}{n_j^2 2\pi d_j} & \sin \frac{2\pi}{\lambda} n_j d_j \end{array} \right\| \times \\ \times \prod_{k=j+1}^N M(x_k) = \mathbb{M}^{j,2}(x).$$

Из (3.6) нетрудно получить, что производная оператора  $A(x, \lambda)$ , взятая при значении  $\mathbb{M}(x)$  и примененная к матрице

$$\hat{L} = \left\| \begin{array}{cc} l_{11} & i l_{12} \\ i l_{21} & l_{22} \end{array} \right\|,$$

будет иметь вид:

$$A'[\mathbb{M}(x)] \hat{L} = -\frac{A^2(x, \lambda)}{2n_0 n_{N+1}} [(m_{11} n_0 + m_{12} n_{N+1})(l_{11} n_0 + l_{12} n_{N+1}) + \\ + (m_{21} + n_0 n_{N+1} m_{22})(l_{21} + n_0 n_{N+1} l_{22})].$$

Если  $\hat{L}$  - одна из матриц или  $\mathbb{M}^{j,1}$  или  $\mathbb{M}^{j,2}$ , то, используя унимодулярность [13] матрицы  $\mathbb{M}(x)$ , легко получить

$$A'[\mathbb{M}(x)] \hat{L} = -\frac{A^2(x, \lambda)}{2n_0 n_{N+1}} [n_0^2 m_{11} l_{11} + n_{N+1}^2 m_{22} l_{22} + m_{21} l_{21} + n_0^2 n_{N+1}^2 m_{12} l_{12}]. \quad (3.7)$$



Из (3.7) получаем, что градиент первого члена функционала  $\|A(x, \lambda) - \bar{F}(\lambda)\|_{L_2}$  (см. (3.4)) есть  $2N$ -мерный вектор с координатами,

находимыми по формуле

$$-\int_{\lambda_1}^{\lambda_2} \frac{A^2(x, \lambda)}{n_0 n_{N+1}} \varphi(\lambda) \left[ n_0^2 m_{11} m_{11}^{j,1} + n_{N+1}^2 m_{22} m_{22}^{j,1} + m_{21} m_{21}^{j,1} + n_0^2 n_{N+1}^2 m_{12} m_{12}^{j,1} \right]$$

$$\eta_{j+N} = -\int_{\lambda_1}^{\lambda_2} \frac{A^2(x, \lambda)}{n_0 n_{N+1}} \varphi(\lambda) \left[ n_0^2 m_{11} m_{11}^{j,2} + n_{N+1}^2 m_{22} m_{22}^{j,2} + m_{21} m_{21}^{j,2} + n_0^2 n_{N+1}^2 m_{12} m_{12}^{j,2} \right],$$

$$j = 1, 2, \dots, N,$$

где  $m_{ik}^{j,1}$  и  $m_{ik}^{j,2}$  - соответственно элементы матриц  $\bar{M}^{j,1}$  и  $\bar{M}^{j,2}$ ,  $\varphi(\lambda) = A(x, \lambda) - \bar{F}(\lambda)$ .

Таким образом,  $F'(x)$  будет представлять собой  $2N$ -мерный вектор с координатами

$$[F'(x)]_i = \begin{cases} \eta_i + 2\alpha \left( \sum_{k=1}^N d_k \right), i = 1, 2, \dots, N, \\ \eta_i, i = N + 1, N + 2, \dots, 2N. \end{cases} \quad (3.8)$$

### 3.4. Алгоритм минимизации функционала и некоторые результаты расчетов

Как показал предварительный анализ, функционал (3.4) обладает большим количеством локальных минимумов, к которым ведут «глубокие наклонные овраги», в результате чего движение к минимуму по методу градиентного спуска происходит как бы по «ломаной линии», направления и величины отрезков которой через один близки друг к другу; чем ближе к минимуму, тем более заметным становится это свойство. В соответствии с такой структурой функционала был построен алгоритм его минимизации, включающий поиск локальных минимумов по случайно заданным начальным приближениям, отбор «подозрительных» на глобальный

минимум и уточнение их значений. Полученный таким образом глубокий минимум считается глобальным.

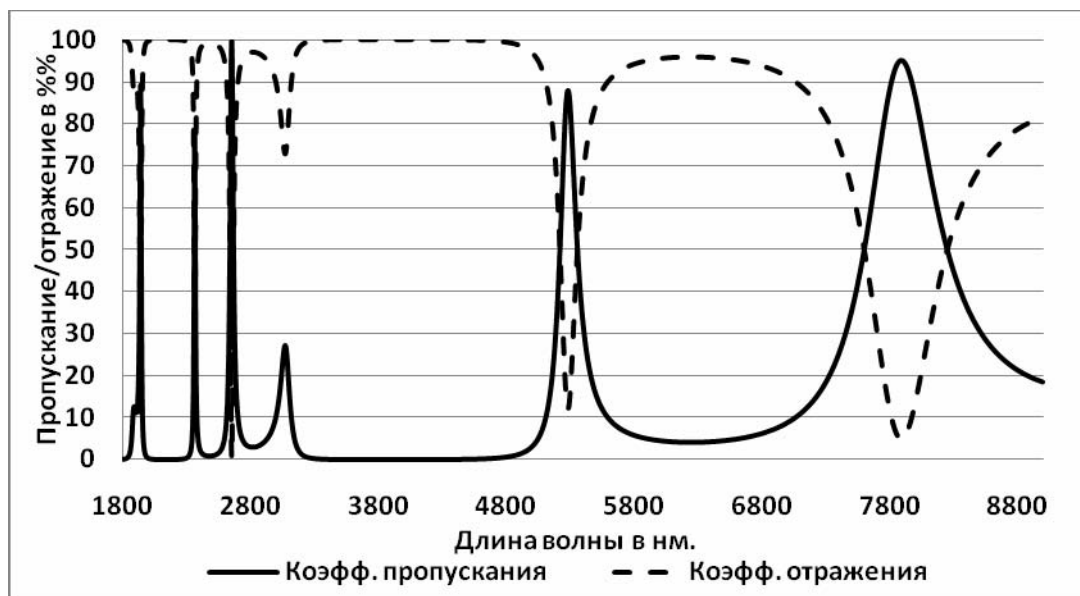
Первоначальный «грубый» поиск локальных минимумов производился по обычному методу проекцией градиента [17], при котором  $n$ -й итерационный шаг осуществляется по формуле

$$x^n = P_{D_{2N}}(x^{n-1} - \gamma_n B(x^{n-1})), \quad (3.9)$$

где  $P_{D_{2N}}$ -проектор на множество  $D_{2N}$ ;  $B(x) = F'(x)$ , т. е. равен градиенту функционала (3.4), определенному формулой (3.8);  $\gamma_n$  - параметр, определяемый из условия минимальности функционала (3.4) по данному направлению градиента.

Процесс уточнения, в соответствии с описанной выше структурой функционала, отличался от (3.9) тем, что для итераций с номерами, кратными трем,

$$B(x^{n-1}) = F'(x^{n-1}) + \frac{\gamma_{n-1}}{\gamma_{n-2}} F'(x^{n-2}).$$



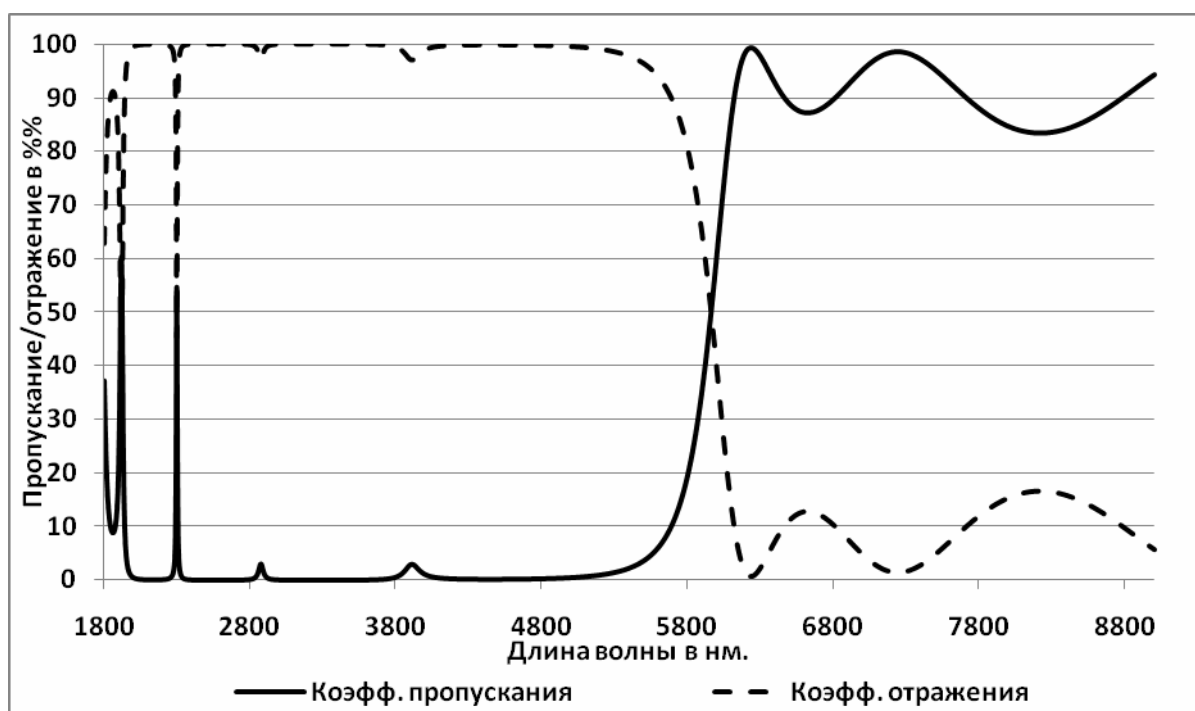
**Рис. 14. Коэффициенты пропускания и отражения для 11-слойной модели при начальных толщинах**

Как нетрудно заметить, эта поправка к формуле (3.9) отвечает направлению движения вдоль оврага, а не от «стенки к стенке», как

обычно, что позволяет в нашей задаче значительно сократить время счета и существенно ближе подойти к минимуму.

При практической реализации вначале рассматривалась постановка задачи, при которой показатели преломления слоев фиксированы и принимают чередующиеся значения:  $n_1$  - для нечетных слоев,  $n_2$  - для четных.

На рис. 14 и 15 изображены коэффициенты пропускания и отражения в широком диапазоне длин волн (расчеты проводились с помощью программы MorphoVision [18]) для так называемой отрезающей системы (заданная характеристика, которую необходимо приблизить, такова:  $T = 0$ , при  $\lambda < 6 \mu\text{к}$ ,  $T = 1$  при  $\lambda > 6 \mu\text{к}$ ) в инфракрасной области. Сплошная линия – коэффициент пропускания для оптической системы, штриховая – для коэффициента отражения. Требуемой точности (для синтезированной системы  $\delta \approx 10^{-2}$ ) удалось достичь на 11-слойной системе.



**Рис. 15. Коэффициенты пропускания и отражения для 11-слойной модели. Оптимизированные толщины**

Минимизация по суммарной толщине при этом не производилась. Толщины слоев полученной системы ( $d_k$ ), соответствующего ей начального приближения ( $d_H$ ) и показатели преломления слоев приведены в табл. 1. Основные трудности получения таких систем заключаются в том, что интересующая нас область длин волн весьма широка (от 2-4 до 10 мк), и существующие полуэмпирические методы, достаточно полный обзор которых приведен в [19], не позволяют задать начальное приближение, по которому можно было бы с требуемой точностью получить нужный коэффициент пропускания.

Таблица 1

Номер слоя	Показатель преломления	$d_H$ , мк	$d_K$ , мк	Номер слоя	Показатель преломления	$d_H$ , мк	$d_K$ , мк
1	4.00	0.086	0.142	7	4.00	0.332	0.216
2	1.34	0.423	0.640	8	1.34	0.305	0.186
3	4.00	0.369	0.334	9	4.00	0.203	0.173
4	1.34	0.798	0.611	10	1.34	0.585	0.455
5	4.00	0.284	0.238	11	4.00	0.367	0.164
6	1.34	0.488	0.630				

Таблица 2

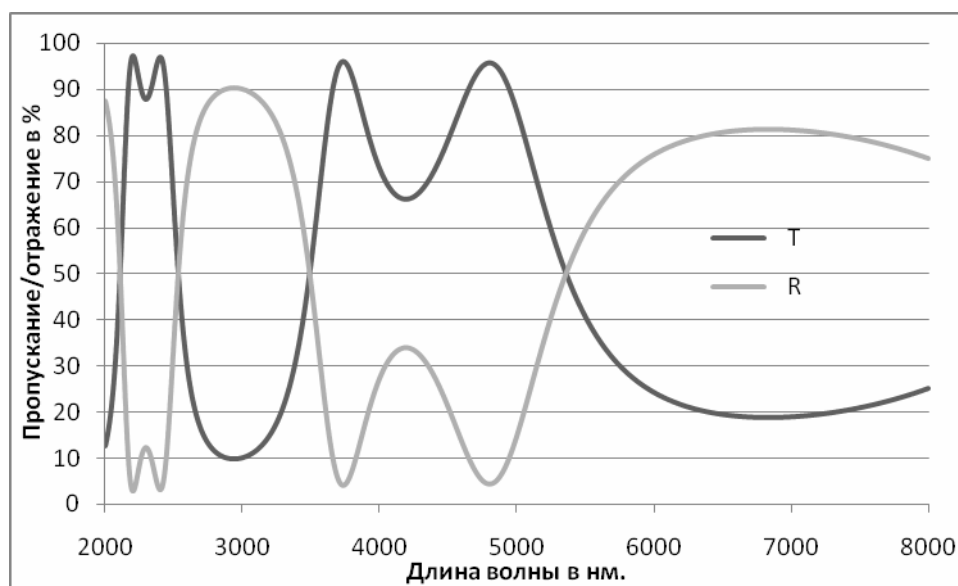
$\alpha$	Толщины слоев					$\ d\ ^2$	$\delta^2$
	1-й	2-й	3-й	4-й	5-й		
0	0.133	0.749	0.269	0.691	0.149	3.95	0.103
$10^{-2}$	0.135	0.737	0.274	0.680	0.150	3.90	0.103
$2 \cdot 10^{-2}$	0.137	0.724	0.279	0.669	0.151	3.85	0.103
$10^{-1}$	0.163	0.578	0.356	0.520	0.173	3.20	0.106
$2 \cdot 10^{-1}$	0.177	0.533	0.365	0.505	0.164	3.05	0.118

Тем самым предлагаемый метод оказывается достаточно эффективным (однако в нем не учтены дисперсия и поглощение среды). Для синтезированной системы значение  $\delta$  уменьшилось при минимизации в 165 раз – от 1.98 для начальной системы до 0.012 для синтезированной.

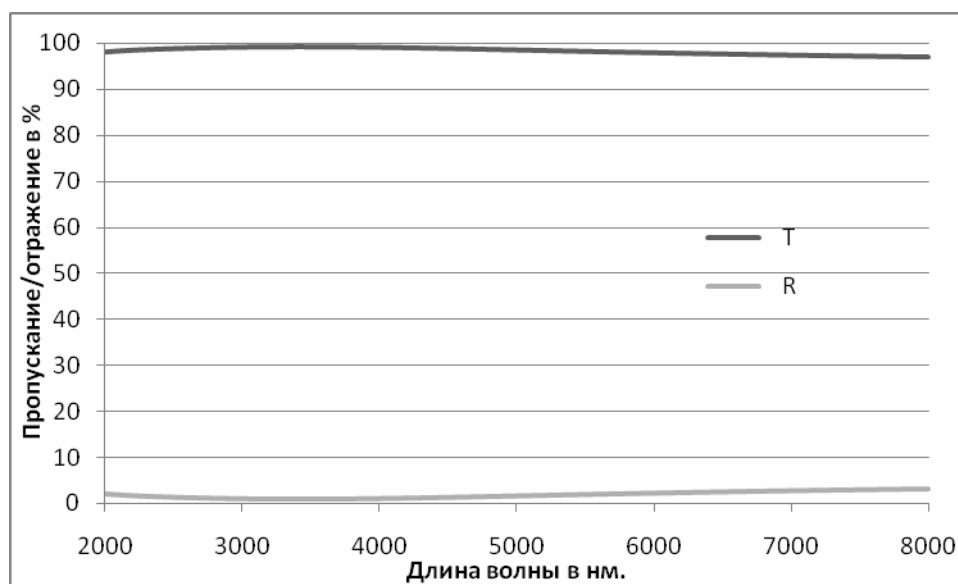
При фиксированном числе слоев ( $N = 5$ ) исследовалось поведение составляющих функционала (3.7) в зависимости от величины  $\alpha$ . Заданная характеристика, которую необходимо приблизить, такова: в области от 3 до 6 мк должно быть полное отражение, в области от 6 до 10 мк должно быть 100%-ное пропускание. В табл. 2 приведены значения  $\delta^2$ ,  $\|d\|^2$  и толщины слоев синтезированных систем для различных значений  $\alpha$ . Показатели преломления для четных слоев 4.00, для нечетных — 1.34. Видно, что с ростом  $\alpha$  величина  $\delta^2$  возрастает, а  $\|d\|^2$  уменьшается, при этом  $\|d\|^2$  в области до  $\alpha = 10^{-1}$  уменьшается довольно значительно при небольшом увеличении  $\delta^2$  (т. е. можно существенно уменьшить суммарную толщину, практически не увеличивая погрешности приближения). Заметим, что при дальнейшем увеличении  $\alpha$  погрешность, наоборот, быстро возрастает, в то время как суммарная толщина уменьшается незначительно. Таким образом, рассматриваемая методика оказывается эффективной и при решении задачи «оптимального» (в указанном выше смысле) синтеза.

Было исследовано влияние на точность реализации требуемой характеристики пропускания варьирования показателей преломления наряду с варьированием толщин. Для этого был проведен синтез 9-слойной отрезающей системы с областью отражения 3-6 мк и областью пропускания 6-10 мк. При этом сначала показатели преломления были фиксированы и принимали для нечетных слоев максимально возможное в этой области длин волн значение  $n_1 = 4.00$ , а для четных – минимально возможное  $n_2 = 1.34$  (первое вещество – германий, второе – криолит

являются основными веществами, используемыми при практическом конструировании систем, работающих в микронном диапазоне). Варьировались только толщины. Затем был проведен поиск при варьировании также и показателей преломления в пределах, определяемых величинами  $n_1$  и  $n_2$ . Хотя при этом значения показателей преломления для некоторых слоев полученной системы и приняли значения, отличные от  $n_1$  или  $n_2$  (отметим, что для четырех из девяти слоев  $n$  приняло одно из крайних значений:  $n_1$  или  $n_2$ ), точность приближения повысилась лишь очень незначительно (величина  $\delta$  уменьшилась лишь в 1.07 раза). Видимо, это связано с тем, что требуемая характеристика имеет широкую область отражения, а для получения широкой области отражения, как это замечено на практике (в случае систем со слоями равной оптической толщины это показано аналитически), необходимо, чтобы показатели преломления соседних слоев отличались друг от друга как можно больше.



**Рис. 16. Коэффициенты пропускания и отражения просветляющей 3-слойной системы. Начальная система слоев**



**Рис. 17. Коэффициенты пропускания и отражения просветляющей системы. Оптимизированные толщины слоев**

Но при синтезе систем с характеристиками другого вида варьирование показателей преломления дает значительный эффект. На рис. 16, 17 изображены коэффициенты пропускания  $T$  и отражения  $R$  начальной (рис. 16) просветляющей системы и коэффициенты пропускания  $T$  и отражения  $R$  соответствующей ей полученной (рис. 17) в области длин волн от 2 до 8 мк. Просветление в столь широкой области обеспечивает 2-слойное покрытие с показателями преломления и толщинами 1-го и 2-го слоя  $n_1 = 1.34$ ,  $d_1 = 0.552$ ;  $n_2 = 1.51$ ,  $d_2 = 0.390$  (показатель преломления среды  $n_0 = 1.00$ , подложки  $n = 1.56$ ). Данная система была получена при синтезе 3-слойных систем. Толщины слоев начальной системы (см. рис. 16) задавались случайным образом, показатели преломления 1-го и 3-го слоев – равными 4.00, а второго 1.34. Интересно отметить, что при минимизации толщина первого слоя обратилась в нуль (система стала двухслойной), а показатель преломления 3-го слоя (для полученной системы – второго) изменился с 4.00 до 1.51.

Последний пример говорит о том, что одновременное варьирование и показателей преломления, и толщин позволит получить системы с

интересными характеристиками при небольшом общем числе слоев системы, что особенно важно для практики.



## **Тема 4. Теоретические аспекты задачи синтеза оптических покрытий**

### ***4.1. Качественные закономерности структуры оптических покрытий***

В течение многих десятилетий актуальной является проблема создания композиционных материалов с новыми свойствами. Наиболее распространенный тип неоднородных структур – слоистые среды, свойства которых меняются вдоль одной координаты. При волновом воздействии на слоистую структуру возникает система отраженных и преломленных волн, которые, взаимодействуя с падающей волной, образуют сложную интерференционную картину. Меняя структуру слоистой среды, можно в широких пределах управлять волновой картиной процесса. На этом основана работа многих приборов и устройств в различных областях физики и техники. Велика роль слоистых структур, применяемых для управления энергетикой волнового процесса.

Слоистые структуры широко применяются в оптике (просветление оптических поверхностей, создание интерференционных зеркал, светоделителей, отрезающих фильтров и т.п.), радиофизике (для преобразования волн в СВЧ- и радиодиапазонах), квантовой электронике, акустике (в ультразвуковой технике, для решения задач звукоизоляции, в качестве согласующих устройств), в теплофизике (для решения задач температурной стабилизации). Упругие слоистые среды применяются в качестве фильтров продольных и поперечных волн, для решения задач вибрационной защиты, создания эффективных демпфирующих устройств. В физике твердого тела значительное внимание уделяется изучению волновых процессов в периодических структурах. Многие природные структуры моделируются слоистой средой, например, земная кора, состоящая из пород с различными физическими свойствами.

Сейсмологические методы разведки земной коры базируются на изучении явлений, происходящих при распространении упругих волн в слоистых средах.

Наиболее важными характеристиками волнового процесса являются энергетические коэффициенты пропускания и отражения, характеризующие в виде функции от частоты доли энергии волны, прошедшей через конструкцию (или отраженной от нее). При исследовании волновых процессов в слоистых средах существенный интерес представляет проблема конструирования структуры с требуемыми свойствами. Данная проблема заключается в таком выборе структуры неоднородной среды, при котором ее энергетические характеристики будут наиболее близки к заданным зависимостям. От решения этой проблемы зависит не только эффективность применяемых неоднородных структур, но и расширение возможностей их эффективного использования в различных областях физики и техники.

Широкое распространение неоднородных структур, общность их математического описания приводят к необходимости создания единого подхода к задачам оптимизации структуры неоднородных сред. Возникающие при этом задачи оптимизации обладают целым рядом специфических особенностей. К ним относятся: дискретность области значений для ряда управляющих параметров, являющихся физическими свойствами материалов слоев (так как набор материалов при проектировании, как правило, конечен), разрывность решений краевых задач, описывающих волновые процессы в неоднородных средах, существенная многоэкстремальность волновых задач синтеза.

Задачи синтеза как обратные задачи математической физики обладают свойствами некорректности по А.Н. Тихонову. Анализ существующих методов оптимизации и синтеза неоднородных структур при волновых воздействиях показывает, что большинство из них малоэффективно или

вовсе непригодно для решения рассматриваемых задач. Многоэкстремальность волновых задач синтеза существенно затрудняет исследование предельных возможностей и характеристик конструкций. Дискретность варьируемых переменных приводит к невозможности построения вариаций, малых в равномерной метрике, к неэффективности применения большинства известных методов. Разрывность решений вариационных задач вызывает дополнительные трудности при поиске оптимального решения. Из сказанного вытекает необходимость разработки новых методов исследования и расчета задач оптимального синтеза неоднородных структур. Это позволит повысить эффективность применения неоднородных структур в различных областях физики и техники.

При решении задач оптимального проектирования центральным является вопрос о том, каких предельных возможностей по управлению энергетикой волновых процессов различной физической природы можно достичь на основе направленного выбора структуры композиционной конструкции, а именно: физических свойств материалов слоев, толщин слоев, числа слоев, а также порядка взаимного расположения слоев с различными физическими свойствами в конструкции. Предельные возможности соответствуют тому предельному уровню, которого можно достичь на основе направленного управления структурой конструкции.

Изучение предельных возможностей неоднородных структур при воздействии волновых процессов различной физической природы является актуальной научной задачей, имеющей важное теоретическое и прикладное значение.

Пусть элемент  $g$  определяет интересующую исследователя функциональную характеристику волнового процесса, взаимодействующего со слоистой средой, структура которой задается

элементом  $\mathbf{u} \in U$ . Связь между структурой слоистой среды и ее волновыми свойствами определяется дифференциальным оператором  $\mathbf{L}$ :

$$\mathbf{L}\mathbf{u} = \mathbf{g}. \quad (4.1)$$

Пусть  $\tilde{\mathbf{g}}$  - заданная функциональная характеристика волнового поля. Необходимо подобрать такую структуру слоистой среды (физические свойства материалов слоев, толщины слоев, число слоев, а также порядок взаимного расположения слоев с различными физическими свойствами в конструкции), при которой функциональная характеристика волнового поля будет наиболее близка к требуемой. В математической постановке задача оптимального синтеза заключается в минимизации критерия:

$$J(\mathbf{u}) = \|\mathbf{L}\mathbf{u} - \tilde{\mathbf{g}}\| \Rightarrow \min_{\mathbf{u} \in U} \quad (4.2)$$

на решениях операторного уравнения (4.1) при ограничении:

$$\mathbf{u} \in U. \quad (4.3)$$

При этом в задачах оптимального синтеза представляет интерес выделение всего множества решений  $U^*$ , реализующих предельные возможности по управлению энергетиков волнового процесса:

$$U^* = \left\{ \mathbf{u} \in U : \|\mathbf{L}\mathbf{u} - \tilde{\mathbf{g}}\|^2 = \min_{\mathbf{u} \in U} \|\mathbf{L}\mathbf{u} - \tilde{\mathbf{g}}\|^2 \right\}. \quad (4.4)$$

Для определенности в дальнейшем будем считать, что функциональная характеристика  $g$  является зависимостью энергетического коэффициента пропускания от частоты:

$$g = g(\omega), \quad \omega_{\min} \leq \omega \leq \omega_{\max}. \quad (4.5)$$

Здесь  $\omega_{\min}$ ,  $\omega_{\max}$  - нижняя и верхняя границы фильтруемого диапазона частот. Наиболее подходящей для большинства задач синтеза слоистых структур является среднеквадратическая метрика оценки близости функциональных характеристик волнового поля:

$$\|g - \tilde{g}\|_{L_2}^2 = \int_{\omega_{\max}}^{\omega_{\min}} (g(\omega) - \tilde{g}(\omega))^2 d\omega. \quad (4.6)$$

Создание конструкций с уникальными свойствами тесно связано с исследованием их предельных возможностей. Предельные возможности соответствуют тому предельному уровню, которого можно достичь на основе направленного управления структурой конструкции. В настоящее время отсутствуют методологические основы, позволяющие достоверно находить глобальный экстремум даже в относительно простых задачах. В многоэкстремальных задачах, к которым относятся волновые задачи синтеза, ограничены возможности предсказания поведения целевой функции. Локального же предсказания недостаточно для построения эффективных процедур поиска решений.

Решение проблемы исследования предельных возможностей предполагает существование возможности эффективного полного перебора всех допустимых вариантов конструкций, количество которых чрезвычайно велико. Однако реализация полного перебора невозможна даже с применением быстродействующих компьютеров. Известные подходы к решению данной проблемы связаны с пересмотром части множества, несравнимой со множеством всех допустимых вариантов и поэтому, с одной стороны, не гарантируют нахождение варианта, реализующего предельные возможности, а, с другой стороны, не позволяют оценить, насколько существенно найденный вариант структуры отличается от варианта, реализующего предельные возможности.

В соответствии с этим в рамках существующих подходов:

- 1. Отсутствует возможность объективно оценить, насколько возможности созданных структур, функционирующих в различных областях физики и техники, отличаются от предельно достижимых.
- 2. Отсутствует возможность эффективного конструирования структур, реализующих предельные возможности.

Общим для существующих подходов является то, что они не учитывают качественную структуру исследуемых задач, в частности,

качественные закономерности взаимосвязи параметров в структурах, реализующих предельные возможности.

В соответствии с этим была поставлена гипотеза о возможности существования фундаментальных закономерностей, которым подчиняется взаимность параметров в структурах, реализующих предельные возможности. То есть было предложено существование общих закономерностей, которые присущи структурам, реализующим предельные возможности.

Тогда, если бы эти общие закономерности были выделены, то они могли быть использованы для упрощения структуры исходной задачи. Эффективное упрощение структуры исходной задачи достигается на основе оптимального включения установленных качественных закономерностей структуры оптимального решения в структуру задачи.

На основе конструктивного анализа необходимых условий оптимальности в волновых задачах синтеза установлены уникальные свойства оптимальных структур [20, 21].

Открыто свойство внутренней симметрии во взаимосвязях параметров в структурах, реализующих предельные возможности. Показано, что установленное свойство внутренней симметрии для волновых процессов различной физической природы имеет аналогичную структуру. Изучение различных моделей распространения волновых процессов в композиционных структурах показало, что свойство внутренней симметрии сохраняется при достаточно существенном усложнении математических моделей, описывающих волновые процессы в композиционных конструкциях [21].

Получены условия, при выполнении которых в оптимальную конструкцию может входить только два материала допустимого набора независимо от количества материалов, составляющих исходный набор.

Установленные закономерности позволяют априори существенно уменьшить количество допустимых вариантов конструкций, анализируемых на оптимальность. Априорное знание таких закономерностей позволяет разрабатывать на их основе эффективные методы исследования предельных возможностей композиционных конструкций по управлению энергетикой волновых процессов различной физической природы в таких областях, как теплофизика, электродинамика, механика деформируемого твердого тела, радиоэлектроника, солнечная энергетика, акустика, оптика, космическая техника.

В настоящее время значительное внимание уделяется изучению волновых задач синтеза слоистых структур. При распространении волновых процессов в слоистых структурах возникает система отраженных и преломленных волн, которые при взаимодействии с падающей волной образуют сложную интерференционную картину. Меняя структуру слоистой среды (физические свойства материалов слоев, толщины слоев, число слоев, а также порядок взаимного расположения слоев с различными физическими свойствами), можно в широких пределах управлять энергетическими характеристиками волновых процессов. При этом одно из центральных мест при исследовании таких задач занимает проблема исследования предельных возможностей конструкций по достижению заданного комплекса свойств. В вариационной постановке данная проблема заключается в построении решений, доставляющих глобальный минимум функционалу качества, оценивающему меру близости функциональных характеристик проектируемой конструкции к требуемым [21].

В общей постановке задача оптимального синтеза слоистой структуры с требуемыми свойствами может быть сформулирована следующим образом:

$$\begin{aligned} \mathbf{L}\mathbf{u} &= \mathbf{g}, \quad \mathbf{u} \in U \\ J(\mathbf{u}) &= \|\mathbf{L}\mathbf{u} - \tilde{\mathbf{g}}\|_{L_2}^2 \rightarrow \min_{\mathbf{u} \in U}, \end{aligned} \quad (4.7)$$

где  $\mathbf{u}$  - вектор-функция, описывающая структуру оптимизируемой слоистой среды;  $U$  - функциональное множество вектор-функций  $\mathbf{u}$ , соответствующих возможным допустимым параметрам в конструкции;  $\mathbf{g}$  - вектор-функция, компонентами которой является совокупность функциональных характеристик, интересующих проектировщика.

Связи между  $\mathbf{u}$  и  $\mathbf{g}$  в общем виде описываются операторным уравнением вида (4.7). Критерий качества характеризует меру близости функциональных характеристик проектируемой структуры к требуемым  $\tilde{\mathbf{g}}$ .

Задача оптимального проектирования в рассматриваемой постановке заключается в поиске множества решений  $U^*$ , доставляющих глобальный минимум функционалу (4.2):

$$U^* = \left\{ \mathbf{u}^* \in U : \|\mathbf{L}\mathbf{u}^* - \tilde{\mathbf{g}}\|_{L_2}^2 = \inf_{\mathbf{u} \in U} \|\mathbf{L}\mathbf{u} - \tilde{\mathbf{g}}\|_{L_2}^2 \right\}, \quad (4.8)$$

т.е. в нахождении всех квазирешений уравнения:  $\mathbf{L}\mathbf{u} = \tilde{\mathbf{g}}$ .

Вследствие сложности построения конструктивных условий, позволяющих отличать глобальный минимум от множества локальных, существующие подходы не могут осуществлять эффективное исследование предельных возможностей конструкций.

В соответствии с этим возникает чрезвычайно важная проблема выделения таких качественных закономерностей структуры оптимального решения, которые позволили бы осуществлять эффективное априорное сжатие множества допустимых вариантов, анализируемых на оптимальность.

Данную проблему исследуем для случая взаимодействия электромагнитных волн с неоднородными структурами. Будем



рассматривать ситуацию, близкую к реальному проектированию, когда в распоряжении проектировщика имеется конечный набор материалов.

## **4.2. Наклонное падение волны на слоистую структуру**

Будем считать, что пространственная неоднородность достигается за счет набора слоев из различных однородных и изотропных материалов. Основой для описания распространения электромагнитных волн в неоднородных средах служат уравнения Максвелла.

Будем рассматривать случай наклонного падения электромагнитной волны под углом  $\theta_0$  на многослойную конструкцию общей толщины  $l$ , состоящую из  $N$  плоскопараллельных слоев с различными физическими свойствами. Плоскость падения совпадает с плоскостью  $xz$ . Ось  $z$  направлена в сторону падения, наружная поверхность многослойника, на которую падает волна, совпадает с плоскостью  $z=0$ . При таком выборе системы координат электрический и магнитный векторы не будут зависеть от координаты  $y$ .

Уравнения Максвелла допускают два независимых решения. В одном из них электрический вектор  $\mathbf{E}$  перпендикулярен плоскости падения (электрическая волна с горизонтальной поляризацией, ТЕ-волна), в другом электрический вектор  $\mathbf{E}$  параллелен плоскости падения (электромагнитная волна с вертикальной поляризацией, ТМ-волна). Любое решение системы уравнений Максвелла можно представить в виде линейной комбинации (с комплексными коэффициентами) ТЕ- и ТМ-волн. Отношение коэффициентов этой комбинации определяет поляризацию волны. В соответствии с этим достаточно рассмотреть случай ТЕ- и ТМ-волн.

### **4.2.1. Наклонное падение ТЕ-волны на систему непоглощающих магнитоэлектрических слоев**

ТЕ-волна общего вида может быть представлена суперпозицией плоских монохроматических волн, т. е. в виде интеграла Фурье:

$$E(x, z, t) = \int_{-\infty}^{\infty} f(z, \omega) \exp(i\Delta_0 x - i\omega t) d\omega,$$

где  $f(z, \omega), (0 \leq z \leq l)$  - спектральная плотность ТЕ-волны;  
 $\Delta_0 = k_{in} \sin \theta_0 k_{in} = \omega \sqrt{\varepsilon_{in}} / c$ ;  $\varepsilon_{in}$  - диэлектрическая проницаемость полупространства, из которого приходит волна,  $c$  - скорость света. Можно показать, что задача распространения ТЕ-волны в системе магнитоэлектрических слоев сводится к нахождению решения следующей краевой задачи [8,9]:

$$\begin{aligned} j(z) &= \mu(z) g(z), \\ g(z) &= -k_0^2(\omega) u(z) f(z), 0 \leq z \leq l, \\ g(0) &= ik_{in}(\omega) \cos \theta_0 (2 - f(0)), \\ g(l) &= ik_{out}(\omega) \cos \theta_{out} f(l). \end{aligned} \quad (4.9)$$

Здесь  $\theta_{out}$  - угол, под которым электромагнитная волна выходит из слоистой структуры:

$$\cos \theta_{out} = \left( 1 - \frac{\varepsilon_{in} \sin^2 \theta_0}{\varepsilon_{out}} \right)^{1/2},$$

$k_0(\omega) = \omega / c$  - волновое число в вакууме;  $k_{out}(\omega) = \omega \sqrt{\varepsilon_{out}} / c$ ;  $\varepsilon_{out}$  - диэлектрическая проницаемость последнего полупространства,  
 $u(z) = (\varepsilon(z)\mu(z) - \varepsilon_{in} \sin^2 \theta_0) / \mu(z)$ ;  $\varepsilon(z), \mu(z)$  - распределение диэлектрической и магнитной проницаемости по толщине слоистой структуры. Физические параметры слоистой структуры считаем связанными между собой функциональной зависимостью, позволяющей однозначно восстановить магнитную проницаемость допустимого материала по его известной диэлектрической проницаемости. В рассматриваемом случае независимым физическим параметром будет являться только  $\varepsilon$ . Пусть задан дискретный набор материалов, множество диэлектрических проницаемостей которого обозначим через  $\Lambda$ . Для

каждого  $z \in [0,1]$  выполнено включение

$$\varepsilon(z) = \Lambda. \quad (4.10)$$

Энергетический коэффициент пропускания ТЕ-волны определяется через решение краевой задачи (4.9):

$$T^{TE} = \frac{\cos \theta_{out}}{\cos \theta_0} \sqrt{\frac{\varepsilon_{out}}{\varepsilon_{in}}} \text{mod}^2(f(l, \omega)).$$

Требуется сконструировать слоистую структуру, обладающую низким отражением в одних участках спектра и высоким — в других. Математически данная задача формулируется следующим образом: необходимо определить распределение диэлектрической проницаемости по толщине структуры  $\varepsilon^*(z)$ , при котором принимает наименьшее значение критерий

$$J^{TE} = \int_{\omega_{min}}^{\omega_{max}} \tau(\omega) \text{mod}^2(f(l, \omega)) d\omega. \quad (4.11)$$

Здесь  $\tau(\omega), (-l \leq \tau(\omega) \leq l)$  - весовая функция.

Функционал качества (4.11) определяется зависимостью энергетического коэффициента пропускания от частоты в заданном диапазоне частот  $[\omega_{min}, \omega_{max}]$ . Пусть  $\varepsilon^*(z), (0 \leq z \leq l)$  - решение задачи синтеза (4.9)—(4.11):

$$J(\varepsilon^*(\cdot)) = \inf_{\varepsilon(\cdot)} J(\varepsilon(\cdot)).$$

Возникает интересный вопрос: можно ли априори установить качественную структуру оптимального решения в задачах оптимального синтеза неоднородных покрытий? Тогда, если бы качественные закономерности структуры оптимального решения были установлены, их можно было использовать для эффективного априорного сжатия множества допустимых вариантов конструкций, анализируемых на оптимальность. Применение таких качественных закономерностей может

открыть новые перспективные возможности для повышения эффективности различных методов поиска оптимального решения.

Оказывается, что конструктивный анализ необходимых условий оптимальности в волновых задачах оптимального синтеза позволяет установить ряд важных качественных закономерностей структуры оптимальных конструкций. А именно:

1) получить верхние оценки для числа различных материалов дискретного набора, которые могут входить в оптимальную конструкцию;

2) получить оценки оптимального числа слоев конструкции;

3) построить систему рекуррентных соотношений, позволяющих априори, до проведения численных расчетов, выделить именно те материалы допустимого набора, которые могут входить в оптимальную конструкцию;

4) установить характер сочленения слоев с различными физическими свойствами в оптимальной конструкции;

5) получить условия, при выполнении которых в оптимальную конструкцию может входить только два материала допустимого набора независимо от количества материалов, составляющих исходный дискретный набор;

6) установить свойство внутренней симметрии во взаимосвязи параметров в оптимальных структурах.

Поэтому справедливо следующее утверждение.

**Утверждение 1.** Число различных материалов, составляющих оптимальную многослойную систему, не может превышать  $p + q$ , где  $p$  и  $q$  - число точек разрыва функций  $\varepsilon^+(\tau)$  и  $\varepsilon^-(\tau)$ .

Тогда непосредственным следствием того, что экстремум функции Гамильтона в координатах границ раздела слоев оптимальной конструкции

достигается одновременно на двух материалах допустимого набора, является следующее свойство.

**Утверждение 2.** Физические свойства материалов слоев оптимальной конструкции удовлетворяют следующей системе рекуррентных соотношений:

$$\begin{aligned} L(\varepsilon^{j_{r-1}^+}, \varepsilon^{j_r^+}) &= \min L(\varepsilon^{j_{r-1}^+}, \varepsilon), \\ (r = 1, \dots, p; j_0^+ = 1, j_p^+ = m). \end{aligned} \quad (4.12)$$

$$\begin{aligned} L(\varepsilon^{j_{r-1}^-}, \varepsilon^{j_r^-}) &= \min L(\varepsilon^{j_{r-1}^-}, \varepsilon), \\ (r = 1, \dots, p; j_0^- = 1, j_p^- = m). \end{aligned} \quad (4.13)$$

В этих обозначениях

$$M_r^+ = \left\{ \varepsilon \in \Lambda : u(\varepsilon) > u(\varepsilon^{j_{r-1}^+}) \right\}, (r = \overline{1, p}),$$

$$M_r^- = \left\{ \varepsilon \in \Lambda : u(\varepsilon) > u(\varepsilon^{j_{r-1}^-}) \right\}, (r = \overline{1, p}).$$

Введем множества

$$\begin{aligned} \Lambda^+ &= \left\{ \varepsilon^{j_0^+}, \varepsilon^{j_1^+}, \dots, \varepsilon^{j_p^+} \right\}, \\ \Lambda^- &= \left\{ \varepsilon^{j_0^-}, \varepsilon^{j_1^-}, \dots, \varepsilon^{j_p^-} \right\} \\ \Lambda^* &= \Lambda^+ \cup \Lambda^-. \end{aligned} \quad (4.14)$$

Соотношения (4.12) и (4.13) являются системой рекуррентных соотношений для определения материалов допустимого набора  $\Lambda$ , которые могут составлять оптимальную многослойную систему. Те материалы, физические параметры которых не удовлетворяют системе рекуррентных соотношений (4.12) и (4.13), заведомо не могут входить в состав оптимальной структуры. Поэтому данная система рекуррентных соотношений позволяет проводить эффективное априорное сужение исходного набора материалов  $\Lambda$  до множества  $\Lambda^+$  (4.14). Кроме того, открываются возможности поиска материалов с новыми свойствами,

включение которых в исходный набор может приводить к существенному улучшению характеристик синтезируемых структур.

Рассмотрим последовательность элементов

$$\dots, \epsilon^{j_0^+}, \epsilon^{j_1^+}, \dots, \epsilon^{j_p^+}, \epsilon^{j_q^-}, \epsilon^{j_{q-1}^-}, \dots, \epsilon^{j_0^-}, \epsilon^{j_0^+}, \epsilon^{j_1^+}, \dots \quad (4.15)$$

**Утверждение 3.** Физические параметры материалов, входящих в состав оптимальной многослойной структуры, являются соседними в последовательности (4.15).

Справедливость данного утверждения следует из того, что экстремального значения в координатах границ раздела слоев оптимальной конструкции одновременно на двух материалах допустимого набора функция Гамильтона может достигать тогда и только тогда, когда данные материалы являются соседними в последовательности (4.15).

Таким образом, порядок взаимного сочленения материалов слоев с различными физическими свойствами в оптимальной структуре может быть установлен заранее.

Установленные закономерности позволяют осуществлять эффективное сужение множества допустимых вариантов. Поэтому их применение открывает новые возможности для повышения эффективности различных методов поиска оптимального решения.

**Утверждение 4.** В случае наклонного падения монохроматической электромагнитной волны с горизонтальной поляризацией на систему непоглощающих магнитодиэлектрических слоев оптимальная конструкция состоит не более чем из двух материалов независимо от количества материалов, составляющих допустимый набор. При этом в состав оптимальной конструкции входят материалы, диэлектрические проницаемости которых являются соседними в последовательности (4.15).

#### 4.2.2. Наклонное падение ТМ-волны на систему непоглощающих магнитоэлектрических слоев

Конструктивный анализ необходимых условий оптимальности позволяет для случая наклонного падения ТМ-волны получить результаты, аналогичные результатам, полученным для случая наклонного падения ТЕ-волны. ТМ-волна общего вида также может быть представлена в виде суперпозиции плоских монохроматических волн, т. е. в виде ряда Фурье:

$$E(x, z, t) = \int_{-\infty}^{\infty} q(z, \omega) \exp(i\Delta_0 x - i\omega t) d\omega,$$

где  $q(z, \omega)$  - спектральная плоскость ТМ-волны. Также можно показать, что задача о распространении ТМ-волны в системе магнитоэлектрических слоев сводится к нахождению решения следующей краевой задачи [22]:

$$\begin{aligned} q(z, \omega) &= \varepsilon(z)h(z, \omega), \\ h(z, \omega) &= -k_0^2(\omega)v(z)q(z, \omega), 0 \leq z \leq l, \\ q(0, \omega) &= ik_{in}(\omega) \cos \theta_0 (2 - f(0, \omega)), \\ q(l, \omega) &= ik_{out}(\omega) \cos \theta_{out} f(l, \omega). \end{aligned} \quad (4.16)$$

Здесь  $v(z) = (\varepsilon(z)\mu(z) - \varepsilon_{in} \sin^2 \theta_0) / \varepsilon(z)$ . Энергетический коэффициент пропускания ТМ-волны определяется через решение краевой задачи (4.16) следующим образом:

$$T^{TM} = \frac{\cos \theta_{out}}{\cos \theta_0} \sqrt{\frac{\varepsilon_{in}}{\varepsilon_{out}}} \text{mod}^2(q(l, \omega)).$$

Задача оптимального синтеза в вариационной постановке заключается в минимизации критерия

$$J^{TM} = \int_{\omega_{min}}^{\omega_{max}} \tau(\omega) \text{mod}^2(q(l, \omega)) d\omega \quad (4.17)$$

на решениях системы (4.16).

Введем множества

$$\Lambda^+ = \{\epsilon^{j_0^+}, \epsilon^{j_1^+}, \dots, \epsilon^{j_p^+}\}, \Lambda^- = \{\epsilon^{j_0^-}, \epsilon^{j_1^-}, \dots, \epsilon^{j_q^-}\}, \quad (4.18)$$

$$\Lambda^* = \Lambda^+ \cup \Lambda^-.$$

**Утверждение 5.** Физические параметры материалов, входящих в состав оптимальной многослойной структуры, могут быть только элементами множества  $\Lambda^*$  (4.18).

Рассмотрим следующую последовательность элементов:

$$\dots, \epsilon^{j_0^+}, \epsilon^{j_1^+}, \dots, \epsilon^{j_p^+}, \epsilon^{j_q^-}, \epsilon^{j_{q-1}^-}, \dots, \epsilon^{j_0^-}, \epsilon^{j_1^-}, \dots. \quad (4.19)$$

**Утверждение 6.** Физические параметры соседних слоев оптимального многослойного покрытия являются соседними в последовательности (4.19).

#### **4.2.3. Наклонное падение монохроматической электромагнитной волны с вертикальной поляризацией (ТМ-волны) на систему непоглощающих магнитодиэлектрических слоев**

Для этого случая может быть доказано утверждение, аналогичное утверждению 4.

**Утверждение 7.** В случае наклонного падения монохроматической электромагнитной волны с вертикальной поляризацией на систему непоглощающих магнитодиэлектрических слоев оптимальная конструкция состоит не более чем из двух материалов независимо от количества материалов, составляющих исходный набор. При этом в состав оптимальной конструкции входят материалы, диэлектрические проницаемости которых являются соседними в последовательности (4.19).

Таким образом, для данного случая так же, как и в случае монохроматической электромагнитной волны с горизонтальной поляризацией, полученные результаты позволяют полностью решить проблему синтеза, т. е. эффективно выделить полную совокупность



решений, доставляющих глобальный минимум функционалу качества (4.17).

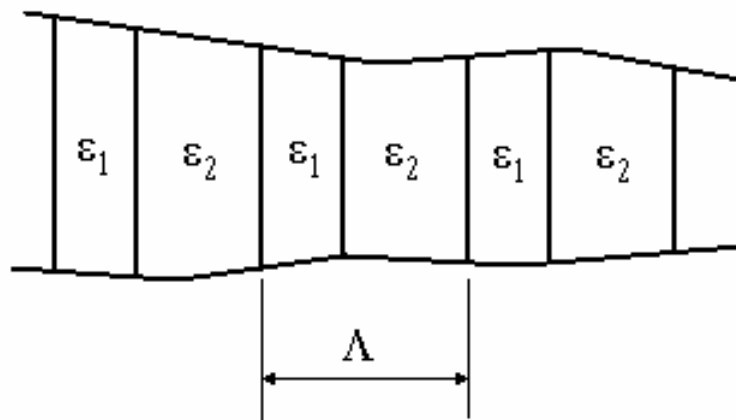
Качественные закономерности оптимальных покрытий установлены для случая поляризованных магнитных электромагнитных волн ТЕ- и ТМ-типа, а также для случая монохроматических волн.

Тем не менее полученные результаты могут быть использованы и для более общего случая неполяризованных электромагнитных волн, а также для случая, когда покрытие должно обладать требуемыми свойствами (например, обеспечить максимальное отражение или пропускание) в интервале длин волн. Результирующее покрытие, рассчитанное с учетом установленных закономерностей, может служить достаточно хорошим начальным приближением при применении методов оптимального проектирования для более общих случаев. В зависимости от специфических особенностей конкретных задач такие начальные приближения по совокупности варьируемых параметров могут несущественно отличаться от наилучших решений.

## Тема 5. Фотонные кристаллы

Фотонными кристаллами принято называть среды, у которых диэлектрическая проницаемость периодически меняется в пространстве с периодом, допускающим брэгговскую дифракцию света. Фотонные кристаллы могут быть естественного происхождения (например, опал), но в большинстве случаев это искусственно изготовленные материалы.

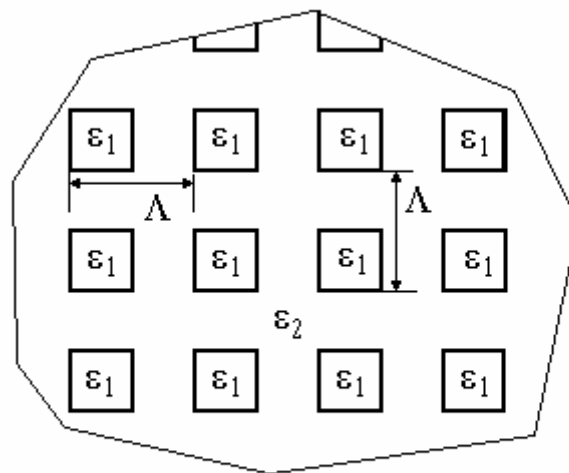
Фотонные кристаллы по характеру изменения диэлектрической проницаемости можно разделить на три основных класса.



**Рис. 17. Схематическое представление одномерного фотонного кристалла**

1. Одномерные, в которых диэлектрическая проницаемость периодически изменяется в одном пространственном направлении (рис. 17). На рис. 17 символом  $\Lambda$  обозначен период изменения диэлектрической проницаемости,  $\epsilon_1$  и  $\epsilon_2$  – диэлектрические проницаемости двух материалов (но в общем случае может присутствовать любое число материалов). Такие фотонные кристаллы состоят из параллельных друг другу слоев различных материалов с разными диэлектрическими проницаемостями и могут проявлять свои

свойства в одном пространственном направлении, а именно - перпендикулярно слоям.



**Рис. 18. Схематическое представление двумерного фотонного кристалла**

2. Двумерные, в которых диэлектрическая проницаемость периодически изменяется в двух пространственных направлениях (рис. 18). На рис. 18 фотонный кристалл создан прямоугольными областями с диэлектрическими проницаемостями  $\epsilon_1$ , которые находятся в среде с диэлектрической проницаемостью  $\epsilon_2$ . При этом области с диэлектрической проницаемостью  $\epsilon_1$  упорядочены в двумерной кубической решетке. Такие фотонные кристаллы могут проявлять свои свойства в двух пространственных направлениях, и форма областей с диэлектрической проницаемостью  $\epsilon_1$  не ограничивается прямоугольниками, как на рис. 18, а может быть любой (окружности, эллипсы, произвольная и т.д.). Кристаллическая решётка, в которой упорядочены эти области, также может быть другой, а не только кубической, как на приведённом рис. 18.

3. Трёхмерные, в которых диэлектрическая проницаемость периодически изменяется в трех пространственных направлениях.

Такие фотонные кристаллы могут проявлять свои свойства в трех пространственных направлениях, и их можно представить как массив объемных областей (сфер, кубов и т.д.), упорядоченных в трехмерной кристаллической решётке.

### ***5.1. Аналитические и численные методы теоретического исследования фотонных кристаллов***

Фотонные кристаллы позволяют проводить манипуляции с электромагнитными волнами оптического диапазона; для этого размерные постоянные фотонных кристаллов ( $\Lambda$ ) должны быть близки к длине волны света. Поэтому к ним не применимы методы лучевой теории, а используется волновая теория, основанная на решении уравнений Максвелла. Уравнения Максвелла могут быть решены аналитически или численно, но именно численные методы решения используются для исследования свойств фотонных кристаллов наиболее часто по причине их доступности и легкой подстройки под решаемые задачи. Уместно также упомянуть, что имеются два основных подхода к рассмотрению свойств фотонных кристаллов: методы для временной области (которые позволяют получить решение задачи в зависимости от времени) и методы для частотной области (которые предоставляют решение задачи в виде функции от частоты).

Методы для временной области удобны в отношении динамических задач, которые предусматривают временную зависимость электромагнитного поля от времени. Они также могут быть использованы для расчета зонных структур фотонных кристаллов, однако практически сложно выявить положение разрешенных и запрещенных зон в выходных данных таких методов. Кроме того, при расчете зонных диаграмм фотонных кристаллов используется преобразование Фурье, частотное

разрешение которого зависит от общего времени расчета метода. То есть для получения большего разрешения в зонной диаграмме нужно потратить больше времени на выполнение расчетов. Есть еще и другая проблема - временной шаг таких методов должен быть пропорционален размеру пространственной сетки метода. Требование увеличения частотного разрешения зонных диаграмм требует уменьшения временного шага, а, следовательно, и размера пространственной сетки, увеличения числа итераций, необходимой оперативной памяти компьютера и времени расчета.

Методы для частотной области удобны прежде всего тем, что решение уравнений Максвелла происходит сразу для стационарной системы, и непосредственно из решения определяются частоты оптических мод системы, это позволяет быстрее рассчитывать зонные диаграммы фотонных кристаллов, чем с использованием методов для временной области. К их достоинствам можно отнести число итераций, которое практически не зависит от разрешения пространственной сетки метода, и то, что ошибка численного метода спадает экспоненциально с числом проведенных итераций. Недостатком метода является, естественно, невозможность описания динамики развития оптических колебаний в системе.

Безусловно, теоретические исследования фотонных кристаллов не ограничиваются только расчетом зонных диаграмм, а требуют и знаний о стационарных процессах при распространении электромагнитных волн через фотонные кристаллы. Примером может служить задача исследования спектра пропускания фотонных кристаллов. Для таких задач можно использовать метод матрицы распространения излучения, на котором вкратце и остановимся.

## 5.2. Распространение света в фотонных кристаллах

### 5.2.1. Спектр отражения и пропускания одномерных фотонных кристаллов

**Метод матриц распространения для расчета отражения и пропускания света в одномерных фотонных кристаллах**

Рассмотрим плоскую линейно-поляризованную волну  $E_0^+ \exp(i(k_0^\omega r - \omega t))$  с волновым вектором  $k_0^\omega$  и частотой  $\omega$ , падающую под углом  $\theta$  на многослойную среду, состоящую из  $N$  слоев, характеризующихся толщиной  $d_j$  и, в общем случае, комплексной диэлектрической проницаемостью  $\varepsilon_j''$  или коэффициентом преломления  $n_j = \sqrt{\varepsilon_j}$ ,  $j = 1 \dots N$ . Пусть ось  $z$  задает нормаль к слоям, а плоскость  $xz$  является плоскостью падения волны. Прошедшее и отраженное от многослойной структуры электромагнитное поле является результатом многолучевой интерференции в каждом из ее слоев. Однако в силу линейности волнового уравнения в каждом из слоев поле является суперпозицией двух волн, распространяющихся соответственно в положительном и отрицательном направлениях оси  $z$  (далее называемыми прямой и обратной волнами):

$$\begin{aligned} E_j^\omega(z, t) = & \\ & E_j^+ \exp\left[ (+ik_{z,j}^\omega(z - z_{ij}) + (ik_x^\omega x - i\omega t)) \right] + \\ & E_j^- \exp\left[ (-ik_{z,j}^\omega(z - z_{ij}) + (ik_x^\omega x - i\omega t)) \right]. \end{aligned} \quad (5.1)$$

Тангенциальная компонента волнового вектора  $k_x^\omega = |k_0^\omega| \sin \theta_0$  сохраняется постоянной внутри многослойной структуры, что является следствием трансляционной симметрии вдоль слоев, а нормальная компонента  $k_{z,j}^\omega = |k_j^\omega| \sin \theta_j$  определяется дисперсией  $j$ -го слоя.

Амплитуды прямой и обратной волн  $E_j^+$  и  $E_j^-$  в уравнении (5.1) являются комплексными величинами вследствие многолучевой интерференции.

Ограничиваясь рассмотрением стационарного случая, множителями  $\exp(ik_x^\omega x - i\omega t)$  далее будем пренебрегать.

Электрическая компонента электромагнитного поля на границе раздела  $i$ -го и  $j$ -го слоев ( $i < j$ ) с координатой  $z_{ij}$  записывается в виде

$$\begin{aligned} E_j^\omega(z_{ij} + 0) &= E_j^+ + E_j^- \\ E_i^\omega(z_{ij} - 0) &= E_i^+ + E_i^- . \end{aligned} \quad (5.2)$$

Соотношение между  $E_i^\omega$  и  $E_j^\omega$  определяется граничными условиями для тангенциальных компонент электрических и магнитных составляющих, которое может быть записано в виде следующего матричного уравнения:

$$\begin{pmatrix} E_i^+ \\ E_i^- \end{pmatrix} = \begin{pmatrix} 1/t_{ij} & r_{ij}/t_{ij} \\ r_{ij}/t_{ij} & 1/t_{ij} \end{pmatrix} \begin{pmatrix} E_j^+ \\ E_j^- \end{pmatrix}, \quad (5.3)$$

где  $r_{ij}$  и  $t_{ij}$  — коэффициенты Френеля для отражения и прохождения поля, падающего из  $i$ -го слоя, через  $ij$ -ю границу раздела. Совокупность амплитуд  $E_j^+$  и  $E_j^-$  будем рассматривать как двухкомпонентный вектор

$$E_j = \begin{pmatrix} E_j^+ \\ E_j^- \end{pmatrix}, \quad (5.4)$$

преобразование которого на  $ij$ -ой границе раздела задается матрицей  $M_{ij}$ :

$$M_{ij} = \begin{pmatrix} 1/t_{ij} & r_{ij}/t_{ij} \\ r_{ij}/t_{ij} & 1/t_{ij} \end{pmatrix} \quad (5.5)$$

в виде тензорной свертки

$$E_i = M_{ij} \cdot E_j. \quad (5.6)$$

$M_{ij}$  является матрицей преобразования для  $ij$ -той границы раздела.

Поля в точках  $z_j$  и  $z_j + \zeta$  внутри  $j$ -го слоя связаны матрицей распространения  $\Phi_j(\zeta)$ :

$$\Phi_j(\zeta) = \begin{pmatrix} \exp(ik_{z,j}^\omega \zeta) & 0 \\ 0 & \exp(-ik_{z,j}^\omega \zeta) \end{pmatrix}, \quad (5.7)$$

в виде

$$E_j(z_j + \zeta) = \Phi_j(\zeta) \cdot E_j(z_j). \quad (5.8)$$

Полная  $2 \times 2$  матрица распространения через многослойную среду имеет вид:

$$T \equiv T_{(N+1)0} = M_{(N+1)N} \Phi_N \dots M_{10}, \quad (5.9)$$

где  $\Phi_m = \Phi_m(d_m)$ . Таким образом, в предположении, что в последнем слое (заднем полупространстве)  $N + 1$  бегущая назад волна отсутствует, а на границу раздела  $0 - 1$  падает единичная волна, можно связать поля в первом и последнем слоях:

$$\begin{pmatrix} T \\ 0 \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} 1 \\ R \end{pmatrix}. \quad (5.10)$$

Коэффициент отражения от многослойной среды  $R$  задается как

$$R = -T_{21} / T_{22}, \quad (5.11)$$

где  $T_{21}, T_{22}$  - соответствующие матричные элементы  $T$ . Коэффициент пропускания многослойной среды  $T$  запишется в виде:

$$T = \frac{T_{11}T_{22} - T_{12}T_{21}}{T_{22}}. \quad (5.12)$$

Если амплитуда падающей волны полагается равной единице, то двухкомпонентный вектор поля перед многослойной средой  $E_0^\omega = (1, R)$ , а пространственное распределение поля внутри структуры задается следующим выражением:



$$E_j^\omega(z) = T_j(z)E_0^\omega = \begin{pmatrix} \exp(ik_{z,j}^\omega z) & 0 \\ 0 & \exp(-ik_{z,j}^\omega z) \end{pmatrix} \cdot M_{j(j-1)} \Phi_{(j-1)} \dots M_{10} \begin{pmatrix} 1 \\ R \end{pmatrix}. \quad (5.13)$$

Метод матриц распространения применим для любых многослойных структур. В случае одномерных фотонных кристаллов, где толщины  $d_j$  и диэлектрические проницаемости  $\varepsilon_j$  слоев изменяются периодически, выражения для коэффициента пропускания и коэффициента отражения имеют аналитическое выражение. Дисперсионное соотношение для волны с волновым числом  $q$ , распространяющейся внутри одномерного фотонного кристалла, состоящего из чередующихся слоев толщиной  $d_1$  и  $d_2$ , и коэффициентами преломления  $n_1$  и  $n_2$ , имеет вид:

$$\cos(qd) = \cos(k_1 d_1) \cos(k_2 d_2) - \frac{1}{2} \left( \frac{n_1}{n_2} + \frac{n_2}{n_1} \right) \sin(k_1 d_1) \sin(k_2 d_2). \quad (5.14)$$

Парциальные волновые числа  $k_{1(2)} = \sqrt{\varepsilon_{1(2)} \frac{\omega}{c} - k_x}$  характеризуют распространение электромагнитной волны частотой  $\omega$  в каждом из слоев,  $k_x$  - проекция волнового вектора на плоскость слоев, сохраняющаяся при прохождении волны через границы раздела слоев;  $d = d_1 + d_2$  - период фотонного кристалла. Обозначим френелевские коэффициенты пропускания и отражения от границы раздела слоев как  $t_1$  и  $r_1$ . Коэффициенты пропускания и отражения света от фотонного кристалла из  $N$  слоев, когда вне его диэлектрическая постоянная принята равной  $\varepsilon_1$ , имеют вид:

$$t_N = \cos(Nqd) - H \frac{\sin(Nqd)}{\sin(qd)}, \quad r_N = \frac{r_1 \sin(Nqd)}{t_1 \sin(qd)} t_N, \quad (5.15)$$

где

$$H = \frac{1}{2t_1} ((t_1^2 - r_1^2 - 1)\cos(qd) + i(t_1^2 - r_1^2 + 1)\sin(qd)). \quad (5.16)$$

Если диэлектрическая проницаемость вне фотонного кристалла не равна  $\varepsilon_1$ , то в выражении (5.15) необходимо учесть коэффициенты пропускания (отражения) света реальными передней и задней диэлектрическими границами конечной периодической структуры,  $t_I(r_I)$  и  $t_{II}(r_{II})$ . В этом случае пропускание фотонного кристалла запишется в виде:

$$T_N = \left| \frac{t_N t_I t_{II}}{\Delta_N} \right|^2, \quad (5.17)$$

где знаменатель

$$\Delta_N = 1 + (r_I - r_{II})r_N + r_I r_{II} (t_N^2 - t_N^2). \quad (5.18)$$

Численный расчет спектра коэффициента пропускания света  $T_N(\omega)$  показывает, что при падении света на фотонный кристалл в спектре имеется полоса частот, в которой пропускание отсутствует даже в случае пренебрежимо малых потерь (действительных  $n_1$  и  $n_2$ ). Вне провала наблюдаются осцилляции, обусловленные интерференцией света на внешних границах фотонного кристалла. Положение и ширина провала в спектре пропускания обусловлены запрещенной зоной в одномерном законе дисперсии электромагнитных волн, соответствующей комплексным значениям волнового вектора  $q$  в уравнении (5.14).

## 5.2.2. Дифракция света в двумерных фотонных кристаллах

### Расчет отражения и пропускания двумерных фотонных кристаллов методом разложения по плоским волнам

Рассмотрим плоскую электромагнитную волну с волновым вектором  $k_i$ , падающую из среды 1 под углом падения  $\theta$  на фотонный кристалл, занимающий слой 2, в плоскости, перпендикулярную осям цилиндров,

составляющих двумерный кристалл (рис.19). Введем систему координат с осью  $z$ , совпадающей с направлениями цилиндров, и плоскостью  $xz$ , задающей плоскость падения. В выбранной системе координат волновой вектор падающего излучения имеет координаты  $k_i = (k_x, k_{1y}, 0) = (k_1 \sin \theta, k_1 \cos \theta, 0)$  с волновым числом  $k_1 = \sqrt{\epsilon_1} \omega / c$ . В силу периодичности фотонного кристалла вдоль направления  $x$ , отраженное и прошедшее через фотонный кристалл электромагнитное поле является суперпозицией плоских волн, называемых брэгговскими волнами, имеющих волновые вектора  $k_r^{(n)}$  и  $k_t^{(n)}$  ( $n = 0, \pm 1, \pm 2 \dots$ ), соответственно лежащие в плоскости  $xu$ . Брэгговские волны с  $k_r^{(0)}$  и  $k_t^{(0)}$  определяют зеркальное отражение и прямое пропускание, волны с  $k_r^{(\pm 1)}$   $k_t^{(\pm 1)}$  и т.д. характеризуют дифрагированные отраженные (прошедшие) волны порядка  $\pm 1 \dots$ . Тангенциальные компоненты волновых векторов брэгговских волн порядка  $n$  задаются в виде:

$$k_{r,x}^{(n)} = k_{t,x}^{(n)} = k_x^{(n)} = k_x + G_n, \quad (5.19)$$

где

$$G_n = 2\pi n / a, \quad n = \pm 1, \pm 2 \dots \quad (5.20)$$

являются векторами обратной решетки фотонного кристалла с периодом  $a$ . Выражение (5.19) является, по сути дела, утверждением о сохранении тангенциальной компоненты волнового вектора, что следует из трансляционной симметрии рассматриваемой системы в плоскости  $xz$  с дополнительным учетом периодичности вдоль направления  $x$ . Нормальные компоненты волновых векторов брэгговских волн порядка  $n$  находятся из условия сохранения волновых чисел  $k_1$  и  $k_3 = \sqrt{\epsilon} \omega / c$ . Для отраженных волн выражения для  $k_{r,y}^{(n)}$  имеет вид:

$$k_{r,y}^{(n)} = \begin{cases} -\sqrt{k_1^2 - (k_x^{(n)})^2}, & k_1 \geq |k_x^{(n)}| \\ -i\sqrt{(k_x^{(n)})^2 - k_1^2}, & k_1 < |k_x^{(n)}|. \end{cases} \quad (5.21)$$

Аналогично, для прошедших волн нормальные компоненты волновых векторов брэгговских волн задаются в виде

$$k_{t,y}^{(n)} = \begin{cases} -\sqrt{k_3^2 - (k_x^{(n)})^2}, & k_3 \geq |k_x^{(n)}| \\ -i\sqrt{(k_x^{(n)})^2 - k_3^2}, & k_3 < |k_x^{(n)}|. \end{cases} \quad (5.22)$$

В дальнейшем достаточно рассмотреть случаи двух характерных поляризаций падающей электромагнитной волны —  $E \parallel z$  и  $E \perp z$ . Случай произвольной поляризации будет представлять собой их линейную суперпозицию.

Рассмотрим подробнее случай поляризации электрической компоненты вдоль осей цилиндров —  $E \parallel z$ .

Поле в среде 1 является суперпозицией падающей волны и отраженных брэгговских волн:

$$E_{1z}(r) = E_0 e^{ik_r \cdot r} + \sum_{n=-\infty}^{\infty} R_n e^{ik_r^{(n)} \cdot r}, \quad (5.23)$$

где  $E_0$  - амплитуда падающей волны;  $R_n$  - амплитуда отраженной брэгговской волны порядка  $n$ ;  $r$  - двумерный радиус-вектор в плоскости, перпендикулярной осям цилиндров фотонного кристалла. Поле в среде 3 является суперпозицией прошедших брэгговских волн:

$$E_{3z}(r) = \sum_{n=-\infty}^{\infty} T_n e^{ik_r^{(n)} \cdot (r-L)}, \quad (5.24)$$

где  $T_n$  - амплитуда прошедшей брэгговской волны порядка  $n$ ,  $L = (0, L)$  а  $L$  - толщина фотонного кристалла. Амплитуды полей  $R_n$  и  $T_n$  находятся из решения волнового уравнения для волны, распространяющейся в фотонном кристалле (среда 2), которое записывается в виде:

$$\Lambda_E E_{2z}(r) \equiv \left\{ \frac{1}{\varepsilon(r)} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) + \frac{\omega^2}{c^2} \right\} E_{2z}(r) = 0, \quad (5.25)$$

где  $\omega$  - частота волны. Для решения волнового уравнения (5.25) введем функцию  $f_E(x, y)$  специального вида:

$$f_E(x, y) \equiv \frac{1}{L} \sum_{n=-\infty}^{\infty} \{yT_n + (L-y)(\delta_{n0}E_0 + R_n)\} e^{ik_x^{(n)}x}, \quad (5.26)$$

где  $\delta_{n0}$  - символ Кронекера. Функция  $f_E(x, y)$  сконструирована таким образом, чтобы на границах фотонного кристалла задавать поля  $E_{1z}$  и  $E_{3z}$ :

$$f_E(x, 0) = E_{1z}(x, 0), \quad f_E(x, L) = E_{3z}(x, L). \quad (5.27)$$

Наконец, введя разностную функцию

$$\psi_E(x, y) = E_{2z}(x, y) - f_E(x, y), \quad (5.28)$$

удовлетворяющую нулевым граничным условиям

$$\psi_E(x, 0) = \psi_E(x, L) = 0, \quad (5.29)$$

волновое уравнение (5.25) запишется в виде:

$$\Lambda_E \psi_E(x, y) = -\Lambda_E f_E(x, y). \quad (5.30)$$

Разложим функции  $\psi_E(x, y)$  и  $1/\varepsilon(x, y)$  в ряд Фурье:

$$\psi_E(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=1}^{\infty} A_{nm} e^{ik_x^{(n)}x} \sin \frac{m\pi}{L} y, \quad (5.31)$$

и

$$\frac{1}{\varepsilon(x, y)} = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} k_{nm} e^{i(G_n x + \frac{m\pi}{L} y)}. \quad (5.32)$$

Подставим в выражение (5.26) для  $f_E(x, y)$  формальные разложения в ряд Фурье

$$\frac{y}{L} = \frac{2}{\pi} \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m} \sin \frac{m\pi}{L} y, \quad 1 = \frac{2}{\pi} \sum_{m=1}^{\infty} \frac{1 - (-1)^m}{m} \sin \frac{m\pi}{L} y. \quad (5.33)$$

С учетом выражений (5.31) и (5.32) волновое уравнение (5.30) для фиксированных чисел  $n$  и  $m$  примет вид:

$$\begin{aligned}
& \frac{\omega^2}{c^2} A_{nm} + \sum_{n'=-\infty}^{\infty} \sum_{m'=-1}^{\infty} \left\{ (k_x^{(n')})^2 + \left( \frac{m'\pi}{L} \right)^2 \right\} (k_{n-n',m+m'} - k_{n-n',|m-m'|}) A_{n'm'} = \\
& = -\frac{2\omega^2}{\pi c^2} \frac{(-1)^{m-1} T_n + R_n + \delta_{n0} E_0}{m} + \\
& + \frac{2}{\pi} \sum_{n'=-\infty}^{\infty} (k_x^{(n')})^2 \sum_{m'=-1}^{\infty} (k_{n-n',|m-m'|} - k_{n-n',m+m'}) \frac{(-1)^{m'-1} T_{n'} + R_{n'} + \delta_{n'0} E_0}{m'}.
\end{aligned} \tag{5.34}$$

Дополнительными уравнениями для нахождения коэффициентов  $A_{nm}$ ,  $R_n$  и  $T_n$  являются граничные условия непрерывности тангенциальной ( $x$ ) компоненты магнитного поля, принимающие вид:

$$\pi \sum_{m=1}^{\infty} m A_{nm} = (iLk_{r,y}^{(n)} + 1) R_n - T_n + \delta_{n0} E_0 (iLk_{1,y} + 1) \tag{5.35}$$

для плоскости  $y = 0$  и

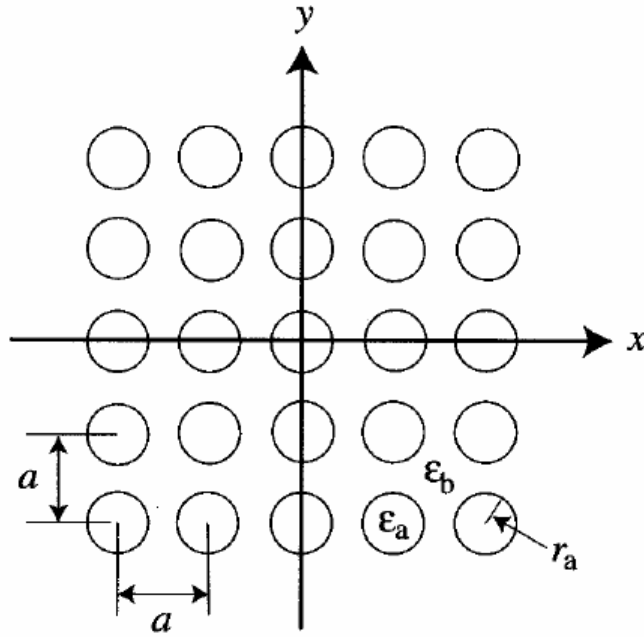
$$\pi \sum_{m=1}^{\infty} m (-1)^m A_{nm} = R_n + (iLk_{t,y}^{(n)} - 1) T_n + \delta_{n0} E_0 \tag{5.36}$$

для плоскости  $y = L$ . Система уравнений для  $p$ -поляризованного излучения может быть получена аналогично.

### Случай двумерных фотонных кристаллов с квадратной решеткой

Решение системы уравнений (5.19) для конкретного фотонного кристалла определяется фурье-амплитудами  $k_{nm}$  разложения (5.32) обратной диэлектрической проницаемости  $\varepsilon(x, y)$ . Получим выражения для  $k_{nm}$  в случае двумерного фотонного кристалла с квадратной решеткой. В задаче об отражении и пропускании фотонного кристалла важным фактором является его ограниченность, которая приводит к модификации выражений для  $k_{nm}$ . Обратная диэлектрическая проницаемость двумерного фотонного кристалла, состоящего из  $N$  слоев и занимающего интервал  $0 \leq y \leq L$ , записывается в виде:

$$\frac{1}{\varepsilon(x, y)} = \frac{1}{\varepsilon_b} + \left( \frac{1}{\varepsilon_a} - \frac{1}{\varepsilon_b} \right) \sum_{j=1}^2 \sum_{l=-\infty}^{\infty} S(r - u_j(l, l')). \tag{5.37}$$



**Рис. 19. Квадратная решетка двумерного фотонного кристалла**

Функция  $S(r)$  задается следующим образом

$$S(r) = \begin{cases} 1 & |r| \leq r_a \\ 0 & |r| > r_a \end{cases}, \quad (5.38)$$

а двумерные вектора  $u_j(l, l')$  целочисленных аргументов  $l$  и  $l'$  задаются в виде:

$$\begin{aligned} u_1(l, l') &= (al, al' + r_a + d), \\ u_2(l, l') &= (al, -al' - r_a - d). \end{aligned} \quad (5.39)$$

Здесь и далее для простоты вычислений будем полагать, что толщина фотонного кристалла  $L$  и расстояние от поверхностей до осей цилиндров первого (последнего) ряда  $d$  соразмерны периоду  $a$  фотонного кристалла:

$$r_a + d = \frac{a}{2}, \quad L = Na. \quad (5.40)$$

Фурье-амплитуды  $k_{nm}$  разложения (5.32) определяются в виде двойных интегралов:

$$k_{nm} = \frac{1}{2aL} \int_0^a dx \int_{-L}^L dy \frac{1}{\epsilon(x, y)} e^{-i(G_n x + \frac{m\pi}{L} y)}. \quad (5.41)$$

Симметричные пределы интегрирования по  $y$  подразумевают, что для удобства вычисления фотонный кристалл был периодически продолжен на интервал  $-L \leq y \leq 0$ , что не меняет выражения для  $k_{nm}$  в интервале  $0 \leq y \leq L$ . Окончательные выражения для  $k_{nm}$  имеют вид:

$$k_{nm} = \begin{cases} \frac{f}{\varepsilon_a} + \frac{1-f}{\varepsilon_b} & n=0, m=0 \\ 2f \left( \frac{1}{\varepsilon_a} - \frac{1}{\varepsilon_b} \right) \frac{J_1(G_{nm}r_a)}{G_{nm}r_a} & n \neq 0, m=0 \\ 2f \left( \frac{1}{\varepsilon_a} - \frac{1}{\varepsilon_b} \right) (-1)^j \frac{J_1(G_{nm}r_a)}{G_{nm}r_a} & m=2jN \\ 0 & . \end{cases} \quad (5.42)$$



## **Тема 6. Проектирование зеркал: многослойное покрытие, дифракционная однослойная решетка**

Ранее рассматривалось, как с помощью однородных изотропных слоев с разными показателями преломления (чередую их) можно синтезировать просветляющие и/или отражающие многослойные покрытия высокого качества. Первые примеры таких многослойников предложены довольно давно: в работах Тихонравова и его учеников дан устойчивый воспроизводимый алгоритм решения обратной задачи (синтеза) таких покрытий.

В 90-е годы появились первые экспериментальные образцы дифракционных оптических элементов, структура которых отличалась от стопки однородных слоев. Для математического моделирования такого типа устройств и решения обратных задач потребовалась разработка новых алгоритмов и методов решения. Среди тех, кто внес заметный вклад в построение этих методов, необходимо назвать имена таких исследователей, как М. Невьер, Е. Попов, Ф. Лаланн, Л. Ли и многих других. Наиболее существенный с вычислительной точки зрения вклад в разработку устойчивых методов численного решения задач рассеяния света на многослойных оптических решетках внес М. Мохарам с соавторами. В России этими проблемами активно занимались, в частности, В.А. Сойфер с соавт. [23].

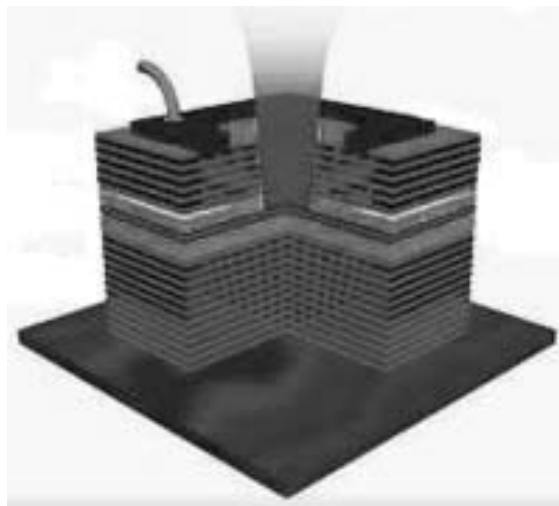
При проектировании по методу Тихонравова зеркал в виде стопки однородных слоев высокая эффективность в широковолновом диапазоне достигается при большой толщине устройств. В то же время создание зеркал на основе дифракционных решеток [24] позволяет достичь таких же параметров, но при более скромных (более чем на порядок меньших толщинах) размерах устройства.

Во втором учебном пособии "Методы спаренных волн расчета

оптических покрытий" подробно рассмотрена модель прохождения монохроматической электромагнитной волны через однослойные и многослойные дифракционные решетки.

Примером технологического устройства, для изготовления которого требуется проектирование оптического дифракционного покрытия, является лазер для чтения и записи информации на компакт-дисках. Лазеры типа VCSEL, благодаря своей эффективности, нашли широкое применение в электронике, а также в лабораторных экспериментах, например, по анализу биологических клеток.

Поясним, VCSEL – это vertical-cavity surface-emitting lasers, т. е. поверхностно-излучающий лазер с вертикальным резонатором (рис. 20) – высокоэффективная разновидность лазерного диода.



**Рис. 20. Поверхностно-излучающий лазер с вертикальным резонатором**

Ранние версии полупроводниковых лазеров использовали в качестве зеркал естественные сколы кристаллов, которые обеспечивали коэффициент отражения в 30%. Это не слишком много, если учесть, что зеркала в лазере обеспечивают многократный пробег фотонов через рабочую среду, где они вызывают генерацию новых фотонов. Вся эта лавина накапливается и в конечном счёте выходит через одно из зеркал

(полупрозрачное) в виде лазерного луча. Для лазеров в своё время были разработаны многослойные зеркала с отражением от 99 до 99,9%.

Зеркалами для VCSEL служит пара так называемых распределённых рефлекторов Брэгга. Состоит такое зеркало из десятков чередующихся чрезвычайно тонких слоёв полупроводников двух типов: арсенида галлия (GaAs) с индексом преломления 3,6 и арсенида алюминия-галлия (AlGaAs) с индексом преломления 3.

Такая комбинация, которая может быть синтезирована по методу Тихонравова, при условии достаточного числа слоёв (порядка 80) и создаёт зеркало с необычайно высоким коэффициентом отражения. Каждая пара соседних слоёв возвращает назад лишь небольшую часть упавшего на них света, а 80 слоев отражают практически весь свет.

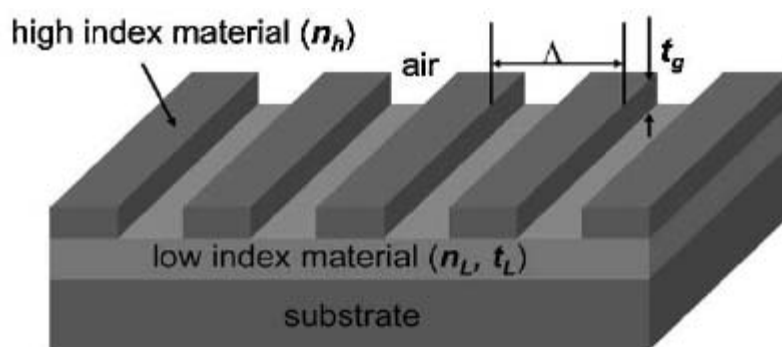
Однако общая толщина такого пирога может достигать 5 мкм, что для ряда перспективных применений очень много. Но главное в том, что высокая точность, необходимая для формирования этих слоёв, требует сложного производственного процесса, что отражается и на цене такого зеркала.

Изобретение американцев называется «Контрастная решётка с высоким индексом преломления и шагом меньше длины волны» (high-index contrast sub-wavelength grating — HCG). Это новое зеркало, построенное физиками США, отражает свыше 99,9% падающего излучения, а толщина его составляет всего 0,23 мкм.

В HCG (рис. 21) имеется всего два рабочих слоя: один из них с низким показателем преломления, а другой — с высоким, например, арсенид алюминия-галлия. Но не сплошной, а сформированный в виде решётки, с углублениями, разделёнными расстоянием, меньшим, чем длина волны падающего света. Свет этот направляется в углубления, а там, сталкиваясь с границей раздела полупроводник-воздух, отбрасывается

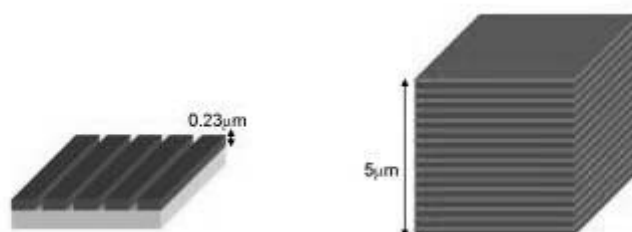
назад.

Исследователи отметили, что воздух, как компонент с низким индексом преломления, могли бы заменить другие материалы. Например, диоксид кремния, с индексом преломления, равным 1,5.



**Рис. 21. Зеркало на основе HCG-рефлектора**

Зеркало HCG в 20 раз тоньше рефлектора Брэгга, при этом показывает отражающую способность больше 99,9%, и изготавливать его намного проще и дешевле.



**Рис. 22. Схематическое сравнение зеркала HCG (слева) и распределённого рефлектора Брэгга.**

На разрезе HCG (рис. 22) тёмным цветом показана полупроводниковая решётка, светлым— воздух (иллюстрация Michael Huang, UC Berkeley)

Поскольку лазеры типа VCSEL используются в оптических коммуникациях, оптических мышах и других системах, требующих низкого расхода энергии, высокоэффективные и при том дешёвые и

технологичные зеркала HCG смогли бы обеспечить если не революцию, то существенный скачок вперёд при создании DVD-приводов, лазерных принтеров, оптических компьютеров.

Новое зеркало работает в более широком диапазоне частот, чем зеркало Брэгга. А это важно, поскольку оптические технологии всё больше полагаются на сине-фиолетовые лазеры. Более короткая длина волны даёт возможность применить более высокую плотность упаковки, к примеру, питов (углубления, кодирующие биты информации) на лазерных дисках.

Исследователи также работают над мобильным зеркалом HCG для микроэлектромеханических систем (MEMS), вроде лазеров с перестраиваемой длиной волны. Сокращение размера зеркала лазера означает существенное сокращение веса, что является особенно важным для быстродействующих устройств MEMS.

По мнению авторов нового зеркала, их детище можно напечатать на той или иной поверхности, что позволит создавать по новой технологии тонкие и лёгкие органические полимерные дисплеи.

Тонкие полупроводниковые полоски, которые и визуально не видны, в перспективе станут основой для массы новых устройств, извлекающих выгоду из впечатляющего параметра отражения HCG в 99,9%.

Стремление к микроминиатюризации устройств и составляющих их элементов и открывшиеся новые возможности привели к необходимости разработки устойчивых методов проектирования дифракционных оптических элементов (ДОЭ). В числе лучших инноваций 2007 г. в области нанотехнологий вошли решётки из наноантенн, отпечатанные на тонкой и гибкой подложке.

Исследователи из национальной лаборатории Айдахо (Idaho National Laboratory — INL) в сотрудничестве со специалистами из американской

компании MicroContinuum и университета Миссури (University of Missouri) создали удивительный прототип солнечной батареи, за которую получили в 2007 г. одну из премий Nano 50.

О том, что для электромагнитных волн видимого диапазона можно строить антенны, наподобие радиоантенн (только в соответствующем масштабе), учёные знают сравнительно давно. Исследователи из INL предложили для инфракрасного излучения следующую конструкцию: на подложке из изолятора разместили систему спиралей. Экспериментаторы пишут, что каждая такая спираль имеет поперечник в  $1/25$  диаметра человеческого волоса (т. е. получается, что диаметр антенны составляет порядка 2-4 мкм). Это сопоставимо с длинами волн инфракрасной части спектра.

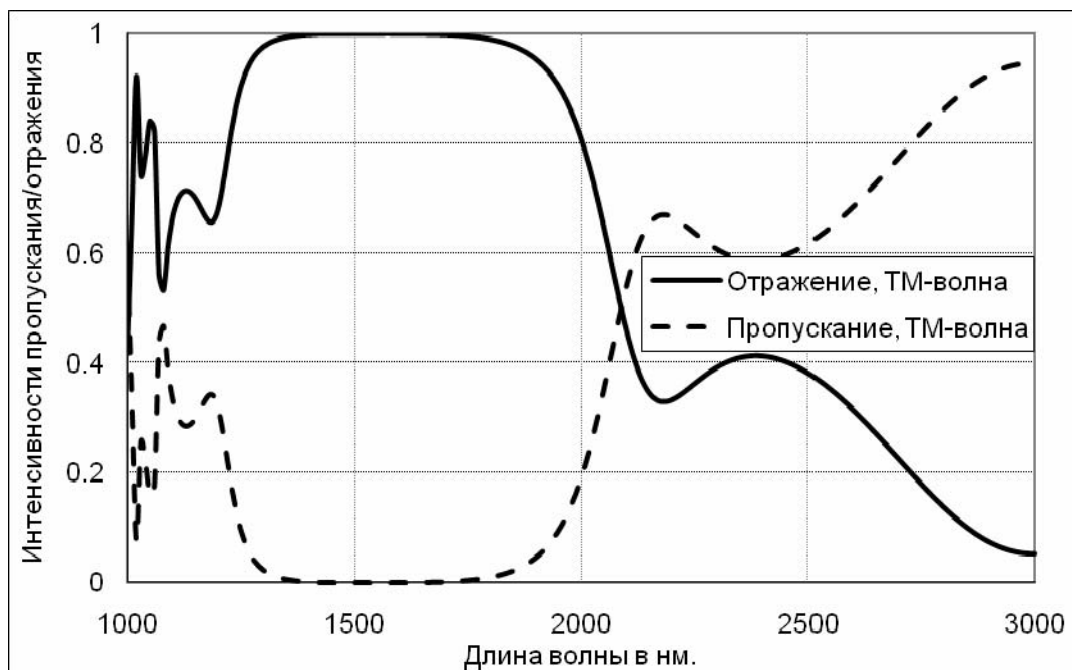
Падение ИК-лучей на такую спираль наводит в ней ток, но этот ток получается не за счёт фотоэффекта, а по принципу СВЧ-антенны.

Ток в каждой такой СВЧ-антенне переменный, и его частота очень велика (в прототипе — десять терагерц). Найти экономичный способ преобразования переменного тока в постоянный непросто, но весьма перспективно. По оценке INL, каждая наноантенна может обращать в электричество целых 80% доступной ей световой энергии.

Эти примеры демонстрируют актуальность и необходимость решения задач проектирования многослойных оптических покрытий, состоящих из одномерных либо двумерных решеток. Толщины слоев и характерные размеры решеток в таких покрытиях должны быть меньше длины волны излучения.

Во втором учебном пособии "Методы спаренных волн расчета оптических покрытий" приведены расчетные формулы решения прямой задачи дифракции монохроматических электромагнитных волн на многослойных дифракционных покрытиях. Воспользуемся ими для

решения задачи синтеза на примере расчета практически стопроцентного отражающего покрытия для лазерного излучения с опорной длиной волны в 1.5 мкм.



**Рис. 23. Пример расчета оптимальной геометрии дифракционного зеркала**

Расчетная модель изображена на рис. 21. Индекс преломления материала дифракционной решетки (3.48), индекс преломления промежуточного слоя (1.47) и индекс преломления кремниевой подложки (3.48) считаются заданными. Оптимизация геометрических параметров решетки с целью получения максимального отражения р-поляризованной волны излучения в окрестности длины волны в 1.5 мкм приводит к решению со следующими параметрами:

период решетки равен 700 нм;

высота решетки равна 460 нм;

фактор заполнения решетки составляет 75%;

высота промежуточного слоя равна 830 нм.

В результате получено зеркало с коэффициентом отражения свыше 99.9% в диапазоне длин волн от 1410нм до 1680нм.

Задача решена с помощью минимизации целевого функционала методом нулевого порядка, при котором на каждом шаге решалась прямая задача вычисления интенсивностей прошедшего и отраженного р-поляризованного света.



## **Литература**

1. *Lagarias J. C., Reeds J. A., Wright M. H., Wright P. E.* Convergence properties of the Nelder-Mead simplex method in low dimensions // SIAM J. OPTIM. —1998. —Vol. 9.— p. 112-147
2. *Nelder J. A., Mead R.* A simplex method for function minimization// Computer Journal. 1965, 7, 308-313.
3. <http://www.cc.gatech.edu/people/home/ananth/docs/lmtut.pdf>
4. *Химмельблау Х.* Нелинейное программирование. - М.: Мир, 1969.
5. *Levenberg K.* A method for the solution of certain problems in least squares// Quart. Appl. Math. – 1944. – Vol. 2. – p. 164–168.
6. *Marquardt D.* An algorithm for least-squares estimation of nonlinear parameters// SIAM J. Appl. Math. – 1963.– Vol. 11. – p. 431–441.
7. *Сотин В. Е.* Брэгговское отражение в периодических структурах. – М.: УДН, 1981.
8. *Тихонов А. Н., Васильева А. Б., Свешников А. Г.* Дифференциальные уравнения. – М.: Наука, 1980.
9. *Тихонов А.Н., Гласко В.Б., Тихонравов А.В.* О синтезе многослойных покрытий // Журн. вычислит. матем. и матем. физики. – 1974. – Т.14. – №1. – С.135-144.
10. *Тихонов А. Н.* О регуляризации некорректно поставленных задач// Докл. АН СССР. – 1963. – 153. – № 1. – 49-52.
11. *Тихонов А. Н., Дмитриев В. И.* О методах решения обратной задачи теории антенн// Вычисл. Методы и программирование. Вып. XIII. – М.: ВЦ МГУ, 1969, – С. 209-215.
12. *Бахрах Л. Д., Кременецкий С. Д.* Некорректно поставленные задачи и

- теория синтеза излучающих систем// Докл. АН СССР. – 1968. – 178. – № 4. – С. 825-828.
13. *Розенберг Г. В.* Оптика тонкослойных покрытий. – М.: Физматгиз, 1958.
14. *Умеров Р. И., Шкляревский И. Н., Пономарева Г. И.* Изготовление двухслойных ахроматических просветляющих покрытий // Оптика и спектроскопия. – 1969. – 26. – Вып. 3. – С. 1027-1030.
15. *Гласко В. Б., Кулик Н. И., Тихонов А. Н.* Об определении геоэлектрического разреза на основе метода регуляризации// Журнал вычисл. матем. и матем. физ. – 1972. – 12. – № 1. – С. 139-149.
16. *Люстерник Л. А., Соболев В. И.* Элементы функционального анализа. – М.: Наука, 1965.
17. *Будак Б. М., Васильев Ф. П.* Приближенные методы решения задач оптимального уравнения. Вып. II. Ротапринт ВЦ МГУ, 1969.
18. *Ловецкий К.П., Жуков А.А.* Методы расчета рефракционных индексов тонких кристаллических пленок// Вестник РУДН. Сер. Прикладная и компьютерная математика. – 2005. – Т.4. – №1. – С. 56-66.
19. *Телен А.* Конструирование многослойных интерферированных светофильтров// Физика тонких пленок. Т. 5. – М.: Мир. – 1972.
20. *Гусев Е.Л.* Априорное сужение области поиска в волновых задачах синтеза неоднородных структур// Математическое моделирование. – 2000. – № 4. – Т.12.
21. *Гусев Е.Л.* Качественные закономерности структуры оптимальных композиционных конструкций при волновых воздействиях. // Математическое моделирование. – 2000. – № 7.– Т.12. – С.7-10.
22. *Гусев Е.Л.* Математические методы синтеза слоистых структур. // Новосибирск: Наука, 1993.

23. Методы компьютерной оптики. / Под ред. В.А. Сойфера.– М.: Физматлит, 2003.
24. *Carlos F. R. Mateus, Michael C. Y. Huang, Lu Chen, Connie J. Chang-Hasnain, Yuri Suzuki.* Broad-Band Mirror (1.12–1.62  $\mu$ m) Using a Subwavelength Grating. // IEEE PHOTONICS TECHNOLOGY LETTERS. 2004. – Vol. 16. – N. 7. – P. 1676-1678.
25. *Лаврентьев М.М., Романов В.Г., Шишатский С.П.* Некорректные задачи математической физики и анализа. – М.: Наука. 1987 (английский перевод: AMS, 1987)
26. *Paukshto M., Lovetskiy K., Zhukov A.* Dielectric Constants of Display Optical Components// SID Symposium Digest of Technical Papers. – May 2007. – Vol. 38. – Issue 1. P. 410-413.

## ОПИСАНИЕ КУРСА И ПРОГРАММА

---

### **Цели и задачи магистерской программы «Оптика наноструктур»**

Целью учебно-методического комплекса магистерской программы «Оптика наноструктур» является формирование у студентов четкого представления об основных принципах функционирования современных дифракционных оптических элементов и устройств, тонкопленочных многослойных покрытий; о законах взаимодействия электромагнитного излучения видимого диапазона с материалом. Целью является также изучение способов и возможностей математического синтеза и компьютерного проектирования дифракционных оптических покрытий. Полученные знания закрепляются в оптической лаборатории и дисплейном классе на примерах изучения конкретных дифракционных оптических элементов и многослойных покрытий со сложной геометрией.

Задачей учебно-методического комплекса магистерской программы «Оптика наноструктур» является обучение студентов навыкам самостоятельного анализа технических заданий на проектирование дифракционных оптических элементов и устройств. Они должны научиться выбирать из имеющихся в наличии алгоритмов и программ математического синтеза или разрабатывать их самостоятельно. В результате обучения обретут навыки ориентации в научной и бизнес информации с целью выбора нужной функции или нужного инструмента для реализации известной функции в области проектирования и создания дифракционных оптических наноструктур.

## **Цели и задача курса «Математический синтез оптических наноструктур»**

Курс «Математический синтез оптических наноструктур» является составной частью магистерской программы «Оптика наноструктур». Магистерская программа «Оптика наноструктур» реализуется в рамках направления «Прикладная математика и информатика» и направления «Прикладная математика и физика», а возможно и других направлений. В составе магистерской программы «Оптика наноструктур» курс «Математический синтез оптических наноструктур» является обязательным, привязанным к семестру. Для других магистерских программ этот курс может быть курсом по выбору без привязки к семестру или факультативным на усмотрение методической комиссии программы. Курс носит теоретический и практический характер.

Целью курса является подробное ознакомление студентов с устойчивыми современными методами численного решения математических задач, возникающих при проектировании дифракционных нанометровых оптических элементов и устройств, разработке тонкопленочных покрытий с характерными размерами порядка длины волны излучения. Задачи проектирования стали особенно актуальны в последние годы в связи с широким применением наноэлементов и тонких (менее одного микрометра толщиной) пленок в производстве жидкокристаллических дисплеев, солнечных батарей на основе диэлектриков, фотоэмиссионных диодов, просветляющих покрытий, поляризаторов, миниатюрных лазеров, управляемых оптических элементов. Задачи оптики наноструктур практически не поддаются аналитическому решению, поэтому важным является изучение эффективных численных методов, используемых при решении задач математического синтеза, приобретение

навыков создания программного обеспечения для компьютерного дизайна оптических наноструктур.

Задачей курса «Математический синтез оптических наноструктур» является формирование у студентов навыков работы на современной измерительной аппаратуре. Задачей курса является также обучение студентов использованию строгих методов связанных волн при решении задач моделирования современных оптических устройств на основе тонкопленочных покрытий и дифракционных оптических элементов. Это позволит им при необходимости разрабатывать новое программное обеспечение. Безусловной задачей курса является также освоение существующего программного обеспечения, ориентированного на расчет и проектирование оптических покрытий. В результате обучения они получают умение и навыки правильно оценить сложность научно-исследовательских и конструкторских заданий на разработку дифракционных оптических элементов и устройств, аргументированно выбирать метод решения конструкторской задачи, а затем экономично и эффективно выполнять компьютерный дизайн требуемого дифракционного оптического покрытия или устройства.

Трудоемкость курса составляет 3 кредита; 2 часа лекций и 2 часа лабораторных занятий в дисплейном классе в неделю.

### **Инновационность курса.**

Курс является инновационным по содержанию и по литературе, он включает в себя последние научные достижения в области решения задач дифракционной оптики, когда характерные размеры исследуемых объектов не превышают либо сравнимы с длиной волны оптического излучения. Эта область знаний интенсивно развивалась в последнее время, но лишь недавно были созданы устойчивые алгоритмы и разработаны

численные методы решения задач для многослойных решеток. Следует отметить, что для оптических однослойных и многослойных решеток с характерными размерами больше длины волны оптического излучения устойчивые методы решения известны с середины прошлого века. Сейчас алгоритмы решения оптических задач в субволновой области распространяются на объекты со сложной геометрией, такие как двумерные решетки с произвольным профилем, трехмерные решетки (фотонные кристаллы) и на анизотропные материалы. Они востребованы, поскольку позволяют создавать математические модели взаимодействия излучения с веществом в наномасштабах, а затем с их помощью проектировать новые эффективные устройства в высокотехнологичных областях медицины, энергетики, инфокоммуникаций и приборостроения.

В ходе проведения занятий по этому курсу разработчики предполагают использование традиционных методик преподавания, принятой в странах болонской системы образования, то есть с использованием кредитной системы оценки знаний.

Наряду с традиционными элементами преподавания математических методов решения прикладных задач, разработчики курса предполагают воспользоваться хорошо зарекомендовавшим себя опытом МФТИ и подобных вузов. Для этого в рамках подпрограммы «Оптика наноструктур» осуществляется закупка уникального измерительного и аналитического оборудования для выполнения измерений разнообразных характеристик оптических наноустройств с целью использования этого оборудования в учебном процессе и для проведения научно-исследовательских работ преподавателями, аспирантами и студентами.

По окончании магистратуры по направлению «Оптика наноструктур» выпускники Российского университета дружбы народов станут конкурентно-способными специалистами в области проектирования

современных оптических устройств, которые не будут испытывать затруднений при последующем трудоустройстве.

Данное направление научно-практических разработок сформировалось лишь в последние 10 – 15 лет. Поэтому наблюдается сильный дефицит учебно-методической литературы не только в России, но и во всем мире. Разрабатываемые в рамках инновационной программы «Оптика наноструктур» учебные пособия восполнят в некоторой степени этот пробел и составят основной список литературы для слушателей курсов. Вместе с ними следует использовать несколько учебников и монографий, вышедших в свет к настоящему времени и перечисленные в списке литературы. Курс базируется на публикациях научных статей мировых лидеров исследований в данной области в научной периодике, диссертационных работах их учеников, включающих работы по непосредственному моделированию, дизайну и последующему изготовлению лабораторных образцов оптических элементов и устройств. В список дополнительной и рекомендуемой литературы включены все научно-исследовательские публикации, положенные в основу предлагаемого курса.

В качестве практических заданий, курсовых работ и тем рефератов слушателям магистерской программы будут предложены актуальные проблемы и задачи, решение которых востребовано современным уровнем развития высокотехнологичных отраслей промышленности и научно-исследовательских лабораторий.



## **Структура курса (с указанием количества часов аудиторных/самостоятельной работы на темы)**

### **Темы лекций**

Тема 1. История возникновения задач проектирования в оптике. Цветные витражи готических соборов это сверхрешетка, возникающая в результате случайно проведенных экспериментов. Переход к осознанному проектированию оптических устройств на основе геометрической оптики – бинокли, телескопы, параболические зеркала и пр. (1 пара)

Тема 2. Скалярный подход к расчету дифракционных оптических элементов. Обзор итеративных методов решения обратных задач скалярной теории дифракции применительно к синтезу ДОЭ. Решение задачи синтеза ДОЭ методом решения нелинейного интегрального уравнения, нелинейность которого связана с операцией взятия модуля от комплексной амплитуды когерентного света. (3 пары)

Тема 3. Градиентные методы решения задачи синтеза бинарных дифракционных решеток. Расчет параметров отражающих решеток со ступенчатым профилем для формирования заданных значений интенсивностей дифракционных порядков. (3 пары)

Тема 4. Строгий подход к анализу двумерных дифракционных решеток из изотропных диэлектрических материалов. Проектирование двумерных дифракционных решеток из изотропных диэлектрических материалов с заданными энергетическими характеристиками пропускания и отражения.

Строгий подход к анализу двумерных дифракционных решеток из анизотропных диэлектрических материалов. Проектирование двумерных дифракционных решеток из анизотропных диэлектрических материалов с заданными энергетическими характеристиками пропускания и отражения. (4 пары)

Тема 5. Точный метод анализа и синтеза дифракционных оптических элементов (RCWA). Расчет многослойной дифракционной решетки, поворот ориентации слоев которой поворачивает поляризацию проходящего света, строгим методом связанных волн RCWA. (2 пары)

Тема 6. Проектирование зеркала с 99% отражения заданной линейной поляризации в наиболее широком (в области видимого света) диапазоне.

Проектирования и оптимизация параметров ячеек солнечных батарей с учетом диапазона поглощаемых волн, сезонных колебаний освещенности и углов падения солнечного света. (2 пары)

Тема 7. Основы моделирования и проектирования фотонных кристаллов. (1 пара).

### **Темы семинарских и практических занятий**

- Проектирование оптических устройств методами геометрической оптики.
- Основы проектирования оптических устройств методами скалярной волновой оптики.

- Проектирование оптических устройств методом решения нелинейного интегрального уравнения.
- Построение целевой функции ошибок для градиентного метода для расчета отражающих решеток со ступенчатым профилем.
- Построение целевой функции ошибок для градиентного метода для расчета диэлектрических бинарных решеток для ТМ-поляризации.
- Построение целевой функции ошибок для градиентного метода для расчета диэлектрических бинарных решеток для ТЕ-поляризации.
- Модальный метод Фурье для задачи дифракции на многослойной двумерной решетке из изотропного диэлектрического материала
- Система обыкновенных дифференциальных уравнений для задачи дифракции на многослойной двумерной решетке из изотропного диэлектрического материала
- Задача оптимизации геометрических параметров диэлектрической решетки с целью максимизации коэффициента отражения.
- Задача оптимизации геометрических параметров диэлектрической решетки с целью максимизации коэффициента поглощения.
- Расчетные формулы моделирования и проектирования фотонных кристаллов.

## **Темы коллоквиумов и контрольных работ**

Проектирование оптических устройств методом решения нелинейного интегрального уравнения.

Построение целевой функции ошибок для градиентного метода для расчета диэлектрических бинарных решеток для ТМ-поляризации.

Построение целевой функции ошибок для градиентного метода для расчета диэлектрических бинарных решеток для ТЕ-поляризации.

Задача оптимизации геометрических параметров диэлектрической решетки с целью максимизации коэффициента отражения.

Задача оптимизации геометрических параметров диэлектрической решетки с целью максимизации коэффициента поглощения.

### **Описание системы контроля знаний:**

### **Общие правила выполнения контрольных заданий;**

#### ***Требования к оформлению работы***

##### *Постановка задачи.*

1. Краткая формулировка задачи.
2. Развернутая постановка задачи с указанием основных режимов работы и их сценариев.

##### *Алгоритм решения.*

1. Математическое описание алгоритма.
2. Структура алгоритма ядра программы (укрупненная блок-схема).

### *Тестирование.*

1. Описание основных режимов тестирования алгоритма и программы и результатов работы программы.
2. Список возможных ошибок и аномалий, описание реакции программы на них.

### *Заключение.*

Содержит общие комментарии и замечания исполнителя о выполненной работе.

### *Приложение.*

Приложение может содержать текст программы (полная распечатка или распечатка алгоритма ядра программы).

Работа должна быть представлена в виде распечатанного текста и на дискете.

### ***Рекомендации к составлению отчета***

#### *Оформление.*

отчет по работе должен быть оформлен в форме Word-файла.

#### *Содержание отчета.*

Каждый пункт задания вычислительного эксперимента должен найти свое отражение в отчете.

Каждый раздел отчета должен содержать:

формулировку цели эксперимента

описание исходных данных - приближаемая функция, интервал и порядок приближения, метод приближения и т.п.

результаты эксперимента, представленные в форме таблиц, гистограмм и графиков

иллюстрационный материал в виде копий экрана с графиками зависимостей погрешности приближения, вида приближаемой функции и т.п.

выводы, следующие из результатов эксперимента в контексте его цели.

### **Шкала оценок, итоговые оценки (методика выставления)**

Бально-рейтинговая методика оценки уровня знаний по обязательной дисциплине «Методы связанных волн расчета оптических покрытий», привязанной к семестру

Порядок начисления баллов за семестр.

Контрольная работа № 1: 0 – 40 баллов

    Теоретические вопросы: 0 – 10 баллов

    Практические задания: 0 – 30 баллов

Контрольная работа № 2: 0 – 40 баллов

    Теоретические вопросы: 0 – 10 баллов

    Практические задания: 0 – 30 баллов

Контрольная работа № 3: 0 – 20 баллов

    Теоретические вопросы: 0 – 20 баллов

**Шкала бально-рейтинговой системы.**

Баллы за семестр	Автоматическая оценка		Баллы за итоговый контроль знаний	Общая сумма баллов	Итоговая оценка
	Итоговая оценка	Дополнительные баллы			
78 – 80	зачет	по 5 баллов за каждый свыше 76**	0 – 20*	86 – 100	зачет
41 – 77	Нет	Нет	0 – 20	51 – 97	зачет
			0 – 20	41 – 50	незачет
< 41	незачет	Нет	Нет	Нет	незачет

\* студент имеет право не проходить итоговый контроль знаний.

\*\* дополнительные баллы начисляются автоматически:

за 86 баллов, набранных в семестре, начисляется дополнительно 6 баллов

(общая сумма баллов – 92);

за 87 баллов – 12 баллов (99);

за 88 баллов – 18 баллов (106);

за 89 баллов – 24 балла (113);

за 90 баллов – 30 баллов (120).

### **Академическая этика, соблюдение авторских прав.**

Все имеющиеся в тексте сноски тщательно выверены и снабжены «адресами». Авторы не включали в свою работу выдержки из работ других авторов без указания на это, не пересказывали чужих работ близко к тексту без отсылки к ним. Авторы также не использовали чужих идей без указания первоисточников. Это касается и источников, найденных в интернете. В необходимых случаях указан полный адрес сайта.

## **Программа курса УМК:**

### Аннотированное содержание курса.

Первый модуль трудоемкостью в 1 кредит составляют:

- теоретический материал, излагаемый в первых трех темах,
- практическими занятиями в дисплейном классе в течение 14 академических часов,
- самостоятельные занятия.

В конце этого модуля проводится промежуточный контроль знаний.

Второй модуль трудоемкостью в 1 кредит составляют:

- теоретический материал, излагаемый в четвертой теме.
- практическими занятиями в дисплейном классе в течение 8 академических часов,
- самостоятельные занятия.

В конце этого модуля проводится промежуточный контроль знаний.

Третий модуль трудоемкостью в 1 кредит составляют:

- теоретический материал, излагаемый в пятой, шестой и седьмой темах,
- практическими занятиями в дисплейном классе в течение 10 академических часов,
- самостоятельные занятия.

В конце этого модуля проводится итоговый контроль знаний.

Список обязательной и дополнительной литературы с указанием соответствия разделов источника (постранично) разделам читаемого курса

### **Список обязательной литературы.**

1. Математический синтез оптических наноструктур /Под ред.К.П.



- Ловецкого: Учебное пособие. - М.: Изд-во РУДН (готовится к печати).
2. Методы компьютерной оптики/Под ред. В.А. Соифера: Учеб. для вузов. — 2-е изд., испр. - М.: ФИЗМАТЛИТ, 2003. - 688 с.
  3. Ильинский А.С., Кравцов В.В., Свешников А.Г. Математические модели электродинамики. М.: Высшая школа. 1991. 224с.
  4. Doskolovich L. L., Kharitonov S. I., Petrova O. I., Soifer V. A. Opt. And Lasers in Eng. V.29 p249 1998.
  5. Л.Л. Досколович Расчет бинарных диэлектрических решеток и одномерных ДОЭ в рамках электромагнитной теории.// Компьютерная оптика, Выпуск 19, p.21-28 1999
  6. Jianhua Jiang. PhD. Thesis. Rigorous analysis and design of diffractive optical elements. The University of Alabama in Huntsville, 2000.
  7. Panfilo C. Deguzman. PhD. Thesis. Stacked subwavelength gratings for imaging polarimetry. The University of Alabama in Huntsville, 2000.
  8. M. Neviere, E. Popov. Light Propagation in Periodic Media: Differential Theory and Design Marcel Dekker Inc, 2002, 432 p.
  9. Li, Lifeng. New formulation of the Fourier modal method for crossed surface-relief gratings. // J.O.S.A. A/v.14, p.2758-2767.
  10. Carlos Fernando Rondina Mateus. PhD. Thesis. Tunable Optoelectronic Devices, UNIVERSITY of CALIFORNIA at BERKELEY, 2004, 217 p.

**Список дополнительной литературы и источников в интернете.**

11. Lifeng Li. 'Reformulation of the Fourier modal method for surface-relief gratings made with anisotropic materials', Journal of Modern Optics, (1998), v.45, p.1313 – 1334
12. Lifeng Li. Fourier modal method for crossed anisotropic gratings with arbitrary permittivity and permeability tensors. // J. Opt. A: Pure Appl.

Opt. 2003 v.5 p.345-355

13. Lifeng Li. A modal analysis of lamellar diffraction gratings in conical mountings, Journal of Modern Optics, (1993), v.40, p.553 – 573.
14. Т.В. Амочкина. Алгоритм синтеза многослойных оптических покрытий, основанный на теории эквивалентных слоев. Вычислительные методы и программирование, т.6, 2005.
15. Е.Л. Гусев. Априорное сужение области поиска в волновых задачах синтеза неоднородных структур. Математическое моделирование, № 4, т.12, 2000.
16. [www.en.wikipedia.org/wiki/Solar\\_cell](http://www.en.wikipedia.org/wiki/Solar_cell)
17. [http://nobelprize.org/nobel\\_prizes/physics/laureates/2000/alferov-lecture.pdf](http://nobelprize.org/nobel_prizes/physics/laureates/2000/alferov-lecture.pdf)
18. Nanotechnology in energy applications  
<http://www.im.isu.edu.tw/seminar/2005.11.16.pdf>

#### Темы рефератов, курсовых работ, эссе

##### ***Темы рефератов.***

- Обзор литературы по задаче проектирования четвертьволновых пластинок.
- Обзор литературы по задаче проектирования полуволновых пластинок.
- Обзор литературы по задаче проектирования поляризаторов. Из однородных изотропных слоев, анизотропных слоев, дифракционных решеток.
- Обзор литературы по задаче проектирования солнечных батарей.

##### ***Темы курсовых работ***

1. Проектирование тонкопленочных фильтров с заданными свойствами по пропусканию.

2. Проектирование фильтров с заданными свойствами по отражению.
3. Проектирование поляризационных фильтров с заданными свойствами.
4. Проектирование многослойных зеркал
5. Проектирование зеркал на одномерных дифракционных решетках
6. Задача проектирования солнечных батарей.

Учебный тематический план курса УМК (календарный план,  
структурированный по видам учебных занятий)

***Календарный план (20 недель) учебных занятий по обязательной дисциплине «Математический синтез оптических наноструктур», привязанный к третьему семестру магистратуры.***

Виды и содержание учебных занятий				
Неделя	Лекции	Число часов	Лабораторные занятия	Число часов
1	История возникновения задач проектирования в оптике. Переход к проектированию оптических устройств на основе геометрической оптики – бинокли, телескопы, параболические зеркала	2	Проектирование оптических устройств методами геометрической оптики.	2
2	Скалярный подход к расчету дифракционных оптических элементов. Обзор итеративных методов решения обратных задач скалярной теории дифракции	2	Основы проектирования оптических устройств методами скалярной волновой оптики.	2

	применительно к синтезу ДОЭ.			
3	Решение задачи синтеза ДОЭ методом решению нелинейного интегрального уравнения.	2	Проектирование оптических устройств методом решения нелинейного интегрального уравнения.	2
4	Решение задачи синтеза ДОЭ методом решению нелинейного интегрального уравнения. Продолжение.	2	Проектирование оптических устройств методом решения нелинейного интегрального уравнения.	2
5	Строгий подход к анализу одномерных бинарных дифракционных оптических элементов.	2	Построение целевой функции ошибок для градиентного метода для расчета отражающих решеток со ступенчатым профилем.	2
6	Дифракция на отражающих решетках со ступенчатым профилем.	2	Построение целевой функции ошибок для градиентного метода для расчета	2

			диэлектрических бинарных решеток для ТМ-поляризации.	
7	Дифракция на пропускающих диэлектрических решетках.	2	Построение целевой функции ошибок для градиентного метода для расчета диэлектрических бинарных решеток для ТЕ-поляризации.	2
8	Промежуточный контроль знаний			2
9	Строгий подход к анализу двумерных дифракционных решеток из изотропных диэлектрических материалов.	2	Модальный метод Фурье для задачи дифракции на двумерной решетке из изотропного диэлектрического материала	2
10	Проектирование двумерных дифракционных решеток из изотропных диэлектрических материалов с заданными энергетическими характеристиками пропускания и отражения.	2	Система обыкновенных дифференциальных уравнений для задачи дифракции на двумерной решетке из изотропного диэлектрического материала	2

11	Строгий подход к анализу двумерных дифракционных решеток из анизотропных диэлектрических материалов.	2	Модальный метод Фурье для задачи дифракции на двумерной решетке из анизотропного диэлектрического материала	2
12	Проектирование двумерных дифракционных решеток из анизотропных диэлектрических материалов с заданными энергетическими характеристиками пропускания и отражения.	2	Система обыкновенных дифференциальных уравнений для задачи дифракции на двумерной решетке из анизотропного диэлектрического материала	2
13	Промежуточный контроль знаний			2
14	Точный метод анализа и синтеза дифракционных оптических элементов (RCWA). Расчет многослойной дифракционной решетки, поворот ориентации слоев	2	Модальный метод Фурье для задачи дифракции на многослойной двумерной решетке из изотропного диэлектрического материала	2

	<p>которой поворачивает поляризацию проходящего света, строгим методом связанных волн RCWA.</p>			
15	<p>Расчет многослойной дифракционной решетки, поворот ориентации слоев которой поворачивает поляризацию проходящего света, строгим методом связанных волн RCWA. Продолжение.</p>		<p>Система обыкновенных дифференциальных уравнений для задачи дифракции на многослойной двумерной решетке из изотропного диэлектрического материала</p>	
16	<p>Проектирование зеркала с 99% отражения заданной линейной поляризации в наиболее широком (в области видимого света) диапазоне.</p>	2	<p>Задача оптимизации геометрических параметров диэлектрической решетки с целью максимизации коэффициента отражения.</p>	2
17	<p>Проектирования и оптимизация параметров ячеек солнечных батарей с учетом диапазона поглощаемых</p>	2	<p>Задача оптимизации геометрических параметров диэлектрической решетки с целью</p>	2



	волн, сезонных колебаний освещенности и углов падения солнечного света.		максимизации коэффициента поглощения.	
18	Основы моделирования и проектирования фотонных кристаллов.	2	Расчетные формулы моделирования и проектирования фотонных кристаллов.	2
19	Заключительный обзор курса. Консультации по подготовке к итоговому контролю знаний.	2	Заключительный обзор курса. Консультации.	2
20	Итоговый контроль знаний			2