ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ имени академика С.П. КОРОЛЕВА»

ИНСТИТУТ СИСТЕМ ОБРАБОТКИ ИЗОБРАЖЕНИЙ
РОССИЙСКОЙ АКАДЕМИИ НАУК

Н.Л. Казанский

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ОПТИЧЕСКИХ СИСТЕМ

САМАРА 2005

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ имени академика С.П. КОРОЛЕВА»

ИНСТИТУТ СИСТЕМ ОБРАБОТКИ ИЗОБРАЖЕНИЙ
РОССИЙСКОЙ АКАДЕМИИ НАУК

Н.Л. Казанский

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ОПТИЧЕСКИХ СИСТЕМ

Учебное пособие

САМАРА 2005

УДК 535.4, 535.8

Казанский Н.Л. **Математическое моделирование оптических систем:** Учеб. пособие / Самар. гос. аэрокосм. ун-т. - Самара, 2005. - 240 с.

ISBN 5-7883-0379-6

В учебном пособии систематически изложены основные физические подходы, применяемые для моделирования оптических систем: геометрическая оптика, скалярная теория дифракции и строгая электромагнитная теория. Отдельно сформулированы методы моделирования дифракционных оптических элементов и оптических систем с ними. Дается краткая характеристика программных продуктов, предназначенных для моделирования оптических систем и расчета дифракционных оптических элементов.

Учебное пособие предназначено для студентов специальностей и направлений «Прикладные математика и физика», «Прикладная математика и информатика», а также аспирантов и докторантов, обучающихся по специальностям 01.04.05 «Оптика» и 05.13.18 «Математическое моделирование, численные методы и комплексы программ».

Подготовлено и издано при поддержке гранта Президента РФ НШ-1007.2003.01, программы Президиума РАН «Поддержка молодых ученых» и российско-американской программы "Фундаментальные исследования и высшее образование" (BRHE).

Печатается по решению редакционно-издательского совета Самарского государственного аэрокосмического университета.

Рецензенты: д-р ф.-м. наук В.П.Захаров;
д-р ф.-м. наук В.В.Котляр

ISBN 5-7883-0379-6

ВВЕДЕНИЕ

К настоящему времени сложились основы новой методологии научных исследований - математического моделирования и вычислительного эксперимента. Сущность этой методологии состоит в замене исходного объекта его математической моделью и исследовании современными вычислительными средствами математических моделей. Методология математического моделирования бурно развивается, охватывая все новые сферы - от разработки больших технических систем и управления ими до анализа сложнейших экономических и социальных процессов. Поэтому современные программы подготовки по направлениям 010500 (Прикладная математика и информатика) и 010600 (Прикладные математика и физика) обязательно включают курсы математического моделирования. В остальных курсах студентам-математикам приходится изучать весьма формализованные теории конкретных разделов математики и осваивать ряд математических (в том числе, численных) методов. Эти стороны математической подготовки выпускников, несомненно, важны и входят в арсенал средств прикладного математика. Например, изучение ньютоновской механики позволяет увидеть, как принятая модель разворачивается в стройную теорию. Однако при таком изучении студенту не видно, что пришлось преодолеть Ньютону, прежде чем он выдвинул и развил идеи, составляющие теперь общепризнанную теорию. Кроме того, при этом остается в тени вопрос о том, какое отношение имеет эта модель к практическим проблемам, возникающим в повседневной работе физиков и инженеров на производстве. Существующая в России система подготовки специалистов по прикладной математике основывается на сочетании фундаментального математического и прикладного компьютерного образования, что ориентирует студентов на решение точно сформулированных математических задач. Так, может создаться впечатление, что приложение математики сводится просто к подбору подходящих формул, подстановке в них некоторых чисел и взмаху волшебной палочки, в результате чего получается «ответ». При этом формируется определенная отчужденность от важнейших этапов и сторон фактического решения прикладных задач, таких, как осмысление и глубокое проникновение в существо конкретной задачи, и совершенно упускается из виду математическая подготовка, включающая упрощения, принимаемые идеализации и составление математической модели, ее коррекцию и сопоставление результатов исследования с реальным объектом. Без таких навыков «приложение» математики превращается просто в демонстрацию известных математических приемов.

Таким образом, существенное и важнейшее умение прикладного математика заключается в переводе нашего так называемого «реального мира» на язык математики, что позволяет получить более точное представление о существенных свойствах изучаемого объекта (явления, процесса) и в некотором смысле предсказать будущие события. Это обстоятельство как раз и отражает термин «математическое моделирование».

ГЛАВА 1
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ - ИСКУССТВО ПРИМЕНЕНИЯ МАТЕМАТИКИ

1.1. ОСНОВНЫЕ ХАРАКТЕРНЫЕ ЧЕРТЫ МОДЕЛИРОВАНИЯ

Холл (1963) сказал, что целью прикладной математики является математическое осмысление действительности. С другой стороны, инженеру-практику, пожалуй, более важно знать, выдержит ли его мост предполагаемую нагрузку, а администратору больницы — найти способ сократить время, затрачиваемое пациентами местной амбулатории на ожидание, — иными словами, получить конкретные ответы на конкретные вопросы, а не стремиться к более возвышенным целям.

С чего начинает математик-прикладник? На практике исходным пунктом часто является некоторая эмпирическая ситуация, выдвигающая перед исследователем «задачу», на которую требуется найти «ответ». Однако употребление таких слов, как «задача» и «ответ», может ввести в заблуждение. Прежде всего, необходимо установить, в чем именно заключается «задача». Это замечание связано с тем, что реальные ситуации редко бывают четко очерченными, а сложное взаимодействие с окружающей средой часто делает точное описание ситуации затруднительным. Процесс выделения «задачи», поддающейся математическому анализу, часто бывает продолжительным и требует владения многими навыками, не имеющими отношения к математике (например, беседы с коллегами-нематематиками, работающими в данной области, и чтение всевозможной литературы, имеющей отношение к делу, являются важным элементом процесса моделирования).

Часто (но не всегда) параллельно с этой стадией постановки задачи идет процесс выявления основных или существенных особенностей явления (см. рис. 1.1).

Рис. 1.1. Схема математического моделирования

В частности, для физических явлений этот процесс *схематизации* или *идеализации* играет решающую роль, поскольку в реальном явлении участвует множество процессов, и оно чрезвычайно сложно. Некоторые черты явления представляются важными, многие другие — несущественными. Возьмем, к примеру, движение маятника, образованного тяжелым грузом, подвешенным на конце нити. В этой «ситуации» существенным является регулярный характер колебаний маятника, а несущественным обстоятельством — то, что нить белая, а груз черный. После того как существенные факторы выявлены, следующий шаг состоит в переводе этих факторов на язык математических понятий и величин и постулировании соотношений между этими величинами. Как правило, это самая трудная стадия процесса моделирования, причем здесь невозможно дать никаких общих рекомендаций!

После построения модели ее следует подвергнуть проверке. В действительности адекватность модели до некоторой степени проверяется обычно в ходе постановки задачи. Уравнения или другие математические соотношения, сформулированные в модели, постоянно сопоставляются с исходной ситуацией. Так, в упомянутом случае маятника математическое уравнение движения маятника можно подвергнуть проверке, сравнивая физические размерности величин, входящих в это уравнение.

Существует несколько аспектов проверки адекватности. Во-первых, сама математическая основа модели (которая и составляет ее существо) должна быть непротиворечивой и подчиняться всем обычным законам математической логики. Во-вторых, справедливость модели зависит от ее способности адекватно описывать исходную ситуацию. Однако ответ на вопрос о том, успешно ли проходит предложенная модель такую проверку, в значительной степени субъективен. Модель можно заставить отражать действительность, однако она не есть *сама* действительность. Наш маятник вполне реален, но его часто называют математическим маятником, и вот здесь-то нас поджидает ловушка. Дело в том, что объект, известный физикам, а также математикам-прикладникам под названием простого маятника, это всего лишь математическая идеализация реального объекта, и ничего больше. Это становится очевидным, когда мы замечаем, как размах колебаний нашего маятника уменьшается и, в конце концов, маятник останавливается. Модель математического маятника не предсказывает такого поведения. Означает ли это, что указанная модель неверна? Не обязательно: ведь до полного затухания колебаний может пройти больше часа, а нас, возможно, интересуют лишь события, происходящие в первые пять минут после начала движения.

Адекватность модели проявляется и в других формах. Например, описывая функционирование системы записи на прием в амбулатории, Шахани (1974) применил стандартный подход, основанный на теории очередей, что позволило ему прийти к некоторым заключениям относительно времени ожидания, как пациентов, так и врача-консультанта. Его выводы не были основаны на принятии его модели в качестве *истинной*, т. е. правильно отражающей работу системы на всех этапах (такую модель на самом деле было бы трудно обосновать). Они просто основывались на решении вопроса об ее *адекватности*. Иначе говоря, достаточно ли хорошо *для целей рассматриваемой задачи*, результаты, полученные на основе этой модели, отражают положение дел. Таким образом, «решение» (даже одной и той же задачи, но в другой раз) зависит от критериев, выдвинутых автором модели, в такой же степени, как и от установления физических, экономических или любых других характеристик исходной ситуации.

Можно, потратить много времени на такое улучшение решения для данной модели, которое не оправдано самой постановкой задачи. Это связано, в частности, со степенью точности опытных данных. Так, если имеющиеся исходные данные известны с погрешностью, скажем, 5%, то, разумеется, бессмысленно предлагать «решения», обеспечивающие погрешность, не превышающую 1%. Поэтому нужно подчеркнуть, что ответ, который невозможно реализовать на практике (хотя он и получен с помощью тонкого математического анализа), оказывается бесполезным для данной задачи. Как сказал однажды один инженер, «всякое уравнение длиной более двух дюймов, скорее всего, неверно!» Кроме того, можно сказать, что приближенный ответ, который получается быстрее, может оказаться более эффективным, чем более точный ответ, на получение которого уходит больше времени. Это часто свидетельствует в пользу непосредственного численного приближенного решения, позволяющего избежать затрат времени на поиски наиболее изящного аналитического решения.

Ситуации моделируют для разных целей. Главная из них — необходимость предсказывать новые результаты или новые свойства явления. Эти предсказания могут быть связаны с распространением уже существующих результатов или иметь более принципиальный характер. Часто они относятся к условиям, которые, по всей вероятности, будут иметь место в некоторый момент в будущем. С другой стороны, предсказания могут относиться к событиям, непосредственное экспериментальное исследование которых неосуществимо; наиболее важный пример такого рода дают многочисленные прогнозы, которые делались на основе математических моделей в программе космических исследований. Однако для этой цели моделируются не все ситуации: в некоторых случаях достаточно уметь описывать математическими средствами работу системы для того, чтобы добиться более глубокого понимания явления (именно эту роль и играют многие выдающиеся физические теории, хотя на их основе делаются также и прогнозы). Обычно при таком математическом описании не учитывается элемент контроля, однако в моделях, построенных, например, для исследования работы сетей, таких, как схемы движения поездов или самолетов, контроль часто является важным фактором. Действительно, многие модели в исследовании операции или технике имеют целью облегчить администрации процесс принятия *решений*. Другие модели строятся для того, чтобы сделать более удобным масштаб измерения. Так, линейная шкала для температуры является математической моделью, которой пользуются при изготовлении измерительного устройства, а именно термометра. Можно привести и много других примеров.

Математическая модель представляет собой *упрощение* реальной ситуации. Ощутимое упрощение наступает тогда, когда несущественные особенности ситуации отбрасываются, и исходная сложная задача сводится к идеализированной задаче, поддающейся математическому анализу. Именно при таком подходе в классической прикладной математике возникли блоки без трения, невесомые нерастяжимые нити, невязкие жидкости и многие другие понятия подобного рода. Эти понятия не существуют в реальной действительности, они являются абстракциями, составной частью идеализации, предпринятой автором модели. И, тем не менее, их часто можно с успехом считать хорошим приближением к реальным ситуациям.

Это всего лишь одна сторона упрощения. Другая сторона связана со сравнением порядка различных величин, фигурирующих в модели. Например, изменение некоторой величины x с течением времени можно описать уравнением вида

$$a\frac{d^2x}{dt^2}+b\frac{dx}{dt}+cx=0.$$

Можно было бы сразу приступить к решению этого уравнения, однако допустим, что в результате наблюдения было замечено, что член $b\mathrm{d}x/\mathrm{d}t$ гораздо больше по величине, чем cx. В таком случае можно сэкономить много времени, упростив уравнение (отбросив член cx), в результате чего решение получится *быстрее*, но, несмотря на это, оно будет правильно отражать ситуацию. Решение же исходного уравнения, будучи в математическом отношении более точным, может в действительности привести к неверным заключениям.

Описанный образ действий при построении математических моделей не является единственным, и этому совсем не стоит удивляться. В другом возможном подходе первым шагом является построение простой модели нескольких наиболее характерных особенностей явления; это часто делается для того, чтобы «почувствовать» данную задачу, причем делается еще до того, как сама задача окончательно сформулирована. Затем эта простая модель обобщается, чтобы охватить другие факторы, пока не будет найдено «приемлемое» или «адекватное» решение. Есть еще подход, при котором с самого начала вводится в рассмотрение одновременно большое число факторов. Он часто применяется в исследовании операций, и такие модели обычно изучают имитационными методами с использованием компьютеров.

Важнейшее решение, которое часто принимается в самом начале процесса моделирования, касается природы рассматриваемых математических переменных. По существу они делятся на два класса. В один из них входят известные характеристики, т. е. величины, поддающиеся (по крайней мере, теоретически) точному измерению и управлению; они называются *детерминированными* переменными. В другой класс входят неизвестные характеристики, т. е. величины, которые никогда не могут быть точно измерены и имеют случайный характер; они называются *стохастическими* переменными. Модель, содержащая стохастические переменные, должна по определению описываться математическим аппаратом теории вероятностей и статистики; детерминированные переменные часто, но отнюдь не всегда, требуют привлечения обычного математического анализа. Природа некоторых ситуаций бывает ясна не сразу, другие ситуации характеризуются переменными обоих типов. Для построения модели чрезвычайно важно, чтобы природа переменных была правильно установлена.

Наконец, переходим к вопросу об интерпретации вытекающих из модели выводов. Работа математика-прикладника не заканчивается в тот момент, когда после многочисленных выкладок и математических манипуляций получается формула или иной результат. Ему еще предстоит совершить обратный перевод с математического языка на язык, на котором первоначально формулировалась исходная задача: ведь сомнительно, чтобы на коллег-нематематиков (например, из администрации компании) произвели впечатление ответы на языке, который для них непонятен. Следует отчетливо осознавать как математический смысл полученных решений, так и то, что они означают на языке реального мира, который математика призвана описывать. Искусством построения моделей можно овладеть только в результате собственной практики, однако почувствовать, в чем состоит это искусство, можно, разбирая примеры, которые тем или иным образом иллюстрируют различные особенности процесса моделирования. Однако сначала обратимся к авторитетному мнению академика Александра Андреевича Самарского, корифея отечественной школы математического моделирования, приведя его взгляд (пункты 1.2 – 1.7) на математическое моделирование и вычислительный эксперимент.

1.2. МЕСТО МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ В СОВРЕМЕННОМ МИРЕ

Широкое применение математических методов позволяет поднять общий уровень теоретических исследований, дает возможность проводить их в более тесной связи с экспериментальными исследованиями. Математическое моделирование сейчас рассматривается как новый метод познания, конструирования, проектирования, который сочетает в себе многие достоинства как теории, так и эксперимента. Работа не с самим объектом (явлением, процессом), а с его моделью дает возможность безболезненно, относительно быстро и без существенных затрат исследовать его свойства и поведение в любых мыслимых ситуациях (преимущества теории). В то же время вычислительные (компьютерные, имитационные) эксперименты с моделями объектов позволяют, опираясь на мощь современных вычислительных методов и технических инструментов информатики, подробно и глубоко изучать объекты в достаточной полноте, недоступной чисто теоретическим подходам (преимущества эксперимента).

Технические, экологические, экономические и иные системы, изучаемые современной наукой, больше не поддаются исследованию (в нужной полноте и точности) обычными теоретическими методами. Прямой натурный эксперимент над ними долог, дорог, часто либо опасен, либо попросту невозможен. Вычислительный эксперимент позволяет провести исследование быстрее и дешевле. Математическое моделирование является в настоящее время одной из важнейших составляющих научно-технического прогресса. Без применения этой методологии в развитых странах не реализуется ни один крупномасштабный технологический, экологический или экономический проект.

Рождение и становление методологии математического моделирования пришлось на конец 40-х - начало 50-х годов XX века и было обусловлено, по крайней мере, двумя причинами. Первым, но не основным, побудительным мотивом послужило появление компьютеров, которые избавили исследователей от огромной по объему рутинной вычислительной работы. Второй, более важной, причиной явился беспрецедентный социальный заказ - выполнение национальных программ СССР и США по созданию ракетно-ядерного щита. Эти сложнейшие научно-технические проблемы не могли быть реализованы традиционными методами без широкого использования вычислительных средств. Ядерные взрывы и полеты ракет и спутников были промоделированы сначала на компьютерах и лишь затем претворены на практике.

Основу вычислительного эксперимента составляет триада **«модель - алгоритм -- программа»**. Математические модели реальных исследуемых процессов сложны и включают системы нелинейных функционально-дифференциальных уравнений. Ядро математической модели составляют уравнения с частными производными.

На первом этапе вычислительного эксперимента выбирается (или строится) модель исследуемого объекта, отражающая в математической форме важнейшие его свойства - законы, которым он подчиняется, связи, присущие составляющим его частям, и т.д. Математическая модель (ее основные фрагменты) исследуется традиционными аналитическими средствами прикладной математики для получения предварительных знаний об объекте.

Второй этап связан с выбором (или разработкой) вычислительного алгоритма для реализации модели на компьютере. Необходимо получить искомые величины с заданной точностью на имеющейся вычислительной технике. Вычислительные алгоритмы должны не искажать основные свойства модели и, следовательно, исходного объекта, они должны быть адаптирующимися к особенностям решаемых задач и используемых вычислительных средств. Изучение математических моделей проводится методами вычислительной математики, основу которых составляют численные методы решения задач математической физики - краевых задач для уравнений с частными производными.

На третьем этапе создается программное обеспечение для реализации модели и алгоритма на компьютере. Программный продукт должен учитывать важнейшую специфику математического моделирования, связанную с использованием ряда (иерархии) математических моделей, многовариантностью расчетов. Это подразумевает широкое использование комплексов и пакетов прикладных программ, разрабатываемых, в частности, на основе объектно-ориентированного программирования.

Успех математического моделирования определяется одинаково глубокой проработкой всех основных звеньев вычислительного эксперимента. Опираясь на триаду «модель - алгоритм – программа», исследователь получает в руки универсальный, гибкий и недорогой инструмент, который вначале отлаживается, тестируется и калибруется на решении содержательного набора пробных задач. После этого проводится широкомасштабное исследование математической модели для получения необходимых качественных и количественных свойств и характеристик исследуемого объекта.

Вычислительный эксперимент по своей природе носит междисциплинарный характер, невозможно переоценить синтезирующую роль математического моделирования в современных научно-технических разработках. В совместных исследованиях участвуют специалисты в прикладной области, прикладной и вычислительной математике, по прикладному и системному программному обеспечению. Вычислительный эксперимент проводится с опорой на широкое использование самых разных методов и подходов - от качественного анализа нелинейных математических моделей до современных языков программирования.

Моделирование в том или ином виде присутствует почти во всех видах творческой деятельности. Математическое моделирование расширяет сферы точного знания и поле приложений рациональных методов. Оно базируется на четкой формулировке основных понятий и предположений, апостериорном анализе адекватности используемых моделей, контроле точности вычислительных алгоритмов, квалифицированной обработке и анализе данных расчетов.

Решение проблем жизнеобеспечения на современном этапе основывается на широком использовании математического моделирования и вычислительного эксперимента. Вычислительные средства (компьютеры и численные методы) традиционно хорошо представлены в естественнонаучных исследованиях, прежде всего в физике и механике. Идет активный процесс математизации химии и биологии, наук о земле, гуманитарных наук и т.д.

Наиболее впечатляющие успехи достигнуты при применении математического моделирования в инженерии и технологии. Компьютерные исследования математических моделей в значительной степени заменили испытания моделей летательных аппаратов в аэродинамических трубах, взрывы ядерных и термоядерных устройств на полигонах.

Современные информационные технологии используются в медицине. Сбор и анализ диагностических данных позволяет провести своевременную диагностику заболеваний. Например, компьютерный томограф является примером того, как использование математических методов обработки больших массивов данных позволило получить качественно новый медицинский инструментарий.

Здесь изложены основные подходы к построению и анализу математических моделей, общие для различных областей знания, не зависящие от конкретной специфики. Окружающий людей мир един, что проявляется, в частности, в универсальности математических моделей, в использовании одних и тех же математических конструкций для описания различных явлений и объектов.

Указаны общие черты вычислительного эксперимента с теоретическими и экспериментальными методами в научных исследованиях. Ниже приводится краткое описание различных типов вычислительного эксперимента. Вычислительный эксперимент рассматривается как наиболее высокая ступень математического моделирования, порожденная преобладающим использованием компьютеров и численных методов для изучения математических моделей.

1.3. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Математизация научного знания, под которой понимается применение математических понятий в естественных и гуманитарных науках, технике, является приметой нашего времени. Часто и уровень развития той или иной науки характеризуется по степени использования математических методов. Известный афоризм "Во всяком знании столько науки, сколько в ней математики" отражает это мнение.

1.3.1. Математизация знаний

На эмпирическом уровне развития науки описываются наблюдаемые явления, проводятся опыты, собираются и классифицируются экспериментальные данные. Для теоретического уровня характерно введение новых абстракций и идеализаций, понятий, формулировка основных законов, образующих ядро теории. При этом достигается целостный взгляд на исследуемый объект, дается единое истолкование всей совокупности экспериментальных данных.

Большая эвристическая роль теории проявляется в том, что она позволяет предсказать новые, ранее не известные характеристики объекта, явления или процесса. История развития науки содержит блестящие иллюстрации этого: открытие Нептуна, открытие позитрона и т.д. Математические идеи и методы служат не просто математическими украшениями, а действенными средствами количественного и качественного анализа.

Различные науки имеют разный уровень математизации. Для наук, в которых превалирующее значение имеют качественные математические модели, характерен невысокий (более точно, относительно невысокий) уровень математизации. Степень математизации можно характеризовать по тому, какие математические модели используются и насколько широко. Например, применение математики в механике базируется на использовании систем уравнений с частными производными. Причем такие математические модели используются не от случая к случаю, а во всех разделах механики, таких как теория упругости, гидроаэродинамика и т.д. Большой уровень математизации характерен и для физики, хотя в различных ее разделах математические методы пока используются в разной степени.

В настоящее время отмечается все возрастающий уровень математизации химии. Например, химическая кинетика базируется на системах обыкновенных дифференциальных уравнений, химическая гидродинамика - на уравнениях в частных производных и т.д. Повышается и уровень математизации биологии. В этой связи достаточно сослаться на классические работы В.Вольтерра по моделированию системы «хищник – жертва», выполненные еще в начале двадцатого века.

Мы являемся свидетелями все более широкого использования математических идей в экономике, истории и других гуманитарных науках.

Процесс математизации наук идет чрезвычайно быстро благодаря опыту, накопленному при математизации механики и физики, благодаря достигнутому уровню развития самой математики. Применение математики в химии и биологии в большой степени базируется на уже разработанном ранее математическом аппарате. Поэтому темпы математизации этих наук в значительной степени сдерживаются только уровнем развития самой химии, самой биологии. Здесь важное значение имеет и психологический фактор боязни математики. Без развития экспериментальных и теоретических исследований существенное продвижение за счет только математических методов невозможно. Успешное применение математических методов требует прежде всего глубокого овладения содержанием исследуемого процесса или явления, необходимо быть прежде всего специалистом в прикладной области, а потом уже математиком.

Единство природы проявляется в том, что для описания различных физических, химических, биологических и т.д. процессов и явлений применяются одни и те же математические модели. Это свойство конечного числа математических моделей отражает прежде всего их абстрактность. Одно и то же математическое выражение (понятие) может описывать совершенно различные процессы, характеристики. Так, например, уравнение Лапласа описывает движение несжимаемой жидкости в гидродинамике, электростатическое поле вне заряженных тел, стационарное тепловое поле, прогиб мембраны в теории упругости и т.д. Как отмечал А.Пуанкаре "Математика - это искусство давать разным вещам одно наименование". Это позволяет, в частности, при исследовании одного конкретного явления или процесса использовать результаты, полученные при исследовании другого явления или процесса. В такой общности, единстве математических моделей проявляется интегрирующая роль (ее наддисциплинарный характер) математики, ее методов.

1.3.2. Использование математических моделей

При математизации научных знаний выделяется этап абстрагирования от конкретной природы явления, идеализации и выделения его математической формы (строится математическая модель). Именно абстрактность математической модели порождает определенные трудности для ее применения к описанию конкретного явления или процесса. Сейчас, благодаря накопленному опыту, процесс идеализации, абстрагирования проходит значительно спокойнее и быстрее в различных науках.

Вторым этапом математизации является исследование математических моделей как чисто математических (абстрактных) объектов. С этой целью используются средства самой математики как уже созданные, так и специально построенные. В настоящее время большие возможности для исследования математических моделей предоставляют вычислительные средства: компьютеры и численные методы.

Третий этап применения математики в прикладных исследованиях характеризуется интерпретацией - приданием конкретного прикладного содержания математическим абстракциям. Специалист по прикладному математическому моделированию, работая бок о бок со специалистами в прикладной области, всегда за математическими абстракциями видит конкретное прикладное содержание.

Математические модели могут изучаться в традициях чистой математики. В этом случае математические модели изучаются сами по себе, без какой-либо связи с прикладным содержанием. Они исследуются на принятом в математике уровне строгости, что обеспечивает им универсализм и необходимую общность. Здесь уместно сослаться на мнение крупных математиков: Д.Гильберта, А.М.Ляпунова и др. Эта точка зрения сводится к следующему. После математической формулировки прикладной проблемы ее нужно рассматривать на уровне чистой математики. Несомненно, что исследование математических моделей является одним из самых мощных стимулов развития самой математики.

Эвристическая роль математического моделирования проявляется в том, что вместо натурного эксперимента проводится математический эксперимент. Вместо исследования проявления того или иного воздействия на исследуемый объект используется параметрическое изучение математической модели, устанавливается зависимость решения от того или иного параметра. Такой эксперимент, дополняя натурный, позволяет значительно глубже исследовать явление или процесс.

1.4. ИСПОЛЬЗОВАНИЕ КОМПЬЮТЕРОВ ПРИ МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ

Появление электронных вычислительных машин, быстрое развитие вычислительной математики, повсеместное использование вычислительной техники чрезвычайно расширило возможности математического моделирования

1.4.1. Новые возможности математики

Вычислительные средства, под которыми мы понимаем компьютеры и вычислительные методы, позволили решить с приемлемой точностью и за разумное время задачи, которые ранее были недоступны для исследования, дали возможность реализовать крупнейшие научно-технические проекты.

В качестве примеров отметим использование компьютеров при запуске и управлении полетом космических кораблей, при обработке данных сейсмической разведки полезных ископаемых, полное численное моделирование аэродинамики реальной конфигурации самолета, и т.д.

Даже в чистой математике компьютеры нашли достойное применение: доказательные вычисления на компьютере, решение знаменитой проблемы четырех красок и т.д.

Идет быстрое формирование новых научных дисциплин, новых научных направлений, основанных на широком использовании вычислительных средств при теоретическом исследовании прикладных проблем. Отметим в этой связи прежде всего вычислительную физику, вычислительную гидродинамику, вычислительную геометрию, вычислительную алгебру, вычислительную теплофизику и другие.

Исследование математических моделей подразумевает прежде всего качественное изучение математических моделей и получение точного или приближенного решения. Компьютер предоставляет новые возможности не только для нахождения приближенного решения численными методами, но и для качественного исследования математической модели.

1.4.2. Аналитические методы исследования математических моделей

Качественное исследование начинается с размерностного анализа задачи. Приведение задачи к безразмерному виду позволяет сократить число определяющих параметров задачи. Выделение малых или больших безразмерных параметров дает возможность в ряде случаев существенно упростить исходную математическую модель, учесть особенности задачи при разработке численных методов ее решения.

Сама математическая модель может быть достаточно сложной, нелинейной. Это зачастую делает невозможным ее качественное исследование традиционными методами прикладной математики. Именно поэтому в громадном большинстве случаев проводиться качественное исследование на более простых, но обязательно содержательных, по отношению к исходной математической модели задачах. В этом случае мы должны говорить о модельных (упрощенных) задачах для основной математической модели (моделей для модели). Так, например, особенности модели потенциального течения с дозвуковыми и сверхзвуковыми подобластями течения в плане качественного исследования передаются уравнением Трикоми, которое в математической физике относится к классу уравнений смешанного типа.

Большое внимание при качественном исследовании математических моделей (или модельных задач для них) уделяется вопросам корректности. Прежде всего рассматривается проблема существования решения. Соответствующие строгие результаты (теоремы существования) дают уверенность в корректности математической модели. Кроме того, конструктивные доказательства теорем существования могут быть положены в основу приближенных методов решения поставленной задачи.

При прикладном математическом моделировании важным является вопрос об устойчивости решения относительно малых возмущений входных данных. Неустойчивость (неограниченный рост решения при малых возмущениях) наиболее характерна для обратных задач и должна учитываться при построении приближенного решения.

Для нелинейных математических моделей может быть характерна множественность, неединственность решения. При качественном исследовании математических моделей изучаются точки ветвления, бифуркации решения, вопросы выделения нужного искомого решения и т.д.

Методы качественного исследования для различных типов математических моделей разработаны с неодинаковой полнотой. Среди моделей, где качественные методы принесли наиболее впечатляющие результаты, отметим обыкновенные дифференциальные уравнения. В теории уравнений с частными производными качественные методы также используются, хотя и не в такой большой степени. В качестве содержательного примера отметим принцип максимума для параболических и эллиптических уравнений второго порядка, который позволяет провести качественное исследование математических моделей, основанных на уравнениях с частными производными.

Точное или приближенное решение находится с использованием аналитических и численных методов. В этой связи среди классических примеров аналитических методов отметим методы разделения переменных, интегральных преобразований для линейных задач математической физики.

Для нелинейных математических моделей особое значение имеют методы линеаризации, различные варианты методов возмущений. Теория возмущений базируется на использовании асимптотических разложений по выделенному малому параметру. Особое внимание этим методам, несмотря на их ограниченность, уделяется при рассмотрении сингулярно возмущенных задач.

Качественное поведение решения нелинейной задачи может хорошо передаваться некоторыми частными решениями. Поиск частных решений нелинейных задач основывается на использовании автомодельных переменных, на результатах группового анализа уравнений, лежащих в основе математической модели.

Сложные нелинейные многопараметрические модели могут быть исследованы на компьютере численными методами. В отличие от аналитического решения, которое может давать явную параметрическую зависимость решения от тех или иных условий задачи, при численном решении требуется многократное решение задачи при изменении того или иного параметра. Но ведь численное решение может быть получено и для тех задач, для которых аналитического решения нет.

1.4.3. Использование компьютеров

Перейдем теперь к характеристике основных этапов использования компьютеров при математическом моделировании. Мы будем основное внимание обращать на использование вычислительных средств при нахождении приближенного решения задачи. Необходимо однако отметить и возможности применения компьютеров и на этапе качественного исследования математической модели, этапе отыскания аналитических решений модельных задач. Например, компьютер можно использовать для нахождения автомодельных решений. При выделении автомодельной переменной исходная задача для уравнения в частных производных сводится, например, к обыкновенному дифференциальному уравнению, происходит понижение размерности. Общее решение последнего находится на основе использования систем аналитических вычислений на компьютере (методов вычислительной алгебры), широко представленных в современных математических пакетах.

В применении компьютеров при математическом моделировании можно выделить, по крайней мере, два этапа, два уровня. Первый из них характеризуется исследованием достаточно простых математических моделей. На этом этапе (уровне) применения компьютеров вычислительные средства используются наряду и наравне с другими методами (чисто математическими) прикладной математики.

Выделенный этап применения компьютеров при математическом моделировании характеризуется условной цепочкой «заказчик (теоретик) - исполнитель (прикладной математик)». Заказчик ставит задачу, анализирует результаты, а исполнитель обеспечивает решение задачи с применением компьютеров. В этом случае речь идет о решении конкретной (достаточно узкой) задачи с определенным набором входных данных.

Для этого уровня применения компьютеров в прикладном математическом моделировании характерен лозунг Р. Хеминга: «Цель расчетов - понимание, а не числа». Это отражает традиции работы заказчика-теоретика, который больше всего ценит качественный анализ. Для современного этапа научных исследований и разработок одного понимания мало. Для выхода на эксперимент, реальную конструкцию требуются точные количественные зависимости и характеристики.

Второй этап (уровень) применения компьютеров характеризуется исследованием сложных нелинейных математических моделей. В этих условиях вычислительные средства становятся основными, абсолютно преобладающими. Традиционные средства прикладного математического моделирования выполняют вспомогательную, обслуживающую роль (качественное исследование задачи в сильно упрощенных постановках - модельные задачи, тестирование вычислительных алгоритмов и т.д.).

Именно возможность исследования сложных математических моделей на основе численных методов и компьютеров позволяет с новых позиций рассмотреть методологию научных исследований. Мощные компьютеры, высокоэффективные вычислительные алгоритмы, современное программное обеспечение позволяют в настоящее время организовать научные исследования в рамках единой технологии вычислительного эксперимента, который включает в себя теоретические и экспериментальные исследования.

1.4.4. Обработка экспериментальных данных

Экспериментатор, в самой общей схеме своего исследования, воздействует на исследуемый объект, получает информацию о результатах этого воздействия и обрабатывает ее. Эти данные зашумлены случайными погрешностями измерений. В силу этого при первичной обработке экспериментальных данных основной математический аппарат базируется на теории вероятностей и математической статистике. Экспериментальные исследования все чаще ведутся с помощью измерительно-вычислительных комплексов, которые позволяют получать, хранить и обрабатывать экспериментальные данные.

В каждом экспериментальном исследовании проводится статистическая обработка опытных данных. Количественная оценка влияния отдельных факторов (параметров) проявляется в построении эмпирических зависимостей, интерполирующих с той или иной точностью экспериментальные данные. В этом случае можно говорить об использовании аппроксимационных математических моделей, в которых содержательные математические модели как таковые просто отсутствуют. Выбор числа и условий проведения опытов для решения той или иной проблемы осуществляется на этапе планирования эксперимента. Здесь привлекаются результаты математической теории оптимального эксперимента, математической теории планирования эксперимента.

1.4.5. Математическая модель прибора

Настоящий уровень развития экспериментальных исследований характеризуется возрастающим применением все более совершенных приборов. Сами приборы с неизбежностью вносят возмущения в исследуемое явление или процесс. С целью избавления от этих погрешностей строится математическая модель прибора.

При проведении экспериментов необходимо иметь в виду две принципиально различные ситуации. Первая из них связана с ситуацией, когда для исследуемого явления или объекта нет теоретического описания, нет математической модели, и ставится задача накопления экспериментального материала с тем, чтобы в последующем дать теоретическое описание. В этом случае математические методы используются для хранения и переработки информации, в частности, для получения эмпирических зависимостей.

При построении аппроксимационных математических моделей типичной является ситуация с определением параметров эмпирических формул, подборе самой формулы. По массе экспериментальных данных необходимо подобрать параметры аппроксимационных моделей так, чтобы с приемлемой точностью можно было описать экспериментальные данные. В этом случае мы сталкиваемся с необходимостью приближенного решения соответствующих задач минимизации.

Второй класс экспериментов проводится в условиях, когда есть теоретическое описание исследуемого объекта. Структура математической модели определена и ставится задача определения параметров модели. Сам натурный эксперимент направлен на то, чтобы определить те или иные свойства объекта, на конкретизацию математической модели объекта.

При обработке опытных данных таких экспериментов часто приходится иметь дело с обратными задачами. Такие задачи могут быть некорректными в классическом смысле и поэтому трудными для численного исследования. На стадии обработки и интерпретации данных экспериментальных исследований вычислительные средства находят все более широкое применение с использованием различных классов математических моделей.

1.5. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Теоретические и экспериментальные исследования обладают большой степенью автономности. В условиях, когда фундаментальные модели известны, апробированы, может быть поставлена проблема более тесного координирования и связи теоретических и экспериментальных исследований. Речь идет о новой объединяющей технологии научных исследований, которой является математическое моделирование и вычислительный эксперимент.

1.5.1. Основные этапы вычислительного эксперимента

Изложим вначале общую схему вычислительного эксперимента, а затем дадим краткую характеристику его основных этапов. Понимая вычислительный эксперимент в узком смысле, как создание и изучение математических моделей исследуемого объекта с помощью вычислительных средств, можно выделить в качестве основы триаду «модель - алгоритм – программа». В широком (методологическом) смысле под вычислительным экспериментом мы понимаем новую технологию научных исследований. Основные этапы вычислительного эксперимента прослеживаются на рис.1.

Для исследуемого объекта сначала строится математическая модель. Она базируется на известных фундаментальных моделях. Вычислительный эксперимент, по своей сути, предусматривает исследование группы близких моделей. Вначале строится простая, но достаточно содержательная и полная с точки зрения описания исследуемых процессов, с точки зрения близости к экспериментальным данным модель.

В процессе проведения вычислительного эксперимента, на его последующих циклах модель уточняется, учитываются новые факторы и т.д. Поэтому мы всегда можем говорить (более того, должны говорить) о наборе, упорядоченном наборе (об иерархии) математических моделей, каждая из которых с той или иной точностью описывает действительность. И в рамках наиболее простой модели необходимо добиваться согласия с экспериментом. Это и является, в конце концов, целью вычислительного эксперимента.

Рис.1.2. Схема вычислительного эксперимента

После построения математической модели традиционными средствами прикладной математики проводится предварительное исследование математической модели. Суть вычислительного эксперимента, его содержательное зерно состоит в исследовании на компьютере математических моделей численными методами. Здесь же речь идет только о предварительном исследовании математической модели. На этом этапе с доступной полнотой, на принятом в математике уровне строгости решаются вопросы о корректности полной задачи в узком математическом смысле.

Основное содержание предварительного исследования математической модели состоит в выделении более простых (модельных) задач и их всестороннем исследовании, так как полная математическая модель слишком сложна. Модельные математические задачи в цикле вычислительного эксперимента строятся для двух различных целей: во-первых, для качественного исследования полной задачи (а опосредовано и исследуемого объекта), во-вторых - для проверки, тестирования вычислительных алгоритмов приближенного решения полной задачи.

При качественном исследовании модельных (упрощенных) задач изучаются вопросы множественности решения, его устойчивости и т.д. Большое значение имеют также точные частные решения существенно нелинейных задач, асимптотические решения и т.д. Таким образом, здесь применяется обычный математический арсенал теоретического исследования проблемы.

На следующем этапе вычислительного эксперимента строится дискретная задача и численный метод решения этой дискретной задачи. Сама математическая модель включает в себя, как правило, уравнения с частными производными (ядро математической модели), системы дифференциальных и алгебраических уравнений. Построение вычислительных алгоритмов и их исследование является прерогативой вычислительной математики.

При прикладном математическом моделировании наблюдаются две тенденции научных исследований. В традициях (парадигме) чистой математики одни исследователи изучают дискретные модели и численные методы их исследования вне связи их с прикладным математическим моделированием, реализацией на компьютере в контексте решения прикладной проблемы. Проводятся строгие доказательства существования решения дискретной задачи, получают теоретические оценки погрешности приближенного решения, сходимости итерационного процесса. Это уместно прежде всего при разработке методов решения базовых задач, при разработке вычислительного арсенала исследователя.

Представители прикладного направления в вычислительной математике работают на несколько другом ("физическом") уровне строгости, для которого характерны такие нестрогие понятия как "практическая сходимость", "реальные сетки" и т.д. Безусловное требование полной строгости при прикладном математическом моделировании ни к чему хорошему не приводит.

Вычислительный эксперимент характеризуется двумя особенностями, которые необходимо учитывать при создании адекватного ему программного обеспечения. Это, во-первых, многовариантность расчетов в рамках фиксированной математической модели и, во-вторых, многомодельность. Здесь уже нельзя обойтись одной программой на компьютере, нужно иметь возможность легко менять ее для решения близких задач (задач для набора моделей).

Программное обеспечение вычислительного эксперимента базируется на использовании комплексов и пакетов прикладных программ. Комплекс программ предназначен для решения близких по своей математической природе задач из одной предметной области. Он включает в себя библиотеку программных модулей (в большой или меньшей степени независимых), из которых комплектуются рабочие программы. В комплексах прикладных программ сборка программ из модулей осуществляется вручную.

В пакетах прикладных программ для сборки используются системные средства компьютера, что позволяет в значительной степени автоматизировать этот процесс. Пакеты прикладных программ, рассматриваемые как технология решения задач в рамках вычислительного эксперимента, позволяют наиболее эффективно использовать накопленный программный продукт, резко поднять производительность труда программистов.

В наибольшей степени основные особенности вычислительного эксперимента учитывается при использовании объектно-ориентированного программирования и современных языков программирования.

Затем в цикле вычислительного эксперимента проводится серия расчетов на компьютерах при изменении тех или иных параметров задачи. Полученные данные анализируются и интерпретируются с участием специалистов в прикладной области. Обработка результатов проводится с учетом имеющихся теоретических представлений и экспериментальных данных. Она осуществляется, во многом, в традициях классического натурного эксперимента. Сами опытные данные представляются в виде таблиц, графиков, фотографий с дисплея, кинофильмов и т.д.

Надо только всегда иметь в виду, что объем обрабатываемой информации, детализация полученных результатов в вычислительном эксперименте несравненно больше. В вычислительном эксперименте проблемы хранения и обработки информации имеют все возрастающее значение.

На этапе анализа результатов становиться ясным, удачно ли выбрана математическая модель, ее вычислительная реализация. Если есть необходимость, модели и численные методы уточняются, и весь цикл вычислительного эксперимента повторяется, то есть совершается новый виток спирали в познании истины.

1.5.2. Основные особенности новой технологии научных исследований

Характеризуя вычислительный эксперимент в целом, чрезвычайно важно отметить его универсальность, которая позволяет легко переносить эту технологию на исследование других объектов. Это обстоятельство характерно вообще для математического моделирования и порождено тем, что многие явления и процессы имеют одни и те же математические модели.

Отмеченная многоцелевая направленность и методологическая универсальность вычислительного эксперимента позволяет на основе накопленного опыта математического моделирования, банка вычислительных алгоритмов и программного обеспечения быстро и эффективно решать новые задачи.

Второй особенностью вычислительного эксперимента, как технологии научных исследований, является его междисциплинарный характер. Мы постоянно подчеркиваем это обстоятельство, говоря о том, что прикладной математик объединил теоретика и экспериментатора для более быстрого достижения общей цели. Вычислительный эксперимент может рассматриваться как удобная форма кооперации умственного труда, повышения его производительности. В едином цикле вычислительного эксперимента работает и теоретик, и экспериментатор, и прикладной математик, и программист.

Можно отметить следующие отличительные особенности и преимущества вычислительного эксперимента перед натурным экспериментом.

Во-первых, вычислительный эксперимент проводится даже тогда, когда натурный эксперимент невозможен. Такая ситуация имеет место с крупномасштабными экологическими экспериментами. Отметим в этой связи моделирование глобальных климатических изменений при использовании атомного оружия. Другой пример - исследование процессов при термоядерных параметрах (кроме взрыва атомной бомбы пока нет других возможностей достичь их)

Во-вторых, при использовании вычислительного эксперимента резко снижается стоимость разработок и экономится время. Это обеспечивается многовариантностью выполняемых расчетов, простотой модификации математических моделей для имитации тех или иных реальных условий.

В качестве иллюстрации отметим то, что расчеты на компьютерах в большой степени заменили эксперименты в аэродинамических трубах при создании космического корабля многоразового использования «Шатл». Создание новых изделий и технологий с необходимостью связано с тяжелой, дорогостоящей и длительной доводкой. Вычислительные средства позволяют в значительной степени сэкономить время и деньги именно на этой стадии.

Данные экспериментальных исследований используются для калибровки математических моделей, контроля точности приближенного решения задачи. В традициях экспериментального исследования мы воздействуем на математическую модель и обрабатываем результаты (вот почему мы говорим об эксперименте, хотя и вычислительном). И лишь изредка мы контролируем точность своего "прибора", сравнивая его с эталоном. В традициях теоретического исследования в вычислительном эксперименте мы имеем дело с математической моделью, а не с самим объектом. Эти общие черты мы рассматриваем как дополнительные аргументы в пользу интерпретации вычислительного эксперимента в широком (методологическом) смысле как интегрирующей технологии научных исследований.

Вычислительный эксперимент необходимо рассматривать как новую технологию научных исследований в перспективе, как тенденцию, как логику развития организации научных исследований. В настоящее время он, зачастую, реализуется в узком смысле по цепочке "заказчик - прикладной математик". Более тесная увязка теоретических и экспериментальных исследований в единой технологии научных исследований является ярко выраженной тенденцией нашего времени. И примечательно, что основным связующим звеном этой методологии является математическое моделирование и вычислительный эксперимент.

1.6. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ В НАУКЕ И ТЕХНОЛОГИИ

Остановимся теперь на краткой характеристике основных областей применения математического моделирования. Основное внимание уделим классификации видов вычислительного эксперимента по применениям и по типам используемых математических моделей. Отмеченная взаимоувязанная классификация позволяет ориентировать исследователя на использование адекватного математического аппарата исследования математических моделей. Такая методологическая проблема зачастую затушевывается и сдерживает интеграционные процессы в самой прикладной математики, не говоря уже о трудностях математического моделирования.

1.6.1. Области применения вычислительного эксперимента

Математическое моделирование традиционно развивается в недрах фундаментальных наук: механике и физике, для которых отмечается наивысший уровень теоретических исследований (другими словами, уровень математизации). В этих науках с внедрением современных математических методов, в том числе и численных, относительно благополучно. Для механики, например, характерно наличие устоявшихся математических моделей, существует банк основных задач. Поэтому здесь основное внимание уделяется построению вычислительных алгоритмов и созданию достаточно гибкого программного обеспечения. В биологии и химии фронт работ по математическому моделированию проходит на первой части триады вычислительного эксперимента «модель – алгоритм – программа». Хотя и в разной степени, на разном уровне, но вопросы применения математических методов в фундаментальных науках решаются.

Значительно менее совершенен математический арсенал инженера и технолога. В технике до настоящего времени традиционным является путь опосредованного внедрения научного знания. Прежде всего новые идеи становятся достоянием фундаментальных наук, затем трансформируются в той или иной прикладной области и лишь затем - в конкретных технических проектах и разработках. Это относится прежде всего к применению современных математически методов теоретического исследования, математическому моделированию и вычислительному эксперименту. Такой путь превращения идеи в конкретное научно-техническое решение, новую технологию неоправданно долог и расточителен.

В современных условиях необходимо обеспечить повсеместное непосредственное внедрение математических методов в науку и технологию. Математическое моделирование технологических процессов сулит огромную выгоду, переход на новый качественный уровень самой технологии. Наиболее благодатное поле для приложения методов математического моделирования и вычислительного эксперимента - техника и промышленность, технология. Особое внимание заслуживают отрасли определяющие научно-технический прогресс сегодня, и прежде всего микроэлектроника. Численное моделирование в этом случае обеспечивает подъем своей технической базы - компьютеров.

Отметим еще один аспект в применении вычислительного эксперимента. В настоящее время мировая общественность совершенно справедливо обеспокоена экологическими последствиями крупномасштабных проектов, обеспечением безопасности функционирования работающих установок и проектируемых объектов. Вычислительный эксперимент на базе адекватных моделей позволяет испытать модель экологически опасного объекта в мыслимых и немыслимых условиях, дать практические рекомендации обеспечения условий безопасной работы, дать, если хотите, гарантии такой работы.

1.6.2. Различные типы вычислительного эксперимента

При исследовании нового процесса или явления обычный подход связан с построением той или иной математической модели и проведением расчетов при изменении тех или иных параметров задачи. В этом случае мы имеем «поисковый вычислительный эксперимент». Если основу математической модели составляют уравнения с частными производными, то в цикле вычислительного эксперимента исследуется и решается численными методами прямая задача математической физики.

В результате проведения поискового вычислительного эксперимента дается описание наблюдаемым явлениям, прогнозируется поведение исследуемого объекта в тех или иных условиях, возможно и не достижимых в реальных условиях. Такой тип вычислительного эксперимента характерен при проведении теоретических исследований в фундаментальных науках.

С другой стороны, при математическом моделировании технологических процессов в качестве основного может быть выбран «оптимизационный вычислительный эксперимент». Для него характерно решение задачи оптимизации по уменьшению затрат, облегчению конструкции и т.д. Для сформулированной математической модели ставится соответствующая задача оптимального управления, задача оптимизации.

Характерным примером могут служить задачи оптимального управления для уравнений математической физики, например, граничного управления, когда граничные условия подбираются так, чтобы минимизировать соответствующий функционал (функционал качества). В этом случае многовариантные расчеты проводятся с целью подобрать управляющие параметры, а результатом является решение в том или ином смысле оптимальное.

При обработке данных натурных экспериментов используется «диагностический вычислительный эксперимент». По дополнительным косвенным измерениям делается вывод о внутренних связях явления или процесса. В условиях, когда структура математической модели исследуемого процесса известна, ставится задача идентификации модели, например, определяются коэффициенты уравнений. Диагностическому вычислительному эксперименту обычно ставится в соответствие обратная задача математической физики.

Часто приходится сталкиваться с положением, когда математической модели исследуемого процесса или явления нет и создать ее не представляется возможным. Такая ситуация характерна, в частности, при обработке данных натурного эксперимента. Тогда обработка проводится в режиме "черного ящика" и мы имеем дело с аппроксимационными моделями. При отсутствии математических моделей на основе широкого использования компьютеров проводится имитационное моделирование.

1.7. ПЕРСПЕКТИВЫ ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА

Современный этап развития прикладной математики характеризуется исследованием математических моделей на основе широкого использования возможностей вычислительных средств (компьютеров и соответствующего аппарата самой математики - численных методов).

Возможность исследования сложнейших математических моделей позволила подойти по-новому и к вопросу организации прикладных научных исследований, более тесной координации экспериментальных и теоретических исследований в рамках новой методологии научных исследований - вычислительного эксперимента. Она основана на качественно новом уровне научных исследований, который связан с использованием сложных прикладных математических моделей, которые не могут быть изучены стандартными аналитическими методами прикладной математики.

Суть вычислительного эксперимента наиболее полно отражается в триаде «модель - алгоритм – программа». Для исследуемого объекта строится математическая модель (набор моделей), которые исследуются численными методами на компьютере. Для предварительного исследования математических моделей используются традиционные методы прикладной математики. Данные расчетов анализируются, сопоставляются с данными экспериментальных исследований, проводится уточнение математической модели и т.д.

Методология вычислительного эксперимента сложилась при решении наиболее крупных научно-технических проблем, которые ставит перед нами жизнь. Активное внедрение идей математического моделирования, которое является интеллектуальным ядром информатизации общества, позволит поднять уровень научных исследований в естественнонаучных и гуманитарных областях.

Охарактеризовав общие проблемы и возможности математического моделирования и вычислительного эксперимента, перейдем к применению полученных знаний для автоматизированного проектирования и исследования оптических систем.

1.8. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В ОПТИКЕ

Оптические системы и оптические элементы предназначены для преобразования электромагнитных волн оптического диапазона. Идеальным математическим описанием распространения электромагнитных волн являются уравнения Максвелла, однако решение системы уравнений для математического исследования конкретной оптической схемы крайне трудоемко и не всегда возможно. Поэтому обычно приходится применять различные приближения. Для моделирования оптических элементов и систем с ними используются три основных оптических подхода: методы геометрической оптики, методы скалярной теории дифракции и, наконец, в самых сложных случаях - векторной теории дифракции (уравнения Максвелла) [1]. В режиме геометрической оптики лучи используются, чтобы описать распространение волнового фронта, и этот подход в основном пренебрегает дифракцией, которая имеет место во время распространения (за исключением выходного зрачка), поляризационными эффектами, отраженными назад волнами (т.е. Френелевскими отражениями и обратными порядками дифракции), изменениями в амплитуде волнового фронта, раздвоением внутри оптического элемента, эффектами когерентности и интерференции. Геометрическая оптика описывает только направление каждого дифракционного порядка и не описывает относительную силу каждого порядка. Следовательно, это должно быть учтено дополнительными средствами. Несмотря на эти ограничения, даже если оптическая система, проектируемая обычно с помощью лучевого расчета ("ray tracing"), содержит дифракционные оптические элементы, то и тогда лучевой расчет вместе с оценками эффективности будут достаточными для моделирования изображающих систем, коллимирующей, освещающей и фокусирующей оптики, лазерных трансляционных систем, лазерных сканеров и т.п.

Скалярная теория дифракции используется, когда проблемы дифракции, интерференции и когерентности делают невозможным применение приближения геометрической оптики. Типичные приложения охватывают решеточные расщепители пучков, лазерную внутрирезонаторную оптику, диффузоры, формирователи лазерных пучков, многофокусные линзы, некоторые типы микрооптики, цифровые голограммы. Существующие коммерческие программы создавались для моделирования общего случая скалярной теории дифракции, поэтому большинство проектировщиков вынуждены использовать свои собственные численные приложения соответствующего скалярного дифракционного интеграла.

Наиболее значимы среди недостатков скалярной теории дифракции: пренебрежение эффектами поляризации, отраженными назад и бесконечно малыми дифракционными порядками; неадекватное обхождение с конечной толщиной элемента и неперпендикулярным освещением; невозможность работы со структурами, имеющими размеры порядка или меньше длины волны. В этом случае для получения более точной картины требуется численное решение уравнений Максвелла. Возможно численное решение уравнений Максвелла для произвольной геометрии (т.е. метод конечной разности). Однако этот подход до последнего времени ограничивался в использовании из-за повышенных требований к скорости вычислений и памяти компьютера. Проще решение уравнений Максвелла для периодических структур (т.е. проблемы дифракции плоской волны на дифракционной решетке). К приложениям, которые целесообразно анализировать, используя этот метод, относятся антиотражающие структуры, фильтры для длины волны и некоторые компоненты оптической памяти.

ВЫВОДЫ

Сущность методологии вычислительного эксперимента состоит в замене исходного объекта его математической моделью и исследовании современными вычислительными средствами математической модели.

Математическое моделирование может рассматриваться как новый метод познания, конструирования и проектирования, который сочетает в себе достоинства, как теории, так и эксперимента.

Основу вычислительного эксперимента составляет триада «модель - алгоритм – программа».

Математизация знаний, использование компьютеров при математическом моделировании являются приметами наших дней.

Искусством построения моделей можно овладеть только в результате собственной практики, однако почувствовать, в чем состоит это искусство, можно разбирая примеры, которые тем или иным образом иллюстрируют различные особенности процесса моделирования.

При моделировании оптических систем доминируют три подхода: геометрическая оптика, скалярная теория дифракции и решение уравнений Максвелла. Выбор используемого приближения определяется характером и параметрами решаемой задачи и носит нетривиальный характер.

ГЛАВА 2
ЭЛЕКТРОМАГНИТНОЕ ПОЛЕ И ЕГО СВОЙСТВА

2.1. ПОНЯТИЕ ЭЛЕКТРОМАГНИТНОГО ПОЛЯ И ЕГО ОПИСАНИЕ

Электромагнитное поле — это возбужденное состояние пространства, характеризующее взаимодействие электрически заряженных частиц.

Частота колебаний поля может быть различной, например (4—8) 10^{14} Гц, что соответствует видимой человеком области электромагнитного спектра. Если расширить этот диапазон, включив в него инфракрасную область спектра (вплоть до 10^{13} Гц) и ультрафиолетовую (вплоть до 10^{16} Гц), то получим оптический диапазон электромагнитного поля, который и является предметом нашего рассмотрения.

Электромагнитное поле описывается двумя векторами **E** и **B**, называемыми соответственно электрическим вектором (или вектором напряженности электрического поля) и магнитной индукцией, а также парой дополнительных векторов — электрического смещения **D** и магнитным вектором **H** (вектором напряженности магнитного поля), связанными с параметрами среды. Обычно в качестве пары базовых векторов используют электрический и магнитный векторы **E** и **H**, что также вполне допустимо. В соответствии с определением электромагнитного поля необходимо также ввести две величины, характеризующие состояние системы зарядов в материальной среде. — вектор плотности электрического тока **J** и скалярную величину плотности электрических зарядов ρ. Теперь система параметров, необходимых для описания электромагнитного поля является полной и можно перейти к системе уравнений, которым оно подчиняется.

Эта система уравнений, связывающая векторы, описывающие поле, и их производные, носит название системы уравнений Максвелла и состоит из двух векторных и двух скалярных дифференциальных уравнений:

$$\left.\begin{aligned} &\operatorname{rot}\mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J};\\ &\operatorname{rot}\mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0;\\ &\operatorname{div}\mathbf{D} = \rho;\\ &\operatorname{div}\mathbf{B} = 0.\end{aligned}\right| \qquad (2.1)$$

Последние два уравнения говорят о наличии в природе электрических зарядов и отсутствии магнитных. Данное утверждение является верным для макромира, в то время как в микромире говорить об отсутствии магнитных зарядов не совсем корректно, так как имеются косвенные доказательства наличия монополя.

Первое из этих уравнений может быть преобразовано к виду, аналогичному уравнению непрерывности в гидродинамике. Для этого необходимо взять div от обеих частей с учетом условия, что div rot = 0, а также с учетом возможности изменения порядка дифференцирования по времени и координатам. Тогда будем иметь

$$\operatorname{div}\mathbf{J} = -\operatorname{div}\frac{\partial \mathbf{D}}{\partial t},$$

и, воспользовавшись третьим уравнением из (2.1), получим

$$\frac{\partial \rho}{\partial t} + \operatorname{div}\mathbf{J} = 0. \qquad (2.2)$$

Из уравнения (2.2) следует, что изменение заряда (его плотности) в окрестности любой точки может происходить только при появлении дополнительных токов.

Если все величины, связанные с полем, не зависят от времени и отсутствуют токи ($\mathbf{J} = 0$), то такое поле называется статическим, если же все величины не зависят от времени, но присутствуют токи, то такое поле называется стационарным.

Уравнения Максвелла допускают единственное решение в среде, если к ним добавлены соотношения, описывающие поведение вещества под действием поля. Эти соотношения называются *материальными уравнениями*, и для нерелятивистского случая они могут быть записаны в следующем виде:

$$\left.\begin{aligned} &\mathbf{J} = \sigma \mathbf{E}; \\ &\mathbf{D} = \varepsilon_0 \varepsilon\, \mathbf{E}; \\ &\mathbf{B} = \mu_0 \mu\, \mathbf{H}. \end{aligned}\right\} \qquad (2.3)$$

Здесь σ — удельная проводимость; ε — диэлектрическая проницаемость; μ — магнитная проницаемость; ε_0 и μ_0 — электрическая и магнитная постоянные. Величины σ, ε и μ в общем случае представляют собой тензоры для анизотропных сред (скаляры — для изотропных).

Первое уравнение из (2.3) является дифференциальной формой закона Ома. В зависимости от величины σ вещества подразделяются на проводники (значение σ велико) и диэлектрики (значение σ мало). При промежуточных значениях вещества называются полупроводниками.

Параметр μ, определяющий магнитные свойства веществ, может быть равен единице (немагнитное вещество), больше единицы — парамагнетик (μ>>1 — ферромагнетик) и меньше единицы — диамагнетик. В дальнейшем в основном будем рассматривать немагнитные вещества, считая, что для оптического интервала $\mu=1$, но каждый раз будем оговаривать значение этого параметра.

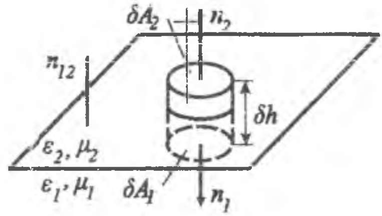

Рис. 2.1. К выводу граничных условий для нормальных компонентов B и D

Следует также отметить, что соотношения (2.3) в ряде случаев носят нелинейный характер, что приводит к новым интересным эффектам, но сильно затрудняет анализ явлений. Примером подобной нелинейности, появляющейся во втором материальном уравнении, служит поведение веществ в сверхсильных оптических полях, например в фокусе мощного лазера. В этом случае наблюдается и параметрическая генерация, и самофокусировка, и многое другое.

Уравнения Максвелла были сформулированы лишь для областей пространства, в которых физические свойства среды непрерывны. Часто, однако, на поверхности раздела двух сред электрические и магнитные параметры резко меняются. Тогда и векторы **E**, **D**, **H** и **B** могут претерпевать разрыв. В таких случаях необходимо задавать граничные условия на различные компоненты полей.

Рассмотрим поверхность, разделяющую первую и вторую среду (рис. 2.1). Заменим эту поверхность тонким слоем, внутри которого ε и μ плавно меняются от ε_1, μ_1 до ε_2, μ_2. Вырежем из этого слоя небольшой цилиндр с боковой поверхностью, нормальной к поверхностям разделительного слоя. Во всем цилиндре компоненты полей меняются также плавно вместе с производными. С помощью теории Гаусса перейдем от интегрирования по объему цилиндра к интегрированию по его поверхности и с учетом (2.1) получим

$$\int_V \operatorname{div} \mathbf{B} \, dV = \int_S \mathbf{Bn} \, dS = 0. \tag{2.4}$$

Здесь **n** — единичный вектор внешней нормали к поверхности цилиндра, второй интеграл берется по этой поверхности.

Площадки δA_1 и δA_2 малы, так как сам цилиндр мал. Здесь нужно сделать небольшое отступление и обсудить вопрос о малости параметра. Этот вопрос всегда носит относительный характер. Если мы говорим о малости какой-либо величины, то всегда сравниваем ее с какой-то другой величиной. В данном случае параметром малости является скорость изменения в пространстве вектора **B**. Мы выбираем размеры площадок такими, чтобы внутри них можно было считать этот вектор постоянным. Тогда можно считать, что на площадке δA_1 магнитная индукция $\mathbf{B}=\mathbf{B}_1$, а на площадке δA_2 магнитная индукция $\mathbf{B}=\mathbf{B}_2$.

В этих предположениях интеграл (2.4) можно заменить следующим выражением:

$$\mathbf{B}_1\mathbf{n}_1\delta A_1 + \mathbf{B}_2\mathbf{n}_2\delta A_2 + \text{интеграл по боковой поверхности} = 0, \quad (2.5)$$

где $\mathbf{n}_1$ и $\mathbf{n}_2$ — нормали к торцевым поверхностям.

Если толщину переходного слоя δh, а значит, и высоту цилиндра устремить к нулю, то третье слагаемое в выражении (2.5) также устремится к нулю, если отсутствует поверхностный поток магнитной индукции. Такой поток никогда не наблюдается, и, следовательно, выражение (2.5) в пределе $\delta h \to 0$ значительно упрощается:

$$(\mathbf{B}_1\mathbf{n}_1 + \mathbf{B}_2\mathbf{n}_2)\delta A = 0.$$

Здесь δA — площадь пересечения цилиндра с границей раздела двух сред. Если теперь ввести единичный вектор нормали к поверхности раздела $\mathbf{n}_{12}$, направленный из первой среды во вторую, то получим необходимое граничное условие

$$\mathbf{n}_{12}\left(\mathbf{B}_2 - \mathbf{B}_1\right) = 0, \quad (2.6)$$

которое формулируется следующим образом: *нормальная компонента вектора магнитной индукции непрерывна на поверхности раздела двух сред.*

Другие граничные условия выводятся аналогично и имеют вид:

$$\mathbf{n}_{12}\left(\mathbf{D}_2 - \mathbf{D}_1\right) = \rho, \quad (2.7)$$

т.е. *нормальная компонента вектора электрического смещения испытывает скачок на границе раздела двух сред, пропорциональный поверхностной плотности заряда* ρ;

$$\mathbf{n}_{12} \times \left(\mathbf{E}_2 - \mathbf{E}_1\right) = 0, \quad (2.8)$$

т.е. *тангенциальная компонента электрического вектора непрерывна на границе раздела двух сред;*

$$\mathbf{n}_{12} \times \left(\mathbf{H}_2 - \mathbf{H}_1\right) = \mathbf{J}, \quad (2.9)$$

т.е. тангенциальная компонента магнитного вектора испытывает скачок на границе раздела двух сред, пропорциональный плотности поверхностного тока **J**.

Таким образом, математически сформулирована основная задача теории электромагнитного, в нашем случае оптического, поля: есть система уравнений для поля, есть система соотношений, описывающих взаимодействие его со средой, есть граничные условия, определяющие изменение полевых векторов при переходе через границы раздела сред.

Другой задачи в теоретической оптике нет, и все дальнейшие усилия будут направлены на решение этой задачи. К сожалению, в общем виде она не решается, так как описывается системой нелинейных дифференциальных уравнений в частных производных для четырех полевых векторов. Приходится строить модели, решать систему уравнений в различных приближениях и находить частные или приближенные решения.

Вся физическая оптика представляет собой решение задачи Максвелла путем последовательных приближений, постепенного перехода от геометрической оптики к скалярной теории дифракции и более сложным теориям.

В этом случае важнейшее значение имеют те модельные предположения, которые делаются в условиях различных приближений. Эти предположения определяют границы применимости моделей и степень адекватности получаемых теоретических результатов экспериментальным данным.

В следующих параграфах данной главы, не решая уравнений Максвелла, обсудим вид решений этих уравнений и некоторые их свойства.

2.2. ВОЛНОВОЕ УРАВНЕНИЕ И СКОРОСТЬ СВЕТА

Уравнения Максвелла могут быть преобразованы к виду, удобному для предварительного анализа их решения. Для этого продифференцируем по времени третье из материальных уравнений (2.3) и подставим результат во второе уравнение (2.1), тогда получим

$$\left(\frac{1}{\mu_0\mu}\right)\mathrm{rot}\,\mathbf{E}+\frac{\partial\mathbf{H}}{\partial t}=0.$$

Взяв rot от обеих частей, будем иметь

$$\mathrm{rot}\left(\frac{1}{\mu_0\mu}\mathrm{rot}\,\mathbf{E}\right)+\mathrm{rot}\left(\frac{\partial\mathbf{H}}{\partial t}\right)=0. \qquad (2.10)$$

Далее, продифференцировав первое уравнение из (2.1) по времени и второе из материальных уравнений (2.3) дважды по времени и подставив их в уравнение (2.10), получим

$$\mathrm{rot}\left(\frac{1}{\mu_0\mu}\mathrm{rot}\,\mathbf{E}\right)+\frac{\varepsilon_0\varepsilon\,\partial^2\mathbf{E}}{\partial t^2}+\frac{\partial\mathbf{J}}{\partial t}=0. \qquad (2.11)$$

Считаем, что токов нет ($\mathbf{J}=0$), и пользуемся правилом взятия оператора rot от произведения скалярной u и векторной $\mathbf{v}$ функций

$$\mathrm{rot}\,u\mathbf{v}=u\,\mathrm{rot}\,\mathbf{v}+(\mathrm{grad}\,u)\times\mathbf{v},$$

а также правилом взятия операции rot rot = grad div - Δ, в результате чего получаем уравнение вида

$$\Delta\mathbf{E}-\left(\frac{\varepsilon\mu}{c^2}\right)\frac{\partial^2\mathbf{E}}{\partial t^2}+\left[\operatorname{grad}\left(\ln\mu_0\mu\right)\right]\times\operatorname{rot}\mathbf{E}+\operatorname{grad}\left[\mathbf{E}\left(\ln\varepsilon_0\varepsilon\right)\right]=0, \quad (2.12)$$

Далее, используя третье уравнение из (2.1) и тождество div $u\mathbf{v}$ = $= u$ div $\mathbf{v}$ + $\mathbf{v}$ grad u, получаем, что

$$\varepsilon_0\varepsilon\operatorname{div}\mathbf{E}+\mathbf{E}\operatorname{grad}\varepsilon_0\varepsilon=0.$$

Подставляя это уравнение в (2.12), получаем уравнение для $\mathbf{E}$:

$$\Delta\mathbf{E}-\left(\frac{\varepsilon\mu}{c^2}\right)\frac{\partial^2\mathbf{E}}{\partial t^2}+\left[\operatorname{grad}\left(\ln\mu_0\mu\right)\right]\times\operatorname{rot}\mathbf{E}+\operatorname{grad}\left[\mathbf{E}\left(\ln\varepsilon_0\varepsilon\right)\right]=0, \quad (2.13)$$

где c — скорость света.

Уравнение (2.13) упрощается, если среда однородна и производные от ε и μ равны нулю, и принимает вид хорошо известного волнового уравнения

$$\Delta\mathbf{E}-\left(\frac{\varepsilon\mu}{c^2}\right)\frac{\partial^2\mathbf{E}}{\partial t^2}=0, \quad (2.14)$$

где учтено, что $c=1/\sqrt{\varepsilon_0\mu_0}$

Аналогичное уравнение может быть получено и для вектора напряженности магнитного поля $\mathbf{H}$.

Появление подобных уравнений в теории электромагнитного поля подтверждает гипотезу Максвелла о наличии электромагнитных волн, распространяющихся в среде со скоростью

$$v=c/\sqrt{\varepsilon\mu}$$

Для прозрачных веществ ε обычно больше 1, $\mu\approx1$, а скорость света в среде v меньше скорости света в вакууме. Величина $n=\sqrt{\varepsilon\mu}$ носит название показателя преломления среды.

2.3. СКАЛЯРНЫЕ ВОЛНЫ

Простейшие свойства волнового уравнения, и, следовательно, электромагнитных волн, которые ему подчиняются, можно получить, если рассмотреть случай однородной среды, свободной от токов и зарядов. Тогда имеем возможность рассматривать не векторный вариант волнового уравнения, а скалярный, когда каждая составляющая V векторов, описывающих поле, удовлетворяет однородному волновому уравнению вида

$$\nabla^2 V-\frac{1}{v^2}\frac{\partial^2 V}{\partial t^2}=0. \quad (2.15)$$

Простейшим решением данного уравнения является плоская волна.

Пусть $\mathbf{r}(x, y, z)$ — радиус-вектор точки P, а $\mathbf{s}(s_x, s_y, s_z)$ — единичный вектор с фиксированным направлением (рис. 2.2). Тогда любое решение вида $V(\mathbf{rs}, t)$ представляет собой плоскую волну, так как в каждый момент времени эта функция постоянна в плоскостях $\mathbf{rs}$ = const, перпендикулярных к единичному вектору $\mathbf{s}$.

Для того чтобы сделать анализ более наглядным, перейдем к новой переменной $\zeta=\mathbf{rs}$, которая отсчитывается по оси, перпендикулярной плоскости $\mathbf{rs}$ = const. Тогда производные по координатам приобретают вид

$$\frac{\partial}{\partial x}=s_x\frac{\partial}{\partial\zeta};\quad \frac{\partial}{\partial y}=s_y\frac{\partial}{\partial\zeta};\quad \frac{\partial}{\partial z}=s_z\frac{\partial}{\partial\zeta},$$

а само уравнение (2.15) становится весьма простым:

$$\frac{\partial^2 V}{\partial\zeta^2}-\frac{1}{v^2}\frac{\partial^2 V}{\partial t^2}=0,$$

так как

$$\Delta V=\frac{\partial^2 V}{\partial\zeta^2},$$

а сумма квадратов компонент вектора $\mathbf{s}$ равна единице.

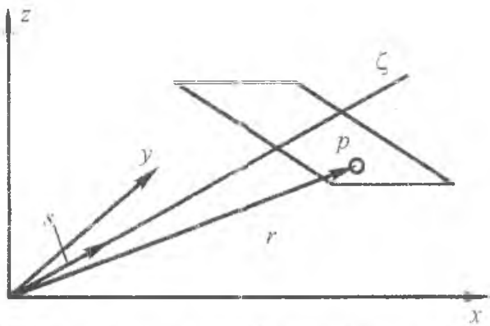

Рис. 2.2. Распределение плоской волны (скалярный вариант)

Если произвести замену и обозначить $\zeta-vt=p$ и $\zeta+vt=q$, то вышеприведенное уравнение еще более упростится и примет вид

$$\frac{\partial^2 V}{\partial p\,\partial q}=0.$$

Общим решением этого уравнения служит функция

$$V=V_1(p)+V_2(q)=V_1(\mathbf{rs}-vt)+V_2(\mathbf{rs}+vt),$$

где V_1, V_2 — произвольные функции.

Видно, что V_1 представляет собой возмущение, движущееся со скоростью v против направления оси ζ, а V_2 — такое же возмущение, но движущееся в противоположном направлении, т.е. решение уравнения представляет собой суперпозицию плоских волн, движущихся в противоположных направлениях.

Аналогичная ситуация складывается и в том случае, когда решение ищется с учетом сферической симметрии, т.е. в виде сферических волн вида

$$V = V(r, t),$$

где $r = \left(x^2 + y^2 + z^2\right)^{1/2}$.

Переход к сферическим координатам при дифференцировании преобразует уравнение (2.15) к виду

$$\frac{\partial^2}{\partial r^2}(rV) - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}(rV) = 0.$$

Решение этого уравнения также имеет вид суммы двух волн, распространяющихся в противоположных направлениях,

$$V = \frac{V_1(r - vt)}{r} + \frac{V_2(r + vt)}{r},$$

первая — расходящаяся от начала координат, а вторая — сходящаяся.

2.4. ГАРМОНИЧЕСКИЕ ВОЛНЫ И ИХ КОМПЛЕКСНАЯ ЗАПИСЬ

Рассмотрим оптическое поле, зависящее от времени по гармоническому закону. Если это плоская волна, распространяющаяся вдоль некоторого единичного вектора s, то она может быть представлена выражением вида

$$V(\mathbf{r}, t) = a\cos\left[\omega\left(t - \frac{\mathbf{rs}}{v}\right) + \delta\right],$$

где δ — начальная фаза.

Подобная волна распространяется в среде со скоростью v и круговой частотой ω. Поверхности постоянной фазы в данный момент времени носят название волновых фронтов.

Расчеты, связанные с гармоническими волнами, значительно упрощаются, если перейти от тригонометрических функций к экспоненциальным, т.е. записать волновое поле в виде

$$V(\mathbf{r}, t) = \mathrm{Re}\left\{U(\mathbf{r})\exp[-j\omega t]\right\}, \qquad (2.17)$$

где

$$U(\mathbf{r}) = a(\mathbf{r})\exp\left[jg(\mathbf{r})\right]. \qquad (2.18)$$

Последняя функция носит название комплексной амплитуды, а Re обозначает действительную часть. Функция фазы $g(\mathbf{r})$ в частном случае плоской волны имеет вид

$$g(\mathbf{r}) = \omega\left(\frac{\mathbf{rs}}{v}\right) - \delta = \mathbf{kr} - \delta,$$

где $\mathbf{k}$ — волновой вектор, численно равный

$$k = \frac{2\pi}{\lambda} = \frac{n\omega}{c} = \frac{\omega}{v}.$$

Если подставить функцию поля (2.17) в волновое уравнение (2.14), то последнее преобразуется к виду

$$\Delta U + k^2 U = 0. \qquad (2.19)$$

Это уравнение носит название уравнения Гельмгольца.

Следует отметить ограничения, возникающие при использовании комплексного представления световых полей. Они связаны с тем, что в реальном мире могут существовать только действительные величины. Поэтому он должен описываться гармоническими полями в тригонометрическом представлении либо действительными частями комплексных функций [см. (2.17)]. Однако в том случае, когда операции, проводимые над полями, линейны, можно забыть о различии между комплексной функцией и ее действительной частью и проводить все вычисления с функциями в комплексном виде, переходя к действительным их частям только в самом конце преобразований. Подобный подход позволяет существенно упростить вычисления и сделать их более наглядными.

2.5. БИЕНИЯ ГАРМОНИЧЕСКИХ ВОЛН

Гармонические сигналы очень наглядны и просты для расчетов, однако они идеализированы, поскольку в природе не бывает строго монохроматических волн. Если бы такой сигнал существовал, то его спектр описывался δ-функцией и производная по спектру была бы бесконечной. Но на практике любой сигнал имеет отличную от нуля ширину спектра.

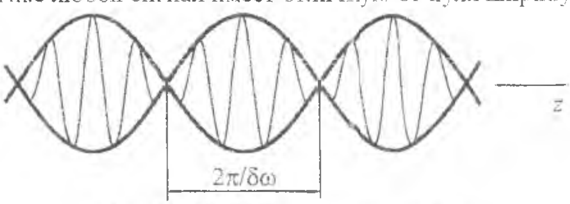

Рис. 2.3. Биение гармонических волн

Рассмотрим простейшую модель сигнала, спектр которого представляет собой две монохроматические компоненты, имеющие одинаковую амплитуду и распространяющиеся вдоль оси Z:

$$V(z,t) = a\exp\left[-j(\omega t - kz)\right] + a\exp\left\{\left[-j(\omega + \delta\omega)t - (k + \delta k)z\right]\right\}. \qquad (2.20)$$

Уравнение (2.20) можно привести к более наглядному виду:

$$V(z,t)=a\left\{\exp\left[\frac{1}{2}j\left(t\delta\omega-z\delta k\right)\right]+\exp\left[-\frac{1}{2}j\left(t\delta\omega-z\delta k\right)\right]\right\}\exp\left(-\bar{\omega}k-\bar{k}z\right)=$$

$$=2a\cos\left[\frac{1}{2}\left(t\delta\omega-z\delta k\right)\right]\exp\left[-j\left(\bar{\omega}t-\bar{k}z\right)\right], \qquad (2.21)$$

где $\bar{\omega}=\omega-1/2\delta\omega$; $\bar{k}=k+1/2\delta k$ — средняя частота и среднее волновое число соответственно.

Можно считать, что данное выражение описывает плоскую волну с частотой $\bar{\omega}$ и волновым вектором $\bar{k}$, распространяющуюся вдоль оси z. Причем амплитуда этой волны меняется во времени и пространстве от нуля до $2a$, что вызывает хорошо известное явление биений (рис. 2.3).

Из формулы (2.21) вытекает, что плоскости постоянной амплитуды распространяются со скоростью

$$v^{(g)}=\frac{\delta\omega}{\delta k},$$

которая носит название групповой скорости, а плоскости постоянной фазы распространяются со скоростью

$$v^{(p)}=\bar{\omega}/\bar{k},$$

носящей название фазовой скорости.

В недиспергирующей среде, т.е. в такой среде, где нет зависимости показателя преломления n от частоты ω, фазовая и групповая скорости равны и могут быть определены как c/n. Однако если в среде присутствует дисперсия, то эти скорости различны, а если среда является еще и анизотропной, то векторы этих скоростей, как правило, неколлинеарны.

Именно групповая скорость является скоростью передачи информации в оптической системе. Это связано с тем, что информация, переносимая волной, в любом случае изменяет несущую частоту, что всегда может быть представлено как ее модуляция: амплитудная, частотная, фазовая или поляризационная. Скорость распространения модуляции, как это следует из выражения (2.21), есть групповая скорость, т.е. она и есть скорость передачи информации.

2.6. ВЕКТОРНЫЕ ВОЛНЫ

В общем случае решением системы уравнений Максвелла является волна, состоящая из нескольких компонент, т.е. имеющая векторный, а не скалярный характер. Рассмотрим простейшее решение системы уравнений (2.1) в виде плоской волны. В этом случае, как это следует из § 2.4, каждая из компонент поля зависит от пространственных и временных переменных только через их комбинацию $u = \mathbf{rs} - vt$, т.е.

$$\mathbf{E} = \mathbf{E}(\mathbf{rs} - vt), \quad \mathbf{H} = \mathbf{H}(\mathbf{rs} - vt),$$

где $\mathbf{s}$ — как и раньше, единичный вектор в направлении распространения волны.

Получим выражения для производных по времени (обозначаются точкой) и по комбинации u (обозначаются штрихом) при такой комбинации переменных:

$$\dot{\mathbf{E}} = -v\mathbf{E}';$$

$$(\operatorname{rot}\mathbf{E})_x = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = E'_z s_y - E'_y s_z = (\mathbf{s} \times \mathbf{E}')_x.$$

Подставим теперь производные в этом виде в уравнения Максвелла и, воспользовавшись материальными уравнениями, получим

$$\left.\begin{aligned} \mathbf{s} \times \mathbf{H} + \varepsilon_0 \varepsilon v \mathbf{E}' = 0; \\ \mathbf{s} \times \mathbf{E} + \mu_0 \mu v \mathbf{H}' = 0. \end{aligned}\right\} \qquad (2.22)$$

Считая постоянную интегрирования равной нулю, т.е. пренебрегая постоянным полем, и учитывая, что $v/c = 1/\sqrt{\varepsilon\mu}$, получаем решения:

$$\left.\begin{aligned} \mathbf{E} = -\sqrt{\frac{\mu_0\mu}{\varepsilon_0\varepsilon}}(\mathbf{s} \times \mathbf{H}); \\ \mathbf{H} = -\sqrt{\frac{\varepsilon_0\varepsilon}{\mu_0\mu}}(\mathbf{s} \times \mathbf{E}). \end{aligned}\right\} \qquad (2.23)$$

Умножаем скалярно полученные уравнения на векторы $\mathbf{s}$ и получаем условие поперечности электромагнитной волны

$$\mathbf{Es} = \mathbf{Hs} = 0,$$

которое показывает, что электрический и магнитный векторы лежат в плоскости, перпендикулярной направлению распространения.

2.7. ПОЛЯРИЗАЦИЯ ЭЛЕКТРОМАГНИТНОЙ ВОЛНЫ

Для того чтобы наглядно ввести понятие поляризации волны, рассмотрим случай гармонической плоской волны. Для нее каждая из декартовых компонент полей меняется косинусоидально:

$a\cos(\tau + \delta)$, где τ переменная часть фазового множителя,

$$\tau = \omega\left(t - \frac{\mathbf{rs}}{v}\right) = \omega t - \mathbf{kr}$$

Пусть распространение волны идет вдоль оси z. Тогда вследствие поперечности электромагнитной волны у нее будут только x- и y- компоненты (вектор $\mathbf{s}$ направлен вдоль оси z).

Рассмотрим кривую, которую описывает конец вектора E в произвольной точке пространства. Эта кривая является геометрическим местом точек, координаты которых равны

$$\left.\begin{aligned} E_x &= a_1\cos(\tau+\delta_1); \\ E_y &= a_2\cos(\tau+\delta_2). \end{aligned}\right\} \tag{2.24}$$

Преобразуем уравнения (2.24):

$$\frac{E_x}{a_1} = \cos\tau\cos\delta_1 - \sin\tau\sin\delta_1;$$

$$\frac{E_y}{a_2} = \cos\tau\cos\delta_2 - \sin\tau\sin\delta_2,$$

следовательно,

$$\left.\begin{aligned} \frac{E_x}{a_1}\sin\delta_2 - \frac{E_y}{a_2}\sin\delta_1 &= \cos\tau\sin(\delta_2-\delta_1); \\ \frac{E_x}{a_1}\cos\delta_2 - \frac{E_y}{a_2}\cos\delta_1 &= \cos\tau\sin(\delta_2-\delta_1). \end{aligned}\right\} \tag{2.25}$$

Возводим уравнения (2.25) в квадрат и складываем

$$\left(\frac{E_x}{a_1}\right)^2 + \left(\frac{E_y}{a_2}\right)^2 - 2\frac{E_xE_y}{a_1a_2}\cos\delta = (\sin\delta)^2, \tag{2.26}$$

где $\delta = \delta_2 - \delta_1$.

Уравнение (2.26) носит название канонического сечения. Геометрическое место точек концов вектора напряженности электрического (магнитного) поля в общем случае представляет собой эллипс, который вписан в прямоугольник со сторонами $2a_1$ и $2a_2$. В этом случае говорят, что волна эллиптически поляризована (рис. 2.4, а).

В частном случае эллипс канонического сечения может выродиться либо в прямую линию, либо в окружность (рис. 2.4, б, в). В этих случаях имеет место линейная или круговая поляризация (в зависимости от направления вращения по окружности круговая поляризация также может быть правой и левой).

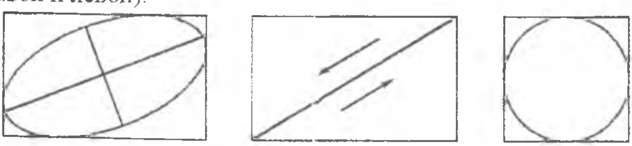

Рис. 2.4. Световая волна эллиптической поляризации при разных значениях δ
а — $0<\delta<\pi/2$; б — $\delta=0$; в — $0<\delta<\pi/2$, но $a_1 = a_2$

2.8. ОТРАЖЕНИЕ И ПРЕЛОМЛЕНИЕ СВЕТА (СЛУЧАЙ ПЛОСКОЙ ВОЛНЫ)

Анализ уравнений Максвелла на границе раздела двух сред (2.6)—(2.9) позволяет определить, что происходит с компонентами векторов поля. Рассмотрим случай плоской волны, пересекающей границу раздела двух однородных сред, и получим формулы, связывающие отраженную, преломленную и падающие волны.

Плоская волна, распространяющаяся в направлении единичного вектора $\mathbf{s}^i$, известна во всем пространстве, если она известна хотя бы в одной его точке. На границе раздела вторичные поля (отраженное и преломленное) будут меняться во времени, как и первичное поле падающей волны. Поэтому, если $\mathbf{s}^r$ и $\mathbf{s}^t$ единичные векторы в направлении распространения отраженной и преломленной волн, то, приравнивая аргументы трех волновых функций в точке с координатами (x, y, z) на границе раздела, получаем

$$t-\frac{\mathbf{rs}^i}{v_1}=t-\frac{\mathbf{rs}^r}{v_1}=t-\frac{\mathbf{rs}^t}{v_2},$$

где v_1, v_2 — скорость волны в первой и второй средах (рис. 2.5).

Выбрав границей раздела двух сред плоскость $z = 0$, находим, что предыдущее выражение преобразуется к виду

$$\frac{xs_x^i+ys_y^i}{v_1}=\frac{xs_x^r+ys_y^r}{v_1}=\frac{xs_x^t+ys_y^t}{v_2}. \qquad (2.27)$$

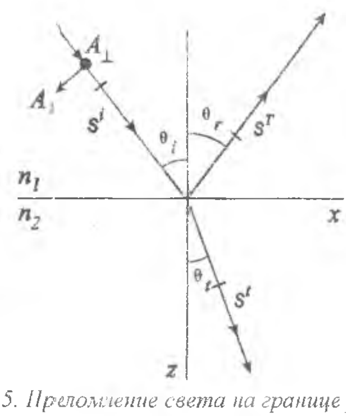

Рис. 2.5. Преломление света на границе раздела

Это равенство должно выполняться в произвольной точке границы, т.е. для всех x и y, поэтому

$$\left.\begin{aligned} \frac{s_x^i}{v_1} = \frac{s_x^r}{v_1} = \frac{s_x^t}{v_2}; \\ \frac{s_y^i}{v_1} = \frac{s_y^r}{v_1} = \frac{s_y^t}{v_2} \end{aligned}\right\} \qquad (2.28)$$

Соотношение (2.28) показывает, что направляющие векторы отраженной и преломленной волн $\mathbf{s}^r$ и $\mathbf{s}^t$ лежат в одной плоскости, определяемой направляющим вектором падающей волны $\mathbf{s}^i$ и нормалью к границе раздела двух сред и называемой плоскостью падения.

Считаем xz плоскостью падения и вводим углы θ_i, θ_r и θ_t, тогда

$$s_x^i = \sin\theta_i, \quad s_y^i = 0, \quad s_z^i = \cos\theta_i;$$

$$s_x^r = \sin\theta_r, \quad s_y^r = 0, \quad s_z^r = \cos\theta_r;$$

$$s_x^t = \sin\theta_t, \quad s_y^t = 0, \quad s_z^t = \cos\theta_t.$$

Если волна распространяется из первой среды во вторую, то компонента вектора $\mathbf{s}$ вдоль оси z положительна, если наоборот, то отрицательна, т.е.

$$s_z^i = \cos\theta_i \ge 0, \quad s_z^r = \cos\theta_r \le 0; \quad s_z^t = \cos\theta_t \ge 0. \qquad (2.29)$$

Подставляя выражения (2.29) в выражение (2.28), получаем

$$\frac{\sin\theta_i}{v_1} = \frac{\sin\theta_r}{v_1} = \frac{\sin\theta_t}{v_2}, \text{ или } \frac{\sin\theta_i}{\sin\theta_t} = \frac{v_1}{v_2} = \frac{n_1}{n_2} \text{ или } n_1\sin\theta_i = n_2\sin\theta_t. \qquad (2.30)$$

Данное соотношение вместе с условием, что все волновые векторы лежат в одной плоскости, составляет *закон преломления Снеллиуса — произведение показателя преломления первой среды на синус угла падения равно произведению показателя преломления второй среды на синус угла преломления.*

Если выполняется условие $n_2 \ge n_1$, то оптическая плотность второй среды больше, чем первой. В этом случае

$$\sin\theta_t = \frac{n_1}{n_2}\sin\theta_i \le \sin\theta_i$$

и для каждого угла падения существует вещественный угол преломления θ_t. Но если вторая среда оптически менее плотная, чем первая, то вещественное значение угла θ_t, можно получить только для таких θ_i, для которых выполняется условие

$$\frac{n_1}{n_2}\sin\theta_i \le 1. \qquad (2.31)$$

Это условие определяет эффект полного внутреннего отражения, а соответствующий ему угол носит название угла полного внутреннего отражения. В случае падения волны под углом, большим угла полного внутреннего отражения, во второй среде исчезает однородная преломленная волна, и коэффициент отражения границы раздела двух сред становится равным единице (по модулю). Как показывает анализ уравнений Максвелла, во второй среде в этом случае распространяется неоднородная электромагнитная волна.

Явление полного внутреннего отражения широко используется в науке и технике, в том числе для реализации оптических волноводов, измерительных приборов и т.д.

2.9. ФОРМУЛЫ ФРЕНЕЛЯ

Остановимся теперь на соотношении амплитуд падающего, отраженного и прошедшего через границу раздела двух сред света. Предположим, что обе среды прозрачные, однородные и изотропные (т.е. поглощения и рассеяния нет, $\mu_1 = \mu_2 = 1$).

Вернемся к рис. 2.5 и проанализируем векторные соотношения для комплексных амплитуд рассматриваемых волн. Пусть A — комплексная амплитуда электрического вектора падающей волны. Переменная часть фазы электрического вектора записывается в следующем виде:

$$\tau_i = \omega\left(t - \frac{\mathbf{rs}'}{v_1}\right) = \omega\left(t - \frac{x\sin(\theta_i + z\cos\theta_i)}{v_1}\right). \tag{2.32}$$

Разложим электрический вектор $\mathbf{E}$ на две компоненты — параллельную и перпендикулярную (см. рис. 2.5) плоскости падения ($A_{||}$ и $A_{\perp}$):

$$\left.\begin{aligned} E_x^i &= -A_{||}\cos\theta_i\exp(-j\tau_i); \\ E_y^i &= A_{\perp}\exp(-j\tau_i); \\ E_z^i &= A_{||}\sin\theta_i\exp(-j\tau_i). \end{aligned}\right\} \tag{2.33}$$

Воспользовавшись соотношением (2.23) и выражениями (2.33), можно получить соответствующие компоненты магнитного вектора $\mathbf{H}$:

$$\left.\begin{aligned} H_x^i &= -A_{\perp}\cos\theta_i\sqrt{\varepsilon_0\varepsilon/\mu_0}\exp(-j\tau_i); \\ H_y^i &= -A_{||}\sqrt{\varepsilon_0\varepsilon/\mu_0}\exp(-j\tau_i); \\ H_z^i &= A_{\perp}\sin\theta_i\sqrt{\varepsilon_0\varepsilon/\mu_0}\exp(-j\tau_i). \end{aligned}\right\} \tag{2.34}$$

Если обозначить комплексные амплитуды отраженной и преломленных волн соответственно R и T, то можно аналогичным образом получить формулы по типу (2.33), (2.34) для их компонент. При этом необходимо учесть, что преломленная волна распространяется во второй среде и в формулах для компонент ее полей будут стоять ε_2 и v_2. В качестве при мера запишем выражения для компонент магнитного вектора преломленной волны:

$$\left.\begin{aligned} H_x^t &= -T_\perp \cos\theta_t \sqrt{\varepsilon_0\varepsilon/\mu_0}\exp(-j\tau_t);\\ H_y^t &= -T_\parallel \sqrt{\varepsilon_0\varepsilon/\mu_0}\exp(-j\tau_t);\\ H_z^t &= T_\perp \sin\theta_t \sqrt{\varepsilon_0\varepsilon/\mu_0}\exp(-j\tau_t). \end{aligned}\right\} \quad (2.35)$$

В соответствии с выведенными в данной главе граничными условиями для векторов поля необходимо, чтобы тангенциальные составляющие векторов **E** и **H** были непрерывны на границе раздела двух сред:

$$E_x^i + E_x^r = E_x^t; \quad E_y^i + E_y^r = E_y^t;$$

$$H_x^i + H_x^r = H_x^t; \quad H_y^i + H_y^r = H_y^t.$$

Для нормальных компонент векторов граничные условия будут выполняться автоматически, так как нет зарядов и токов. Используя тот факт, что $\cos\theta_r = \cos(\pi - \theta_i) = -\cos\theta_i$, получаем:

$$\left.\begin{aligned} &\cos\theta_i \left(A_\parallel - R_\parallel\right) = \cos\theta_t T_\parallel;\\ &A_\perp + R_\perp = T_\perp;\\ &\sqrt{\varepsilon_1}\cos\theta_i\left(A_\perp - R_\perp\right) = \sqrt{\varepsilon_2}\cos\theta_t T_\perp;\\ &\sqrt{\varepsilon_1}\left(A_\parallel - R_\parallel\right) = \sqrt{\varepsilon_2}\, T_\parallel. \end{aligned}\right\} \quad (2.36)$$

Решая систему (2.36) и используя равенство $n = \sqrt{\varepsilon}$, получаем формулы, связывающие компоненты комплексных амплитуд, участвующих в процессе отражения волн. Эти формулы и носят название формул Френеля:

$$\left.\begin{aligned} T_\parallel &= \frac{2n_1\cos\theta_i}{n_2\cos\theta_i + n_1\cos\theta_t} A_\parallel;\\ T_\perp &= \frac{2n_1\cos\theta_i}{n_1\cos\theta_i + n_2\cos\theta_t} A_\perp;\\ R_\parallel &= \frac{n_2\cos\theta_i - n_1\cos\theta_t}{n_2\cos\theta_i + n_1\cos\theta_t} A_\parallel;\\ R_\perp &= \frac{n_1\cos\theta_i - n_2\cos\theta_t}{n_1\cos\theta_i + n_2\cos\theta_t} A_\perp. \end{aligned}\right\} \quad (2.37)$$

Так как рассматривается обычное отражение (без полного внутреннего отражения), то значения углов действительны и тригонометрические функции, стоящие в правых частях уравнений также действительны. Поэтому фаза компонент отраженной и преломленной волн либо совпадает с фазой падающей волны, либо отличается от нее на π

Так как знаки параллельной и перпендикулярной составляющих преломленной волны совпадают со знаками таких же компонент падающей волны, то их фазы совпадают. Другая ситуация складывается с фазами компонент отраженной волны. Либо третье, либо четвертое уравнение системы (2.37), но не одновременно, имеют знак минус по отношению к компонентам падающей волны. Это означает, что фаза одной из компонент отраженной волны всегда отличается на π от фазы соответствующей компоненты падающей волны.

Возвращаясь к вопросу о полном внутреннем отражении, можно показать, что при подстановке формулы (2.31) для соответствующих углов в формулы Френеля получаем тот же вывод, что и ранее:

$$|R_{\parallel}| = |A_{\parallel}|, \quad |R_{\perp}| = |A_{\perp}|,$$

а амплитуды составляющих преломленной волны равны нулю. Энергия волны в общем случае не проникает во вторую среду, а распространяется вдоль границы раздела двух сред в плоскости падения.

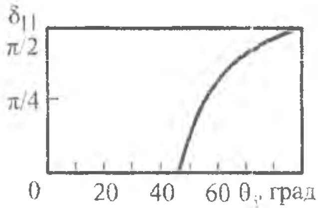

Рис. 2.6. Изменение фазы для $R_{\parallel}$ на границе раздела плавленого кварца и воздуха

Следует отметить еще один важный момент относительно комплексных амплитуд, связанный с фазами отраженных волн за границей полного внутреннего отражения. Можно показать, что фазы отраженных компонент волны меняются экспоненциально относительно компонент падающей волны:

$$\left.\begin{aligned} \frac{R_{\parallel}}{A_{\parallel}} &= \exp\left(j\delta_{\parallel}\right); \\ \frac{R_{\perp}}{A_{\perp}} &= \exp\left(j\delta_{\perp}\right), \end{aligned}\right| \qquad (2.38)$$

а величины $\delta_{\perp}$ и $\delta_{\parallel}$ могут быть определены исходя из следующих формул, которые легко выводятся из формул Френеля и условий полного внутреннего отражения:

$$\left.\begin{aligned} \operatorname{tg}\frac{\delta_{\parallel}}{2} &= -\frac{\sqrt{\sin^2\theta_i - n^2}}{n^2\cos\theta_i}; \\ \operatorname{tg}\frac{\delta_{\perp}}{2} &= -\frac{\sqrt{\sin^2\theta_i - n^2}}{\cos\theta_i} \end{aligned}\right\} \tag{2.39}$$

В качестве примера приведем график изменения фазы $\delta_{\parallel}$ для $R_{\parallel}$ в случае границы раздела плавленого кварца SiO_2 и воздуха при отношении показателей преломления 1,46 (рис. 2.6).

Кроме угла полного внутреннего отражения существует и еще один критический угол при отражении, который связан с изменением поляризации отраженного света, — это *угол Брюстера*. Он соответствует случаю, когда отраженный и преломленный лучи перпендикулярны. При этом выполняется условие

$$\operatorname{tg}\theta_i = n_2/n_1. \tag{2.40}$$

Если свет падает под этим углом, то электрический вектор отраженной волны не имеет составляющей в плоскости падения, т.е. отраженный свет линейно поляризован. Это явление широко используется в технике, в частности, полировка торцов активных элементов твердотельных лазеров под углом Брюстера позволяет получать в них линейно поляризованное излучение.

Объяснение этого эффекта довольно простое, и связано с характером возбуждения колебаний в среде под действием падающего излучения. Падающая во вторую среду электромагнитная волна возбуждает в ней колебания электронов в атомах в направлении изменения электрического вектора, т.е. перпендикулярно направлению падения. Колеблющиеся относительно тяжелого ядра атома электроны формируют излучающие диполи, которые, как известно, излучают в направлении, перпендикулярном оси диполя, т.е. в направлении колебаний диполя поток энергии отсутствует. Поэтому в отраженном луче энергия колебаний в плоскости падения равна нулю, если отраженный и прошедший лучи перпендикулярны.

2.10. ФОРМУЛА ЛОРЕНТЦ-ЛОРЕНЦА (СВЯЗЬ ОПТИЧЕСКИХ И МЕХАНИЧЕСКИХ ПАРАМЕТРОВ СРЕДЫ)

До сих пор, говоря в общем виде о взаимодействии электромагнитного поля и среды, мы не рассматривали механизмы этого взаимодействия. На самом деле, электрически заряженные частицы среды должны реагировать на поле, а также должно быть и обратное воздействие частиц на поле, поэтому необходимо найти форму описания этих процессов, причем такую, чтобы она не противоречила уравнениям Максвелла.

Модель вещества может быть построена как совокупность взаимодействующих частиц в вакууме. Подобные частицы образуют локальные поля, сильно меняющиеся по мере удаления от частиц, кроме того, на эти поля накладывается еще и внешнее поле. Необходимо также учитывать, что оптика рассматривает макрособытия, поэтому измеряются некоторые усредненные значения и свойства вещества находятся путем усреднения по полному полю внутри некоторого объема. Но надо помнить, что воздействие на конкретную частицу осуществляет все-таки локальное поле, а не усредненное. Имеет смысл ввести понятие эффективного поля, определяющее взаимодействие поля и частицы в среде.

Связь между полем и средой ранее описывалась материальными уравнениями (2.3), однако более удобны соотношения вида

$$\left.\begin{aligned} \mathbf{D} &= \varepsilon_0\mathbf{E} + \mathbf{P}; \\ \mathbf{B} &= \mu_0\mathbf{H} + \mathbf{M}, \end{aligned}\right\} \tag{2.41}$$

где **P** — вектор электрической поляризации, а **M** — вектор намагничивания. Суть этих векторов определяется тем, что поведение электронов в среде, колеблющихся относительно неподвижного тяжелого ядра под действием прикладываемого электромагнитного поля, схоже с поведением электрических и магнитных диполей. Векторные суммы электрических и магнитных дипольных моментов единицы объема среды определяют дипольные моменты единицы объема, которые с точностью до коэффициента равны векторам электрической поляризации и намагничивания.

В приближении слабых полей (малых по сравнению с внутриатомными) можно предположить линейную зависимость поляризации и намагничивания от напряженности электрического и магнитного полей:

$$\left.\begin{aligned} \mathbf{P} &= \varepsilon_0\eta\mathbf{E}; \\ \mathbf{M} &= \mu_0\chi\mathbf{H}. \end{aligned}\right\} \tag{2.42}$$

Величины η и χ носят название диэлектрической и магнитной восприимчивости. Ситуации, когда линейность не соответствует реально наблюдаемым явлениям, рассматриваются в нелинейной оптике.

Объединение формул (2.41) и (2.42) позволяет найти связь между проницаемостями и восприимчивостями:

$$\left.\begin{aligned} \varepsilon &= 1 + \eta; \\ \mu &= 1 + \chi. \end{aligned}\right\} \tag{2.43}$$

Молекулы среды могут быть разными — они могут обладать, а могут и не обладать электрическим и магнитным дипольным моментом. Кроме того, даже если дипольный момент у молекулы есть, вещество, состоящее из них, может обладать, а может не обладать спонтанной поляризацией или намагничиванием. Однако если прикладывается внешнее поле, происходит либо поляризация молекул, либо поворот уже имеющихся дипольных моментов по полю, то у единицы объема среды появляется средний дипольный момент или, как мы уже говорили, вектор поляризации или намагничивания. В общем случае, векторы поляризации среды и напряженности электрического поля неколлинеарны.

Рассмотрим изотропную и немагнитную среду и получим зависимости электрических постоянных от плотности среды для вещества, состоящего из одинаковых молекул.

Как уже было отмечено, необходимо различать эффективные поля **E'**, **H'**, действующие на молекулу, и поля **E**, **H**, полученные усреднением по объему. Необходимо найти разность **E'**-**E**, **H'**-**H**. Для этого построим модель взаимодействия молекулы с полем и соседними молекулами. Предположим, что молекула окружена небольшой сферой, но большей по сравнению с размерами молекулы. Определим воздействие на данную молекулу вещества внутри сферы и снаружи её. Из-за произвольного выбора размера сферы мы можем считать, что вне её вещество непрерывно, а значит, и поляризация **P**, создаваемая усредненным полем, постоянна. Внутри сферы вещество не воздействует на молекулу, так как воздействия с разных сторон (сфера большая относительно размеров молекул и межмолекулярных расстояний) уравновешивают друг друга. Для случая хаотического расположения молекул это доказывается относительно просто. Таким образом, можно считать, что молекула находится в вакууме внутри сферы, снаружи которой равномерно поляризованная среда. На границе сферы поляризация меняется от значения **P** до нуля, следовательно на этой границе должны быть свободные заряды. Потенциал поля такой конфигурации обозначим φ.

Введем компенсирующий потенциал φ, создаваемый однородно заряженной сферой, находящейся в вакууме. Сложение двух этих конфигураций дает среду без границ с однородной поляризацией, т.е.

$$\varphi + \overline{\varphi} = 0. \tag{2.44}$$

Для конфигурации с однородно заряженной сферой можно доказать, что потенциал создаваемого ею поля

$$\overline{\varphi} = \mathbf{P}\int \mathrm{grad}'\frac{1}{R}\mathrm{d}V', \tag{2.45}$$

где координатам со штрихом, по которым проводится интегрирование и взятие операции градиента (grad'), соответствуют точки объема V', ограниченного сферой.

Учитывая, что

$$R = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2},$$

можем брать градиент по нештрихованным координатам, одновременно взяв интеграл со знаком минус,

$$\bar{\varphi} = -\mathbf{P}\,\mathrm{grad}\int\frac{\mathrm{d}V'}{R} = \mathbf{P}\,\mathrm{grad}\,\varphi_0 = -\varphi. \tag{2.46}$$

где φ_0 — потенциал однородно заряженной сферы с плотностью заряда, равной 1. Следовательно, он должен удовлетворять уравнению Пуассона

$$\nabla^2\varphi_0 = 1/\varepsilon_0.$$

В явном виде производные от потенциала φ запишутся следующим образом ($\mathbf{P}$ = const):

$$-\frac{\partial\varphi}{\partial x} = \frac{\partial}{\partial x}\left[P_x\frac{\partial\varphi_0}{\partial x} + P_y\frac{\partial\varphi_0}{\partial y} + P_z\frac{\partial\varphi_0}{\partial z}\right] = P_x\frac{\partial^2\varphi_0}{\partial x^2} + P_y\frac{\partial^2\varphi_0}{\partial x\partial y} + P_z\frac{\partial^2\varphi_0}{\partial x\partial z}. \tag{2.47}$$

Из условия симметрии в центре поля имеем

$$\frac{\partial^2\varphi_0}{\partial x\partial y} = \frac{\partial^2\varphi_0}{\partial x\partial z} = \frac{\partial^2\varphi_0}{\partial y\partial z} = 0; \quad \frac{\partial^2\varphi_0}{\partial x^2} = \frac{\partial^2\varphi_0}{\partial y^2} = \frac{\partial^2\varphi_0}{\partial z^2}.$$

Используя уравнение Пуассона, находим, что каждый член последнего равенства равен $1/3\varepsilon_0$, тогда из выражения (2.47) следует, что

$$\nabla\varphi = \frac{1}{3\varepsilon_0}\mathbf{P}.$$

Эффективное поле, действующее на молекулу, представляет собой сумму среднего поля и вклада поляризации среды:

$$\mathbf{E}' = \mathbf{E} + \frac{1}{3\varepsilon_0}\mathbf{P}. \tag{2.48}$$

Теперь, определив эффективное поле, действующее на молекулы среды, выведем формулу Лоренц—Лоренца. Для этого рассмотрим отдельную молекулу в поле. Эффективное поле вызывает перераспределение заряда в молекуле таким образом, что образуется диполь, характеризуемый дипольным моментом

$$P = \alpha\mathbf{E}', \tag{2.49}$$

где α — средняя поляризуемость.

Полный дипольный момент единицы объема будет складываться из дипольных моментов составляющих его частиц N:

$$\mathbf{P} = N\mathbf{p} = N\alpha\mathbf{E}'. \tag{2.50}$$

Пользуясь формулами (2.50) и (2.48), можно получить связь между макроскопическими и микроскопическими параметрами поля,

$$\left.\begin{aligned} &\mathbf{P} = \varepsilon_0 \eta \mathbf{E}; \\ &\eta = \frac{N\alpha}{\varepsilon_0 - \frac{1}{3} N\alpha}. \end{aligned}\right\} \tag{2.51}$$

Подставим (2.51) в формулу (2.43) и получим аналогичное выражение для ε:

$$\varepsilon = \frac{1 + \frac{2}{3\varepsilon_0} N\alpha}{1 - \frac{1}{3\varepsilon_0} N\alpha}. \tag{2.52}$$

Изменив эту формулу с заменой диэлектрической проницаемости на квадрат показателя преломления, получим формулу Лоренц—Лоренца

$$\alpha = \frac{3\varepsilon_0}{N} \frac{\varepsilon - 1}{\varepsilon + 2} = \frac{3\varepsilon_0}{N} \frac{n^2 - 1}{n^2 + 2}. \tag{2.53}$$

Формула (2.53) практически очень важная. Если считать, что N определяет плотность вещества, то получается зависимость показателя преломления от плотности. В частности, для веществ с показателем преломления близким к 1

$$\rho \approx n^2 - 1,$$

т.е. чем выше плотность, тем выше и показатель преломления.

2.11. ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ДИСПЕРСИИ

Экспериментальные факты однозначно свидетельствуют, что существует зависимость оптических характеристик среды от частоты электромагнитного поля, т.е. существует дисперсия. Наиболее яркими и известными проявлениями дисперсии являются разложение света в спектр в призме и радуга. Для того чтобы объяснить подобное явление, необходимо предположить, что отклик среды на падающее поле частотно зависим, и рассмотреть сам процесс взаимодействия поля с частицами среды.

В данной главе уже несколько раз рассматривалось поведение атомов и молекул среды в электромагнитном поле. Предполагалось, что действующие электромагнитные потенциалы поля вызывают смещение электронов относительно тяжелых ядер, образуя диполи, совершающие вынужденные колебания. Поэтому, для того чтобы выяснить частотную зависимость этого взаимодействия, необходимо рассмотреть вопрос движения электронов в поле.

Известно, что на движущийся в поле заряд действует сила Лоренца

$$\mathbf{F} = e\left(\mathbf{E} + \left[v, \mathbf{B}\right]\right),$$

где e — заряд электрона, а v — его скорость.

Однако вследствие того, что $|v| \ll c$, можно считать, что движение электрона осуществляется только под действием электрической компоненты поля, т.е. внешняя сила $\mathbf{F} = e\mathbf{E}'$.

С другой стороны, на электрон действует электростатическая сила притяжения положительно заряженного ядра, можно предположить, что электрон в атоме представляет собой пружинный маятник с жесткостью q и массой m, совершающий вынужденные колебания. Уравнение подобных колебаний

$$m\ddot{\mathbf{r}} + q\mathbf{r} = e\mathbf{E}'. \tag{2.54}$$

Будем искать решение этого уравнения в виде

$$\mathbf{r} = \mathbf{r_0} \exp(-j\omega t), \tag{2.55}$$

поскольку действующее электромагнитное поле с частотой ω имеет гармонический характер и может быть представлено в виде

$$\mathbf{E'} = \mathbf{E'_0} \exp(-j\omega t).$$

Подставляя выражения для векторов поля и смещения электрона $\boldsymbol{r}$ относительно ядра в уравнение (2.54), получаем его стационарное решение в виде

$$\mathbf{r} = \frac{e\mathbf{E}'}{m\left(\omega_0^2 - \omega^2\right)}, \tag{2.56}$$

где $\omega_0 = \sqrt{q/m}$ — резонансная частота атома.

Дипольный момент каждого атома в этом случае составляет $\mathbf{P}_a = e\mathbf{r}$ а вектор поляризации, соответствующий дипольному моменту единицы объема,

$$\mathbf{P} = N\mathbf{P}_a = Ne\mathbf{r} = N\frac{e^2}{m}\frac{\mathbf{E}'}{\omega_0^2 - \omega^2}, \tag{2.57}$$

где N — число атомов в единице объема. С другой стороны, в соответствии с формулой (2.50) $\mathbf{P} = N\alpha\mathbf{E}'$, поэтому

$$N\alpha = N\frac{e^2}{m\left(\omega_0^2 - \omega^2\right)}, \tag{2.58}$$

что говорит о частотной зависимости поляризуемости и о наличии дисперсии в среде.

Подставив формулу (2.53) в (2.58), получим зависимость показателя преломления от частоты

$$\frac{n^2-1}{n^2+2}=\frac{1}{3\varepsilon_0}\frac{Ne^2}{m\left(\omega_0^2-\omega^2\right)}, \qquad (2.59)$$

которая для показателей преломления, близких к единице, имеет вид

$$n^2 \approx 1+\frac{Ne^2}{\varepsilon_0 m\left(\omega_0^2-\omega^2\right)} \qquad (2.60)$$

На рис. 2.7 представлена зависимость квадрата показателя преломления от частоты, имеющая характерный резонансный вид. Показатель преломления растет с увеличением частоты, что соответствует *нормальной* дисперсии.

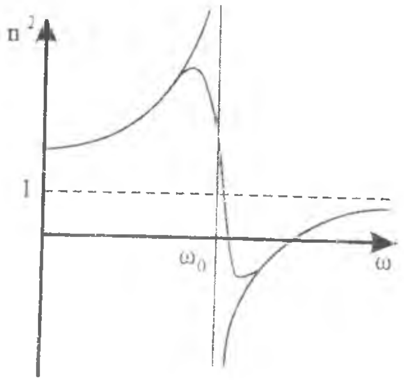

Рис. 2.7. Зависимость квадрата показателя преломления от частоты (действительная часть)

Однако стремление показателя преломления к бесконечности на резонансной частоте есть эффект абсолютно нефизический, так как расходимости в природе не существует. Видимо, при построении модели поведения электрона в электромагнитном поле мы не учли какой-то важный фактор, что и привело к разрыву кривой на рис. 2.7. Этим фактором является диссипация энергии или трение в среде.

Уравнение гармонического осциллятора, совершающего вынужденные колебания, с учетом потерь энергии выглядит следующим образом:

$$m\ddot{\mathbf{r}}+g\dot{\mathbf{r}}+q\mathbf{r}=e\mathbf{E}', \qquad (2.61)$$

где g — коэффициент сопротивления.

Стационарное решение этого уравнения имеет вид, схожий с (2.56), но $\boldsymbol{r}$ становится комплексной:

$$\mathbf{r} = \frac{e\mathbf{E}'}{m\left(\omega_0^2 - \omega^2\right) - j\omega g}. \tag{2.62}$$

Теперь необходимо рассматривать по отдельности действительную и мнимую части показателя преломления, который, естественно, тоже стал комплексным. Вещественная часть этой зависимости изображена пунктирной кривой на рис. 2.7. Из нее видно, что расходимости нет, однако появился участок кривой, где показатель преломления уменьшается с ростом частоты. Этот участок носит название *аномальной* дисперсии. На этом участке порядок цветов спектра должен смениться на противоположный. Такие области действительно существуют, но не в видимой области спектра, а в ультрафиолетовой, где лежат резонансные частоты поглощения атомов.

Мнимая часть восприимчивости η определяет поглощение в среде и имеет характерный для формы линий поглощения вид (рис. 2.8).

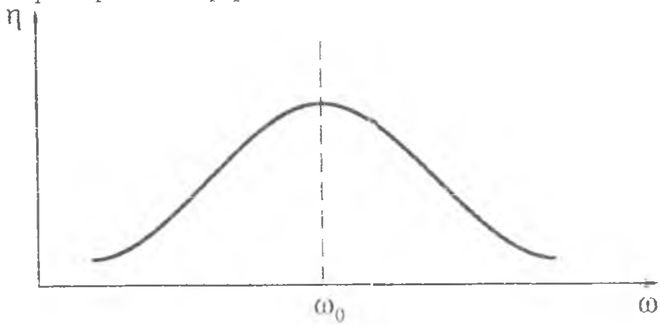

Рис. 2.8. Линия поглощения в веществе (мнимая часть проницаемости)

ВЫВОДЫ

1. Электромагнитное поле, определяющее взаимодействие заряженных частиц в пространстве, описывается нелинейной системой дифференциальных уравнений в частных производных — системой уравнений Максвелла (два векторных и два скалярных уравнения) с соответствующими граничными условиями, а также тремя дополнительными материальными уравнениями для учета влияния реальной среды.

Решение системы уравнений Максвелла может быть найдено точно только в некоторых частных случаях, поэтому, как правило, используется метод последовательных приближений (моделей) — от геометрической оптики к строгой теории дифракции.

Система уравнений Максвелла может быть преобразована в дифференциальное уравнение 2-го порядка, которое в условиях вакуума или в условиях медленно меняющихся параметров среды совпадает с волновым уравнением. Решением этого уравнения являются электромагнитные волны (в простейшем случае плоские и сферические).

При поиске решений системы уравнений Максвелла используется комплексная запись волн и векторов поля с последующим переходом к действительным частям в итоговых формулах. Подобная процедура возможна в связи с предположением о линейном характере операций, производимых над полями.

Электромагнитная волна в среде может быть поляризованной и неполяризованной в зависимости от вида фигуры, являющейся геометрическим местом точек концов вектора напряженности электрического поля в пространстве. Учет поляризации необходим при рассмотрении большинства оптических явлений — от интерференции до дифракции и рассеяния.

Фазовые (угловые) соотношения между падающей, отраженной и преломленной волнами на границе раздела двух сред определяются законом преломления, а амплитудные — формулами Френеля. При некоторых условиях (фазовых) наблюдается наличие линейно поляризованной отраженной волны (угол Брюстера) и отсутствие преломленной волны (угол полного внутреннего отражения).

Существует связь между оптическими (показатель преломления) и механическими (плотность) параметрами среды, которая описывается формулой Лорентц—Лоренца. Как правило, показатель преломления растет с увеличением плотности среды.

Оптические характеристики среды зависят от частоты падающего света (дисперсии), что связано с взаимодействием атомов среды и электромагнитного поля. При этом существует область нормальной дисперсии — рост показателя преломления с увеличением частоты, и аномальной дисперсии — обратная зависимость.

ГЛАВА 3.
ГЕОМЕТРИЧЕСКАЯ ОПТИКА

3.1. ПРИБЛИЖЕНИЕ ГЕОМЕТРИЧЕСКОЙ ОПТИКИ

Электромагнитные колебания, соответствующие видимой области спектра, происходят с частотами порядка $\nu = (4,9—7,9)\cdot 10^{14}$ Гц, что соответствует длинам волн в диапазоне от 0,38 до 0,78 мкм. Поэтому при решении волнового уравнения можно в первом приближении предположить, что длина волны $\lambda_0 \to 0$. Раздел оптики, в котором пренебрегают конечным значением длин волн, носит название геометрической оптики, в этом случае распространение света описывается с помощью луча. Явления, характеризующие волновую природу света (интерференция, дифракция, поляризация), здесь не рассматриваются.

Волновое уравнение для непоглощающей среды с показателем преломления n имеет вид

$$\nabla^2\mathbf{E}_0 + n^2k^2\mathbf{E}_0 = 0, \tag{3.1}$$

где $k = \omega/c$ — волновое число для вакуума. Применяя формулу логарифмического дифференцирования, имеем

$$\frac{\partial \mathbf{E}_0}{\partial x} = \mathbf{E}_0\frac{\partial}{\partial x}\ln\mathbf{E}_0 \text{ и } \frac{\partial^2\mathbf{E}_0}{\partial x^2} = \mathbf{E}_0\left(\frac{\partial}{\partial x}\ln\mathbf{E}_0\right)^2 + \mathbf{E}_0\frac{\partial^2}{\partial x^2}\ln\mathbf{E}_0.$$

Аналогичные выражения получаются для производных по y и z. Тогда

$$\frac{1}{\mathbf{E}_0}\nabla^2\mathbf{E}_0 = \nabla^2(\ln\mathbf{E}_0) + \left[\operatorname{grad}(\ln\mathbf{E}_0)\right]^2.$$

Уравнение (3.1) примет при этом вид

$$\nabla^2(\ln\mathbf{E}_0) + \left[\operatorname{grad}(\ln\mathbf{E}_0)\right]^2 + n^2k^2 = 0. \tag{3.2}$$

По аналогии с решением волнового уравнения в виде плоской однородной волны, распространяющейся в среде с показателем преломления n, представим $\mathbf{E}_0(\mathbf{r})$ в следующем виде:

$$\mathbf{E}_0 = \mathbf{e}(\mathbf{r})\exp\left(jk\varphi(\mathbf{r})\right).$$

Уравнение (3.2) преобразуется тогда следующим образом:

$$\nabla^2(\ln\mathbf{e}) + \left[\operatorname{grad}(\ln\mathbf{e})\right]^2 - k^2\left[\operatorname{grad}(\varphi(\mathbf{r}))\right]^2 +$$
$$+jk\left[2\operatorname{grad}(\ln\mathbf{e})\operatorname{grad}(\varphi(\mathbf{r})) + \nabla^2\varphi(\mathbf{r})\right] + n^2k^2 = 0.$$

Приравнивая нулю вещественные и мнимые части, получаем

$$\nabla^2(\ln \mathbf{e})+\left[\operatorname{grad}(\ln \mathbf{e})\right]^2-k^2\left[\operatorname{grad}\varphi(\mathbf{r})\right]^2+n^2k^2=0; \tag{3.3}$$

$$jk\left[\nabla^2\varphi(\mathbf{r})+2\operatorname{grad}(\ln \mathbf{e})\operatorname{grad}\varphi(\mathbf{r})\right]=0. \tag{3.4}$$

Для малых длин волн $(\lambda_0 \to 0)$ первыми двумя членами в уравнении можно пренебречь. Окончательно получим

$$\left[\operatorname{grad}\varphi(\mathbf{r})\right]^2=n^2. \tag{3.5}$$

Функция $\varphi(\mathbf{r})$ называется функцией эйконала, а выражение (3.5) — уравнением эйконала. Уравнение эйконала представляет собой основное уравнение геометрической оптики, а его вывод определяет ограничения по использованию этого приближения. Они связаны с двумя упрощениями. Первое упрощение связано с тем, что при выводе уравнения эйконала предполагалось, что первые два слагаемых в (3.3) малы по сравнению с последними слагаемым, т.е. не рассматриваются области, в которых велики производные от напряженности поля. Примером такой области (где геометрическая оптика даст неправильный результат) является фокальная плоскость линзы, в которой поле концентрируется в точке и его производная бесконечно велика. Второе упрощение вытекает из условий вывода волнового уравнения из уравнений Максвелла (см. гл. 2), когда были отброшены члены с производными от диэлектрической и магнитной проницаемостей. В этом случае не рассматриваются области, в которых происходят резкие изменения характеристик среды, например край линзы или экрана, где нельзя не учитывать дифракционные явления. Это еще одно подтверждение необходимости учета условий применимости используемой модели явления.

Поверхности, для которых $\varphi(\mathbf{r})$ = const, называются геометрическими волновыми поверхностями, или геометрическими волновыми фронтами. Световые лучи можно определить как траектории, ортогональные к волновым фронтам (рис. 3.1).

Возьмем произвольную точку P, расположенную на луче, и пусть $\mathbf{r}(s)$ — радиус-вектор этой точки, а s — длина луча, отсчитываемая от этой точки. Тогда $\mathrm{d}\mathbf{r}/\mathrm{d}s=\mathbf{s}$, где $\mathbf{s}$ — единичный вектор, направленный по касательной к траектории луча. Уравнение эйконала можно преобразовать к виду

$$n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}=\operatorname{grad}\varphi. \tag{3.6}$$

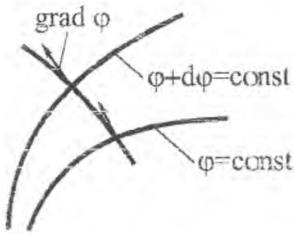

Рис. 3.1. Волновые фронты и световые лучи

Уравнение (3.6) является уравнением луча. Поясним смысл этого уравнения. Рассмотрим два соседних волновых фронта с $\varphi = \text{const}$ и $\varphi + \mathrm{d}\varphi = \text{const}$. Тогда

$$\frac{\mathrm{d}\varphi}{\mathrm{d}s} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\operatorname{grad}\varphi = n, \tag{3.7}$$

т.е. расстояние $\mathrm{d}s$ между точками пересечения луча с волновыми фронтами обратно пропорционально показателю преломления и прямо пропорционально скорости света в среде.

Интеграл $\int n\,\mathrm{d}s$ вдоль кривой называется оптической длиной пути. Обозначим оптическую длину пути между точками P_1 и P_2 как $[P_1P_2]$, тогда

$$\left[P_1P_2\right] = \int\limits_{P_1}^{P_2} n\,\mathrm{d}s = \varphi(P_2) - \varphi(P_1). \tag{3.8}$$

Учитывая, что $n\,\mathrm{d}s = c\,\mathrm{d}t$, получаем

$$\left[P_1P_2\right] = c\int\limits_{P_1}^{P_2} \mathrm{d}t, \tag{3.9}$$

т.е. оптическая длина пути равна произведению скорости света c в вакууме на время распространения света от точки P_1 до P_2.

Дифференциальное уравнение световых лучей. Преобразуем уравнение луча к виду, не содержащему функцию эйконала. Для этого продифференцируем выражение (3.7) по s:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\right) = \frac{\mathrm{d}}{\mathrm{d}s}\left(\operatorname{grad}\varphi\right) = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\left(\nabla\cdot\nabla\varphi\right) = \frac{1}{n}\operatorname{grad}\varphi\left(\nabla\cdot\nabla\varphi\right) =$$

$$= \frac{1}{2n}\operatorname{grad}\left(\operatorname{grad}\varphi\right)^2 = \frac{1}{2n}\operatorname{grad}n^2 = \operatorname{grad}n.$$

Окончательно имеем

$$\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \operatorname{grad} n. \tag{3.10}$$

Уравнение (3.10) представляет собой векторную форму дифференциального уравнения для световых лучей. Если свет распространяется в однородной среде, то n = const, и (3.10) приобретает вид

$$d^2\mathbf{r}/dx^2 = 0,$$

тогда $\mathbf{r} = s\mathbf{a} + \mathbf{b}$, где $\mathbf{a}$, $\mathbf{b}$ — постоянные векторы. Отсюда следует, что в однородной среде световые лучи являются прямыми линиями.

Законы преломления и отражения. Рассмотрим поведение лучей, пересекающих поверхность, разделяющую две однородные среды с различными показателями преломления. Учитывая тождество rot grad = 0 и принимая во внимание, что $n\mathbf{s} = n\,d\mathbf{r}/ds$, получаем

$$\operatorname{rot} n\mathbf{s} = 0.$$

Заменим поверхность раздела двух сред 1 и 2 переходным слоем, в котором величины ε, μ и n меняются быстро, но непрерывно (рис. 3.2). Рассмотрим плоский элемент, стороны которого P_1Q_1 и P_2Q_2 параллельны границе раздела, а P_1P_2 к Q_1Q_2 — перпендикулярны. Пусть $\mathbf{b}$ — нормаль к этому элементу.

Применим теорему Стокса

$$\int(\operatorname{rot} n\mathbf{s})\mathbf{b}\,dS = \oint n\mathbf{s}\,d\mathbf{r} = 0,$$

где второй интеграл берется по контуру $Q_1P_1P_2Q_2$. Переходя к $\delta h \to 0$, получаем

$$\mathbf{n}_{12} \times (n_2\mathbf{s}_2 - n_1\mathbf{s}_1) = 0, \tag{3.11}$$

где $\mathbf{n}_{12}$ — единичный вектор нормали к поверхности раздела, направленный из среды 1 в среду 2 (рис. 3.2 и 3.3). Обозначим углы, которые образуют падающий и преломленный лучи с нормалью к поверхности раздела, как ε_1 и ε_2, тогда

$$n_2(\mathbf{n}_{12} \times \mathbf{s}_2) = n_1(\mathbf{n}_{12} \times \mathbf{s}_1) \tag{3.12}$$

или

$$n_2 \sin\varepsilon_2 = n_1 \sin\varepsilon_1. \tag{3.13}$$

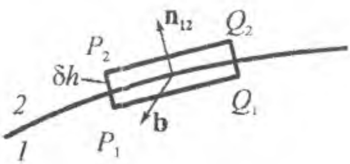

Рис. 3.2. К выводу законов преломления

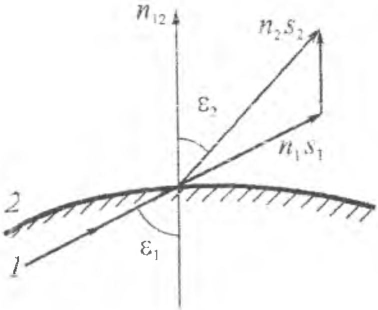

Рис. 3.3. Преломление на границе раздела двух сред

Формулы (3.12) и (3.13) являются математической записью закона преломления (закона Снеллиуса), который был выведен в § 2.8. Однако этот вывод был справедлив для случая падения плоской волны с любым значением λ_0 на плоскую отражающую поверхность. Настоящий вывод относится к случаю, когда $\lambda_0 \to 0$, при этом отражающие и преломляющие поверхности могут быть более общей формы, а именно радиусы кривизны волнового фронта падающей волны и поверхности раздела должны быть велики по сравнению с длиной волны падающего света.

Интегральный инвариант Лагранжа. Предположим, что показатель преломления является непрерывной функцией координат. Тогда из теоремы Стокса следует

$$\oint n\mathbf{s} \cdot \mathrm{d}\mathbf{r} = 0 \tag{3.14}$$

Полученное соотношение называется интегральным инвариантом Лагранжа. Его можно интерпретировать следующим образом. Выберем две произвольные точки поля P_1 и P_2. Интеграл $\int\limits_{P_1}^{P_2} n\mathbf{s} \cdot d\mathbf{r}$ не зависит от пути интегрирования.

Покажем, что выражение (3.14) является справедливым и при пересечении контуром границы раздела двух сред с разными показателями преломления. Разобьем контур интегрирования C на две части C_1 и C_2, расположенные соответственно в первой и второй средах и замыкающиеся линиями K, идущими параллельно границе раздела (рис. 3.4).

Применяя (3.14) к обоим контурам и складывая полученные выражения, получаем

$$\int\limits_{C_1} n_1\mathbf{s}_1 \cdot \mathrm{d}\mathbf{r} + \int\limits_{C_2} n_2\mathbf{s}_2 \cdot \mathrm{d}\mathbf{r} + \int\limits_{K} \left(n_2\mathbf{s}_2 - n_1\mathbf{s}_1\right) \cdot \mathrm{d}\mathbf{r} = 0.$$

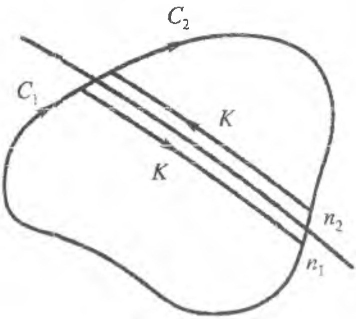

Рис. 3.4. К выводу интегрального инварианта Лагранжа

Из закона преломления $\left(n_2\mathbf{s}_2 - n_1\mathbf{s}_1\right) = 0$, поэтому интеграл вдоль K также равен 0. Отсюда

$$\int\limits_{C_1} n_1\mathbf{s}_1 \cdot \mathrm{d}\mathbf{r} + \int\limits_{C_2} n_2\mathbf{s}_2 \cdot \mathrm{d}\mathbf{r} = \oint n\mathbf{s} \cdot \mathrm{d}\mathbf{r} = 0.$$

Принцип Ферма. Согласно принципу Ферма оптическая длина пути реального луча между любыми двумя точками P_1 и P_2 меньше оптической длины пути вдоль любой другой кривой, соединяющей эти точки и лежащей в некоторой регулярной окрестности луча. Под регулярной окрестностью подразумевается область, через каждую точку которой проходит один и только один луч.

Для доказательства принципа Ферма рассмотрим пучок лучей и сравним оптические длины пути P_1P_2 луча C и произвольной кривой K (рис. 3.5).

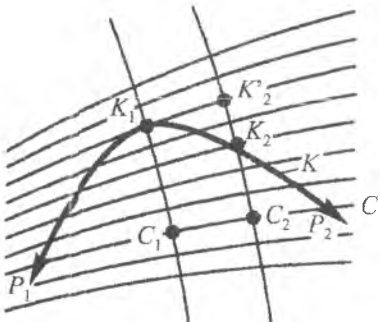

Рис. 3.5. К доказательству принципа Ферма

Рассмотрим два волновых фронта, которые пересекают луч C в точках C_1 и C_2, а кривую K в точках K_1 и K_2. Применяя интегральное соотношение Лагранжа для треугольника $K_1K_2K'_2$, получаем

$$\left(n\mathbf{s}\cdot \mathrm{d}\mathbf{r}\right)_{K_1K_2}+\left(n\mathbf{s}\cdot \mathrm{d}\mathbf{r}\right)_{K_2K'_2}-\left(n\mathbf{s}\cdot \mathrm{d}\mathbf{r}\right)_{K_1K'_2}=0.$$

Поскольку вектор $\mathbf{s}$ перпендикулярен $\mathrm{d}\mathbf{r}$ на волновом фронте, то

$$\left(n\mathbf{s}\cdot \mathrm{d}\mathbf{r}\right)_{K_2K'_2}=0;$$

$$\left(n\mathbf{s}\cdot \mathrm{d}\mathbf{r}\right)_{K_1K_2}=\left(n\mathbf{s}\cdot \mathrm{d}\mathbf{r}\right)_{K_1K'_2}.$$

Из определения скалярного произведения следует

$$\left(n\mathbf{s}\cdot \mathrm{d}\mathbf{r}\right)_{K_2K'_2}\le\left(n\,\mathrm{d}s\right)_{K_1K_2}.$$

С другой стороны, так как точки K_1, K'_2 и C_1, C_2 лежат на пересечении лучей с волновыми фронтами, то

$$\left(n\mathbf{s}\cdot \mathrm{d}\mathbf{r}\right)_{K_1K'_2}=\left(n\,\mathrm{d}s\right)_{K_1K'_2}=\left(n\,\mathrm{d}s\right)_{C_1C_2}.$$

Отсюда получим

$$\left(n\,\mathrm{d}s\right)_{C_1C_2}\le\left(n\,\mathrm{d}s\right)_{K_1K_2},$$

или

$$\int\limits_C n\,\mathrm{d}s\le\int\limits_K n\,\mathrm{d}s. \qquad (3.15$$

В (3.15) знак равенства можно ставить только в том случае, есл направления $\mathbf{s}$ и $\mathrm{d}\mathbf{r}$ совпадают в каждой точке кривой, а это возможн только если кривая K является реальным лучом. Однако мы предположи ли, что через каждую точку проходит единственный луч, следовательн оптическая длина пути луча меньше оптической длины пути вдоль про извольной кривой.

Теорема Малюса. Преломление на любой поверхности вызывае искривление лучей, однако согласно теореме Малюса оптическая ра ность хода между волновыми поверхностями остается неизменной.

Докажем эту теорему для случая единичного преломления. Предп ложим, что лучи преломляются на поверхности T, разделяющей среду показателями преломления n_1 и n_2 (рис. 3.6). Пусть S_1 — один из волновы фронтов в пространстве предметов; A_1 и P — точки пересечения прои вольного луча в пространстве предметов соответственно с волновы фронтом и поверхностью преломления, а A_2 — произвольная точка на пр ломленном луче. Переместим точку A_1 в точку B_1 на том же волново фронте, тогда точка P на поверхности преломления переместится в точк Q. На преломленном в точке Q луче выберем такую точку B_2, чтобы

$$\left[A_1PA_2\right]=\left[B_1QB_2\right]. \tag{3.16}$$

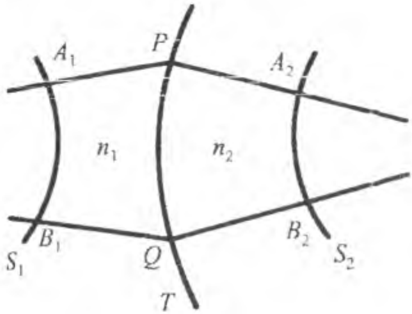

Рис. 3.6. К выводу теоремы Малюса

Если перемещать точку B_1 вдоль поверхности S_1, то точка B_2 будет перемещаться вдоль некоторой поверхности S_2. Докажем, что преломленный луч перпендикулярен этой поверхности.

Применим интегральный инвариант Лагранжа к контуру $A_1PA_2B_2QB_1A_1$

$$\int\limits_{A_1PA_2} n\mathrm{d}s+\int\limits_{A_2B_2} n\mathbf{s}\cdot\mathrm{d}\mathbf{r}+\int\limits_{B_1QB_2} n\mathrm{d}s+\int\limits_{B_1A_1} n\mathbf{s}\cdot\mathrm{d}\mathbf{r}=0. \tag{3.17}$$

Из (3.16) следует, что

$$\int\limits_{A_1PA_2} n\mathrm{d}s+\int\limits_{B_2QB_1} n\mathrm{d}s=0. \tag{3.18}$$

Так как на волновом фронте векторы $\mathbf{s}$ и $\mathrm{d}\mathbf{r}$ перпендикулярны, то

$$\int\limits_{B_1A_1} n\mathbf{s}\cdot\mathrm{d}\mathbf{r}=0.$$

Следовательно $\int\limits_{A_2B_2} n\mathbf{s}\cdot d\mathbf{r}=0$, но это возможно только в том случае, если $s\perp\mathrm{d}\mathbf{r}$ на всей поверхности S_2, т.е. поверхность S_2 является волновым фронтом. Поскольку $\left[A_1PA_2\right]=\left[B_1QB_2\right]$, то оптическая длина пути между двумя волновыми фронтами одинакова для всех лучей.

3.2. ГЕОМЕТРИЧЕСКАЯ ТЕОРИЯ ИЗОБРАЖЕНИЯ

3.2.1. Правило знаков

В приближении геометрической оптики можно считать, что энергия распространяется вдоль световых лучей, а оптические законы можно сформулировать на языке геометрии. Прежде чем приступить к изложению законов геометрической оптики, остановимся на принятом правиле знаков.

Правило знаков для отрезков. Отрезки, отсчитываемые вдоль оптической оси, считаются положительными, если они лежат правее выбранной точки отсчета, и отрицательными, если они лежат левее. Поясним это правило на примере одиночной преломляющей сферической поверхности радиусом r с центром кривизны в точке C (рис. 3.7). На этом рисунке точки A и A' являются соответственно осевыми точками предмета и изображения. Для отрезков $-s$ и s', отсчитываемых вдоль оптической оси, за точку отсчета принимается вершина преломляющей поверхности O. Для отрезков $-t$ и t' вдоль лучей, составляющих некоторый угол с оптической осью, за начало отсчета принимается точка пересечения луча с преломляющей поверхностью (точка M). Отрезки, перпендикулярные оптической оси, считаются положительными, если они расположены над оптической осью (отрезок h), и отрицательными, если они расположены под осью.

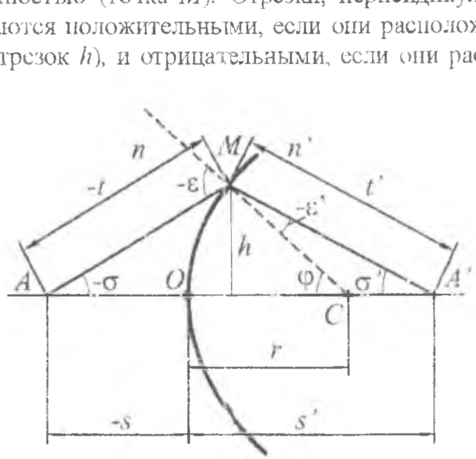

Рис. 3.7. Правило знаков для отрезков и углов

Правило знаков для углов. Для углов $-\sigma$ и σ', которые луч образует с оптической осью, за начало отсчета выбирается ось системы, а для углов падения и преломления $-\varepsilon$ и $-\varepsilon'$ - нормаль к преломляющей поверхности. Углы считаются положительными, если для совмещения выбранной линии отсчета с лучом ее нужно вращать по часовой стрелке (угол σ' на рис. 3.7) и отрицательными при вращении против часовой стрелки (углы $-\sigma$, $-\varepsilon$, $-\varepsilon'$).

3.2.2. Теория идеальной оптической системы

Предположим, что на оптическую систему падает гомоцентрический пучок лучей (пучок называется гомоцентрическим, если его лучи имеют одну точку пересечения). Если, пройдя оптическую систему, пучок остался гомоцентрическим, то такое изображение называется стигматическим. Точки пересечения лучей гомоцентрического пучка называются соответственно точечным предметом и изображением и являются сопряженными относительно оптической системы. Плоскости, перпендикулярные оптической оси и содержащие сопряженные точки, называются сопряженными плоскостями.

Идеальной называется оптическая система, для которой выполняются следующие условия:

1. Гомоцентрический пучок, пройдя оптическую систему, остается гомоцентрическим, т.е. любой точке пространства предметов соответствует единственная точка пространства изображений.

2. Изображение подобно предмету. Для этого отношение размера изображения к размеру предмета, лежащего в плоскости, перпендикулярной оптической оси, должно быть постоянно. Это отношение называется линейным или поперечным увеличением оптической системы и является постоянным для пары сопряженных плоскостей (рис. 3.8):

$$\beta_0 = y'/y = x'/x = \text{const}.$$

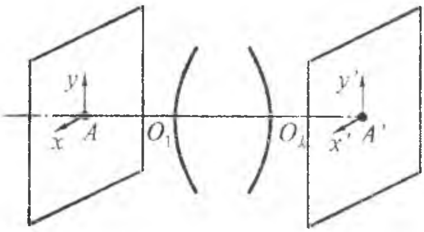

Рис. 3.8. Линейное увеличение идеальной оптической системы

Кардинальные элементы идеальной оптической системы. Рассмотрим идеальную оптическую систему. На рис. 3.9 изображены только первая и последняя преломляющие поверхности такой системы, а O_1 и O_k — вершины этих преломляющих поверхностей. Предположим, что некоторый луч BM, параллельный оптической оси, падает на систему на высоте h. Пройдя оптическую систему, сопряженный луч либо останется также параллельным оптической оси, либо пересечет ее в некоторой точке F'. Продолжим луч BM до пересечения его с лучом $M'F'$ и через полученную точку пересечения N' проведем плоскость, перпендикулярную оптической оси. Точку пересечения этой плоскости с оптической осью обозначим через H'. Рассмотрим луч AO_1, идущий вдоль оптической оси.

Этот луч пройдет систему, не преломляясь. Таким образом, луч AO_1 сопряжен с лучом O_kF', а луч BM сопряжен с лучом $M'F'$. Значит, точка F', лежащая на пересечении лучей $M'F'$ и O_kF', должна быть сопряжена с некоторой точкой, лежащей на пересечении лучей BM и AO_1. Поскольку лучи BM и AO_1 параллельны, то точка F' сопряжена с бесконечно удаленной точкой. Точка F', являющаяся изображением бесконечно удаленной точки, лежащей на оси в пространстве предметов, называется задним фокусом оптической системы. Плоскость Q', содержащая эту точку и перпендикулярная оптической оси, называется задней фокальной плоскостью. Точка H' носит название задней главной точки, а плоскость $N'H'$ — задней главной плоскости.

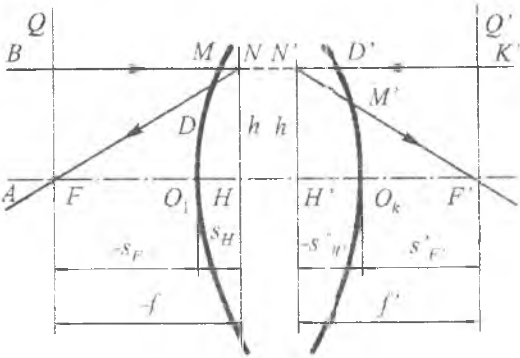

Рис. 3.9. Кардинальные элементы идеальной оптической системы

Проделаем аналогичные построения, направляя луч, параллельный оптической оси, справа налево на той же высоте h. Повторяя все рассуждения, получаем передний фокус системы F, переднюю фокальную плоскость Q, переднюю главную точку H и переднюю главную плоскость NH. Фокусы, фокальные плоскости, главные точки и главные плоскости называются кардинальными элементами идеальной оптической системы.

Можно показать, что точки N и N' сопряжены. Действительно, точка N лежит на пересечении лучей BM и DF, а точка N' — на пересечении лучей $K'D'$ и MF. Но луч BM сопряжен с лучом $M'F'$, а луч DF сопряжен с лучом $K'D'$. Значит, точка N сопряжена с точкой N'. Лучи BM и $K'D'$ проводились на произвольной высоте h. Меняя эту высоту, можно получить другие сопряженные точки, лежащие в главных плоскостях. Отсюда следует, что главные плоскости сопряжены. Так как высоты, на которых лежат точки N и N', равны, то линейное увеличение в главных плоскостях равно единице ($\beta_{0H} = +1$). Таким образом, главными плоскостями оптической системы называется пара сопряженных плоскостей, для которых линейное увеличение равно единице.

Расстояние от задней главной точки до заднего фокуса носит название заднего фокусного расстояния $H'F' = f'$, соответственно $HF = -f$ - переднее фокусное расстояние. Расстояние от вершины последней преломляющей поверхности до заднего фокуса называется задним фокальным отрезком $O_k F' = s'_{F'}$, а расстояние от вершины первой преломляющей поверхности до переднего фокуса $O_1 F = -s_F$ - передним фокальным отрезком. Положения главных точек относительно вершины преломляющих поверхностей характеризуют также отрезками s_H и $-s'_{H'}$.

Задний фокус системы сопряжен с бесконечно удаленной точкой, лежащей на оси пространства предметов. Следовательно, задняя фокальная плоскость сопряжена с бесконечно удаленной плоскостью пространства предметов. Отсюда следует, что пучок параллельных лучей, образующих в пространстве предметов некоторый угол с оптической осью, соберется в точке, лежащей в задней фокальной плоскости (рис. 3.10, *а*). Аналогичные рассуждения позволяют сделать вывод, что пучок параллельных лучей, идущих под некоторым углом в пространстве изображений, выходит из одной точки в передней фокальной плоскости (рис. 3.10, *б*).

Так как линейное увеличение в главных плоскостях равно единице, то можно совместить главные плоскости в одну и рассматривать идеальную оптическую систему как бесконечно тонкую, разделяющую среды с показателями преломления n и n'. В этом случае углы α_H и $\alpha'_{H'}$ - можно рассматривать как углы падения и преломления, причем при малых углах закон преломления можно записать в виде $n \operatorname{tg}\alpha_H = n' \operatorname{tg}\alpha'_{H'}$. Из рис. 3.10 имеем:

$$-f \operatorname{tg}\alpha_H = f' \operatorname{tg}\alpha'_{H'}.$$

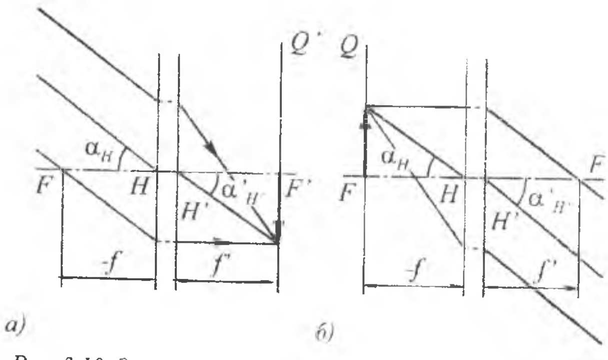

Рис. 3.10. Задняя (а) и передняя (б) фокальные плоскости идеальной оптической системы

Отсюда получаем

$$-f/f' = n/n'. \qquad (3.19)$$

В том случае, когда оптическая система находится в однородной среде, например в воздухе $(n = n')$, $-f = f'$, т.е. абсолютные значения переднего и заднего фокусных расстояний равны.

Основные формулы для сопряженных точек и отрезков. Выберем в качестве предмета отрезок *AB*, перпендикулярный оптической оси, и построим его изображение, даваемое идеальной оптической системой (рис. 3.11). Для этого в пространстве предметов проведем луч *BM* параллельно оптической оси. В пространстве изображений этот луч пересечет оптическую ось в заднем фокусе. Второй луч пройдет через передний фокус системы, а в пространстве изображений пойдет параллельно оптической оси. Пересечение этих лучей даст точку *B'*, являющуюся изображением точки *B*.

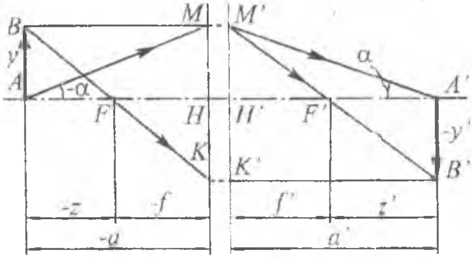

Рис. 3.11. К выводу формул для сопряженных точек и отрезков

Существует несколько соотношений, связывающих положение предмета с положением изображения через параметры оптической системы, которые различаются способом задания положения предмета и изображения. Зададим сначала положение сопряженных точек *A* и *A'* относительно фокусов системы. Это положение определяется соответственно отрезками $-z$ и z' (см. рис. 3.11).

Из подобия треугольников *ABF* и *FHK* и треугольников *M'H'F'* и *F'B'A'* следует $-y/y' = z/f$; $-y/y' = f'/z'$. Отсюда получаем уравнение Ньютона

$$zz' = ff'. \qquad (3.20)$$

Положение точек *A* и *A'* может быть также задано относительно главных плоскостей отрезками $-a$ и a'. Из рис. 3.11 видно, что $z = a - f$, $z' = a' - f'$. Подставляя данные выражения в (3.20), получаем уравнение Гаусса

$$f'/a' + f/a = 1. \qquad (3.21)$$

Задать положение точек *A* и *A'* можно в неявном виде через координаты вспомогательного луча. Проведем луч *AM*, составляющий угол $-\alpha$

оптической осью. Сопряженный луч $M'A'$ пойдет под углом α'. Умножим все члены в уравнении (3.21) на h. Принимая во внимание, что $\operatorname{tg}\alpha = h/a$, $\operatorname{tg}\alpha' = h/a'$, получаем

$$\operatorname{tg}\alpha' = -\frac{f}{f'}\operatorname{tg}\alpha + \frac{h}{f'}. \tag{3.22}$$

Формулы Ньютона, Гаусса и уравнение (3.22) связывают координаты предмета с координатами изображения через параметры оптической системы. Для системы в воздухе эти формулы примут вид:

$$zz' = -f'^2; \quad 1/a' - 1/a = 1/f'; \quad \operatorname{tg}\alpha' = -\operatorname{tg}\alpha + h/f'. \tag{3.23}$$

В этих уравнениях не принимаются во внимание размеры предмета и изображения. Для учета размеров предмета и изображения проделаем следующее. Из рис. 3.11 имеем $h = a\operatorname{tg}\alpha = a'\operatorname{tg}\alpha'$ или $(z+f)\operatorname{tg}\alpha = (z'+f')\operatorname{tg}\alpha'$. Подставляя $-yf/y'$ вместо z, и $-y'f'/y$ вместо z', получаем

$$\left.\begin{aligned} -fy\operatorname{tg}\alpha &= f'y'\operatorname{tg}\alpha'; \\ ny\operatorname{tg}\alpha &= n'y'\operatorname{tg}\alpha'. \end{aligned}\right\} \tag{3.24}$$

Уравнение (3.24) носит название инварианта Лагранжа—Гельмгольца и является критерием идеальности системы. Оно показывает, что для получения идеального изображения произведение показателя преломления на размеры предмета и на тангенс угла наклона луча с оптической осью должно оставаться постоянным при переходе от пространства предметов к пространству изображения.

Линейное увеличение. Как уже упоминалось выше, линейным увеличением называется отношение размера изображения предмета, перпендикулярного оптической оси, к размеру самого предмета. Линейное увеличение не зависит от размера предмета и является постоянным для сопряженных плоскостей

$$\beta_0 = y'/y = -f/z = -z'/f'. \tag{3.25}$$

Рассмотрим частный случай, когда $z = f$, т.е. предмет находится на двойном фокусном расстоянии от системы. Из (3.25) следует, что в этом случае $\beta_0 = -1$, а $z' = f'$, т.е. изображение равно предмету, перевернуто и находится также на двойном фокусном расстоянии.

Легко показать, что $a'/a = z'/f = f'/z$. Отсюда получаем

$$\beta_0 = -\frac{f}{f'}\cdot\frac{a'}{a}. \tag{3.26}$$

Для системы в воздухе $\beta_0 = a'/a$.

На практике часто требуется решить обратную задачу. Если заданы фокусные расстояния системы $(-f = f')$ и требуется получить заданное увеличение, то можно определить положение предмета и изображения. Решая совместно уравнение Гаусса для системы в воздухе и уравнение для линейного увеличения в воздухе, получаем

$$\left.\begin{aligned} a &= \frac{(1-\beta_0)}{\beta_0} f'; \\ a' &= (1-\beta_0) f'. \end{aligned}\right\} \tag{3.27}$$

Если в условиях предыдущей задачи неизвестно также фокусное расстояние, то задача имеет бесконечное число решений. В этом случае необходимо наложить дополнительные условия, например, задать габаритный размер $L = -a + a'$. Тогда получим

$$a = -L/(1-\beta_0); \quad a' = -L\beta_0/(1-\beta_0); \quad f' = -L\beta_0/(1-\beta_0)^2.$$

Угловое увеличение и узловые точки. Угловым увеличением идеальной оптической системы называется отношение тангенса угла наклона луча к оптической оси в пространстве изображений к тангенсу угла наклона сопряженного луча в пространстве предметов:

$$\gamma_0 = \operatorname{tg}\alpha'/\operatorname{tg}\alpha = a/a' = f/z' = z/f'. \tag{3.28}$$

Угловое увеличение не зависит от угла наклона лучей и является постоянным для данной пары сопряженных плоскостей.

Определим положение сопряженных плоскостей, для которых угловое увеличение равно единице. Если $\gamma_0 = 1$, то $z = f'$ и $z' = f$. Данная пара сопряженных плоскостей называется узловыми плоскостями, а точки пересечения их с оптической осью — узловыми точками. Рис. 3.12 иллюстрирует нахождение узловых точек N и N' оптической системы, при этом $\alpha_N = \alpha'_{N'}$.

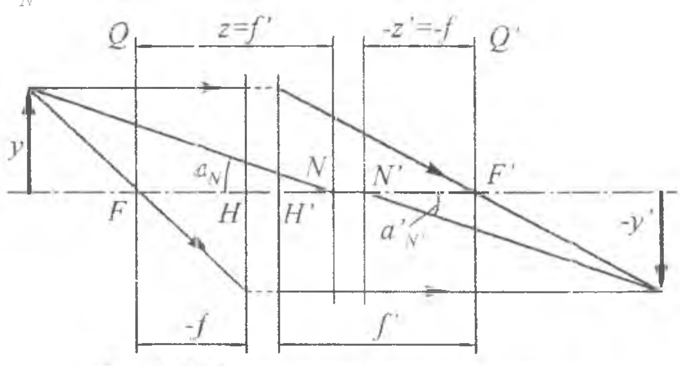

Рис. 3.12. Узловые точки оптической системы

Узловые точки и узловые плоскости также входят в понятие кардинальных элементов идеальной оптической системы.

Сопоставляя выражения (3.28) и (3.26), находим связь между угловым и линейным увеличениями оптической системы

$$\gamma_0 = -\frac{f}{f'}\frac{1}{\beta_0}. \qquad (3.29)$$

Для системы в воздухе $\gamma_0 = 1/\beta_0$. Формула (3.29) доказывает, что для главных плоскостей в общем случае угловое увеличение не равно единице, так как $\gamma_{0H} = -f/f'$. И только в случае однородной среды, в частности для системы в воздухе, $\gamma_{0H} = 1$. В этом случае узловые и главные плоскости совпадают.

Продольное увеличение. Предположим, что бесконечно малому предмету dz, расположенному вдоль оптической оси AA' соответствует изображение dz', также расположенное вдоль оптической оси. Продольным увеличением оптической системы называется отношение изображения бесконечно малого отрезка, расположенного вдоль оптической оси, к этому отрезку (рис. 3.13):

$$\alpha_0 = dz'/dz.$$

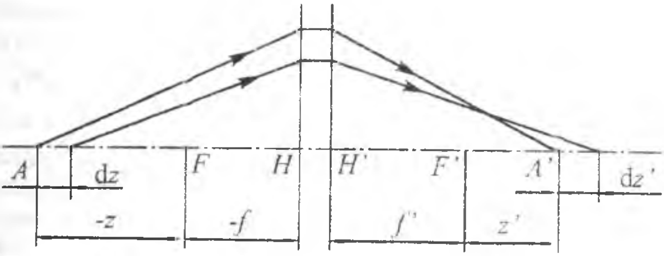

Рис. 3.13. Продольное увеличение оптической системы

Дифференцируя уравнение Ньютона (3.20) по z и z', получаем

$$z\,dz' + z'dz = 0,$$

тогда

$$\alpha_0 = dz'/dz = -z'/z. \qquad (3.30)$$

Найдем связь между продольным и линейным увеличениями. Для этого умножим числитель и знаменатель в выражении (3.30) на ff'. Учитывая, что $z'/f' = f/z = \beta_0$, получаем

$$\alpha_0 = -\frac{f'}{f}\beta_0^2. \qquad (3.31)$$

Для однородной среды $\alpha_0 = \beta_0^2$. Из формулы (3.31) следует, что предметы, имеющие протяженность вдоль оптической оси, даже идеальной оптической системой будут отображаться с искажениями, так как продольное увеличение пропорционально квадрату поперечного увеличения.

Перемножая формулы (3.29) и (3.31), находим связь между увеличениями идеальной оптической системы

$$\gamma_0\alpha_0 = \beta_0 . \tag{3.32}$$

Оптическая система из k тонких компонентов. Сложную оптическую систему можно представить состоящей из нескольких простых систем — компонентов. При этом встает задача нахождения эквивалентной оптической системы, т.е. определения кардинальных элементов системы, которая строила бы изображение так же, как и совокупность простых.

Рассмотрим сложную оптическую систему, состоящую из k тонких компонентов. Компонент называется тонким, если его толщина значительно меньше фокусных расстояний, при этом главные плоскости можно совместить. Предположим, что известны фокусные расстояния каждого компонента и расстояния между ними (рис. 3.14), а также положение размер предмета.

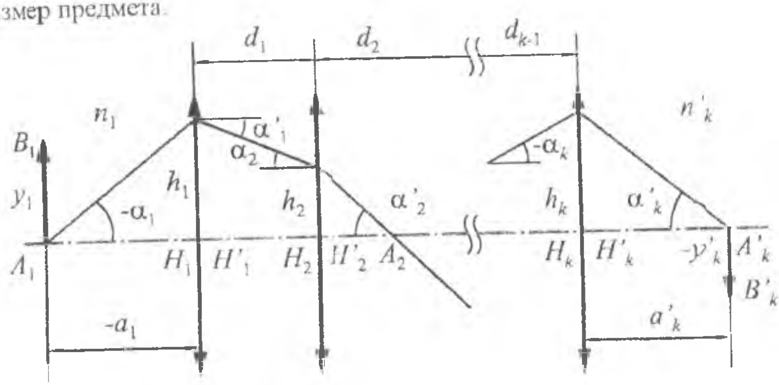

Рис. 3.14. Ход луча через сложную оптическую систему

Для определения положения и размера изображения рассмотрим ход луча, идущего из осевой точки предмета под произвольным углом α_1. Применим последовательно для каждого компонента формулу (3.22):

$$\operatorname{tg}\alpha'_\nu = -\frac{f_\nu}{f'_\nu}\operatorname{tg}\alpha_\nu + \frac{h_\nu}{f'_\nu},$$

при этом высоты пересечения луча с главными плоскостями будут равны $h_\nu = h_{\nu-1} - d_{\nu-1}\operatorname{tg}\alpha_\nu$. Определив для последнего компонента величины h_k и $\operatorname{tg}\alpha'_k$, найдем положение изображения $a'_k = h_k / \operatorname{tg}\alpha'_k$.

Для определения размера изображения воспользуемся связью между линейным и угловым увеличениями системы

$$y'_k = \beta_0 y_1 = -\frac{f}{f'}\frac{1}{\gamma_0}y_1 = -\frac{f}{f'}\frac{\operatorname{tg}\alpha_1}{\operatorname{tg}\alpha'_k}y_1. \tag{3.33}$$

Для расчета эквивалентной оптической системы рассчитывается ход луча, параллельного оптической оси (рис. 3.15).

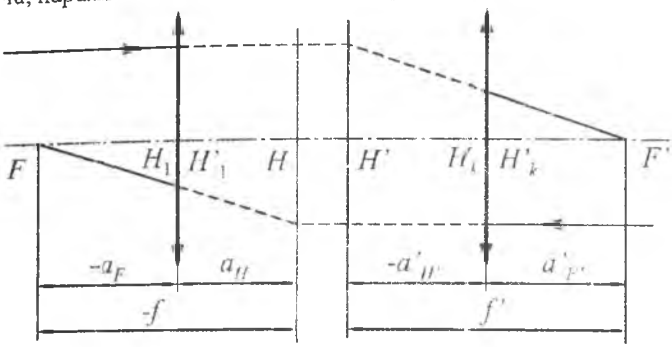

Рис. 3.15. Определение кардинальных элементов эквивалентной системы

Определяя, как и в предыдущем случае, последовательно высоты падения и углы преломления луча для каждого компонента, находим

$$a'_{F'} = h_k/\operatorname{tg}\alpha'_k,\ f' = h_1/\operatorname{tg}\alpha'_k,\ a'_{H'} = a'_{F'} - f',$$

где $a'_{F'}$ и $a'_{H'}$ - расстояния от главной плоскости последнего компонента до заднего фокуса и задней главной плоскости эквивалентной оптической системы.

Для определения величин a_F, a_H и f необходимо рассчитать ход луча, идущего справа налево параллельно оптической оси. На практике для подобного расчета оптическую систему поворачивают на 180° и продолжают считать, что луч распространяется слева направо. Тогда $a_F = -\overleftarrow{a'_{F'}}$, $f = -\overleftarrow{f'}$, $a_H = a_F - f$, где стрелкой обозначены величины, рассчитанные при повороте системы на 180°

В качестве примера рассмотрим оптическую систему, состоящую из двух компонентов (рис. 3.16). В этом случае имеем:

$$\operatorname{tg}\alpha_1 = 0;\quad \operatorname{tg}\alpha'_1 = \frac{h_1}{f_1'} = \operatorname{tg}\alpha_2;\quad h_2 = h_1 - d\operatorname{tg}\alpha_2 = h_1\left(1 - \frac{d}{f_1'}\right);$$

$$\operatorname{tg}\alpha'_2 = -\frac{f_2}{f_2'}\operatorname{tg}\alpha_2 + \frac{h_2}{f_2'}.$$

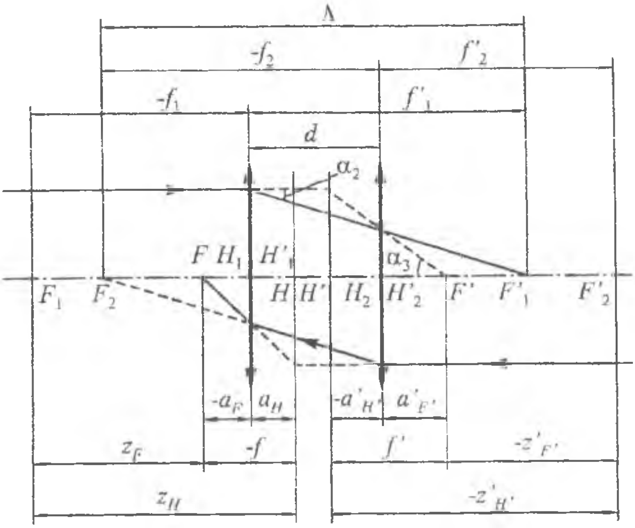

Рис. 3.16. Оптическая система из двух компонентов

Оптической силой компонента называется величина $\Phi = n' / f'$. Тогда

$$\operatorname{tg}\alpha_2 = -\frac{h_1}{n_2}\Phi_1; \quad h_2 = h_1\left(1 - \frac{d\Phi_1}{n_2}\right);$$

$$\operatorname{tg}\alpha'_2 = -\frac{n_2}{n_3}\operatorname{tg}\alpha_2 + \frac{h_2\Phi_2}{n_3} = \frac{h_1\Phi_1}{n_3} + h_1\left(1 - \frac{d\Phi_1}{n_2}\right)\frac{\Phi_2}{n_3} =$$

$$= \frac{h_2}{n_3}\left(\Phi_1 + \Phi_2 - d\Phi_1\Phi_2\right).$$

Учитывая, что оптическая сила эквивалентной оптической системы

$$\Phi = \frac{n_3}{f'} = \frac{n_3 \operatorname{tg}\alpha_3}{h_1},$$

получаем

$$\Phi = \Phi_1 + \Phi_2 - \frac{d\Phi_1\Phi_2}{n_2}. \qquad (3.3$$

Расстояние от второго компонента до заднего фокуса эквивалентной системы и задней главной плоскости будут соответственно равны:

$$a'_{F'} = \frac{h_2}{\mathrm{tg}\,\alpha'_2} = f'\left(1 - \frac{d\Phi_1}{n_2}\right); \quad a'_{H'} = a'_{F'} - f'.$$

Рассчитывая ход луча в обратном направлении, получаем

$$\Phi = -\frac{n_1}{f} = \Phi_1 + \Phi_2 - \frac{d\Phi_1\Phi_2}{n_2}; \quad a_F = -f'(1 - d\Phi_2); \quad a_H = a_F - f.$$

Выражения для фокусных расстояний оптической системы приобретают компактный вид при введении понятия оптического интервала. Оптическим интервалом Δ называется расстояние между задним фокусом первого компонента и передним фокусом второго компонента. Из рис. 3.16 следует $\Delta = -(f'_1 - f_2 + d)$. Заменяя в (3.34) оптические силы компонентов их фокусными расстояниями, получаем

$$f = -f_1 f_2/\Delta; \quad f' = -f_1' f_2'/\Delta. \tag{3.35}$$

Положение фокусов и главных плоскостей в этом случае удобно задать координатами z_F, $z'_{F'}$ и z_H, $z'_{H'}$, отсчитываемыми соответственно от переднего фокуса первого компонента и заднего фокуса второго компонента. Имеем

$$\left.\begin{aligned} z_F &= a_F - f_1 = \frac{f_1 f_1'}{\Delta}; \\ z_H &= z_F - f = \frac{f_1(f_1' - f_2)}{\Delta}; \end{aligned}\right\} \tag{3.36}$$

$$\left.\begin{aligned} z'_{F'} &= a'_{F'} - f_2' = -\frac{f_2 f_2'}{\Delta}; \\ z'_{H'} &= z'_{F'} - f' = \frac{f'_2(f_1' - f_2)}{\Delta}. \end{aligned}\right\} \tag{3.37}$$

Для определения линейного увеличения воспользуемся соотношением $\beta_0 = -f/z = -z'/f'$, где z и z' определяют положения предмета и изображения относительно фокусов эквивалентной системы. Используя полученные соотношения для системы из двух компонентов, имеем

$$\beta_0 = -\frac{f_1 f_2}{z_1\Delta - f_1 f_1'} = -\frac{z_2'\Delta + f_2 f_2'}{f_1' f_2'}, \tag{3.38}$$

где z_1 — расстояние от фокуса первого компонента до предмета; z'_2 — расстояние от фокуса второго компонента до изображения.

3.2.3. Преломление и отражение лучей сферическими поверхностями

В оптических системах используются детали, в большинстве своем имеющие сферические или плоские преломляющие и отражающие поверхности. Остановимся более подробно на преломлении лучей сферической поверхностью. На рис. 3.7 была представлена сферическая преломляющая поверхность радиусом r, делящая среды с показателями преломления n и n'.

Из треугольников AMC и $A'MC$ по закону синусов имеем

$$\left.\begin{aligned} \sin\varepsilon &= \frac{r-s}{r}\sin\sigma; \\ \sin\varepsilon' &= \frac{r-s'}{r}\sin\sigma'. \end{aligned}\right\} \tag{3.39}$$

После ряда преобразований получим

$$\frac{n}{s} = \frac{n}{r} + \frac{n\sin\varepsilon}{r(\sin\delta - \sin\varepsilon)} \tag{3.40}$$

и аналогичное выражение для пространства изображений

$$\frac{n'}{s'} = \frac{n'}{r} + \frac{n'\sin\varepsilon'}{r(\sin\sigma' - \sin\varepsilon')} \tag{3.41}$$

Вычитая из уравнения (3.41) уравнение (3.40), получаем

$$\frac{n'}{s'} - \frac{n}{s} = \frac{n'-n}{r} + \frac{n(r-s)}{rs}\left[\frac{\sin\sigma - \sin\varepsilon - \sin\sigma' + \sin\varepsilon'}{\sin\sigma' - \sin\varepsilon'}\right] \tag{3.42}$$

Уравнение (3.42) носит название уравнения действительного луча меридиональной плоскости. Меридиональной называется любая пло[с]кость, содержащая оптическую ось системы (плоскость, перпендикуля[р]ная меридиональной и содержащая падающий на систему луч, называет[ся] сагиттальной). В дальнейшем в качестве меридиональной плоскости б[у]дем выбирать плоскость чертежа. Из (3.42) следует, что при заданном п[о]ложении предмета (координата s) положение изображения (координата [s']) зависит от угла наклона луча к оптической оси. Это значит, что гомоце[н]трический пучок, выходящий из осевой точки предмета, пройдя чер[ез] сферическую преломляющую поверхность, потеряет свою гомоцентри[ч]ность. В этом случае вместо точечного изображения A'_0 в плоскости н[а]блюдения получаем кружок рассеяния (рис. 3.17).

Параксиальная область. Анализ выражения (3.42) показывает, ч[то] для получения идеального изображения сферической преломляющей п[о]верхностью необходимо устранить зависимость координаты s от углов и ε. Выполнить это условие можно только в том случае, когда $\sigma \to 0$, $\varepsilon \to 0$. При этом (3.42) преобразуется к виду

$$\frac{n'}{s'} - \frac{n}{s} = \frac{n'-n}{r}. \tag{3.43}$$

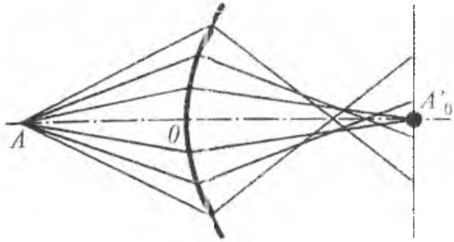

Рис. 3.17. Нарушение гомоцентричности пучка лучей при преломлении на сферической поверхности

Лучи, образующие бесконечно малые углы σ, σ' с оптической осью и углы ε, ε' с нормалью, называются параксиальными лучами, а область, в которой они распространяются, — параксиальной областью. Как видно из (3.43), координата изображения s' является постоянной для любых углов в этой области, а значит, изображение, получаемое в этой области, эквивалентно изображению, даваемому идеальной системой. Уравнение (3.43) называется уравнением параксиального луча в меридиональной области.

Перегруппируем члены в (3.43) следующим образом:

$$Q_s = n\left(\frac{1}{r} - \frac{1}{s}\right) = n'\left(\frac{1}{r} - \frac{1}{s'}\right). \tag{3.44}$$

Уравнение (3.44) называется параксиальным инвариантом Аббе.

Первый и второй параксиальные лучи. При расчете реальной оптической системы использование формул параксиальной оптики затруднительно из-за бесконечно малых углов и высот. В этом случае для расчета используется метод вспомогательных параксиальных лучей. Рассмотрим сферическую преломляющую поверхность радиусом r, разделяющую среды с показателями преломления n и n'. Возьмем на ее оси две пары сопряженных точек (рис. 3.18).

Для сопряженных точек A и A' инвариант Аббе имеет вид

$$\frac{n'}{s'} - \frac{n}{s} = \frac{n'-n}{r}, \tag{3.45}$$

а для точек P и P'

$$\frac{n'}{s'_{p'}} - \frac{n}{s_p} = \frac{n'-n}{r}. \tag{3.46}$$

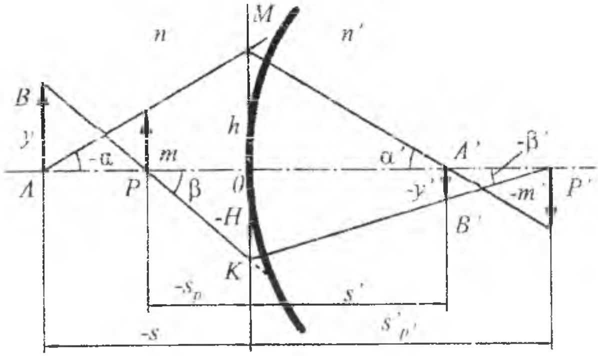

Рис. 3.18. Ход вспомогательных параксиальных лучей через преломляющую поверхность

В параксиальной области главные плоскости преломляющей поверхности совпадают и являются касательными к преломляющей поверхности в ее вершине. Продолжим главные плоскости до конечных размеров и будем считать их фиктивными главными плоскостями преломляющей поверхности. На главной плоскости возьмем точку M, лежащую на высоте h, и соединим ее с точками A и A'. Аналогично точку K, лежащук на высоте $-H$, соединим с точками P и P'

Линии AMA' и PKP' называются соответственно первым и вторы вспомогательными параксиальными лучами. Свойства этих лучей сле дующие:

1) лучи фиктивные, так как преломляются на фиктивных главны плоскостях и в природе существовать не могут;

2) углы α, α', β, β', образуемые лучами с оптической осью, а такж высоты h и H являются конечными величинами, как и для действител ных лучей;

3) отрезки s, s', s_p, $s'_{p'}$, отсекаемые этими лучами вдоль оптич ской оси, подчиняются формулам параксиальной оптики.

Расчет хода вспомогательных лучей через оптическую систему по зволяет определить кардинальные элементы системы, а также положен и размер изображения. Преобразуем (3.45) и (3.46), введя в них углы tg $= h/s$; $\mathrm{tg}\alpha' = h/s'$; $\mathrm{tg}\beta = H/s_p$; $\mathrm{tg}\beta' = H/s'_{p'}$. В формулах для параксиал ных лучей обозначения tg принято опускать. Следуя принятым обознач ниям, получаем для первого луча

$$n'\alpha' - n\alpha = h(n' - n)\rho, \qquad (3.4$$

где $\rho = 1/r$. Для второго луча

$$n'\beta' - n\beta = H(n' - n)\rho. \qquad (3.4$$

Покажем на примере системы из k компонентов порядок расчета с помощью вспомогательных параксиальных лучей. На рис. 3.19 изображена оптическая система, состоящая из k преломляющих поверхностей. Предмет высотой y_1 расположен на расстоянии $-s_1$ от вершины O_1 первой преломляющей поверхности. Для определения расстояния s'_k изображения и высоты y'_k изображения рассмотрим ход первого параксиального луча, идущего из осевой точки предмета под произвольным углом $-\alpha_1$. Применим последовательно (3.47) для каждой преломляющей поверхности. Переход от одной поверхности к другой осуществляется с помощью уравнения

$$h_{v+1} = h_v - d_v \alpha_{v+1}, \tag{3.49}$$

где $v = 1, 2, \ldots, k$.

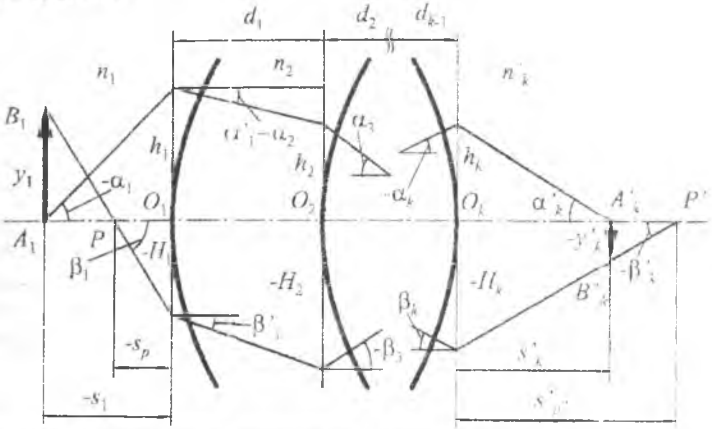

Рис. 3.19. Ход вспомогательных лучей через сложную оптическую систему

Для последней поверхности получаем

$$\alpha'_k = \frac{n_k \alpha_k + h_k \left(n'_k - n_k\right) \rho_k}{n'_k} \tag{3.50}$$

Положение изображения определяется расстоянием $s'_k = h_k \; \alpha'_k$ и высотой $y'_k = \beta_0 y_1$, где β_0 рассчитывается через соотношения между угловым и линейным увеличениями:

$$\beta_0 = n_1 \alpha_1 / \left(n'_k \alpha'_k\right). \tag{3.51}$$

Расчет второго параксиального луча происходит аналогично. Для последней преломляющей поверхности имеем

$$\left.\begin{aligned} H_k &= H_{k-1} - d_{k-1}\beta_k; \\ \beta'_k &= \frac{n_k\beta_k + H_k\left(n'_k - n_k\right)\rho_k}{n'_k}. \end{aligned}\right\} \qquad (3.52)$$

Отсюда получаем

$s'_{F'} = H_k/\beta'_k$; $\beta_{0p} = n_1\beta_1/\left(n'_k\beta'_k\right)$.

Для определения фокусного расстояния эквивалентной оптической системы рассчитывается ход первого параксиального луча, идущего параллельно оптической оси (рис. 3.20), $\alpha_1 = 0$. Задний фокальный отрезок и фокусное расстояние определятся при этом как $s'_{F'} = h_k/\alpha'_k$; $f' = h_1/\alpha'_k$.

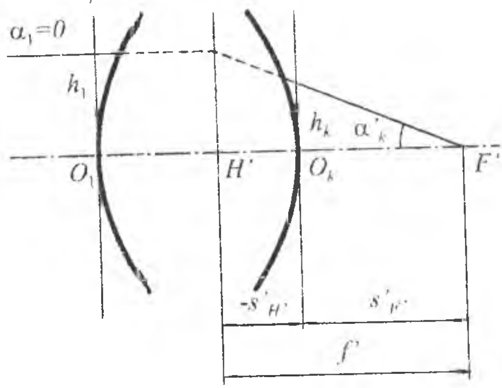

Рис. 3.20. Определение кардинальных элементов системы с помощью первого параксиального луча

3.2.4. Линзы конечной толщины и бесконечно тонкие линзы

Фокусные расстояния преломляющей поверхности. Для определения фокусных расстояний сферической преломляющей поверхности рассмотрим ход параксиальных лучей, параллельных оптической оси (рис. 3.21).

Из уравнения (3.43) для преломляющей поверхности получаем

$$\left.\begin{aligned} n'/f' &= (n'-n)/r, \\ -n/f &= (n'-n)/r; \end{aligned}\right\}$$

$$\left.\begin{aligned} f' &= n'r/(n'-n); \\ f &= -nr/(n'-n); \end{aligned}\right\} \qquad (3.5$$

при этом оптическая сила преломляющей поверхности равна

$$\Phi = -n/f = n'/f' = (n'-n)/r \qquad (3.54)$$

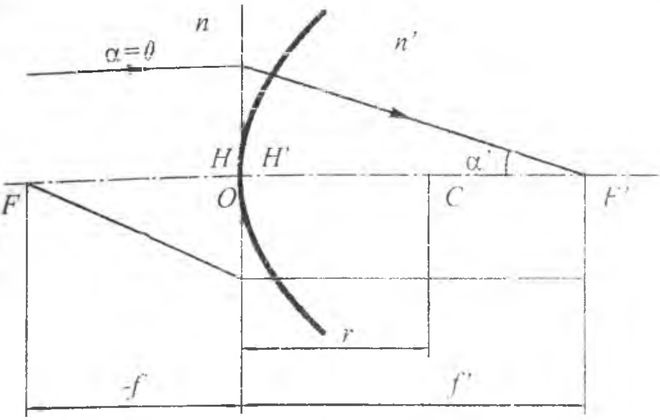

Рис. 3.21. Фокусные расстояния преломляющей поверхности

Кардинальные элементы толстой линзы. Представим толстую линзу как сложную оптическую систему, состоящую из двух компонентов — двух преломляющих поверхностей, находящихся на расстоянии d одна от другой (рис. 3.22). Тогда для определения кардинальных элементов можно воспользоваться формулами (3.35)—(3.37). Определив из (3.54) фокусные расстояния каждой преломляющей поверхности для оптического интервала, получим

$$\Delta = -\left(f'_1 - f_2 + d\right) = \frac{R}{\left(n_2 - n_1\right)\left(n_3 - n_2\right)} \qquad (3.55)$$

где $R = n_2 r_2\left(n_2 - n_1\right) + n_2 r_1\left(n_3 - n_2\right) - d\left(n_2 - n_1\right)\left(n_3 - n_2\right)$

Для фокусных расстояний имеем

$$\left.\begin{aligned} f &= -\frac{n_1 n_2 r_1 r_2}{R}; \\ f' &= \frac{n_2 n_3 r_1 r_2}{R}. \end{aligned}\right| \qquad (3.56)$$

Для определения положения фокусов относительно вершин преломляющих поверхностей s_F, $s'_{F'}$ найдем вначале отрезки z_F, $z'_{F'}$. Подставляя в (3.37) значения фокусных расстояний преломляющих поверхностей, получаем

$$\left.\begin{aligned} z_F &= -f\frac{r_1\left(n_3-n_2\right)}{r_2\left(n_2-n_1\right)}; \\ z'_{F'} &= -f'\frac{r_2\left(n_2-n_1\right)}{r_1\left(n_3-n_2\right)}. \end{aligned}\right\}$$

Из рис. 3.22 видно, что $s_F = z_F - f_1$; $s'_{F'} = z'_{F'} + f'_2$. Тогда

$$\left.\begin{aligned} s_F &= f\left[1-\frac{d\left(n_3-n_2\right)}{n_2 r_2}\right]; \\ s'_{F'} &= f'\left[1-\frac{d\left(n_2-n_1\right)}{n_2 r_1}\right]. \end{aligned}\right\}$$

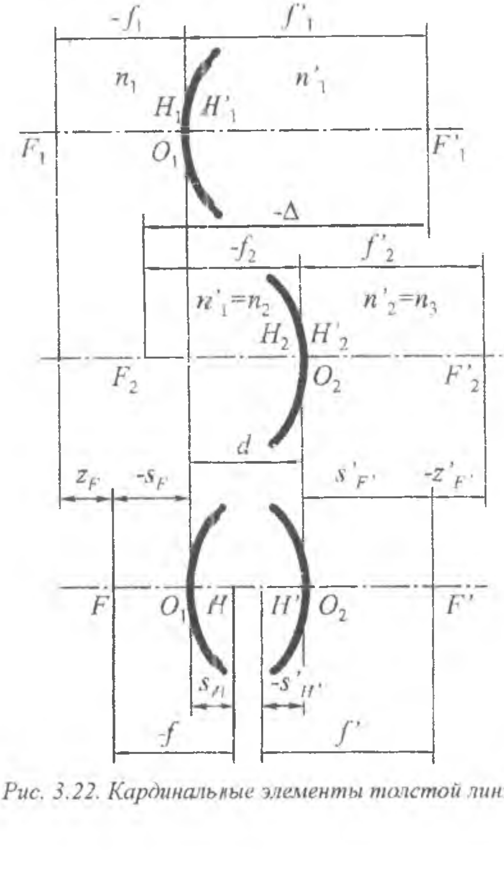

Рис. 3.22. Кардинальные элементы толстой линзы

Положение главных плоскостей относительно вершин преломляющих поверхностей находится как $s_H = s_F - f$; $s'_{H'} = s'_{F'} - f'$, отсюда

$$\left.\begin{aligned} s_H &= -\frac{fd(n_3 - n_2)}{n_2 r_2};\\ s'_{H'} &= -\frac{f'd(n_2 - n_1)}{n_2 r_1}. \end{aligned}\right\} \qquad (3.59)$$

В том случае, когда линза находится в воздухе ($n_3 = n_1 = 1$; $n_2 = n$), формулы (3.56)—(3.59) преобразуются к виду

$$\left.\begin{aligned} -f &= f' = \frac{nr_1 r_2}{(n-1)\left[n(r_2 - r_1) + d(n-1)\right]};\\ s_F &= -f'\left[1 + \frac{d(n-1)}{nr_2}\right];\; s'_{F'} = f'\left[1 - \frac{d(n-1)}{nr_1}\right];\\ s_H &= -f'\frac{d(n-1)}{nr_2}; \quad s'_{H'} = -f'\frac{d(n-1)}{nr_1}. \end{aligned}\right| \qquad (3.60)$$

Бесконечно тонкие линзы. Если толщина линзы намного меньше радиусов кривизны преломляющих поверхностей, т.е. $d \ll r$, то такая линза называется бесконечно тонкой. Формулы (3.60) в этом случае имеют вид:

$$\left.\begin{aligned} -f &= f' = \frac{nr_1 r_2}{(n-1)(r_2 - r_1)};\\ s_F &= f;\; s'_{F'} = f';\; s_H = s'_{H'} = 0. \end{aligned}\right\} \qquad (3.61)$$

3.3. МАТРИЧНЫЕ МЕТОДЫ РАСЧЕТА ОПТИЧЕСКИХ СИСТЕМ

Элементы матричной алгебры находят широкое применение при решении задач геометрической оптики. Известно, что если имеется система алгебраических линейных уравнений

$$x_2 = Ax_1 + By_1,$$

$$y_2 = Cx_1 + Dy_1,$$

то ее можно представить в матричном виде

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}.$$

Применение матричных методов в оптике стало возможным, так как связь между координатами предмета и изображения в пределах параксиальной оптики описывается линейными уравнениями, коэффициенты которых однозначно определяются свойствами оптической системы.

3.3.1. Матрицы преобразования лучей

Рассмотрим прохождение параксиального луча через центрированную оптическую систему. Введем декартову систему координат, ось oz которой совпадает с оптической осью системы, а ось oy направлена вертикально вверх (рис. 3.23).

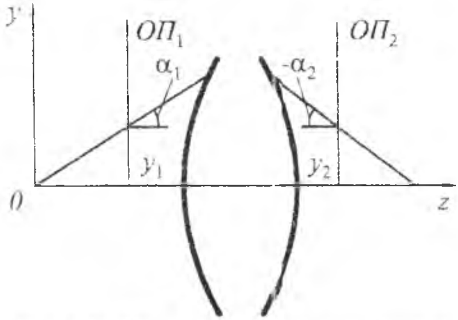

Рис. 3.23. Матрицы преобразования лучей

Выберем произвольно две плоскости, перпендикулярные оптиче ской оси, одну в пространстве предметов, другую в пространстве изобра жений, и назовем их опорными плоскостями $ОП_1$ и $ОП_2$. По отношению опорным плоскостям положение луча можно задать двумя координатами высотой y, на которой луч пересекает опорную плоскость, и углом α, ко торый луч составляет с оптической осью.

В матричной оптике, как и в геометрической оптике, существует правило, определяющее выбор знака угла. Поскольку матричная и гео метрическая оптика развивались независимо, то это правило в матричной оптике диаметрально противоположно принятому в геометрической оп тике, а именно: угол считается положительным, если он соответствует вращению против часовой стрелки от положительного направления оси к направлению, в котором свет распространяется вдоль луча. Будем в дальнейшем для углов придерживаться этого правила знаков, а для отре ков — правила знаков, принятого в геометрической оптике.

Часто в качестве второй координаты используется величина $V = n\alpha$, которая носит название направляющий косинус. Эту величину удобно использовать, поскольку в параксиальной области $n\alpha \approx n \sin\alpha$, а значит направляющий косинус остается постоянным при прохождении грани раздела двух сред. Пусть $V_1 = n_1\alpha_1$, $V_2 = n_2\alpha_2$. Тогда

$$\begin{bmatrix} y_2 \\ V_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \times \begin{bmatrix} y_1 \\ V_1 \end{bmatrix},$$

где коэффициенты матрицы A, B, C, D определяются только параметрами системы.

Выбор в качестве второй координаты направляющего косинуса удобен еще и тем, что в этом случае матрицы, описывающие как отдельные оптические элементы, так и систему в целом, будут унимодулярными, т.е. их определители будут равны единице. Это свойство удобно использовать для проверки правильности составления матриц в случае сложных оптических систем.

Так как распространение луча через оптическую систему сводится в основном к двум процессам: преломлению на преломляющей поверхности (отражение можно рассматривать как частный случай преломления) и распространению по прямой между преломляющими поверхностями, то остановимся более подробно на этих двух процессах.

Матрица перемещения T. Общепринятое обозначение матрицы перемещения буквой **T** происходит от английского слова «transfer» — перемещение. Пусть луч распространяется слева направо в среде с показателем преломления n и проходит путь d между двумя опорными плоскостями (рис. 3.24).

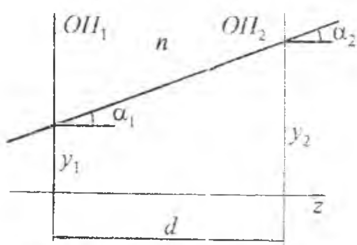

Рис. 3.24. Матрица перемещения

Углы, под которыми луч пересекает опорные плоскости, одинаковы $\alpha_2 = \alpha_1$. Одинаковы и направляющие косинусы: $V_2 = V_1$. Высоты падения на опорные плоскости связаны уравнением $y_2 = y_1 + d\,\mathrm{tg}\,\alpha_1$. Учитывая, что в параксиальной области $\mathrm{tg}\,\alpha \approx \alpha$, имеем $y_2 = y_1 + d\alpha_1 = y_1 + d/nV_1$. Таким образом, матрица перемещения имеет вид

$$\mathbf{T} = \begin{bmatrix} 1 & d/n \\ 0 & 1 \end{bmatrix}.$$

Величина d/n носит название приведенной толщины оптического промежутка. Если бы луч составлял с оптической осью отрицательный угол, матрица перемещения имела бы такой же вид.

Матрица преломления R. Обозначение матрицы преломления буквой **R** происходит от английского слова «refraction» — преломление. Предположим, что луч падает на сферическую преломляющую поверхность с радиусом кривизны r, разделяющую среды с показателями преломления n_1 и n_2 (рис. 3.25). Проведем первую опорную плоскость $ОП_1$ через вершину этой поверхности, а вторую — $ОП_2$ через точку пересечения луча с поверхностью.

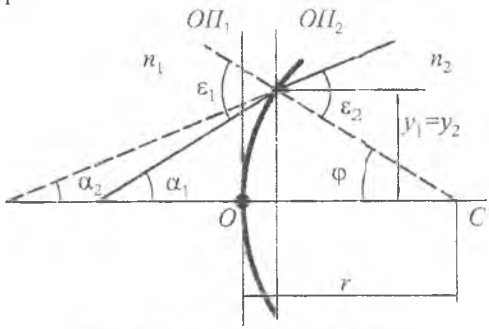

Рис. 3.25. Матрица преломления

Так как луч распространяется в параксиальной области, то координаты пересечения луча с опорными плоскостями практически равны, т.е. $y_2 = y_1$. Из закона преломления $n_1 \sin\varepsilon_1 = n_2 \sin\varepsilon_2$ или в параксиальном приближении $n_1\varepsilon_1 = n_2\varepsilon_2$. Из рис. 3.25 следует $\varepsilon_1 = \varphi + \alpha_1$; $\varepsilon_2 = \varphi + \alpha_2$, где $\varphi = y_1/r$. Тогда закон преломления запишется в виде

$$n_1\left(y_1/r + \alpha_1\right) = n_2\left(y_2/r + \alpha_2\right).$$

Переходя к направляющим косинусам, получаем $V_2 = \left(n_1 - n_2\right) y_1/r + V_1$. Отсюда матрица преломления имеет вид

$$\mathbf{R} = \begin{bmatrix} 1 & 0 \\ \left(n_1 - n_2\right)/r & 1 \end{bmatrix}.$$

Коэффициент C этой матрицы представляет собой оптическую силу преломляющей поверхности, взятую с обратным знаком.

Матрица сложной оптической системы. Предположим, что имеется некоторая оптическая система, состоящая из нескольких прелом ляющих поверхностей, причем известны матрицы каждой преломляющей поверхности в отдельности и матрицы перемещения между поверхностями (рис. 3.26).

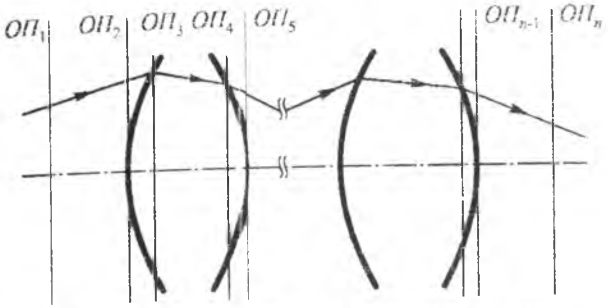

Рис. 3.26. Матрицы сложной оптической системы

Обозначим матрицу перемещения между первой и второй опорными плоскостями через $\mathbf{M}_1$, а матрицу преломления первой поверхности — через $\mathbf{M}_2$ и т.д. Тогда

$$\begin{bmatrix} y_2 \\ V_2 \end{bmatrix} = \mathbf{M}_1 \begin{bmatrix} y_1 \\ V_1 \end{bmatrix}; \quad \begin{bmatrix} y_3 \\ V_3 \end{bmatrix} = \mathbf{M}_2 \begin{bmatrix} y_2 \\ V_2 \end{bmatrix} \text{ и т.д.}$$

Обозначим вектор-столбец

$$\begin{bmatrix} y \\ V \end{bmatrix}$$

через $\mathbf{K}$, тогда получим

$$\mathbf{K}_2 = \mathbf{M}_1 \mathbf{K}_1;$$

$$\mathbf{K}_3 = \mathbf{M}_2 \mathbf{K}_2 = \mathbf{M}_2 \mathbf{M}_1 \mathbf{K}_1;$$

...

$$\mathbf{K}_n = \mathbf{M}_{n-1} \mathbf{K}_{n-1} = \mathbf{M}_{n-1} \mathbf{M}_{n-2} \mathbf{K}_{n-2} =$$
$$= \mathbf{M}_{n-1} \mathbf{M}_{n-2} \mathbf{M}_{n-3} \ldots \mathbf{M}_3 \mathbf{M}_2 \mathbf{M}_1 \mathbf{K}_1.$$

Последнее выражение связывает координаты луча на первой и последней опорных плоскостях. Значит, матрица $\mathbf{M} = \mathbf{M}_{n-1} \mathbf{M}_{n-2} \ldots \mathbf{M}_2 \mathbf{M}_1$ представляет собой матрицу, описывающую всю систему в целом. Таким образом, матрица сложной оптической системы равна произведению матриц отдельных элементов оптической системы, взятых в порядке, обратном расположению элементов вдоль направления распространения луча. Если матрица сложной системы уже известна, то в дальнейшем все промежуточные опорные плоскости можно опустить, а выходную опорную плоскость обозначить как $OП_2$. Тогда матрица **M** будет, как и прежде, связывать координаты луча на двух опорных плоскостях.

Тонкая линза. Рассмотрим прохождение луча через тонкую линзу с радиусами кривизны r_1, r_2 и показателем преломления n. Матрицы преломления на первой и второй поверхностях имеют вид:

$$\mathbf{R}_1 = \begin{bmatrix} 1 & 0 \\ (1-n)/r_1 & 1 \end{bmatrix}; \quad \mathbf{R}_2 = \begin{bmatrix} 1 & 0 \\ (n-1)/r_2 & 1 \end{bmatrix}.$$

Матрица перемещения, соответствующая распространению луча внутри линзы, является единичной матрицей, так как толщину d в этом случае можно принять равной нулю. Тогда

$$\mathbf{M} = \mathbf{R}_2\mathbf{R}_1 = \begin{bmatrix} 1 & 0 \\ \dfrac{(n-1)}{r_2} - \dfrac{(n-1)}{r_1} & 1 \end{bmatrix}.$$

Коэффициент C в данной матрице есть не что иное, как оптическая сила линзы, взятая с обратным знаком. Если несколько тонких линз расположены вплотную одна к другой, то матрица системы будет иметь вид

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -\sum\limits_{i=1}^{k}\Phi_i & 1 \end{bmatrix},$$

где $\sum\Phi_i$ — сумма оптических сил каждой тонкой линзы.

Толстая линза. Рассмотрим матрицу толстой линзы с радиусами кривизны r_1, r_2, толщиной d и показателем преломления n. В качестве опорных плоскостей проведем касательные к вершинам преломляющих поверхностей. Матрицы преломления в этом случае имеют такой же вид, как и в случае тонкой линзы, а матрица перемещения

$$\mathbf{T} = \begin{bmatrix} 1 & d/n \\ 0 & 1 \end{bmatrix}.$$

Перемножая матрицы в требуемом порядке, получаем

$$\mathbf{M} = \mathbf{R}_2\mathbf{T}\mathbf{R}_1 = \begin{bmatrix} 1 + \dfrac{d(1-n)}{r_1} & \dfrac{d}{n} \\ \dfrac{n-1}{r_2} + \left[\dfrac{(n-1)d}{nr_2} + 1\right]\dfrac{1-n}{r_1} & 1 + \dfrac{d(n-1)}{nr_2} \end{bmatrix}.$$

Проведя преобразования, получаем

$$C = \frac{(n-1)\left[n(r_1 - r_2) - (n-1)d\right]}{nr_2 r_1} = -\frac{1}{f'} = -\Phi.$$

Таким образом, как и в случае тонкой линзы, коэффициент C представляет собой оптическую силу линзы со знаком минус. Сопоставляя коэффициенты A и D с формулами (3.60) для толстой линзы, получаем $A = s'_{F'}/f'$, $D = s_F/f$, т.е. коэффициенты A и D равны отношению фокального отрезка к фокусному расстоянию соответственно в пространстве изображений и пространстве предметов.

3.3.2. Матричное описание свойств оптической системы

Предположим, что матрица сложной оптической системы известна. Рассмотрим, какое свойство оптической системы определяет каждый из коэффициентов матрицы.

А. Предположим, что $B = 0$. Уравнение для координаты луча в этом случае будет иметь вид $y_2 = Ay_1$. Это значит, что все лучи, пересекающие точку с координатой y_1 на первой опорной плоскости, пройдя оптическую систему, соберутся в точке с координатой y_2 на второй опорной плоскости (рис. 3.27, *а*), т.е. эти точки являются сопряженными, а опорные плоскости совпадают с плоскостями предмета и изображения. Коэффициент A в этом случае равен линейному увеличения оптической системы: $A = \beta_0 = y_2/y_1$.

Б. Пусть $A = 0$, тогда $y_2 = BV_1$. Это значит, что все лучи, идущие в пространстве предметов под углом $\alpha_1 = V_1/n_1$ соберутся в одной точке с координатой y_2, т.е. вторая опорная плоскость является задней фокальной плоскостью оптической системы (рис. 3.27, *б*).

В. Положим $D = 0$, тогда $V_2 = Cy_1$. Это значит, что все лучи, выходящие из точки с координатой y_1 на первой опорной плоскости, в пространстве изображений будут параллельны и пересекут вторую опорную плоскость под углом $\alpha_2 = V_2/n_2$ (рис. 3.27, *в*). Отсюда следует, что входная плоскость $ОП_1$ является передней фокальной плоскостью системы.

Г. Предположим, что $C = 0$, тогда $V_2 = DV_1$. Это значит, что все параллельные лучи, образующие в пространстве предметов с оптической осью угол $\alpha_1 = V_1/n_1$, на выходе оптической системы дадут также параллельный пучок лучей с углом наклона к оптической оси $\alpha_2 = V_2/n_2$. Такая система называется афокальной, или телескопической. Коэффициент

D в этом случае будет равен угловому увеличению оптической системы $D = V_2 n_1 / (V_1 n_2)$ (рис. 3.27, *г*).

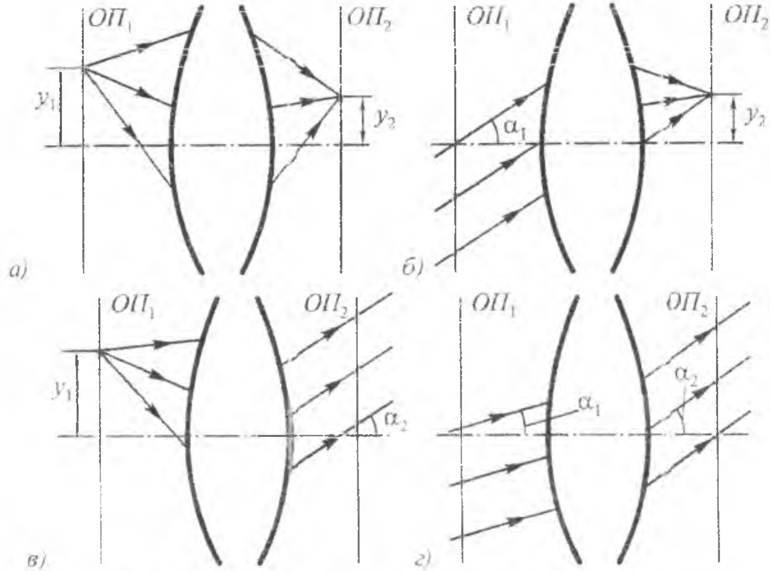

Рис. 3.27. Свойства коэффициентов матрицы оптической системы

3.3.3. Определение кардинальных элементов системы

Предположим, что найдена матрица оптической системы относительно некоторых опорных плоскостей. Определим положение кардинальных элементов системы относительно этих же опорных плоскостей.

Рассмотрим ход луча, идущего параллельно оптической оси на высоте y_1 (рис. 3.28,*а*). Уравнения, связывающие координаты луча на опорных плоскостях, будут иметь вид $y_2 = Ay_1$, $V_2 = Cy_1$, или $\alpha_2 = Cy_1/n$

Тогда получим:

$$f' = -y_1/\alpha_2 = -n_2/C; \; t' = -y_2/\alpha_2 = -n_2 A/C;$$

$$l' = t' - f' = n_2(1-A)/C.$$

За начало отсчета отрезков f', t', l' выбиралась вторая опорная плоскость системы.

Для аналогичных величин в пространстве предметов (рис. 3.28 имеем

$$f = -n_1/C; \; t = -Dn_1/C; \; l = n_1(D-1)/C.$$

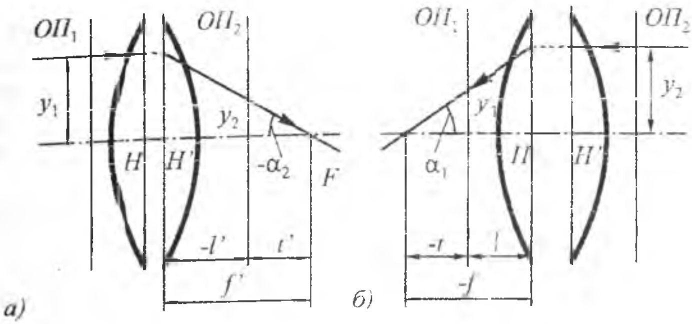

Рис. 3.28. Определение кардинальных элементов системы матричными методами

Рассмотрим случай, когда опорные плоскости являются касательными к вершинам первой и последней преломляющих поверхностей оптической системы. Тогда отрезки l, l' соответствуют отрезкам s_H, $s'_{H'}$, а отрезки t, t' — отрезкам s_F, $s'_{F'}$, (см. рис. 3.9). При этом $s'_{F'} = -n_2/C$, $s_F = Dn_1/C$. Коэффициенты равны: $A = s'_{F'}/f'$ $D = s_F/f$. Аналогичные выражения уже были получены при нахождении матрицы толстой линзы.

Определим положение узловых точек системы. Предположим, что узловые точки N и N' расположены соответственно на расстоянии p и p' от опорных плоскостей (рис. 3.29). Пусть матрица системы относительно опорных плоскостей $ОП_1$ и $ОП_2$

$$\mathbf{M} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

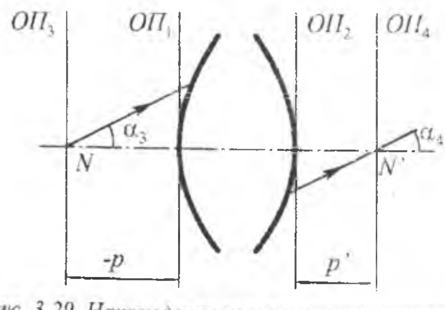

Рис. 3.29. Нахождение узловых точек системы

Введем новую пару опорных плоскостей $ОП_3$ и $ОП_4$, проходящих ерез узловые точки, и найдем новую матрицу системы относительно тих плоскостей. Матрица перемещения между плоскостями $ОП_3$ и $ОП_1$ меет вид

$$\begin{bmatrix} 1 & -p/n \\ 0 & 1 \end{bmatrix},$$

а между плоскостями OH_2 и OH_4

$$\begin{bmatrix} 1 & p'/n \\ 0 & 1 \end{bmatrix}.$$

Тогда матрица всей системы

$$\mathbf{M} = \begin{bmatrix} A + p'C/n_2 & \left(A + p'C/n_2\right)\left(-p/n\right) + B + p'D/n_2 \\ C & -Cp/n_1 + D \end{bmatrix}.$$

Обозначим коэффициенты вновь полученной матрицы как A', B', C D'. Тогда

$$\begin{bmatrix} y_4 \\ V_4 \end{bmatrix} = \begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} \times \begin{bmatrix} y_3 \\ V_3 \end{bmatrix},$$

где y_3, V_3, y_4, V_4 — координаты луча в третьей и четвертой опорных плос костях. Опорные плоскости выбраны таким образом, что $y_3 = y_4 = 0$. Уч тем также, что $\alpha_3 = \alpha_4$. Тогда получим $B' = 0$; $D' = n_2/n_1$ или

$$\left(A + p'C/n_2\right)\left(-p/n_1\right) + B + p'D/n_2 = 0;$$
$$D - pC/n_1 = n_2/n_1.$$

Решая совместно эту систему уравнений, получаем:

$$\left.\begin{aligned} p &= \frac{Dn_1 - n_2}{C}; \\ p' &= \frac{n_1 - An_2}{C}. \end{aligned}\right\}$$

При $n_1 = n_2$ имеем

$$p = \frac{n(D-1)}{C}; \quad p' = \frac{n(1-A)}{C}$$

и узловые точки совпадают с главными точками.

3.3.4. Ограничение пучков лучей в оптической системе

Все оптические детали имеют конечные размеры, которые огран чивают пучки лучей, проходящих через систему. Кроме того, в оптич ских системах устанавливаются специальные преграды. Все элемен оптической системы, ограничивающие размеры световых пучков, прох дящих через оптическую систему, называются диафрагмами.

Апертурная диафрагма. Рассмотрим оптическую систему, состоящую из двух компонентов (рис. 3.30) и установленной между ними диафрагмы P_1P_2. Пучок лучей, вышедший из точки A, вначале ограничивается диаметром первого компонента, а затем срезается оправой диафрагмы P_1P_2. Далее через второй компонент пучок проходит без ограничений.

Диафрагма, максимальным образом ограничивающая пучок лучей, выходящих из осевой точки предмета, называется апертурной диафрагмой. В приведенном примере апертурной является диафрагма P_1P_2.

Параксиальное изображение апертурной диафрагмы, образованное частью оптической системы, установленной перед диафрагмой, носит название входного зрачка системы. Если апертурная диафрагма установлена в пространстве предметов оптической системы, то она совпадает с входным зрачком системы.

Луч, выходящий из осевой точки предмета A и опирающийся на край апертурной диафрагмы, носит название апертурного луча пространства предмета. Угол между апертурным лучом и оптической осью AA' называется апертурным углом пространства предмета (на рис. 3.30 угол σ_a).

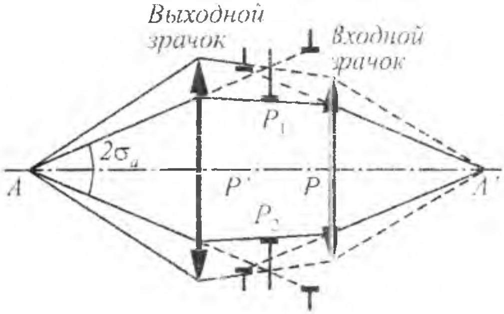

Рис. 3.30. Апертурная диафрагма оптической системы

Параксиальное изображение апертурной диафрагмы, образованное частью системы, расположенной после диафрагмы, называется выходным зрачком системы. Апертурная диафрагма, расположенная в пространстве предметов оптической системы, совпадает с выходным зрачком оптической системы.

Матричные методы расчета оптических систем позволяют определить положение апертурной диафрагмы, входного и выходного зрачков. Предположим, что известна матрица **M** системы, состоящая из **R** и **T** матриц. Рассмотрим влияние некоторой диафрагмы D_i, расположенной в промежуточной опорной плоскости $ОП_i$. Лучевой вектор (вектор-столбец) в данной опорной плоскости можно найти как

$$\mathbf{K}_i = \mathbf{M}_{i-1}\mathbf{M}_{i-2}\ldots\mathbf{M}_3\mathbf{M}_2\mathbf{M}_1 = \mathbf{L}_i\,\mathbf{K}_1,$$

где $\mathbf{L}_i$ — матрица системы между опорными плоскостями $ОП_1$ и $ОП_i$, элементы этой матрицы будем в дальнейшем обозначать как $(L_{11})_i$, $(L_{12})_i$ и т.д.

Рассмотрим луч, выходящий из точки предмета, лежащей на оси на расстоянии R слева от опорной плоскости $ОП_1$, а направляющий косинус для луча равен V. В плоскости $ОП_1$ лучевой вектор этого луча будет иметь вид

$$\mathbf{K}_1 = \begin{bmatrix} 1 & R \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ V \end{bmatrix} = \begin{bmatrix} RV \\ V \end{bmatrix}$$

Высота этого луча в промежуточной опорной плоскости $ОП_i$ задается выражением

$$y_i = \begin{bmatrix} 1 & 0 \end{bmatrix} K_i = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{L}_i \begin{bmatrix} RV \\ V \end{bmatrix} = V \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{L}_i \begin{bmatrix} R \\ 1 \end{bmatrix}$$

После операции умножения получаем

$$y_i = V\left\{R\left(L_{11}\right)_i + \left(L_{12}\right)_i\right\}$$

Поскольку в плоскости $ОП_i$ находится диафрагма диаметром D_i, то максимальная координата y_i должна равняться радиусу диафрагмы, а максимальный угол определится как

$$\alpha_{\max i} n_i = \frac{D_i}{2\left\{R\left(L_{11}\right)_i + \left(L_{12}\right)_i\right\}}$$

Повторив это вычисление для всех диафрагм, найдем диафрагму, для которой значение $\alpha_{\max}$ будет наименьшим. Эта диафрагма и будет являться апертурной диафрагмой.

Если предмет находится в бесконечности, то $R \to \infty$, $\alpha = 0$. В этом случае координату y_i в опорной плоскости $ОП_i$ удобно представить через координату луча y во входной опорной плоскости

$$y_i = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{L}_i \begin{bmatrix} y \\ y/R \end{bmatrix} = y\left\{\left(L_{11}\right)_i + \left(L_{12}\right)_i / R\right\}$$

Тогда наибольшее значение y луча во входной плоскости определится как

$$y_{\max i} = \frac{RD_i}{2\left\{R\left(L_{11}\right)_i + \left(L_{12}\right)_i\right\}}$$

С учетом $R \to \infty$ получаем

$$y_{\max i} = D_i / 2(L_{11})_i.$$

Апертурная диафрагма будет находиться в той плоскости, которой соответствует минимальная координата $y_{\max}$.

Если положение апертурной диафрагмы найдено, то можно определить положение и размер входного и выходного зрачков. Предположим, что $OП_S$ — опорная плоскость, соответствующая положению апертурной диафрагмы, а входной зрачок расположен на расстоянии p слева от опорной плоскости $OП_1$. Тогда матрица преобразования лучей от плоскости апертурной диафрагмы к плоскости входного зрачка будет иметь вид

$$\mathbf{L}_s \begin{bmatrix} 1 & p \\ 0 & 1 \end{bmatrix}.$$

Поскольку плоскости апертурной диафрагмы и входного зрачка сопряжены, получаем

$$p = -\left(\frac{L_{12}}{L_{11}}\right)_s; \quad \frac{D}{D_{AD}} = \frac{1}{(L_{11})_s},$$

где D_{AD} и D — соответственно диаметры апертурной диафрагмы и входного зрачка.

Аналогично для выходного зрачка, расположенного на расстоянии p' от выходной опорной плоскости, матрица преобразования лучей между выходным зрачком и апертурной диафрагмой имеет вид

$$\begin{bmatrix} 1 & p' \\ 0 & 1 \end{bmatrix} \mathbf{Q}_s,$$

где $\mathbf{Q}_s$ - матрица преобразования лучей между апертурной диафрагмой и выходной плоскостью.

Так как плоскости выходного зрачка и апертурной диафрагмы сопряжены, получаем

$$p' = -\left(\frac{Q_{12}}{Q_{22}}\right)_s; \quad \frac{D'}{D_{AD}} = \frac{1}{(Q_{22})_s},$$

где D' — диаметр выходного зрачка.

Полевая диафрагма. Любая диафрагма, расположенная в плоскости предмета, плоскости изображения или в любой другой сопряженной с ними плоскости, и ограничивающая размер изображения, носит название полевой диафрагмы (рис. 3.31). Наибольший размер изображения, лежащего на конечном расстоянии от оптической системы, называется линейным полем пространства изображений $2y'$. Наибольший размер изображаемой части плоскости предмета, расположенной на конечном расстоянии от оптической системы, называется линейным полем пространства предметов $2y$.

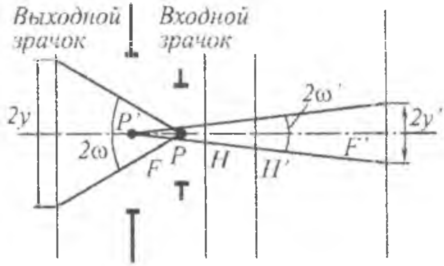

Рис. 3.31. Полевая диафрагма системы

Луч, проходящий через край полевой диафрагмы и центр входного (выходного) зрачка, называется главным лучом пространства предметов (изображений). Удвоенное значение угла между главным лучом пространства предметов (изображений) и оптической осью носит название углового поля пространства предметов (изображений). Угловое поле пространства предметов обозначается как 2ω, пространства изображений $2\omega'$.

Зная размеры линейного поля пространства предметов, можно рассчитать линейное поле пространства изображений $2y' = A \cdot 2y$, где A – коэффициент матрицы, связывающей плоскость предмета с плоскостью изображений.

3.4. АБЕРРАЦИИ ОПТИЧЕСКИХ СИСТЕМ

Падающие на оптическую систему гомоцентрические пучки лучей, пройдя через систему, теряют свою гомоцентричность. Нарушения гомоцентричности пучков лучей называются погрешностями, или аберрациями.

Аберрации оптических систем делят на монохроматические и хроматические. Монохроматические аберрации возникают при освещении светом определенной длины волны. Хроматические аберрации возникают при освещении светом сложного спектрального состава. Вследствие явления дисперсии при прохождении через преломляющие поверхности свет сложного спектрального состава разложится в спектр. При этом изображения для каждой длины волны будут отличаться как по положению, так и по размеру. Изображение становится окрашенным. Это явление носит название хроматических аберраций.

3.4.1. Волновые и лучевые аберрации

Рассмотрим некоторую оптическую систему, конструктивные параметры которой известны (рис. 3.32). Из внеосевой предметной точки проведем луч BG, который в общем случае не лежит в меридиональн плоскости. Пусть G и G' — точки пересечения луча с плоскостями входного и выходного зрачков, а B'_k — точка пересечения с плоскостью параксиального изображения. Предположим, что B'_0 — параксиальное изображение точки B. Отрезок $B'_0 B'_k$ называется лучевой аберрацией.

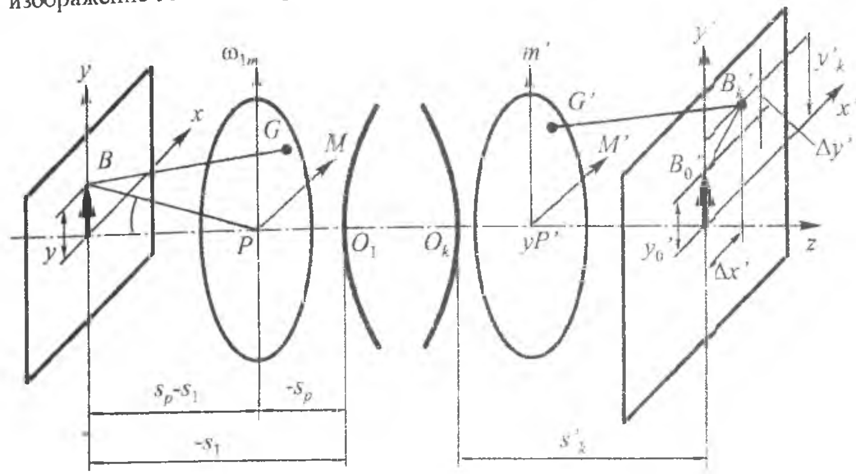

Рис. 3.32. Аберрации внемеридионального луча

Если бы система была свободна от аберраций, то волновая поверхность в пространстве изображений имела бы сферическую форму. При наличии аберраций волновая поверхность деформируется. На рис. 3.33 изображены сферический волновой фронт (сфера) S с центром в точке параксиального изображения B'_0 и реальный волной фронт W, проходящие через центр выходного зрачка P'. Сферу S называют опорной сферой Гаусса. Пусть Q', N'— точки пересечения луча $G'B'$ со сферой S и волновым фронтом W соответственно. Оптическая длина пути $\Phi = [N'Q']$ называется волновой аберрацией.

Для вычисления волновой аберрации и нахождения связи между волновой и лучевой аберрациями введем новую функцию — точечную характеристическую функцию Гамильтона. Пусть (x_0, y_0, z_0) и (x_1, y_1, z_1) — соответственно координаты двух точек P_0 и P_1 в двух прямоугольных системах координат. Точечная характеристика определяется как

$$V(x_0, y_0, z_0; x_1, y_1, z_1) = \int_{P_0}^{P_1} n\,\mathrm{d}s = \varphi(x_1, y_1, z_1) - \varphi(x_0, y_0, z_0). \quad (3.62)$$

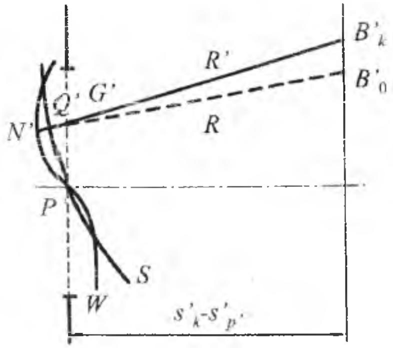

Рис. 3.33. Волновые и лучевые аберрации

Тогда из (3.6) получим

$$\left.\begin{aligned} \operatorname{grad}_0 V &= -n_0 s_0; \\ \operatorname{grad}_1 V &= n_1 s_1, \end{aligned}\right\} \tag{3.63}$$

где индексы 0 и 1 указывают, что оператор grad действует соответственно на координаты (x_0, y_0, z_0) и (x_1, y_1, z_1).

Обозначим углы, образованные вектором $n\mathbf{s}$ с координатными осями, через α, β и γ. Тогда проекции его на оси будут равны: $\lambda = n\cos\alpha$, $\mu = n\cos\beta$, $\nu = n\cos\gamma$, причем $\lambda^2 + \mu^2 + \nu^2 = n^2$. С учетом (3.63) проекции в точках P_0 и P_1 определятся как

$$\lambda_0 = -\frac{\partial V}{\partial x_0}; \quad \lambda_1 = -\frac{\partial V}{\partial x_1}. \tag{3.64}$$

Аналогичные выражения можно записать для μ_0, μ_1 и ν_0, ν_1. Воспользовавшись характеристическими функциями Гамильтона, выражение для волновой аберрации можно записать как (см. рис. 3.32 и 3.33)

$$\Phi = [N'Q'] = [BQ'] - [BN'] = [BQ'] - [BP']. \tag{3.65}$$

Здесь было использовано то обстоятельство, что точки N' и P' лежат на одном волновом фронте.

Введем две прямоугольные системы координат (см. рис. 3.32), начала которых находятся в осевой точке предмета и изображения, а ось z совпадает с осью системы. Точки в пространстве предметов будут рассматриваться в первой системе, а в пространстве изображений — во второй. Согласно (3.65) волновая аберрация выражается через точечную функцию следующим образом:

$$\Phi = V\left(x, y, 0; x'_{Q'}, y'_{Q'}, z'_{Q'}\right) - V\left(x, y, 0; 0, 0, s'_k - s'_{p'}\right), \tag{3.66}$$

где $(x, y, 0)$ — координаты точки B; $x'_{Q'}, y'_{Q'}, z'_{Q'}$ — координаты точки Q' и $\left(0, 0, s'_k - s'_{p'}\right)$ — координаты центра входного зрачка. Координаты точки Q' являются зависимыми, так как точка Q' лежит на опорной сфере, т.е.

$$\left(x'_{Q'} - x'_0\right)^2 + \left(y'_{Q'} - y'_0\right)^2 + \left(z'_{Q'}\right)^2 = R^2, \tag{3.67}$$

где x'_0, y'_0 — координаты точки B'_0 — параксиального изображения точки B. Величину $z'_{Q'}$ в уравнении (3.66) можно исключить с помощью (3.67), тогда

$$\Phi = \Phi\left(x, y, x'_{Q'}, y'_{Q'}\right)$$

Для определения соотношения между волновой и лучевой аберрациями продифференцируем уравнение (3.66). Получим

$$\frac{\partial \Phi}{\partial x'_{Q'}} = \frac{\partial V}{\partial x'_{Q'}} + \frac{\partial V}{\partial z'_{Q'}} \frac{\partial z'_{Q'}}{\partial x'_{Q'}}. \tag{3.68}$$

Пусть α', β' и λ' — углы, которые образует луч $Q'B'_x$ с координатными осями, тогда из (3.64) получаем

$$\left.\begin{aligned} \frac{\partial V}{\partial x'_{Q'}} &= n'\cos\alpha' = n'\frac{x' - x'_{Q'}}{R'}; \\ \frac{\partial V}{\partial z'_{Q'}} &= n'\cos\gamma' = -n'\frac{z'_{Q'}}{R'}, \end{aligned}\right\} \tag{3.69}$$

где

$$R' = \left\{\left(x' - x'_{Q'}\right)^2 + \left(y' - y'_{Q'}\right)^2 + \left(z'_{Q'}\right)^2\right\}^{1/2}$$

— расстояние от точки Q' до точки B'; n' — показатель преломления среды в пространстве изображений. Из (3.67) имеем

$$\frac{\partial z'_{Q'}}{\partial x'_{Q'}} = -\frac{x'_{Q'} - x'_0}{z'_{Q'}}. \tag{3.70}$$

Подставляя (3.70) и (3.69) в выражение (3.68), получаем

$$\left.\begin{aligned} x' - x'_0 &= \frac{R'}{n'}\frac{\partial \Phi}{\partial x'_{Q'}};\\ y' - y'_0 &= \frac{R'}{n'}\frac{\partial \Phi}{\partial y'_{Q'}}.\end{aligned}\right\} \tag{3.71}$$

Формулы (3.71) являются точными соотношениями, связывающими лучевые и волновые аберрации. Однако величина R' сама зависит от лучевых аберраций. В большинстве практических случаев величину R' можно заменить на радиус опорной сферы R, а координаты $x'_{Q'}$, $y'_{Q'}$ — на координаты луча в плоскости выходного зрачка M' и m' (см. рис. 3.32). Получаем

$$\left.\begin{aligned} \Delta x' &= \frac{R'}{n'}\frac{\partial \Phi}{\partial M'};\\ \Delta y' &= \frac{R'}{n'}\frac{\partial \Phi}{\partial m'}.\end{aligned}\right\} \tag{3.72}$$

Величины $\Delta y'$ и $\Delta x'$ называют также меридиональной и сагиттальной составляющими поперечной лучевой аберрации.

Функция Ф зависит от четырех переменных x, y, M', m', где x, y — координаты внеосевой точки предмета, а M', m' — координаты пересечения луча с плоскостью выходного зрачка. Однако эти переменные могут входить в выражение для Ф только в определенных комбинациях, обусловленных осевой симметрией системы. Введем в плоскостях предмета и выходного зрачка полярные координаты, тогда

$$y = r\cos\psi; \quad x = r\sin\psi; \quad m' = r'\cos\varphi; \quad M' = r'\sin\varphi.$$

Предположим, что координатные оси повернулись на некоторый угол, при этом неизменными остались переменные r, r' и $(\varphi - \psi)$. Поскольку функция Ф инвариантна относительно таких поворотов, то в качестве переменных удобно выбрать следующие величины:

$$R_1 = y^2 + x^2 = r^2; \quad R_2 = (m')^2 + (M')^2 = (r')^2;$$

$$u = ym' + xM' = rr'\cos(\varphi - \psi).$$

Отсюда следует, что при разложении функции Ф в ряд по степеням четырех координат нечетные степени будут отсутствовать. Не будет также и членов 2-го порядка, так как согласно (3.71) они соответствуют лучевым аберрациям, линейно зависящим от координат, а это противоречит тому, что B'_0 является параксиальным изображением точки B. Тогда разложение имеет вид

$$\Phi = \Phi_{IV} + \Phi_{VI} + \Phi_{VIII} + \ldots + \Phi_{2k},$$

где член степени $2k$ описывает волновую аберрацию порядка $2k$. Аберрации наинизшего порядка $(2k=4)$ называются первичными аберрациями, или аберрациями Зейделя, составляющие волновых аберраций выше 4-го порядка называются аберрациями высших порядков.

3.4.2. Монохроматические аберрации 3-го порядка

Рассмотрим волновую аберрацию 4-го порядка. Выберем в качестве переменных величины R_1, R_2 и u. В выражение для Φ_{IV} эти переменные могут входить в шести комбинациях, порядок которых равен четырем: $R_1^2,\ R_2^2,\ u^2,\ R_1R_2,\ R_1u,\ R_2u$. Представим

$$\Phi_{IV}=\frac{A_1}{4}R_1^2+\frac{B_1}{4}R_2^2+C_1u^2+\frac{D_1}{2}R_1R_2+E_1R_1u+F_1R_2u, \qquad (3.73)$$

где $A_1, B_1, C_1, D_1, E_1, F_1$ — постоянные, зависящие от конструктивных параметров системы, а коэффициенты 1/4, 1/2 выбраны для удобства дальнейших расчетов.

Выберем плоскость xOy так, чтобы она проходила через предметную точку B, тогда $z=0$. Продифференцируем уравнение (3.73) и с учетом (3.72) получим

$$\left.\begin{aligned}\Delta y'_{III}&=R\Big\{B_1m'\left[(m')^2+(M')^2\right]+F_1\left[3(m')^2+(M')^2\right]y+\\&\quad+\left(2C_1+D_1\right)m'y^2+E_1y^3\Big\};\\ \Delta x'_{III}&=R\left\{B_1M'\left[(m')^2+(M')^2\right]+2m'M'yF_1+D_1M'y^2\right\}\end{aligned}\right\} \qquad (3.74)$$

Величины $\Delta y'_{III}$, $\Delta x'_{III}$ являются меридиональной и сагиттальной составляющими лучевых аберраций 3-го порядка. Каждый член в формуле (3.74) содержит комбинацию произведений величин y, m', M', причем сумма степеней всегда равна трем. Положим $RB_1=A;\quad RF_1=B;$ $\left(2C_1+D_1\right)R=C;\ RD_1=D;\ RE_1=E$, тогда

$$\left.\begin{aligned}\Delta y'_{III}&=Am'\left[(m')^2+(M')^2\right]+By\left[3(m')^2+(M')^2\right]+Cy^2m'+Ey^3;\\ \Delta x'_{III}&=AM'\left[(m')^2+(M')^2\right]+2Bym'M'+Dy^2M',\end{aligned}\right\} \qquad (3.75)$$

где A, B, C, D, E — коэффициенты, зависящие только от параметров оптической системы, от положения плоскостей предмета и выходного зрачка и не зависящие от координат луча. Координаты в плоскости выходного зрачка m', M' можно заменить координатами пересечения луча с входным зрачком m, M, используя соотношения $m' = \beta_p m$ и $M' = \beta_p M$. Коэффициент β_p представляет собой линейное увеличение системы в зрачках. Формулы (3.75) при этом не изменятся, а коэффициенты A, B, C, D, E будут иметь другие численные значения.

Каждый член в формуле (3.75) описывает определенный тип отклонения волнового фронта от правильной сферической формы, при этом в плоскости параксиального изображения вместо четкого изображения наблюдается размытое пятно — пятно рассеяния. По числу коэффициентов в формуле (3.75) различают пять типов монохроматических аберраций 3-го порядка. Коэффициент A определяет сферическую аберрацию, B — кому, C — астигматизм, D — кривизну поля изображения, E — дисторсию.

Коэффициенты A, B, C, D, E выражают не через параметры системы, а через параметры первого и второго параксиальных лучей (рис. 3.34), при этом второй параксиальный луч проходит через центр входного зрачка. Для расчета лучей используются формулы (3.50)—(3.52).

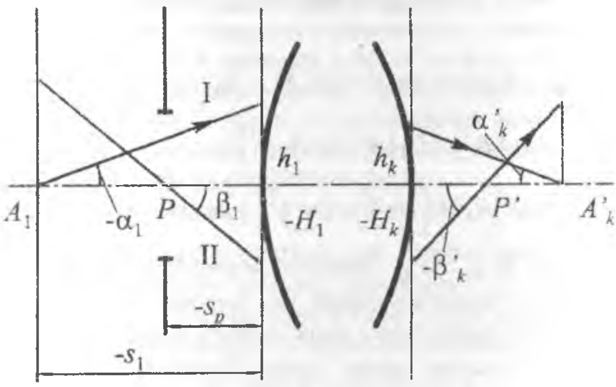

Рис. 3.34. Ход вспомогательных лучей в оптической системе

Вывод коэффициентов A, B, C, D, E подробно изложен в [1]. Здесь же воспользуемся окончательными значениями этих коэффициентов. Тогда выражения для составляющих поперечных аберраций 3-го порядка будут иметь вид:

$$
\left.\begin{aligned}
\Delta y'_{\mathrm{III}} &= -\frac{m\left(m^2+M^2\right)}{2n'_k\left(s_1-s_p\right)^3\alpha_1^3\alpha'_k}S_{\mathrm{I}}+\frac{y\left(3m^2+M^2\right)}{2n'_k\left(s_1-s_p\right)^3\alpha_1^2\alpha'_k\beta_1}S_{\mathrm{II}}-\\
&-\frac{y^2m}{2n'_k\left(s_1-s_p\right)^3\alpha_1\alpha'_k\beta_1^2}\left(3S_{\mathrm{III}}+I^2S_{\mathrm{IV}}\right)+\frac{y^3}{2n'_k\left(s_1-s_p\right)^3\alpha'_k\beta_1^3}S_{\mathrm{V}};\\
\Delta x'_{\mathrm{III}} &= -\frac{M\left(m^2+M^2\right)}{2n'_k\left(s_1-s_p\right)^3\alpha_1^3\alpha'_k}S_{\mathrm{I}}+\frac{ymM}{n'_k\left(s_1-s_p\right)^3\alpha_1^2\alpha'_k\beta_1}S_{\mathrm{II}}-\\
&-\frac{y^2M}{2n'_k\left(s_1-s_p\right)^3\alpha_1\alpha'_k\beta_1^2}\left(S_{\mathrm{III}}+I^2S_{\mathrm{IV}}\right).
\end{aligned}\right\} \qquad (3.76)
$$

Символами $S_{\mathrm{I}}, S_{\mathrm{II}}, S_{\mathrm{III}}, S_{\mathrm{IV}}, S_{\mathrm{V}}$ обозначены суммы Зейделя. Для системы из k компонентов эти суммы имеют вид:

$$
\left.\begin{aligned}
&S_{\mathrm{I}}=\sum_{\nu=1}^{k}h_\nu P_\nu;\quad S_{\mathrm{II}}=\sum_{\nu=1}^{k}H_\nu P_\nu\frac{\delta\beta_\nu}{\delta\alpha_\nu};\quad S_{\mathrm{III}}=\sum_{\nu=1}^{k}h_\nu P_\nu\left(\frac{\delta\beta_\nu}{\delta\alpha_\nu}\right)^2;\\
&S_{\mathrm{IV}}=\sum\frac{\delta(n\alpha)_\nu}{(hnn')_\nu};\quad S_{\mathrm{V}}=\sum\left[h_\nu P_\nu\left(\frac{\delta\beta_\nu}{\delta\alpha_\nu}\right)^2+I^2\frac{\delta(n\alpha)_\nu}{(hnn')_\nu}\right]\frac{\delta\beta_\nu}{\delta\alpha_\nu},
\end{aligned}\right\} \qquad (3.77)
$$

где

$$
P_\nu=\left[\frac{\delta\alpha}{\delta(1/n)}\right]_\nu^2\delta\left(\frac{\alpha}{n}\right)_\nu;\quad \delta\alpha_\nu=\alpha'_\nu-\alpha_\nu;\quad \delta\left(\frac{\alpha}{n}\right)_\nu=\frac{\alpha'_\nu}{n'_\nu}-\frac{\alpha_\nu}{n_\nu};
$$

$I=n_1y\alpha_1=-n_1\alpha_1\left(s_1-s_p\right)\beta_1$ — инвариант Лагранжа—Гельмгольца; h_ν, H_ν — высоты падения первого и второго параксиальных лучей на главные плоскости преломляющих поверхностей.

Выбор начальных данных для расчета вспомогательных лучей не влияет на значения самих аберраций 3-го порядка, однако суммы Зейделя при этом будут разными. Для сравнения различных оптических систем по суммам Зейделя их вычисляют при определенных условиях нормировки вспомогательных лучей.

Если предмет находится на конечном расстоянии, то для параметров вспомогательных лучей принимают

$$\left.\begin{aligned} &\alpha'_k = 1;\ \alpha_1 = \frac{n'_k}{n_1}\beta_0;\ h_1 = s_1\alpha_1;\\ &\beta_1 = 1;\ H_1 = s_p;\ I = -n'_k\left(s_1 - s_p\right)\beta_0, \end{aligned}\right\} \tag{3.78}$$

где β_0 — линейное увеличение системы. Тогда уравнения (3.76) примут вид

$$\left.\begin{aligned} \Delta y'_{\text{III}} = &-\frac{m\left(m^2+M^2\right)}{2n'_k\left(s_1-s_p\right)^3\alpha_1^3}S_{\text{I}} + \frac{y\left(3m^2+M^2\right)}{2n'_k\left(s_1-s_p\right)^3\alpha_1^2}S_{\text{II}} -\\ &-\frac{y^2 m}{2n'_k\left(s_1-s_p\right)^3\alpha_1}\left(3S_{\text{III}}+I^2 S_{\text{IV}}\right) + \frac{y^3}{2n'_k\left(s_1-s_p\right)^3}S_{\text{V}};\\ \Delta x'_{\text{III}} = &-\frac{M\left(m^2+M^2\right)}{2n'_k\left(s_1-s_p\right)^3\alpha_1^3}S_{\text{I}} + \frac{ymM}{n'_k\left(s_1-s_p\right)^3\alpha_1^2}S_{\text{II}} -\\ &-\frac{y^2 M}{2n'_k\left(s_1-s_p\right)^3\alpha_1}\left(S_{\text{III}}+I^2 S_{\text{IV}}\right). \end{aligned}\right\} \tag{3.79}$$

Если предмет находится в бесконечности, то $s_1 \to \infty$, $\alpha_1 \to 0$. Неопределенность в формулах (3.79) раскрывается следующим образом:

$$\left|\left(s_1 - s_p\right)\alpha_1\right|_{\alpha_1 \to 0,\, s_1 \to \infty} = h_1.$$

Для бесконечно удаленного предмета целесообразно также пользоваться угловым размером

$$\frac{y}{s_1 - s_p} = \operatorname{tg}\omega_1. \tag{3.80}$$

Условия нормировки для этого случая имеют вид:

$$\left.\begin{aligned} &\alpha_1 = 0;\ \alpha'_k = 1;\ h_1 = f' = 1;\\ &\beta_1 = 1;\ H_1 = s_p/f';\ I = -n_1. \end{aligned}\right\} \tag{3.81}$$

Выбор $f' = 1$ позволяет исключить влияние фокусного расстояния на суммы Зейделя, при этом все линейные размеры системы делятся на f'. Тогда формулы (3.76) примут вид:

$$\left.\begin{aligned}
\Delta y'_{\text{III}} &= -\frac{m\left(m^2+M^2\right)}{2n'_k\,(f')^2}S_{\text{I}} + \frac{\left(3m^2+M^2\right)\operatorname{tg}\omega_1}{2n'_k\,f'}S_{\text{II}} - \\
&-\frac{m\operatorname{tg}^2\omega_1}{2n'_k}\left(3S_{\text{III}}+I^2S_{\text{IV}}\right)+\frac{\operatorname{tg}^3\omega_1\,f'}{2n'_k}S_{\text{V}}; \\
\Delta x'_{\text{III}} &= -\frac{M\left(m^2+M^2\right)}{2n'_k\,(f')^2}S_{\text{I}} + \frac{mM\operatorname{tg}\omega_1}{n'_k\,f'}S_{\text{II}} - \\
&-\frac{M\operatorname{tg}^2\omega_1}{2n'_k}\left(S_{\text{III}}+I^2S_{\text{IV}}\right).
\end{aligned}\right\} \tag{3.82}$$

Суммы Зейделя определяют различные аберрации оптической системы: S_{I} — сферическую аберрацию, S_{II} — аберрацию комы; S_{III} и S_{IV} — астигматизм и кривизну поля изображения, S_{V} — дисторсию. Рассмотрим более подробно каждый из видов аберрации, полагая, что только один из коэффициентов Зейделя отличен от нуля. При этом необходимо помнить, что в действительности случаи, когда система обладает только одной аберрацией, являются исключительными. Обычно присутствуют все пять видов аберраций.

При исследовании влияния аберраций на качество изображений рассматривают совокупность лучей, выходящих из одной точки предмета и пересекающих плоскость выходного зрачка по окружностям с центром на оси системы. Пересечение этого пучка лучей с плоскостью изображения даст кривую, называемую аберрационной, или характеристической, кривой.

Сферическая аберрация. Предположим, что коэффициенты $S_{\text{II}} = S_{\text{III}} = S_{\text{IV}} = S_{\text{V}} = 0$, а $S_{\text{I}} \neq 0$. В этом случае система обладает только сферической аберрацией и

$$\left.\begin{aligned}
\Delta y'_{\text{III}} &= m\left(m^2+M^2\right)A; \\
\Delta x'_{\text{III}} &= M\left(m^2+M^2\right)A.
\end{aligned}\right\} \tag{3.83}$$

Выразим координаты m и M через полярные координаты ρ и ψ в плоскости входного зрачка (рис. 3.35): $m = \rho\cos\psi$, $M = \rho\sin\psi$. Тогда для составляющих поперечной аберрации получим

$$\Delta y'_{\text{III}} = \rho^3\cos\psi\,A; \quad \Delta x'_{\text{III}} = \rho^3\sin\psi\,A.$$

Возводя в квадрат и складывая выражения, имеем

$$\sqrt{\left(\Delta y'_{\text{III}}\right)^2+\left(\Delta x'_{\text{III}}\right)^2} = \rho^3 A. \tag{3.84}$$

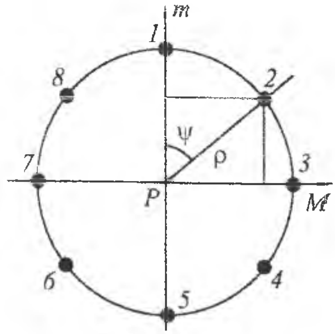

Рис. 3.35. Полярные координаты в плоскости входного зрачка

Данное выражение является уравнением окружности с центром в начале координат. Таким образом, аберрационная кривая в случае сферической аберрации представляет собой окружность, радиус которой пропорционален кубу радиуса входного зрачка.

Нарушение гомоцентричности пучка лучей, прошедшего оптическую систему, при сохранении его симметрии относительно оси пучка называется сферической аберрацией. Из (3.84) следует, что сферическая аберрация не зависит от размера предмета (координаты y), имеет место для любой точки плоскости предмета и зависит от апертуры системы.

На рис. 3.36 изображен ход луча, идущего от осевой точки предмета, при наличии в системе сферической аберрации. Сферическую аберрацию можно также характеризовать разностью продольных отрезков s'_k для действительных лучей и s'_0 для параксиальных лучей:

$$\Delta s' = s'_k - s'_0. \tag{3.85}$$

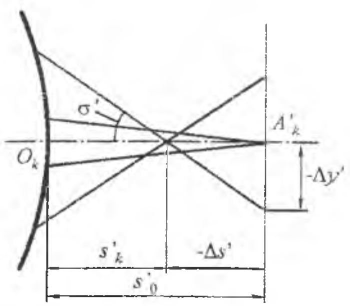

Рис. 3.36. Продольная и поперечная сферические аберрации

Эта разность называется продольной сферической аберрацией. Продольная и поперечная сферические аберрации связаны соотношением

$$\Delta y'_{III} = \Delta x'_{III} \operatorname{tg} \sigma'. \tag{3.86}$$

Тогда для предмета, расположенного на конечном расстоянии в меридиональной плоскости, при $n_1 = n'_k = 1$ (система в воздухе) получаем

$$\left.\begin{aligned}\Delta y'_{III} &= -0{,}5\,\mathrm{tg}^3\sigma' S_I;\\ \Delta s'_{III} &= -0{,}5\,\mathrm{tg}^2\sigma' S_I.\end{aligned}\right\} \qquad (3.87)$$

Для предмета в бесконечности

$$\left.\begin{aligned}\Delta y'_{III} &= -\frac{m^3}{2(f')^2} S_I;\\ \Delta s'_{III} &= -\frac{m^2}{2f'} S_I.\end{aligned}\right\} \qquad (3.88)$$

Кома. Если предположить, что $S_I = S_{III} = S_{IV} = S_V = 0$, а $S_{II} \neq 0$, то выражения для поперечной аберрации примут вид:

$$\left.\begin{aligned}\Delta y'_{III} &= \left(3m^2 + M^2\right) yB = \left(2 + \cos 2\psi\right)\rho^2 yB;\\ \Delta x'_{III} &= 2mMyB = \sin 2\psi \rho^2 yB.\end{aligned}\right\} \qquad (3.89)$$

Переходя к полярным координатам, из (3.89) получаем уравнение аберрационной кривой

$$\left(\Delta y'_{III} - 2\rho^2 yB\right)^2 + \left(\Delta x'_{III}\right)^2 = \left(\rho^2 yB\right)^2.$$

Данное выражение является уравнением окружности с радиусом $r = \rho^2 yB$, центр которой смещен на значение $2r$. Радиус полученной окружности и смещение ее центра пропорциональны квадрату радиуса входного зрачка ρ^2 и размеру предмета.

Для главного луча $m = M = 0$, поэтому $\Delta y'_{III} = \Delta x'_{III} = 0$. Это значит, что главный луч пересекает плоскость параксиального изображения в точке пересечения второго параксиального луча B'_0 (рис. 3.37).

Лучи, проходящие через край входного зрачка в точках 1 и 5 (см. рис. 3.35), в плоскости изображения пересекаются в точке с координатами $\Delta y'_{III} = 3r$ и $\Delta x'_{III} = 0$. Лучи пучка, проходящие через точки 3 и 7, пересекутся также в одной точке с координатами $\Delta y'_{III} = r$ и $\Delta x'_{III} = 0$. Для лучей, проходящих через точки 2 и 6, $\Delta y'_{III} = 2r$, $\Delta x'_{III} = r$, а для лучей, проходящих через точки 4 и 8, $\Delta y'_{III} = 2r$, а $\Delta x'_{III} = -r$.

Если входной зрачок разбить на ряд концентрических окружностей, то при постоянном значении y каждой из них будет соответствовать своя окружность в плоскости параксиального изображения. Таким образом, изображение внеосевой точки будет представлять собой фигуру рассеяния, изображенную на рис. 3.37. Прямые, касательные к окружностям, пересекаются в точке параксиального изображения B'_0 и образуют с осью y угол 30°.

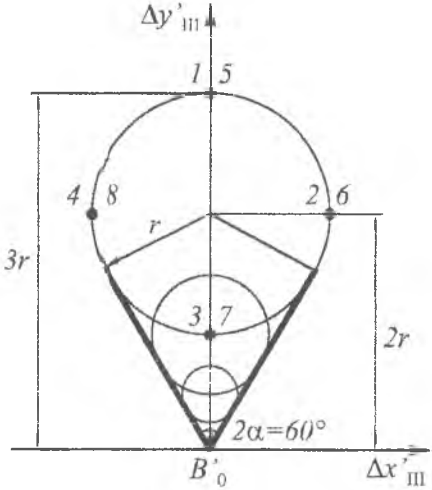

Рис. 3.37. Фигура рассеяния, обусловленная комой 3-го порядка

Нарушение симметрии пучка лучей, вышедших из внеосевой точки предмета, называется аберрацией комы. Несимметрия плоского меридионального пучка лучей называется меридиональной комой. Обозначим ее через K_{III}, тогда $K_{III} = \Delta y'_{III} = 3r$.

Для меридиональной комы 3-го порядка $(\Delta x'_{III} = 0)$ имеем:

- для предмета на конечном расстоянии $\quad K_{III} = -1{,}5\,\mathrm{tg}^2\sigma'\,\mathrm{tg}\,\omega_1 S_{II}$;
- для предмета в бесконечности $\quad K_{III} = -1{,}5\left(m^2/f'\right)\mathrm{tg}\,\omega_1 S_{II}$ $\qquad$ (3.90)

Астигматизм и кривизна поля изображения. Пусть $S_I = S_{II} = S_V = 0$, а S_{III} и $S_{IV} \neq 0$, тогда меридиональная и сагиттальная составляющие аберрации будут равны:

$$\left.\begin{aligned} \Delta y'_{III} &= my^2C = \rho\cos\psi y^2 C; \\ \Delta x'_{III} &= My^2A = \rho\sin\psi y^2 D. \end{aligned}\right\} \qquad (3.91)$$

Преобразуем (3.91) к виду

$$\left[\Delta y'_{III}/\left(\rho y^2 C\right)\right]^2 = \cos^2\psi; \; \left[\Delta x'_{III}/\left(\rho y^2 D\right)\right]^2 = \sin^2\psi.$$

Складывая эти выражения, получаем уравнение для аберрационной кривой

$$\left[\Delta y'_{\text{III}}/\left(\rho y^2 C\right)\right]^2+\left[\Delta x'_{\text{III}}/\left(\rho y^2 D\right)\right]^2=1. \tag{3.92}$$

Формула (3.92) является уравнением эллипса с осями $2a=2\rho y^2 C$ и $2b=2\rho y^2 D$. Таким образом, внеосевой точке предмета B в плоскости параксиального изображения соответствует фигура в виде эллипса (рис. 3.38).

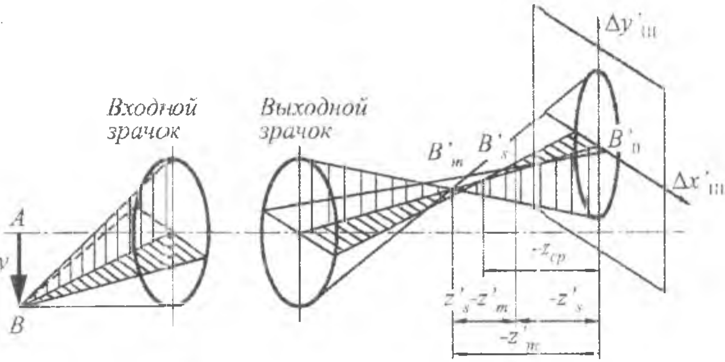

Рис. 3.38. Изображение внеосевой точки астигматическими лучами

При перемещении плоскости изображения вдоль оптической оси ось эллипса $2b$ уменьшается и на некотором расстоянии z'_s фигура рассеяния превращается в отрезок, лежащий в меридиональной плоскости. При дальнейшем перемещении на некотором расстоянии z'_m ось эллипса $2a$ становится равной нулю. Эллипс превращается в отрезок, расположенный в сагиттальной плоскости. Посередине между указанными положениями на расстоянии $z'_{\text{ср}}=\left(z'_m+z'_s\right)/2$ фигура рассеяния представляет собой окружность радиусом $r=\rho y^2 C=\rho y^2 D$.

Аберрация, при которой внеосевые точки предмета изображаются меридиональными и сагиттальными узкими пучками в виде двух взаимно перпендикулярных отрезков, расположенных на разном расстоянии от плоскости параксиального изображения, носит название астигматизма. Эти отрезки называются фокальными линиями.

Расстояние между точками B'_s и B'_m называется астигматической разностью вдоль главного луча. Проекция отрезка $B'_m B'_s$ на оптическую ось называется астигматической разностью вдоль оптической оси.

Координаты, характеризующие положение фокальных линий, следующие:

$$\left.\begin{aligned} z'_m &= -\frac{\operatorname{tg}^2\omega_1}{2n'_k\left(\alpha'_k\right)^2\beta_1^2}\left(3S_{\mathrm{III}}+I^2S_{\mathrm{IV}}\right);\\ z'_s &= -\frac{\operatorname{tg}^2\omega_1}{2n'_k\left(\alpha'_k\right)^2\beta_1^2}\left(S_{\mathrm{III}}+I^2S_{\mathrm{IV}}\right). \end{aligned}\right\} \tag{3.93}$$

Астигматическая разность

$$z'_m - z'_s = \frac{\operatorname{tg}^2\omega_1}{2n'_k\left(\alpha'_k\right)^2\beta_1^2}S_{\mathrm{III}}. \tag{3.94}$$

При изображении протяженных предметов их нужно рассматривать как совокупность точек, каждая из которых изображается астигматическими пучками лучей. На рис. 3.39 кривые M и S являются изображением отрезка AB и имеют форму парабол, касательных к плоскости параксиального изображения в точке A'. При изображении плоских предметов точки B'_m и B'_s будут находиться на поверхностях, представляющих собой параболоиды вращения кривых M и S, которые называются меридиональной и сагиттальной поверхностями вращения.

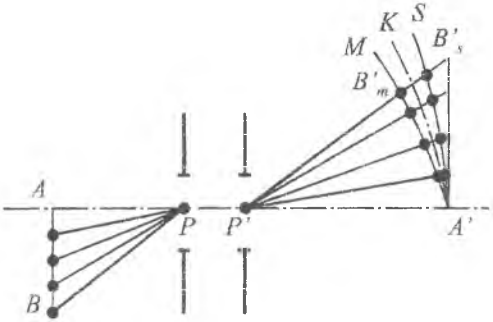

Рис. 3.39. Изображения отрезка прямой астигматическими пучками

Между меридиональной и сагиттальной поверхностями находится поверхность, на которой фигуры рассеяния принимают форму окружностей (поверхность получается при вращении кривой K вокруг оптической оси). Эта поверхность называется поверхностью изображения средней кривизны, а ее координата вычисляется по формуле

$$z'_{\text{ср}} = \frac{z'_s + z'_m}{2} = -\frac{\operatorname{tg}^2\omega_1}{2n'_k\left(\alpha'_k\right)^2\beta_1^2}\left(2S_{\mathrm{III}}+I^2S_{\mathrm{IV}}\right). \tag{3.95}$$

Если в системе устранен астигматизм ($S_{III}=0$), то меридиональная и сагиттальная поверхности совпадают, тогда

$$z'_{ср}=z'_s=z'_m=-\frac{\mathrm{tg}^2\omega_1}{2n'_k\left(\alpha'_k\right)^2\beta_1^2}I^2S_{IV}.$$

Коэффициент S_{IV} называется суммой Пецваля и характеризует кривизну поверхности изображения при отсутствии в системе астигматизма. Для получения плоского изображения необходимо, чтобы $S_{IV}=0$. Системы, у которых устранен астигматизм и поле изображения является плоским, называются анастигматами.

Дисторсия. Предположим, что $S_I=S_{II}=S_{III}=S_{IV}=0$, а $S_V\neq 0$, тогда

$$\left.\begin{aligned}\Delta y'_{III}&=y^3D;\\ \Delta x'_{III}&=0.\end{aligned}\right\}\qquad(3.96)$$

Таким образом, при наличии в системе дисторсии имеет место только меридиональная составляющая, причем значение ее не зависит от координат луча *т. M* в плоскости входного зрачка. Это означает, что все лучи, выходящие из внеосевой точки предмета, в плоскости параксиального изображения пересекутся также в одной точке, отличной от идеального изображения (рис. 3.40). Аберрация, при которой нарушается подобие между предметом и изображением, носит название дисторсии.

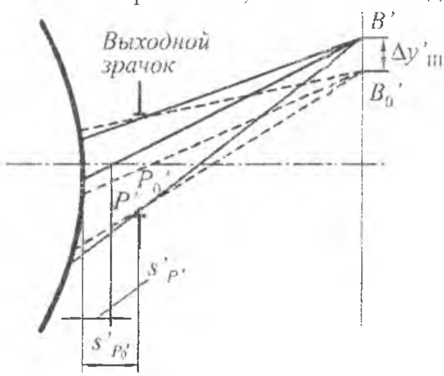

Рис. 3.40. Ход лучей при наличии дисторсии

Нарушение подобия происходит вследствие непостоянства линейного увеличения β для различных y. Дисторсию оптической системы можно оценить в относительных единицах

$$V_{\text{III}} = \frac{\beta - \beta_0}{\beta_0}. \tag{3.97}$$

Если β уменьшается при удалении от оптической оси, то $V_{\text{III}} < 0$ и система имеет отрицательную или бочкообразную дисторсию. Если β увеличивается при удалении от оси, то $V_{\text{III}} > 0$, и система обладает положительной или подушкообразной дисторсией. На рис. 3.41 дано изображение квадрата при наличии положительной (а) и отрицательной (б) дисторсии. Системы, у которых устранена дисторсия, называются ортоскопическими.

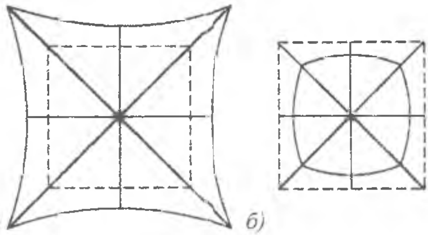

Рис. 3.41. Изображение квадрата:
а — при наличии положительной дисторсии;
б — при наличии отрицательной дисторсии.

Для системы в воздухе с учетом условий нормировки значение дисторсии будет определяться выражением

$$\Delta y'_{\text{III}} = -0{,}5\left(\text{tg}^3\,\omega_1\right)S_{\text{V}}. \tag{3.98}$$

3.4.3. Хроматические аберрации

Хроматические аберрации появляются в системе вследствие дисперсии и имеют место уже в параксиальной области. Различают хроматическую аберрацию положения изображения и хроматическую аберрацию размеров изображения. Хроматические аберрации параксиальной области называют также хроматизмом 1-го порядка. Хроматические аберрации действительных лучей имеют место при конечных значениях апертурных и полевых углов и называются хроматизмом высших порядков.

Хроматическая аберрация положения изображения. Пусть на оптическую систему падает пучок параксиальных лучей сложного спектрального состава, выходящий из осевой точки A (рис. 3.42). После прохождения системы пучок разложится в спектр. Предположим, что точка A'_{λ_1} является изображением для длины волны λ_1, а точка A'_{λ_2} — для длины волны λ_2.

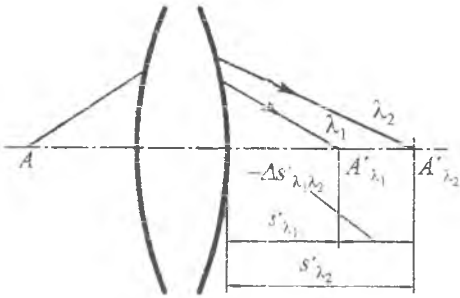

Рис. 3.42. Хроматическая аберрация положения изображения

Аберрация, при которой изображение осевой точки предмета лучами разной длины волны получается в разных точках вдоль оптической оси, называется аберрацией положения изображения. Количественно аберрация характеризуется разностью

$$\Delta s'_{\lambda_1\lambda_2} = s'_{\lambda_1} - s'_{\lambda_2}, \tag{3.99}$$

где $\lambda_1 < \lambda_2$. Часто в качестве λ_1 выбирается длина волны для синего цвета F' ($\lambda = 480{,}0$ нм), а в качестве λ_2 — для красного цвета C' ($\lambda = 643{,}8$ нм). Аберрация считается положительной, если точка A'_{λ_2} лежит правее точки A'_{λ_1}.

Точное значение хроматизма положения вычисляют, рассчитывая ход первого параксиального луча для двух длин волн. Приближенное значение можно вычислить по формуле

$$\Delta s'_{\lambda_1\lambda_2} = \frac{1}{n'_k\left(\alpha'_k\right)^2}\sum_{\nu=1}^{k} h_\nu C_\nu, \tag{3.100}$$

где h_ν — высота первого вспомогательного луча; C_ν — хроматический параметр,

$$C_\nu = \left(\frac{\delta\alpha}{\delta(1/n)}\right)_\nu \delta\left(\frac{\Delta n}{n}\right)_\nu,$$

где Δn — разность показателей преломления при переходе от длины волны λ_1 к длине λ_2; символом δ обозначено изменение соответствующей величины при переходе через преломляющую поверхность. Величина

$$s_{1хр} = \sum_{\nu=1}^{k} h_\nu C_\nu \tag{3.101}$$

называется первой хроматической суммой. Очевидно, что для устранения хроматизма положения необходимо выполнить условие $S_{I хр} = 0$. Поскольку в уравнение (3.100) не входит координата, связанная с положением входного зрачка, то хроматизм положения не зависит от положения входного зрачка.

Хроматическая аберрация размера изображения. Хроматической аберрацией размера, или хроматической аберрацией увеличения, называется аберрация, при которой изображение внеосевых точек лучами различной длины волны получается на разном расстоянии от оптической оси (рис. 3.43). Изображение при этом получается окрашенным.

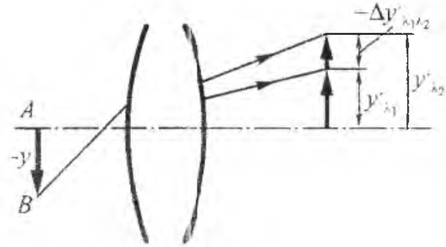

Рис. 3.43. Хроматическая аберрация размера изображения

Мерой хроматической аберрации увеличения является разность размера изображений для двух длин волн

$$\Delta y'_{\lambda_1\lambda_2} = y'_{\lambda_1} - y'_{\lambda_2}.$$

Точное значение хроматической аберрации увеличения можно определить, рассчитав ход второго параксиального луча для соответствующих длин волн. Приближенное значение вычисляют по формуле

$$\Delta y'_{\lambda_1\lambda_2} = \frac{y'_{\lambda_0}}{I} \sum_{v=1}^{k} H_v C_v, \tag{3.102}$$

где H_v — высоты пересечения второго параксиального луча с главными плоскостями преломляющих поверхностей; C_v — хроматический параметр; I — инвариант Лагранжа—Гельмгольца; y'_{λ_0} — размер изображения основной длины волны λ_0. Величина

$$S_{II хр} = \sum_{v=1}^{k} H_v C_v \tag{3.103}$$

называется второй хроматической суммой.

Подбором материалов и фокусных расстояний линз можно устранить хроматизм положения и увеличения для двух длин волн, т.е. провести ахроматизацию системы. Однако если система ахроматизирована для длин волн λ_1 и λ_2, то изображения для остальных длин волн все равно получаются в разных местах вдоль оптической оси и имеют разные размеры. В системе имеют место остаточные хроматические аберрации, которые носят название вторичного спектра. Вторичный спектр характеризуется расстоянием от изображения для основной длины волны λ_0 до изображения, ахроматизированного для длин волн λ_1 и λ_2 (рис. 3.44):

$$\Delta s'_{\text{в.с.}} = s'_{\lambda_1\lambda_2} - s'_{\lambda_0}.$$

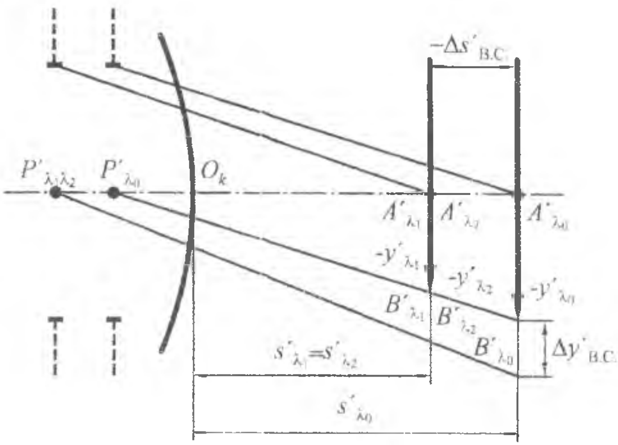

Рис. 3.44. Вторичный спектр положения и размера изображения

Вторичный спектр увеличения равен

$$\Delta y'_{\text{в.с}} = y'_{\lambda_1\lambda_2} - y'_{\lambda_0}.$$

3.4.4. Аберрации типовых оптических систем

Одиночная тонкая линза. Для тонкой линзы в воздухе имеем $h_1 = h_2 = h$; $H_1 = H_2 = H$; $d = 0$, а оптическая сила $\Phi = (\alpha_3 - \alpha_1)/h$. Хроматический параметр C равен сумме хроматических параметров преломляющих поверхностей:

$$C = -h\Phi/\nu, \tag{3.104}$$

где ν — коэффициент дисперсии, $\nu = (n - 1)/\Delta n$.

Из (3.104) видно, что положительные линзы имеют хроматизм положения отрицательный, а отрицательные — положительный. Выбором марок стекол хроматизм положения тонкой линзы может варьироваться.

Рассмотрим сферическую аберрацию для тонкой линзы в воздухе. В этом случае $S_1 = h\left(P_1 + P_2\right)$. Для устранения сферической аберрации необходимо выполнить условие $P = P_1 + P_2 = 0$. Вычисляя P из (3.77), получаем

$$P = \left(\frac{n}{n-1}\right)^2 (\alpha_3 - \alpha_2)\left[\alpha_3^2 + \alpha_1\alpha_3 + \alpha_1^2 - \frac{2n+1}{n}(\alpha_3 + \alpha_1)\alpha_2 + \frac{n+2}{n}\alpha_2^2\right]. \quad (3.105)$$

Приравнивая уравнения нулю и решая квадратное уравнение относительно α_2, получаем два корня. Анализ решения показывает, что корни будут вещественными только в том случае, если α_1 и α_3 имеют одинаковые знаки, т.е. для мнимых изображений.

Так как сферическая аберрация не может быть устранена для действительных изображений, рассмотрим условия, при которых она принимает минимальные значения.

Для предмета на конечном расстоянии $(\alpha_1 = \beta_0, \alpha_3 = 1)$ уравнение (3.105) принимает вид

$$P = \frac{n}{(n-1)^2}\left[(n+2)(1-\beta_0)\alpha_2^2 - (2n+1)\left(1-\beta_0^2\right)\alpha_2 + n\left(1-\beta_0^3\right)\right].$$

Дифференцируя P по α_2, находим условия, соответствующие минимуму сферической аберрации

$$\alpha_{2\min} = \frac{(2n+1)(1+\beta_0)}{2(n+2)}.$$

Радиусы кривизны линз, соответствующих минимуму сферической аберрации, определятся как

$$r_{1\min} = h\frac{n-1}{n\alpha_{2\min} - \beta_0} = h\frac{2(n+2)(n-1)}{n(2n+1) + \beta_0\left[n(2n-1) - 4\right]},$$

$$r_{2\min} = h\frac{n-1}{n\alpha_{2\min} - 1} = h\frac{2(n+2)(n-1)}{n(2n-1) - 4 + \beta_0 n(2n+1)}.$$

Обозначим отношение $\left(r_1/r_2\right)_{\min} = k$, тогда радиусы кривизны, выраженные через фокусное расстояние, будут равны:

$$r_{1\min} = (1-k)(n-1)f'; \quad r_{2\min} = (1-k)(n-1)f'/k.$$

Для предмета в бесконечности $(\alpha_1 = 0, \alpha_3 = 1)$ выражение (3.105) принимает вид

$$P=\frac{n}{(n-1)^2}\left[(n+2)\alpha_2^2-(2n+1)\alpha_2+n\right], \tag{3.106}$$

откуда получаем

$$\alpha_{2\min}=\frac{2n+1}{2(n+2)};\quad P_{\min}=\frac{n(4n-1)}{4(n+2)(n-1)^2}.$$

Отсюда следует, что минимальные значения основных параметров тонкой линзы зависят только от показателей преломления. Радиусы кривизны линз определятся как

$$r_{1\min}=\frac{2(n+2)(n-1)}{n(2n+1)},\quad r_{2\min}=\frac{2(n+2)(n-1)}{n(2n-1)-4};$$

$$\left(\frac{r_1}{r_2}\right)_{\min}=\frac{2(2n-1)-4}{n(2n+1)}.$$

Найдем соотношение радиусов, при котором сферическая аберрация будет минимальной. Выберем для определенности n = 1,5. Тогда $(r_1/r_2)_{\min}=-1/6$. Таким образом, двояковыпуклая линза с соотношением радиусов $r_2=-6r_1$ является наивыгоднейшей линзой в отношении минимума сферической аберрации.

В табл. 3.1 представлены значения параметров α_2 и P для линз разной конфигурации r_1 и r_2, имеющих одно и то же фокусное расстояние f' = 10 см и показатель преломления n =1,5.

Таблица 3.1. Значения параметров для линз разной конфигурации

Параметр	Номер линзы (рис.3.45)					
	1	2	3	4	5	6
r_1	-10	∞	10	5,83	5	3,33
r_2	-3,33	-5	-10	-35	∞	10
f'/r_1	-1	0	1	1,7	2	3
α_2	-0,33	0	0,33	0,56	0,66	1
P	19,2	9	3,36	2,13	2,3	6

На рис. 3.45 представлена зависимость сферической аберрации от формы тонких линз и относительного отверстия. Наибольшей сферической аберрацией обладают мениски, вогнутые поверхности которых обращены к предмету.

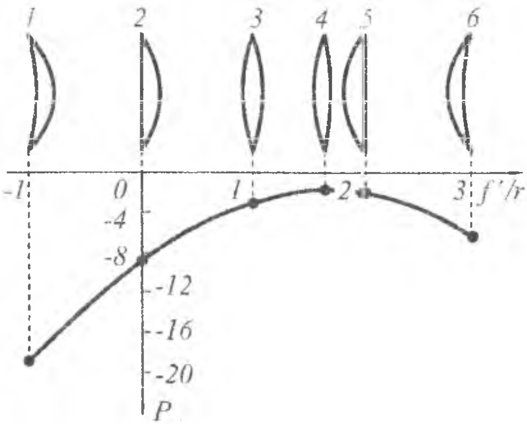

Рис. 3.45. Зависимость сферической аберрации от формы тонкой линзы

Одиночная толстая линза. Рассмотрим возможность ахроматизации толстой линзы. Для устранения хроматизма положения первая хроматическая сумма должна равняться нулю:

$$S_{I\,хр} = h_1 C_1 + h_2 C_2 = h_1 \frac{\alpha_2 - \alpha_1}{(1/n) - 1} \frac{\Delta n}{n} - h_2 \frac{\alpha_3 - \alpha_2}{1 - (1/n)} \frac{\Delta n}{n} = 0. \qquad (3.107)$$

Решая уравнение (3.107) относительно α_3, получаем

$$\alpha_3 = \frac{h_1 \alpha_1 - d\alpha_2^2}{h_2}, \qquad (3.108)$$

где d — толщина линзы.

Уравнение (3.107) является условием ахроматизации толстой линзы. Если известны показатель преломления и толщина линзы (n, d), а также положение предмета s_1, то ахроматизация осуществляется в следующем порядке. Задав произвольно величины α_1, α_2, рассчитывают $h_1 = s_1 \alpha_1$, $h_2 = h_1 - d\alpha_2$. Вычислив α_3, определяют радиусы кривизны ахроматизированной линзы

$$r_1 = \frac{h_1 (n-1)}{n\alpha_2 - \alpha_1}; \quad r_2 = \frac{h_2 (1-n)}{\alpha_3 - n\alpha_2}.$$

Изменяя значение α_2, получаем совокупность ахроматизированных линз с разными радиусами кривизны, т.е. разной формы. Поскольку сферическая аберрация зависит от формы линзы, то можно одновременно выбрать линзу с наименьшим значением сферической аберрации.

Ахроматизацию толстой линзы можно провести из условия $d\Phi/dn=0$. Оптическая сила толстой линзы

$$\Phi = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) + \frac{d(n-1)^2}{nr_1r_2}. \qquad (3.109)$$

Дифференцируя уравнение (3.109) и приравнивая производную нулю, получаем

$$r_1 - r_2 = -\frac{d\left(n^2-1\right)}{n^2}. \qquad (3.110)$$

Подставляя (3.110) в (3.109) для оптической силы линзы, имеем

$$\Phi = -\frac{d(n-1)^2}{n^2 r_1 r_2}. \qquad (3.111)$$

Если r_1 и r_2 имеют разные знаки, то $\Phi > 0$, т.е. ахроматизированная линза представляет собой двояковыпуклую линзу. Однако толщина линзы в этом случае должна быть неприемлемо большой. Если r_1 и r_2 имеют одинаковый знак, то $\Phi < 0$, т.е. условие ахроматизации может быть выполнено для отрицательной линзы в форме мениска.

Двухлинзовый объектив. Рассмотрим условия ахроматизации тонкого двухлинзового объектива, расположенного в воздухе. Оптическая сила объектива

$$\Phi = \Phi_1 + \Phi_2, \qquad (3.112)$$

а хроматический параметр

$$C = -h\left(\Phi_1/v_1 + \Phi_2/v_2\right). \qquad (3.113)$$

Приравнивая (3.112) нулю и решая совместно с (3.113), получаем

$$\Phi_1 = \frac{v_1\Phi}{v_1 - v_2}; \quad \Phi_2 = -\frac{v_2\Phi}{v_1 - v_2}. \qquad (3.114)$$

Таким образом, для исправления хроматизма положения нужна комбинация из двух линз, имеющих фокусное расстояние разных знаков. Линзу, фокусное расстояние которой определяет фокусное расстояние всего объектива, изготовляют из стекла, имеющего больший коэффициент дисперсии. Хроматизм увеличения в этом случае также устраняется.

ВЫВОДЫ

1. Приближение геометрической оптики (первое приближение к решению системы уравнений Максвелла), соответствующее предельному переходу $\lambda \to 0$ (относительно характерных размеров системы), позволяет перейти к уравнению эйконала и понятию геометрических световых лучей.

2. Приближение геометрической оптики не работает там, где происходит быстрое изменение напряженности поля (например, в фокальной плоскости линзы), а также там, где происходят быстрые перепады параметров среды (например, у края линзы).

3. В геометрической оптике вводятся понятия идеальной оптической системы и кардинальных элементов идеальной системы, с помощью которых можно рассчитывать положение и размер изображения.

4. При отражении и преломлении лучей реальными сферическими поверхностями положение и размер изображения зависят от угла наклона лучей к оптической оси. Идеальная система реализуется только в параксиальной области.

5. Для расчета реальных оптических систем удобно использовать вспомогательные первый и второй параксиальные лучи, которые являются фиктивными, поскольку преломляются не на преломляющих поверхностях, а на фиктивных главных плоскостях.

6. Наряду с геометрической оптикой расчет оптической системы можно проводить матричными методами.

7. В реальных оптических системах необходимо учитывать ограничение лучей диафрагмами. Роль диафрагмы может выполнять оправа самой оптической детали.

8. Любая оптическая система имеет аберрации. Аберрации делятся на монохроматические и хроматические.

9. Монохроматические аберрации присущи системе для определенной длины волны. В области Зейделя различают пять монохроматических аберраций.

10. Хроматические аберрации возникают вследствие зависимости показателя преломления среды от длины волны и имеют место уже в параксиальной области.

ГЛАВА 4
ОСНОВЫ СКАЛЯРНОЙ ТЕОРИИ ДИФРАКЦИИ
4.В. ВВЕДЕНИЕ

Применяя метод последовательных приближений, необходимо учитывать влияние перераспределения интенсивности света при его прохождении в средах, имеющих резкие изменения параметров, в частности, при наличии краев, отверстий, ограничений апертуры и т.д. Эти явления носят название дифракционных и могут быть описаны в рамках следующего приближения к точному решению системы уравнений Максвелла.

Явления дифракции играют чрезвычайно важную роль во многих областях науки и техники: в оптике, акустике, радиофизике, радиотехнике и др. Естественно, что при изучении оптики необходимо заниматься дифракционными явлениями только в присущем оптике диапазоне длин волн.

Дифракцию можно определить, следуя за Зоммерфельдом, как любое отклонение световых лучей от прямой линии, если оно не может быть объяснено преломлением на границе раздела двух сред или рефракцией, т.е. изгибанием лучей в средах с непрерывно меняющимся показателем преломления. Первое объяснение дифракции было сделано Гюйгенсом, который предположил, что если окружить все источники излучения замкнутой поверхностью, то каждую точку этой поверхности можно рассматривать как источник вторичных волн, распространяющихся во всех направлениях. Тогда в следующий момент времени волновой фронт можно найти как огибающую вторичных сферических волн (рис. 4.1).

Рис. 4.1. Волновой фронт как огибающая вторичных сферических волн (принцип Гюйгенса)

Затем к принципу Гюйгенса было добавлено Френелем предположение о том, что при нахождении нового светового распределения необходимо учитывать также и интерференцию вторичных волн, принимая во внимание когерентность вторичных источников, так как они возбуждаются одними и теми же первичными источниками. Подобное дополнение привело к значительному уточнению картины дифракции, однако при этом были сделаны достаточно произвольные предположения об амплитудах и фазах вторичных источников. В математическом виде эта теория была сформулирована в 1883 г. Кирхгофом, который показал, что при учете волновой природы света особенности амплитуд и фаз, введенные Френелем, легко объясняются. Но и Кирхгоф допустил некий произвол, связанный с определением граничных условий при дифракции, т.е. компонент поля на границе объекта, на котором происходит дифракция. Окончательно теория была сформулирована Зоммерфельдом на рубеже XIX и XX вв. при использовании метода функций Грина. Ему удалось добиться внутренней непротиворечивости теории. Именно этот вариант теории дифракции Кирхгофа—Зоммерфельда будет рассматриваться.

Следует помнить, что и эта прекрасно работающая теория является приближенной, так как она носит скалярный характер. Теория рассматривает только одну поперечную компоненту поля (электрического или магнитного) и считает, что другие компоненты можно рассматривать независимо. Так как уравнения Максвелла являются векторными, то в общем случае это утверждение несправедливо и следовало бы оперировать с векторными потенциалами. (В конце главы кратко остановимся на особенностях точной теории дифракции.) Теория дает очень хорошие результаты, подтвержденные экспериментами, но при следующих условиях:

1) отверстия в экранах велики, по сравнению с длиной волны λ;

2) дифрагированные волны наблюдаются не слишком близко от экрана.

В качестве примера, когда скалярная теория не работает, можно привести теорию дифракционных решеток высокого разрешения (с количеством штрихов более 1000 на 1 мм). В этом случае угловые соотношения в дифрагированном поле описываются скалярной теорией дифракции, но она дает неверные значения амплитудного распределения.

Ограничимся рассмотрением только скалярной теории дифракции, так как она описывает большинство наблюдаемых явлений.

4.1. ТЕОРИЯ ДИФРАКЦИИ КИРХГОФА-ЗОММЕРФЕЛЬДА

4.1.1. Интегральная теорема Гельмгольца—Кирхгофа

Перед тем как перейти к доказательству центральной теоремы теории дифракции, вспомним формулировку теоремы Грина и определение функций Грина.

Теорема Грина. Пусть $U(P)$ и $G(P)$ — две произвольные комплексные функции пространственных координат в точке P, а S — поверхность, окружающая объем V. Если эти функции, а также их первые и вторые частные производные однозначны и непрерывны внутри указанного объема и на поверхности, его окружающей, то

$$\iiint\limits_V \left(G\nabla^2 U - U\nabla^2 G\right)\mathrm{d}V = \iint\limits_s \left(G\frac{\partial U}{\partial \mathbf{n}} - U\frac{\partial G}{\partial \mathbf{n}}\right)\mathrm{d}s, \tag{4.1}$$

где $\partial/\partial\mathbf{n}$ — обозначает частную производную в каждой точке поверхности по *внешней* нормали в этой точке. Пользуясь этой теоремой, нужно все время помнить об условиях, накладываемых на вид функций и форму поверхности.

Функции Грина. По определению функцией Грина называется решение дифференциального уравнения, в правой части которого стоит дельта-функция. С точки зрения физики это означает, что функция Грина есть решение задачи с точечным источником. В качестве примера можно привести волновое уравнение и его решение. Если в волновом уравнении в правой части стоит $\delta(\mathbf{r})$, то его решением является сферическая волна, так как в однородном пространстве точечный источник излучает волну со сферическим волновым фронтом.

Интегральная теорема Гельмгольца—Кирхгофа. Кирхгоф сформулировал теорему, которая определяет значение поля в произвольной точке пространства через значение поля и его градиента на поверхности, ее окружающей. (Впервые она была доказана в акустике Гельмгольцем.)

Применим (4.1) к объему V, окруженному поверхностью S. Внутри этого объема находится точка P_0 (рис. 4.2).

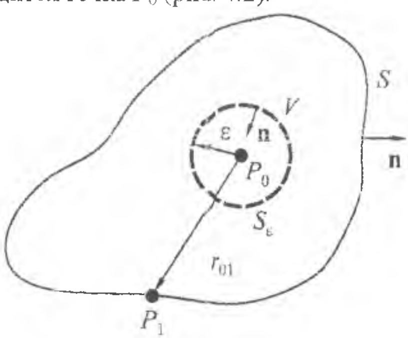

Рис. 4.2. К выводу интегральной теоремы Гельмгольца—Кирхгофа

Сам объем и действующее электромагнитное поле $U(P)$ удовлетворяют условиям теоремы Грина. Выберем функцию Грина $G(P)$ в виде сферической волны, что соответствует нахождению точечного излучателя в точке P_0. Это функция Грина для свободного пространства. В произвольной точке P_1 ее значение равно

$$G(P_1)=\frac{\exp(jkr_{01})}{\mathbf{r}_{01}}, \tag{4.2}$$

где через r_{01} обозначена длина вектора $\mathbf{r}_{01}$ из точки P_0 в точку P_1. В нашем случае не удовлетворяются требования, налагаемые на функцию $G(P)$ условиями теоремы, так как при стремлении к нулю длины r_{01} функция $G(P_1)$ стремится к бесконечности. Окружим точку P_0 сферой малого радиуса ε и рассмотрим объем V', заключенный между поверхностями S и S_ε. В этом объеме условия теоремы Грина выполняются, надо только помнить, что поверхность S', окружающая объем V', стала неодносвязной ($S' = S + S_\varepsilon$) и нормаль к поверхности на ее сферической части направлена внутрь сферы.

Любая электромагнитная волна, распространяющаяся в регулярной области пространства, должна подчиняться волновому уравнению (в данном случае уравнению Гельмгольца (2.19)). Поэтому и искомое решение (функция $U(P)$ и функция Грина свободного пространства $G(P)$) должны подчиняться этому уравнению.

$$\left.\begin{aligned}\left(\nabla^2+k^2\right)U(P)=0;\\ \left(\nabla^2+k^2\right)G(P)=0.\end{aligned}\right\} \tag{4.3}$$

Из этих уравнений следует, что

$$\nabla^2 G(P)=-k^2G(P);\quad \nabla^2U(P)=-k^2U(P).$$

Подставляя эти выражения в левую часть (4.1), получаем

$$\iiint\limits_{V'}\left(G\nabla^2U-U\nabla^2G\right)\mathrm{d}v=-\iiint\limits_{V'}\left(GUk^2-UGk^2\right)\mathrm{d}v=0.$$

Однако это означает, что и правая часть в (4.1) также равна нулю:

$$\iint\limits_{S'}\left(G\frac{\partial U}{\partial \mathbf{n}}-U\frac{\partial G}{\partial \mathbf{n}}\right)\mathrm{d}s=0,$$

или

$$-\iint\limits_{S_\varepsilon}\left(G\frac{\partial U}{\partial \mathbf{n}}-U\frac{\partial G}{\partial \mathbf{n}}\right)\mathrm{d}s=\iint\limits_{S}\left(G\frac{\partial U}{\partial \mathbf{n}}-U\frac{\partial G}{\partial \mathbf{n}}\right)\mathrm{d}s. \tag{4.4}$$

Для дальнейшего рассмотрения необходимо записать явное выражение для производной по нормали от функции Грина $G(P)$:

$$\frac{\partial G}{\partial \mathbf{n}}=\cos(\mathbf{n}\mathbf{r}_{01})\left(jk-\frac{1}{r_{01}}\right)\frac{\exp(jkr_{01})}{r_{01}}, \tag{4.5}$$

где $\cos(\mathbf{n}\mathbf{r}_{01})$ — косинус угла между направлением внешней нормали $\mathbf{n}$ и вектором $\mathbf{r}_{01}$, соединяющим точки P_0 и P_1. В частном случае, если последняя точка лежит на S_ε, то $\cos(\mathbf{n}\mathbf{r}_{01}) = -1$, тогда (4.2) и (4.5) принимают вид

$$G(P_1) = \frac{\exp(jk\varepsilon)}{\varepsilon}; \quad \frac{\partial G(P_1)}{\partial \mathbf{n}} = \frac{\exp(jk\varepsilon)}{\varepsilon}\left(\frac{1}{\varepsilon} - jk\right). \tag{4.6}$$

Если $\varepsilon \to 0$, то в силу непрерывности $U(P)$ и ее производной в точке P_0, можно записать:

$$\iint_{S_\varepsilon}\left(G\frac{\partial U}{\partial \mathbf{n}} - U\frac{\partial G}{\partial \mathbf{n}}\right)\mathrm{d}s = \tag{4.7}$$

$$= 4\pi\varepsilon^2\left[\frac{\partial G(P_0)}{\partial \mathbf{n}}\frac{\exp(jk\varepsilon)}{\varepsilon} - U(P_0)\frac{\exp(jk\varepsilon)}{\varepsilon}\left(\frac{1}{\varepsilon} - jk\right)\right]_{\varepsilon\to 0} = -4\pi U(P_0).$$

При выводе этой формулы учитывалось, что при стремлении ε к нулю площадь сферы S_ε уменьшается, и функцию $U(P)$ можно считать постоянной и равной ее значению в точке P_0.

Подставив (4.7) в (4.4), можно получить решение интегральной теоремы Гельмгольца—Кирхгофа в виде

$$U(P_0) = \frac{1}{4\pi}\iint_S\left\{\frac{\partial U}{\partial \mathbf{n}}\frac{\exp(jkr_{01})}{r_{01}} - U\frac{\partial}{\partial \mathbf{n}}\left[\frac{\exp(jkr_{01})}{r_{01}}\right]\right\}\mathrm{d}s. \tag{4.8}$$

4.1.2. Дифракция на плоском экране

Рассмотрим теперь дифракцию электромагнитной волны на плоском, тонком, бесконечном, непроводящем экране с отверстием, воспользовавшись теоремой, доказанной в предыдущем параграфе (рис. 4.3). Волна падает на экран слева направо, и необходимо найти возмущение за экраном в точке P_0.

Воспользуемся интегральной теоремой и выберем в качестве поверхности интегрирования сферу радиусом R с центром в точке P_0. Там, где сфера касается экрана, она обрезается его поверхностью. Если обозначить часть плоскости, усекающую сферу, как S_1, а оставшуюся часть сферы - как S_2, то при использовании (4.8) необходимо интегрировать по составной поверхности $S = S_1+S_2$:

$$U(P_0) = \iint_{S=S_1+S_2}\left(\frac{\partial U}{\partial \mathbf{n}}G - U\frac{\partial G}{\partial \mathbf{n}}\right)\mathrm{d}s. \tag{4.9}$$

В этой формуле функция G имеет вид, определяемый (4.2).

Увеличиваем радиус сферы. При этом поверхность интегрирования все более и более становится похожей на полусферу. Функции U и G с ростом радиуса сферы уменьшаются пропорционально $1/R$, поэтому можно предположить, что подынтегральное выражение стремится к нулю. В этом случае интегрирование по поверхности полусферы S_2 должно было бы давать нулевой вклад в интеграл (4.9). Однако это неочевидно, так как площадь поверхности полусферы растет как R^2 и неясно, какой же из факторов является определяющим. Необходим дополнительный анализ выражения (4.9).

На поверхности S_2 функция

$$G = \frac{\exp(jkR)}{R}. \tag{4.10}$$

Из выражения (4.6) следует, что производная функции G

$$\frac{\partial G}{\partial \mathbf{n}} = \left(jk - \frac{1}{R}\right)\frac{\exp(jkR)}{R} \approx jkG, \text{ при } R \to \infty. \tag{4.11}$$

Преобразуем (4.9), перейдя от интегрирования по поверхности полусферы к интегрированию по телесному углу, который на нее опирается,

$$\iint\limits_{S_1}\left[G\frac{\partial U}{\partial \mathbf{n}} - U(jkG)\right]\mathrm{d}s = \int\limits_{\Omega} G\left(\frac{\partial U}{\partial \mathbf{n}} - jkU\right)R^2\mathrm{d}\omega, \tag{4.12}$$

где Ω — телесный угол с вершиной в P_0, стягиваемый поверхностью S_2.

Величина $|RG|$ равномерно ограничена на S_2, поэтому полный интеграл будет стремиться к нулю при радиусе сферы, стремящемся к бесконечности, в том случае, если выполняется условие

$$\lim_{R\to\infty} R\left(\frac{\partial U}{\partial \mathbf{n}} - jkU\right) = 0 \tag{4.13}$$

равномерно во всем телесном угле.

Условие (4.13) носит название *требования Зоммерфельда для излучения.* Оно удовлетворяется, если функция U стремится к нулю со скоростью, по меньшей мере равной той скорости, с которой расходится сферическая волна. Так как волна, падающая на отверстие, всегда есть сферическая волна или набор таких волн, то можно считать, что интеграл по полусфере действительно стремится к нулю.

Таким образом, интеграл (4.9) может быть записан в виде

$$U(P_0) = \frac{1}{4\pi}\iint\limits_{S_1}\left[\frac{\partial U}{\partial \mathbf{n}}G - U\frac{\partial G}{\partial \mathbf{n}}\right]\mathrm{d}s. \tag{4.14}$$

Для того чтобы получить окончательную формулу дифракции на плоском экране, необходимо определить граничные условия. Одновременно вспомним, что экран непрозрачен всюду, кроме отверстия Σ. Граничные условия были определены Кирхгофом и носят его имя:

1) в отверстии Σ (см. рис. 4.3) амплитуда поля U и ее производные такие же, как если бы экрана не было вообще;

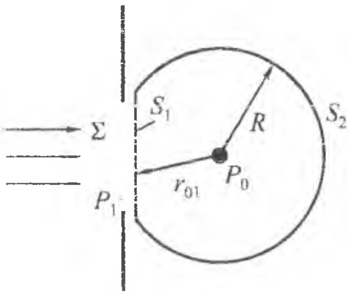

Рис. 4.3. Дифракция на плоском экране

2) на поверхности S_1 вне отверстия амплитуда поля U и ее производные равны нулю.

С учетом граничных условий Кирхгофа (4.14) имеет вид

$$U(P_0) = \frac{1}{4\pi}\iint\limits_{\Sigma}\left[\frac{\partial U}{\partial \mathbf{n}}G - U\frac{\partial G}{\partial \mathbf{n}}\right]\mathrm{d}s. \tag{4.15}$$

Результат достаточно простой и понятный, так как, конечно, основной вклад в поле за экраном будет давать участок отверстия. Однако это не всегда справедливо. Присутствие экрана будет неизбежно вызывать возмущение поля в отверстии, кроме того, тень за экраном никогда не будет резкой, так как поле проникает за экран на расстояние в несколько длин волн. Но если отверстие велико, а длина волны достаточно мала, то оба условия Кирхгофа справедливы, а результаты, рассчитанные с помощью (4.15), хорошо согласуются с данными эксперимента.

4.1.3. Формула дифракции Френеля—Кирхгофа

Дифракционные эффекты слабо сказываются на больших расстояниях от препятствия. Поэтому рассмотрим вначале случай, когда расстояние r_{01} много больше длины волны, т.е. $k >> 1/r_{01}$. В этом случае выражение для производной от функции Грина (4.6) может быть преобразовано к виду

$$\begin{aligned}\frac{\partial G(P_1)}{\partial \mathbf{n}} &= \cos(\mathbf{n},\mathbf{r}_{01})\left(jk - \frac{1}{r_{01}}\right)\frac{\exp(jkr_{01})}{r_{01}} \approx \\ &\approx jk\cos(\mathbf{n},\mathbf{r}_{01})\frac{\exp(jkr_{01})}{r_{01}},\end{aligned} \tag{4.16}$$

где $\cos(\mathbf{n}, \mathbf{r}_{01})$ — косинус угла между вектором $\mathbf{r}_{01}$ и вектором внешней нормали к поверхности интегрирования, т.е. нормали к экрану. Подставим (4.16) и явный вид функции G в (4.15).

$$U(P_0)=\frac{1}{4\pi}\iint_{\Sigma}\frac{\exp(jkr_{01})}{r_{01}}\left[\frac{\partial U}{\partial \mathbf{n}}-jkU\cos(\mathbf{n},\mathbf{r}_{01})\right]\mathrm{d}s. \qquad (4.17)$$

Теперь предположим, что экран с отверстием освещается сферической волной, исходящей из точечного источника, расположенного слева от экрана в точке P_2 (рис. 4.4). Расстояние от источника до экрана равно r_{21}. Тогда поле в отверстии определяется следующим выражением:

$$U(P_1)=\frac{A\exp(jkr_{21})}{r_{21}}, \qquad (4.18)$$

где A — коэффициент пропорциональности.

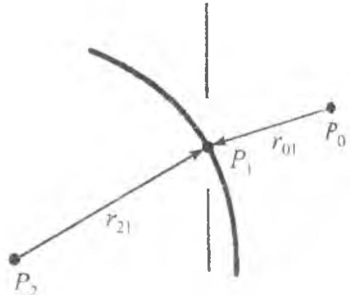

Рис. 4.4 Дифракция сферической волны на плоском экране

Подставив это выражение в (4.17), получим с учетом того, что $r_{21} >> \lambda$

$$U(P_0)=\frac{A}{j\lambda}\iint_{\Sigma}\frac{\exp[jk(r_{21}+r_{01})]}{r_{21}r_{01}}\left[\frac{\cos(\mathbf{n},\mathbf{r}_{01})-\cos(\mathbf{n},\mathbf{r}_{21})}{2}\right]\mathrm{d}s. \qquad (4.19)$$

Этот результат, справедливый только при освещении экрана точечным источником, называется *формулой дифракции Френеля—Кирхгофа*. Отметим, что этот интеграл уже полностью определен, в нем нет неизвестных функций, а те, что есть, легко вычисляются.

Еще одно замечание. Это выражение симметрично по отношению к отражению: если поменять местами точки P_0 и P_2, то результат не изменится. Принцип такой взаимности был сформулирован Гельмгольцем.

Можно интерпретировать формулу дифракции Френеля—Кирхгофа и несколько иным образом, если переписать выражение (4.19):

$$U(P_0)=\iint_{\Sigma}U'(P_1)\frac{\exp(jkr_{01})}{r_{01}}\mathrm{d}s,$$

где

$$U'(P_1) = \frac{1}{j\lambda}\left[\frac{A\exp(jkr_{21})}{r_{21}}\right]\left[\frac{\cos(\mathbf{n},\mathbf{r}_{01}) - \cos(\mathbf{n},\mathbf{r}_{21})}{2}\right].$$

Если исходить из этого выражения, то можно считать, что поле в точке P_0 создается бесконечным множеством сферических волн, исходящих из точечных источников, находящихся в пределах отверстия и имеющих амплитуду $U'(P_1)$. Амплитуда каждого вторичного источника пропорциональна амплитуде падающей волны, но отличается от нее следующим: во-первых, есть множитель $1/\lambda$, во-вторых, происходит уменьшение амплитуды за счет коэффициента наклона, который никогда не превышает единицы и всегда больше нуля, в-третьих, фаза излучаемой волны отличается от фазы падающей волны на $\pi/2$. Этот факт и соответствует предположению Френеля о фазах источников. Таким образом, последняя интерпретация формулы Френеля—Кирхгофа математически описывает принцип Гюйгенса—Френеля для случая дифракции на плоском экране.

4.1.4. Формулировка Зоммерфельда задачи дифракции на плоском экране

Рассмотренная выше теория Кирхгофа дает очень точные результаты, однако она внутренне противоречива. Противоречия кроются в граничных условиях, которые были приняты для поля в плоскости экрана. Дело в том, что условия отсутствия поля и его производной по нормали за экраном вне отверстия приводят к невозможности существования поля вообще с математической точки зрения. Существует известная теорема теории потенциала, которая гласит, что если сама функция и ее производная тождественно равны нулю на каком-то участке поверхности, то эта функция тождественно равна нулю во всем пространстве. Физически это тоже понятно: если в крайней точке этого участка поверхности сама функция и ее производная равны нулю, то почему что-то должно появиться в соседней точке за краем? С другой точки зрения, если учесть волновую природу света, совершенно очевидно, что край отверстия не может не возмутить поле в отверстии, да и край тени не может быть абсолютно резким.

Зоммерфельд предложил обойти это противоречие, изменив граничные условия и налагая требования только на саму функцию или ее производную. При этом все выводы и результаты должны остаться неизменными. С этой целью он видоизменил функцию Грина, сделав ее либо симметричной, либо антисимметричной:

$$G\_ = \frac{\exp(jkr_{01})}{r_{01}} - \frac{\exp(jkr'_{01})}{r'_{01}} \qquad (4.20)$$

Пояснить структуру этой функции можно с помощью рис. 4.5.

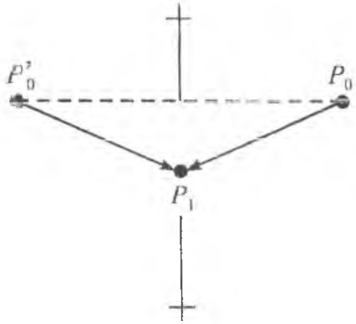

Рис. 4.5. Симметризованная функция

Симметрично точке P_0 относительно экрана расположена точка P'_0, т.е. имеются два точечных источника, формирующих функцию Грина, причем фазы волн, испускаемые этими источниками, одинаковы. Производная новой функции Грина будет сдвинута на 180°:

$$\begin{aligned}\frac{\partial G_-}{\partial \mathbf{n}} &= \cos(\mathbf{n},\mathbf{r}_{01})\left(jk-\frac{1}{r_{01}}\right)\frac{\exp(jkr_{01})}{r_{01}}-\\&-\cos(\mathbf{n},\mathbf{r}'_{01})\left(jk-\frac{1}{r'_{01}}\right)\frac{\exp(jkr'_{01})}{r'_{01}}.\end{aligned} \tag{4.21}$$

Здесь r'_{01} — расстояние от второго источника до поверхности экрана. Если точка P_1 находится на поверхности экрана, то $\mathbf{r}_{01} = \mathbf{r}'_{01}$, $\cos(\mathbf{n}, \mathbf{r}_{01}) = -\cos(\mathbf{n}, \mathbf{r}'_{01})$ и

$$\left.\begin{aligned}&G_-(P_1)=0;\\&\frac{\partial G_-(P_1)}{\partial \mathbf{n}} = 2\cos(\mathbf{n},\mathbf{r}_{01})\left(jk-\frac{1}{r_{01}}\right)\frac{\exp(jkr_{01})}{r_{01}}.\end{aligned}\right\} \tag{4.22}$$

Из (4.22) видно, что отпала необходимость введения жестких граничных условий на функцию и на ее производную одновременно, и имеющиеся противоречия сняты. Следует отметить, что можно было бы ввести симметричную функцию Грина G_+, что привело бы к равенству нулю не функции Грина, а ее производной в плоскости экрана.

Конечно, при использовании видоизмененной функции Грина изменится и формула дифракции Френеля—Кирхгофа, она примет вид

$$U(P_0) = \iint_{\Sigma} \frac{\exp\left[jk(r_{21}+r_{01})\right]}{r_{21}r_{01}}\cos(\mathbf{n},\mathbf{r}_{01})\,ds. \tag{4.23}$$

Формула (4.23) отличается от прежней формулы Френеля—Кирхгофа (4.19) только коэффициентом наклона и называется формулой Зоммерфельда.

До сих пор при анализе дифракции на плоском экране рассматривался вариант, когда он освещается одним точечным источником. В общем случае освещение экрана осуществляется протяженным источником. Однако этот протяженный источник может быть представлен в виде набора точечных источников, каждый из которых формирует свою собственную картину дифракции, которые складываются на выходе, при необходимости с учетом фазы. При таком рассмотрении можно представить (4.21) в виде интеграла суперпозиции

$$U(P_0) = \iint\limits_{\Sigma} h(P_0, P_1) U(P_1) \mathrm{d}s, \qquad (4.24)$$

где

$$h(P_0, P_1) = \frac{1}{j\lambda} \frac{\exp(jkr_{01})}{r_{01}} \cos(\mathbf{n}, \mathbf{r}_{01}).$$

Подобное представление является достаточно наглядным и будет использовано далее. Следует помнить, что интеграл суперпозиции работает только для линейных систем, что справедливо для электромагнитных полей и линейных сред.

До сих пор рассматривались некие точечные источники, на которые разбивались светящиеся поверхности. Однако это только математическая модель, помогающая рассмотреть дифракционные явления. То же самое можно сделать и альтернативным способом. В 1802 г. Юнг предложил формулировку теории дифракции, основывающуюся на интегральном взаимодействии поля с экраном. По его модели картина дифракции формируется как суперпозиция волны, прошедшей через отверстие в экране без взаимодействия с ним, и волны, которая взаимодействовала с краем отверстия. В силу того, что источником дифракции является физический объект — край отверстия, данный подход является очень корректным. Строгая теория дифракции подтверждает это предположение. Однако выводы итоговых формул в этом случае довольно громоздки, хотя, естественно, результат будет тем же самым.

4.2. УГЛОВОЙ СПЕКТР ПЛОСКИХ ВОЛН

Активно применяемые в теории информации, радиотехнике и других областях современной науки и техники преобразования Фурье всюду позволяют построить понятные модели явлений, облегчить синтез сложных систем. В оптике используется преимущественно двумерная форма интегралов Фурье с преобразованием двух поперечных пространственных координат.

Если анализируемая волна имеет вид $U(x, y, z)$, то преобразование Фурье (часто говорят Фурье-образ) может быть записано следующим образом:

$$A_0\left(f_x, f_y, z\right) = \iint_{\infty} U\left(x, y, z\right)\exp\left[-j2\pi\left(f_x x + f_y y\right)\right]\mathrm{d}x\mathrm{d}y \qquad (4.25)$$

и соответственно обратное преобразование Фурье будет

$$U\left(x, y, z\right) = \iint_{\infty} A_0\left(f_x, f_y, z\right)\exp\left[j2\pi\left(f_x x + f_y y\right)\right]\mathrm{d}f_x\,\mathrm{d}f_y. \qquad (4.26)$$

Отметим, что в используемой форме записи прямое и обратное преобразования Фурье отличаются только знаком показателя экспоненты.

Предположим, что волна, созданная некоторой произвольной системой монохроматических источников и распространяющаяся вдоль оси z, достигает плоскости xy ($z = 0$). Комплексное поле в этой плоскости описывается функцией $U(x, y, 0)$. Нашей целью является расчет поля $U(x, y, z)$ в точке (x, y, z).

Проанализируем (4.26). С этой целью запишем математическое выражение для плоской волны, распространяющейся в пространстве в направлении (нормаль к волновому фронту), определяемом направляющими косинусами (α, β, γ):

$$B\left(x, y, z\right) = \exp\left[j\frac{2\pi}{\lambda}\left(\alpha x + \beta y + \gamma z\right)\right], \qquad (4.27)$$

где

$$\gamma = \sqrt{1 - \alpha^2 - \beta^2}.$$

Поэтому в плоскости $z = 0$ экспоненциальную функцию (стоящую под интегралом в выражении (4.24)) можно рассматривать как плоскую волну с направляющими косинусами

$$\left.\begin{array}{l} \alpha = \lambda f_x; \\ \beta = \lambda f_y; \\ \gamma = \sqrt{1 - \left(\lambda f_x\right)^2 - \left(\lambda f_y\right)^2}. \end{array}\right\} \qquad (4.28)$$

При такой записи (4.26) представляет собой суперпозицию плоских волн с комплексными амплитудами вида $A_0(f_x, f_y)\,\mathrm{d}f_x\,\mathrm{d}f_y$ (причем $f_x = \alpha/\lambda$, $f_y = \beta\lambda$), каждая из которых распространяется по своему направлению. Иными словами, любое световое распределение в произвольной плоскости может быть представлено в виде набора плоских волн, каждая из которых распространяется по своему направлению со своей амплитудой.

В новых переменных угловой спектр плоских волн, составляющих световое распределение в плоскости $z = 0$, записывается в виде

$$A_0\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda}\right)=\iint_{\infty} U(x,y,0)\exp\left[-j2\pi\left(\frac{\alpha}{\lambda}x+\frac{\beta}{\lambda}y\right)\right]\mathrm{d}x\,\mathrm{d}y. \tag{4.29}$$

Теперь необходимо переместиться в плоскость $z\neq 0$ и рассмотреть связь между спектром плоских волн в начале координат и в этой плоскости:

$$A_0\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda},z\right)=\iint_{\infty} U(x,y,z)\exp\left[-j2\pi\left(\frac{\alpha}{\lambda}x+\frac{\beta}{\lambda}y\right)\right]\mathrm{d}x\,\mathrm{d}y. \tag{4.30}$$

Запишем обратное преобразование Фурье

$$U(x,y,z)=\iint_{\infty} A\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda},z\right)\exp\left[j2\pi\left(\frac{\alpha}{\lambda}x+\frac{\beta}{\lambda}y\right)\right]\mathrm{d}\frac{\alpha}{\lambda}\,\mathrm{d}\frac{\beta}{\lambda}. \tag{4.31}$$

и воспользуемся уравнением Гельмгольца $\nabla^2 U + k^2 U = 0$ (поскольку в рассматриваемой области пространственных источников света нет), подставив в него (4.31):

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}A\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda},z\right)+\left(\frac{2\pi}{\lambda}\right)^2\left[1-\alpha^2-\beta^2\right]A\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda},z\right)=0. \tag{4.32}$$

Частное решение этого уравнения можно записать в виде

$$A\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda},z\right)=A_0\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda}\right)\exp\left(j\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}\,z\right).$$

Если $\alpha^2+\beta^2\leq 1$, то распространение плоской волны на расстояние z вдоль оси проявляется только в изменении ее фазы относительно других плоских волн, идущих под другими углами. Это понятно, так как они проходят различные пути между плоскостями $z = 0$ и $z \neq 0$. Если же $\alpha^2+\beta^2 > 1$, то волны сильно затухают.

Использование разложения в спектр плоских волн весьма широко распространено в оптике при анализе прохождения света через сложные системы. Как правило, проанализировать прохождение одной плоской волны через них достаточно просто. Поэтому на входе системы разлагают начальное световое распределение в спектр плоских волн, затем находят решение для произвольной плоской волны, а результирующее световое распределение на выходе получается как сумма тех же плоских волн, но с измененными фазами.

4.3. ДИФРАКЦИЯ ФРЕНЕЛЯ И ФРАУНГОФЕРА

Если провести мысленный эксперимент, поместив за экраном с отверстием лист бумаги, то, отодвигая его от экрана, можно наблюдать развитие дифракционных процессов. При близком расположении листа к экрану увидим достаточно четкое изображение отверстия, которое будет постепенно размываться при увеличении расстояния от экрана до листа.

В этой связи есть смысл более внимательно рассмотреть те области, где влияние дифракции уже достаточно велико. Эти области и называются областью дифракции Френеля (дифракция в ближней зоне) и областью дифракции Фраунгофера (дифракция в дальней зоне). На рис. 4.6 показана схема формирования изображения в дифракционной области.

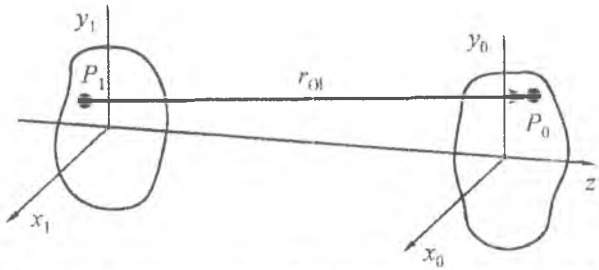

Рис. 4.6. Схема формирования изображения в дифракционной области

Возьмем формулу Зоммерфельда в виде (4.24), заменив в интеграле конечные пределы на бесконечные, считая, что вне отверстия функция $U(x_1, y_1) \equiv 0$:

$$U(x_0, y_0) = \iint\limits_{\infty} h(x_0, y_0, x_1, y_1) U(x_1, y_1) \mathrm{d}x_1 \mathrm{d}y_1,$$

где

$$h(x_0, y_0, x_1, y_1) = \frac{1}{j\lambda} \frac{\exp(jkr_{01})}{r_{01}} \cos(\mathbf{n}, \mathbf{r}_{01}).$$

Наше приближение будет основываться на предположении, что $z >> r_{max}$ (максимального размера отверстия в экране). Область наблюдения в выходной плоскости также лежит вблизи оси z и много меньше, чем z. В этих предположениях $\cos(\mathbf{n}, \mathbf{r}_{01}) = 1$ с погрешностью не более 5 % (если угол не превышает 18°), в знаменателе (4.24) r_{01} тоже будет лишь незначительно отличаться от z. Тогда можно переписать вторую часть (4.24) в виде

$$h(x_0, y_0, x_1, y_1) = \frac{1}{j\lambda z} \exp(jkr_{01}). \tag{4.34}$$

Расстояние r_{01} в показателе экспоненты нельзя просто заменить на z, так как вследствие того, что k велико, даже малое изменение r_{01} приведет к значительному изменению экспоненты.

4.3.1. Приближение Френеля

Можно и далее упростить анализ, если заменить точное выражение для r_{01}:

$$r_{01}=\sqrt{z^2+(x_0-x_1)^2+(y_0-y_1)^2}=z\sqrt{1+\left(\frac{x_0-x_1}{z}\right)^2+\left(\frac{y_0-y_1}{z}\right)^2}$$

его разложением в ряд.

В разложении оставим только два первых члена, считая, что величины, стоящие в скобках, достаточно малы,

$$r_{01}=z\left[1+\frac{1}{2}\left(\frac{x_0-x_1}{z}\right)^2+\frac{1}{2}\left(\frac{y_0-y_1}{z}\right)^2\right]. \tag{4.35}$$

При этом функция h принимает следующий вид:

$$h(x_0,y_0,x_1,y_1)=\frac{\exp(jkz)}{j\lambda z}\exp\left\{j\frac{k}{2z}\left[(x_0-x_1)^2+(y_0-y_1)^2\right]\right\}. \tag{4.36}$$

В том случае, если z достаточно велико, чтобы выражение (4.36) было точным, решение соответствует приближению Френеля.

В приближении Френеля сферические волны заменяются поверхностями 2-го порядка. При этом естественным образом налагаются ограничения на z, размеры отверстия Σ и т.д. Можно провести оценку и показать, что для этого должно выполняться условие

$$z^3>>\frac{\pi}{4\lambda}\left[(x_0-x_1)^2+(y_0-y_1)^2\right]^2. \tag{4.37}$$

Данное условие получается при условии малости следующего члена в разложении (4.35).

Подставим (4.36) в интеграл (4.24):

$$\begin{aligned}U(x_0,y_0)=\frac{\exp(jkz)}{j\lambda z}\iint U(x_1,y_1)\times\\ \times\exp\left\{j\frac{k}{2z}\left[(x_0-x_1)^2+(y_0-y_1)^2\right]\right\}\mathrm{d}x_1\,\mathrm{d}y_1\end{aligned} \tag{4.38}$$

и разложим квадратичные члены в показателе экспоненты

$$\begin{aligned}U(x_0,y_0)=\frac{\exp(jkz)}{j\lambda z}\exp\left[j\frac{k}{2z}(x_0^2+y_0^2)\right]\times\\ \times\iint\left\{U(x_1,y_1)\exp\left[j\frac{k}{2z}(x_1^2+y_1^2)\right]\right\}\exp\left[-j\frac{2\pi}{\lambda z}(x_0x_1+y_0y_1)\right]\mathrm{d}x_1\,\mathrm{d}y_1.\end{aligned} \tag{4.39}$$

Интеграл в (4.39) описывает распределение в ближней зоне дифракции, или зоне дифракции Френеля. Если проанализировать структуру выражения (4.39), то с точностью до амплитудного и фазового множителей, не зависящих от координат выходной плоскости и стоящих перед интегралом, распределение поля в выходной плоскости может быть представлено как преобразование Фурье от функции $U(x_1,y_1)\exp\left[j\frac{k}{2z}\left(x_1^2+y_1^2\right)\right]$ для частот $f_x = x_0/\lambda z$, $f_y = y_0/\lambda z$.

Дополнительные результаты могут быть получены при рассмотрении не самого поля, а его спектра. Для этого проведем преобразование Фурье от функции h (4.36):

$$H\left(f_x,f_y\right)=\exp(jkz)\exp\left[-j\pi\lambda z\left(f_x^2+f_y^2\right)\right]. \tag{4.40}$$

Это выражение описывает эффект распространения волны в пространстве при дифракции Френеля. Первый экспоненциальный множитель определяет общую фазовую задержку при распространении на расстояние z, а второй — фазовую дисперсию, зависящую от пространственной частоты по квадратичному закону.

4.3.2. Приближение Фраунгофера

Можно еще более ужесточить требования к расстоянию r_{01} до выходной плоскости, отодвинувшись в дальнюю зону дифракции — зону дифракции Фраунгофера. Эта область удовлетворяет следующему условию:

$$z \gg \frac{k\left(x_1^2+y_1^2\right)_{\max}}{2}. \tag{4.41}$$

Условие (4.41) определяет параметр малости при разложении величины r_{01} в ряд. При этом квадратичный фазовый множитель в фигурных скобках (4.39) практически равен единице по всему отверстию и амплитуда поля в выходной плоскости имеет вид

$$U\left(x_0,y_0\right)=\frac{\exp(jkz)\exp\left[j\frac{k}{2z}\left(x_0^2+y_0^2\right)\right]}{j\lambda z}\times$$
$$\times\iint\limits_{\infty} U\left(x_1,y_1\right)\exp\left[-j\frac{2\pi}{\lambda z}\left(x_0x_1+y_0y_1\right)\right]dx_1\,dy_1. \tag{4.42}$$

Без учета множителей перед интегралом, которые тем более несущественны при вычислении интенсивности в изображении (при умножении на комплексно-сопряженное выражение), (4.42) представляет собой преобразование Фурье распределения поля в отверстии для пространственных частот $f_x = x_0/\lambda z$ и $f_y = y_0/\lambda z$.

Полученная формула и есть формула дифракции в приближении Фраунгофера. Следует отметить, что в обычных условиях требования по дальности области наблюдения от экрана достаточно жесткие. Так, для отверстия размером 2,5 мм при длине волны света 0,6 мкм расстояние до экрана должно быть более 15 м.

4.3.3. Примеры дифракционных картин Фраунгофера

Рассмотрим ряд примеров дифракционных картин Фраунгофера.

Полученную формулу (4.42) можно непосредственно использовать для расчетов комплексных полей в дальней зоне. Однако надо помнить, что любой детектор, в том числе человеческий глаз, является квадратичным, т. е. чувствует интенсивность, а не амплитуду, поэтому при регистрации дифракционной картины фаза исчезает.

Прямоугольное отверстие. Одной из наиболее важных дифракционных картин, из которых, как из кирпичиков, часто строятся сложные изображения, является дифракционная картина от прямоугольного отверстия размерами w_x, w_y.

Данный объект можно описать как экран, коэффициент пропускания которого $t(x, y)$ равен нулю вне отверстия и единице внутри него:

$$t(x_1, y_1) = \mathrm{rect}(x_1/w_x)\,\mathrm{rect}(y_1/w_y). \qquad (4.43)$$

Пусть отверстие освещается слева плоской монохроматической волной единичной амплитуды, поэтому распределение поля в отверстии описывается коэффициентом пропускания t.

Воспользуемся (4.42) и получим выражение для дифракционной картины

$$U(x_0, y_0) = \frac{\exp(jkz)\exp\left[j\dfrac{k}{2z}\left(x_0^2 + y_0^2\right)\right]}{j\lambda z} F\{U(x_1, y_1)\};\ f_x = \frac{x_0}{\lambda z};\ f_y = \frac{y_0}{\lambda z}.$$

Здесь буквой F обозначен интеграл преобразования Фурье. Учитывая, что

$$F\{U(x_1, y_1)\} = w_x w_y \,\mathrm{sinc}(w_x f_x)\,\mathrm{sinc}(w_y f_y),$$

находим

$$U(x_0, y_0) = \frac{\exp(jkz)\exp\left[j\dfrac{k}{2z}\left(x_0^2 + y_0^2\right)\right]}{j\lambda z} w_x w_y \,\mathrm{sinc}\left(\frac{w_x x_0}{\lambda z}\right)\mathrm{sinc}\left(\frac{w_y y_0}{\lambda z}\right). \qquad (4.44)$$

Выражение (4.44) описывает распределение поля по амплитуде в дальней зоне дифракции. По интенсивности, а именно ее мы и наблюдаем, получим следующее выражение для дифракционной картины:

$$I(x_0, y_0) = \frac{w_x^2 w_y^2}{\lambda^2 z^2}\,\mathrm{sinc}^2\left(\frac{w_x x_0}{\lambda z}\right)\mathrm{sinc}^2\left(\frac{w_y y_0}{\lambda z}\right). \qquad (4.45)$$

На рис. 4.7 приведена фотография дифракционной картины.

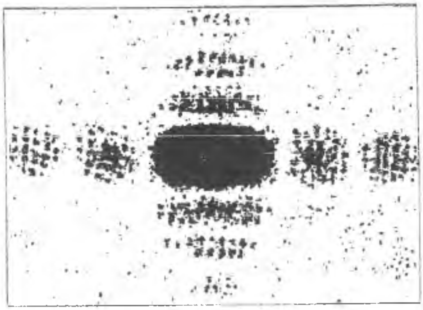

Рис. 4.7. Дифракционная картина от прямоугольного отверстия

Круглое отверстие. Еще одним весьма характерным объектом является круглое отверстие радиусом $w/2$, поскольку соответствующая ему дифракционная картина (картина Эйри) определяет и разрешение в изображении, и минимальное пятно в фокальной плоскости линзы, и многое другое. Если r_1 - радиус-вектор точки в отверстии, то пропускание экрана может быть записано как

$$t(r_1) = \mathrm{circ}\left(\frac{r_1}{w/2}\right).$$

Мы использовали функцию circ(•), которая является аналогом функции rect(•), но в полярных координатах, т.е. она равна нулю, если $r_1 > w/2$, и равна единице, если $r_1 \leq w/2$.

Осевая симметрия этого распределения позволяет заменить преобразование Фурье в формуле (4.42) на преобразование Фурье—Бесселя, обозначаемое $B\{\bullet\}$,

$$U(r_0) = \frac{\exp(jkz)}{j\lambda z}\exp\left(j\frac{kr_0^2}{2z}\right)B\{U(r_0)\},$$

где

$$r_0 = \sqrt{x_0^2 + y_0^2}.$$

В случае освещения экрана нормально падающей плоской волной единичной амплитуды распределение поля в отверстии совпадает с его пропусканием. Учитывая, что

$$B\left\{\mathrm{circ}\left(\frac{r_1}{w/2}\right)\right\} = \left(\frac{w}{2}\right)^2 \frac{J_1(\pi w\rho)}{w\rho/2},$$

где $\rho = r_0/\lambda z$, получаем распределение амплитуды в зоне дифракции Фраунгофера

$$U(r_0)=\exp(jkz)\exp\left(j\frac{kr_0^2}{2z}\right)\frac{kw^2}{j8z}\left[2\frac{J_1(kwr_0/2z)}{kwr_0/2z}\right].$$

Здесь J_1 — функция Бесселя.

Умножив это распределение на комплексно-сопряженное выражение, получим распределение интенсивности

$$I(r_0)=\left(\frac{kw^2}{8z}\right)^2\left[2\frac{J_1(kwr_0/2z)}{kwr_0/2z}\right]^2. \tag{4.46}$$

Вид дифракционной картины Эйри представлен на рис. 4.8. Радиус ее первого темного кольца $\Delta r_0 = 1{,}22\lambda z/w$.

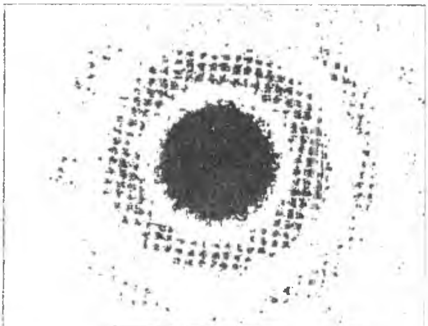

Рис. 4.8. Дифракционная картина от круглого отверстия (картина Эйри)

Синусоидальная амплитудная решетка. Следующий важный случай соответствует синусоидальной амплитудной решетке. Подобный объект интересен как сам по себе, так и в качестве элемента разложения более сложных объектов и изображений.

В отличие от предыдущих примеров объектом является уже не отверстие в экране, которое могло быть описано разного рода ступенчатыми функциями, а более сложная структура с переменным пропусканием. Определим комплексный амплитудный коэффициент пропускания экрана $t(x,y)$ в соответствующих точках как отношение амплитуды поля непосредственно за экраном к амплитуде поля волны, падающей на экран. Координатная зависимость по амплитуде может быть введена диапозитивом, а по фазе — пластинками разной оптической толщины (имеется в виду изменение как реальной толщины, так и показателя преломления).

Возьмем в качестве экрана синусоидальную амплитудную дифракционную решетку.

$$t(x,y)=\left[\frac{1}{2}+\frac{m}{2}\cos(2\pi f_0 x)\right]\mathrm{rect}\left(\frac{x}{w}\right)\mathrm{rect}\left(\frac{y}{w}\right). \quad (4.47)$$

Ее коэффициент пропускания имеет вид (рис. 4.9)

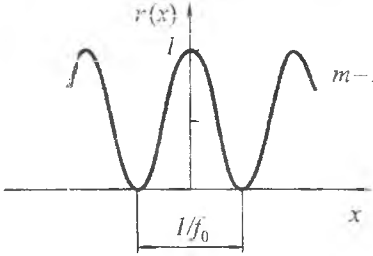

Рис. 4.9. Коэффициент пропускания синусоидальной амплитудной дифракционной решетки вдоль оси x

Для простоты решетка ограничена квадратом со стороной w, о чем и говорят функции rect(•). Первый множитель (4.47) описывает амплитудную синусоидальную модуляцию пропускания внутри квадратного окна. При этом первое слагаемое в нем включено вынужденно, так как не бывает отрицательного пропускания (оно должно меняться от нуля до единицы), а второе определяет пространственные изменения пропускания с глубиной модуляции m, которая меняется от нуля до единицы (f_0 — величина, обратная периоду решетки).

Пусть решетка освещается нормально падающей плоской волной единичной амплитуды. Возьмем преобразование Фурье отдельно для всех сомножителей, входящих в $t(x,y)$, которое, так же как и в предыдущих примерах, должно быть подставлено в интеграл Фраунгофера (4.42) вместо функции $U(x,y)$. Преобразование интеграла свертки есть произведение образов Фурье в нее входящих функций и наоборот:

$$F\left\{\frac{1}{2}+\frac{m}{2}\cos(2\pi f_0 x)\right\}=\frac{1}{2}\delta(f_x,f_y)+$$

$$\frac{m}{4}\delta(f_x+f_0,f_y)+\frac{m}{4}\delta(f_x-f_0,f_y).$$

Здесь также использована теорема смещения преобразования Фурье с разложением косинуса на сумму комплексных экспонент

$$F\left\{\mathrm{rect}\left(\frac{x}{w}\right)\mathrm{rect}\left(\frac{y}{w}\right)\right\}=w^2\,\mathrm{sinc}(wf_x)\,\mathrm{sinc}(wf_y).$$

Подставив все это в дифракционный интеграл для области Фраунгофера, получим распределение поля в ней

$$U(x_0,y_0)=\frac{w^2}{j\lambda z}\exp(jkz)\exp\left[j\frac{k}{2z}\left(x_0^2+y_0^2\right)\right]\mathrm{sinc}\left(\frac{wy_0}{\lambda z}\right)\times$$
$$\times\left\{\mathrm{sinc}\left(\frac{wx_0}{\lambda z}\right)+\frac{m}{2}\mathrm{sinc}\left[\frac{w}{\lambda z}(x_0+f_0\lambda z)\right]+\right. \tag{4.48}$$
$$\left.+\frac{m}{2}\mathrm{sinc}\left[\frac{w}{\lambda z}(x_0-f_0\lambda z)\right]\right\}.$$

Наблюдаемое световое распределение (распределение интенсивности в изображении) получится из (4.48) умножением на комплексно-сопряженное выражение:

$$I(x_0,y_0)=\left(\frac{w^2}{2\lambda z}\right)^2\mathrm{sinc}^2\left(\frac{wy_0}{\lambda z}\right)\left\{\frac{m^2}{4}\mathrm{sinc}^2\left[\frac{w}{\lambda z}(x_0+f_0\lambda z)\right]+\right. \tag{4.49}$$
$$\left.+\frac{m^2}{4}\mathrm{sinc}^2\left[\frac{w}{\lambda z}(x_0-f_0\lambda z)\right]+\mathrm{sinc}^2\left(\frac{wx}{\lambda z}\right)\right\}.$$

Полученное световое распределение (распределение интенсивности) имеет весьма характерный вид (рис. 4.10)

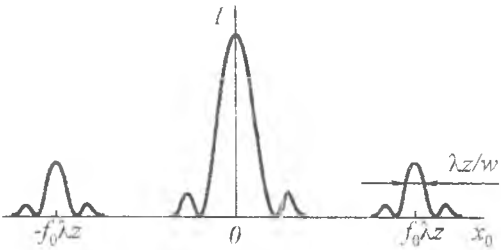

Рис. 4.10. Распределение интенсивности при дифракции на синусоидальной дифракционной решетке

Синусоидальное распределение пропускания по отверстию обуславливает перераспределение энергии проходящего света — некоторая ее доля переходит от центрального максимума к боковым лепесткам. Лепестки находятся на расстоянии $f_0\lambda z$ от центрального максимума (привычный вывод: чем меньше период дифракционной решетки, тем на больший угол отклоняется дифракционный порядок), а ширина лепестков определяется величиной $\lambda z/w$ (чем больше размер решетки, тем более узкими делаются дифракционные максимумы). Объединенный параметр, представляющий собой произведение обратного периода решетки на ее размер, определяет разрешающую способность решетки, т.е. разрешающая способность определяется полным числом штрихов в решетке, а не расстоянием от нее до плоскости наблюдения. Следует также отметить, что исходя из формулы (4.49) при использовании синусоидальной амплитудной дифракционной решетки в 1-й порядок дифракции может попасть не более 25% падающей на решетку энергии.

Проведенный анализ подтверждается экспериментальными результатами, возникает только вопрос о количестве наблюдаемых дифракционных порядков. Обычно их существенно больше трех. Причина данного явления состоит в том, что обычные решетки являются периодическими, но не синусоидальными. Если воспользоваться методом преобразования Фурье для периодических функций, то можно показать, что любая такая функция может быть представлена в виде разложения в гармонический ряд, т.е. несинусоидальная решетка может быть представлена в виде наложения гармонических решеток с кратными периодами. Каждая из них дает свои три порядка дифракции, причем нулевые совпадают, а остальные идут под кратными углами, создавая привычную картину дифракции на решетке.

Синусоидальная фазовая дифракционная решетка. Анализ процесса дифракции на подобном объекте проводится в точности таким же образом, как в предыдущем случае, но с небольшими отличиями.

Функция пропускания имеет вид

$$f(x_1, y_1) = \exp\left[j\frac{m}{2}\sin(2\pi f_0 x_1)\right] rect\left(\frac{x_1}{w}\right) rect\left(\frac{y_1}{w}\right).$$

Она отличается от (4.47) только тем, что синусоидальная модуляция находится в показателе экспоненты.

Воспользуемся известной формулой разложения мнимой гармонической экспоненты в ряд:

$$\exp\left[j\frac{m}{2}\sin(2\pi f_0 x_1)\right] = \sum_{q=-\infty}^{\infty} J_q\left(\frac{m}{2}\right)\exp(j2\pi q f_0 x_1),$$

где J_q — функция Бесселя порядка q, и получим распределение амплитуды поля в зоне дифракции Фраунгофера

$$U(x_0, y_0) = \frac{w^2}{j\lambda z}\exp(jkz)\exp\left[j\frac{k}{2z}(x_0^2 + y_0^2)\right]\times$$
$$\times\sum_{q=-\infty}^{\infty} J_q\left(\frac{m}{2}\right)\mathrm{sinc}\left[\frac{w}{\lambda z}(x_0 - qf_0\lambda z)\right]\mathrm{sinc}\left(\frac{wy_0}{\lambda z}\right). \qquad (4.50)$$

Если считать, что $f_0 >> 2/w$, то перекрытием различных дифракционных порядков можно пренебречь, и тогда распределение интенсивности в области дифракции будет иметь вид

$$I(x_0, y_0) = \left(\frac{w}{\lambda z}\right)^2 \times$$
$$\times \sum_{q=-\infty}^{\infty} J_q^2\left(\frac{m}{2}\right) \operatorname{sinc}^2\left[\frac{w}{\lambda z}\left(x_0 - qf_0\lambda z\right)\right] \operatorname{sinc}^2\left(\frac{wv_0}{\lambda z}\right). \tag{4.51}$$

Естественно, те же угловые зависимости, которые были отмечены для случая амплитудной дифракционной решетки, остались в силе, как для порядков дифракции, так и для их ширины. Существенное отличие заключается в большом числе порядков, а также в возможности перераспределения энергии между ними, в том числе и таким образом, что в нулевой дифракционный порядок попадает меньше энергии, чем в более высокий. Это связано с зависимостью амплитуды дифракционных порядков от бесселевых функций.

Большинство используемых сегодня дифракционных решеток относятся именно к фазовым, так как они работают преимущественно на отражение. Это решетки более сложной конфигурации, что обеспечивает дополнительные возможности спектрального анализа. Все основные принципы их работы уже изложены в этом и предыдущих параграфах.

4.3.4. Некоторые особенности реальных дифракционных решеток

Большинство решеток, которые используются на практике, не являются синусоидальными. В них наблюдается большое число дифракционных порядков. Найдем выражение для зависимости разрешающей силы решетки от номера порядка.

На рис. 4.11 изображена схема щелевой дифракционной решетки. Предположим, что на нее падает свет под углом θ_0 к плоскости решетки. Максимумы интерференционной картины сформируются по направлениям, для которых выполняется условие синфазности прошедших лучей:

$$BL - AK = d\left(\sin\theta - \sin\theta_0\right) = m\lambda, \quad m = 0, \pm1, \pm2, \ldots \tag{4.52}$$

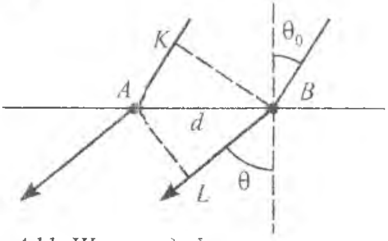

Рис. 4.11. Щелевая дифракционная решетка

В соответствии с принятой терминологией назовем m порядком интерференции. В промежутках между основными максимумами находятся вторичные, которые связаны с дифракцией на краю решетки (4.49), (4.51). Главные максимумы разделяются точками с нулевой интенсивностью, которые определяются соотношением

$$d\left(\sin\theta - \sin\theta_0\right) = \frac{n\lambda}{N}, \quad n = \pm 1, \pm 2 \ldots; \tag{4.53}$$

случай, когда n/N равно целому числу, исключается (N — полное число штрихов в решетке).

Расстояние между главным максимумом порядка m и соседним минимумом определяется в соответствии с (4.52) и (4.53) соотношением

$$\Delta p = \frac{\lambda}{Nd},$$

где $p = \sin\theta - \sin\theta_0$.

Если длина волны меняется на $\Delta\lambda$, максимум порядка m сместится на

$$\Delta' p = \frac{|m|}{d}\Delta\lambda.$$

Предположим, что линии с длинами волн $\lambda \pm \frac{1}{2}\Delta\lambda$ начинают разрешаться, если максимум одной длины волны совпадает с минимумом другой, тогда для предельного разрешения в порядке m получим формулу

$$\frac{\lambda}{\Delta\lambda} = |m|N,$$

т.е. разрешающая сила равна произведению номера порядка на число штрихов в решетке.

Однако возможность повышения спектрального разрешения на высоких порядках решетки может быть реализована только в том случае, если мощность светового потока, попадающего в эти порядки, достаточна для регистрации. В обычных решетках, даже и фазовых, эти два требования совмещаются с трудом. Оптимизация формы штриха решетки позволяет перераспределить свет между порядками с высокой эффективностью. На рис. 4.12 показана отражательная решетка с заданной формой штриха, позволяющая получить так называемый «угол блеска» для нужного порядка («угол блеска» — это угол, под которым концентрируется отраженная энергия).

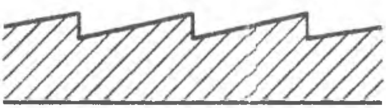

Рис. 4.12. Отражательная дифракционная решетка

В последнее время многие решетки изготавливаются не привычными методами механической обработки поверхности, а голографическими (интерференционными) с последующим изготовлением металлической копии с помощью фотолитографии и травления. Более подробно об этом в последующих главах.

4.4. ЛИНЗА КАК ЭЛЕМЕНТ, ВЫПОЛНЯЮЩИЙ ПРЕОБРАЗОВАНИЕ ФУРЬЕ

До сих пор, занимаясь вопросами дифракции, мы рассматривали ее влияние на формирование изображения достаточно далеко от предмета без учета возможных оптических элементов, участвующих в этом процессе. В этой связи необходимо учесть их возможное влияние (в первую очередь линз) и научиться рассчитывать световые распределения в тех областях, где геометрическая оптика в принципе не работает — в фокальной плоскости линзы и окружающей ее области. Рассмотрим дифракционные эффекты применительно к тонкой линзе. Будем использовать тот же подход, что и при анализе дифракционных картин Фраунгофера, основанный на функции пропускания объекта.

Линза называется тонкой, если луч, входящий в нее в точке с заданными координатами на одной поверхности, выходит из нее в точке с приблизительно такими же координатами на другой поверхности. Таким образом, свет в тонкой линзе проходит оптический путь, пропорциональный толщине линзы в данной точке. Обратимся к рис. 4.13. Проведем две параллельные плоскости, прижатые к линзе с обеих сторон. Луч A проходит через линзу оптический путь $n\Delta_0$ (где n — показатель преломления материала линзы; Δ_0 — ее максимальная толщина), а луч B часть пути между плоскостями проходит в воздухе, а часть — в стекле, поэтому пройденный им оптический путь равен $n\Delta(x, y)$ - (Δ_0 - $\Delta(x, y)$), где $\Delta(x, y)$ — толщина стеклянной части линзы в соответствующей точке.

Этому оптическому пути соответствует изменение фазы световой волны, равное

$$\varphi(x,y) = kn\Delta(x,y) + k\left[\Delta_0 - \Delta(x,y)\right].$$

Тогда коэффициент пропускания линзы как объекта, сформированного двумя параллельными плоскостями,

$$t_L(x,y) = \exp\left(jk\Delta_0\right)\exp\left[jk\left(n-1\right)\Delta(x,y)\right]. \qquad (4.54)$$

При этом мы учли, что линза абсолютно прозрачна, поэтому коэффициент амплитудного пропускания равен единице, а все изменения содержатся в показателе экспоненты.

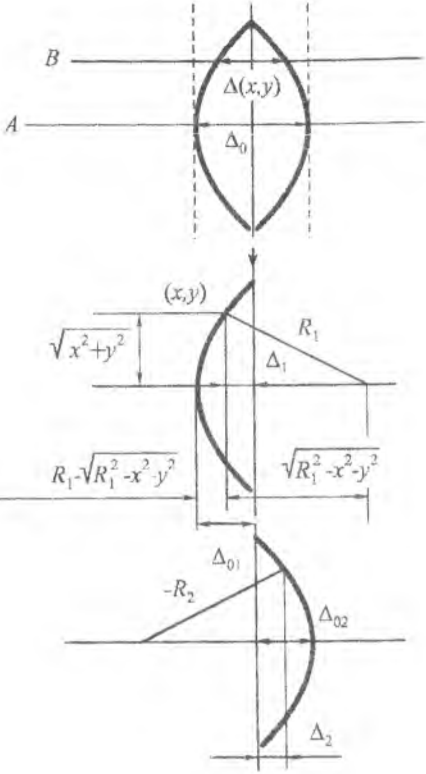

Рис. 4.13. Линза как фазовая задержка

Необходимо определить значение $\Delta(x,y)$. При этом мы будем учитывать правило знаков: радиус кривизны поверхности считается положительным, если эта поверхность выпуклая для луча, идущего слева направо. Разделим линзу на две части (см. рис. 4.13) и проанализируем их по отдельности.

Представим толщину линзы как сумму толщин двух ее половинок $\Delta(x,y) = \Delta_1(x,y)+\Delta_2(x,y)$. Каждая из этих величин может быть выражена через соответствующий радиус кривизны, как это видно из рис. 4.13:

$$\Delta_1(x,y)=\Delta_{01}-\left(R_1-\sqrt{R_1^2-x^2-y^2}\right)=\Delta_{01}-R_1\left(1-\sqrt{1-\frac{x^2+y^2}{R_1^2}}\right),$$

$$\Delta_2(x,y)=\Delta_{02}-\left(-R_2+\sqrt{R_2^2-x^2-y^2}\right)=\Delta_{02}+R_2\left(1-\sqrt{1-\frac{x^2+y^2}{R_2^2}}\right).$$

В уравнениях учитывался знак радиусов кривизны.

Складываем полученные величины и получаем толщину линзы в произвольной точке

$$\Delta(x,y)=\Delta_0-R_1\left(1-\sqrt{1-\frac{x^2+y^2}{R_1^2}}\right)+R_2\left(1-\sqrt{1-\frac{x^2+y^2}{R_2^2}}\right). \qquad (4.55)$$

Будем рассматривать только лучи, идущие под небольшими углами к оси линзы (параксиальное приближение), поэтому в (4.55) корни можно разложить в ряд, оставив только нулевой и первый члены ряда

$$\sqrt{1-\frac{x^2+y^2}{R_1^2}}\approx 1-\frac{x^2+y^2}{2R_1^2}; \qquad \sqrt{1-\frac{x^2+y^2}{R_2^2}}\approx 1-\frac{x^2+y^2}{2R_2^2}.$$

Подставляя это разложение в (4.55), получаем

$$\Delta(x,y)=\Delta_0-\frac{x^2+y^2}{2}\left(\frac{1}{R_1}-\frac{1}{R_2}\right). \qquad (4.56)$$

Теперь преобразуем к явному виду (4.54)

$$t_L(x,y)=\exp(jkn\Delta_0)\exp\left[-jk(n-1)\frac{x^2+y^2}{2}\left(\frac{1}{R_1}-\frac{1}{R_2}\right)\right].$$

Воспользовавшись формулой линзы (3.61), окончательно получим

$$t_L(x,y)=\exp(jkn\Delta_0)\exp\left[-j\frac{k}{2f}\left(x^2+y^2\right)\right].$$

где f - фокусное расстояние.

Посмотрим, как работает линза. Если на нее падает плоский волновой фронт, то она преобразует его в сферический, сходящийся в фокальной плоскости (в представлении геометрической оптики — в точку). Если посмотреть на выражение (4.57) и вспомнить, что при нормальном падении на объект с таким пропусканием плоской волны единичной амплитуды за объектом распределение света будет иметь такой же вид, как (4.57), то станет ясно, что действительно за линзой должна распространяться сферическая волна. Первый же экспоненциальный множитель в этой формуле просто обозначает постоянный фазовый сдвиг. Если фокусное расстояние больше нуля, то мы получаем сходящуюся сферическую волну, если же меньше нуля, — расходящуюся.

Предположение о преобразовании плоской волны в сферическую справедливо только в параксиальном приближении, но в реальной линзе имеющиеся искажения (аберрации) приведут к отклонению от сферичности.

Получив выражение для функции пропускания линзы, необходимо проанализировать распределение поля в фокальной плоскости линзы или рядом с ней.

Рассмотрим модельный эксперимент, схема которого изображена на рис. 4.14. Тонкая линза освещается плоской световой волной через прижатый к ней тонкий объект (транспарант) с пропусканием $t_0(x, y)$. Тогда, если амплитуда плоской волны равна A, на линзу будет падать волна с амплитудой

$$U_L(x,y) = At_0(x,y). \tag{4.58}$$

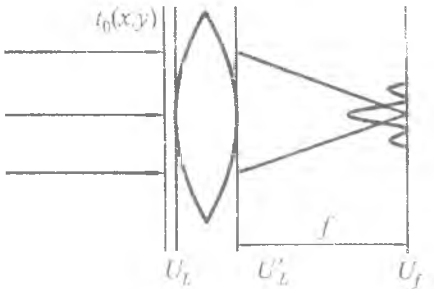

Рис. 4.14. Линза как элемент, осуществляющий преобразование Фурье (транспарант прижат к линзе)

Конечный размер линзы задается с помощью функции зрачка $P(x,y)$, определяющей ее края, которая равна единице внутри апертуры линзы и нулю вне неё. Теперь мы можем записать выражение для распределения амплитуды поля сразу за линзой с учетом функции пропускания линзы и функции зрачка

$$U'_L(x,y) = U_L(x,y)P(x,y)\exp\left[-j\frac{k}{2f}\left(x^2+y^2\right)\right] \tag{4.59}$$

Постоянная фазовая задержка, вносимая линзой, опущена.

Найдем амплитуду поля в фокальной плоскости линзы $U_f(x_f, y_f)$. Для этого воспользуемся формулой дифракции для френелевской области (4.39) с заменой z на f

$$U_f(x_f,y_f) = \frac{\exp\left[j\frac{k}{2f}\left(x_f^2+y_f^2\right)\right]}{j\lambda f}\iint\limits_{\infty} U'_L(x,y)\exp\left[j\frac{k}{2f}\left(x^2+y^2\right)\right]\times \tag{4.60}$$
$$\times\exp\left[-j\frac{2\pi}{\lambda f}\left(xx_f+yy_f\right)\right]\mathrm{d}x\mathrm{d}y.$$

При получении этой формулы все постоянные фазовые множители опущены. Осталось только подставить выражение (4.59) в формулу (4.60) и получить выражение для распределения поля в фокальной плоскости (при этом некоторые квадратичные фазовые множители сокращаются)

$$U_f(x_f, y_f) = \frac{\exp\left[j\frac{k}{2f}\left(x_f^2 + y_f^2\right)\right]}{j\lambda f} \iint_{\infty} U_L(x, y) P(x, y) \times \\ \times \exp\left[-j\frac{2\pi}{\lambda f}\left(xx_f + yy_f\right)\right] \mathrm{d}x\,\mathrm{d}y. \quad (4.61)$$

Таким образом, распределение поля в фокальной плоскости линзы пропорционально двумерному Фурье-образу части исходного поля, вырезанного отверстием линзы. Если размер предмета меньше диаметра линзы, то конечные размеры апертуры можно не учитывать и распределение поля в фокальной плоскости имеет вид

$$U_f(x_f, y_f) = \frac{A\exp\left[j\frac{k}{2f}\left(x_f^2 + y_f^2\right)\right]}{j\lambda f} \times \\ \times \iint_{\infty} t_0(x, y) \exp\left[-j\frac{2\pi}{\lambda f}\left(xx_f + yy_f\right)\right] \mathrm{d}x\,\mathrm{d}y. \quad (4.62)$$

Из формулы (4.62) видно, что распределение поля в фокальной плоскости линзы представляет собой (с точностью до фазового множителя) преобразование Фурье функции, записанной в виде изменения пропускания транспаранта, прижатого к линзе. Очевидно, что амплитуда и фаза волны в точке с координатами (x_f, y_f) определяются амплитудой и фазой в Фурье-образе предмета, имеющими пространственные частоты $f_x = x_f/\lambda_f$, $f_y = y_f/\lambda_f$.

Преобразование Фурье, содержащееся в выражении (4.62), позволяет использовать линзу в качестве простого и очень эффективного спектроанализатора, в том числе в динамическом режиме.

Несколько слов следует сказать о дифракционном пределе, накладывающем ограничение на минимальный размер точки в фокусе линзы. Если обратиться к (4.61) и рассмотреть случай, когда никакого транспаранта, прижатого к линзе, нет, то становится ясно, что все дифракционные эффекты связаны с функцией зрачка. В простейшем случае круглой апертуры в фокальной плоскости образуется картина Эйри. Если считать диаметром пятна диаметр кольца первого ее минимума, то фокальное пятно будет тем меньше, чем меньше длина волны, больше диаметр линзы и меньше фокусное расстояние.

Еще одно замечание, связанное с (4.62), заключается в совпадении с точностью до масштаба дифракционной картины, которая получается в дальней зоне дифракции (область Фраунгофера) и фокальной плоскости линзы.

Возвращаясь к вопросу о регистрации распределения поля в фокальной плоскости, следует напомнить, что регистрируется интенсивность поля, поэтому выражение (4.62) следует умножить на комплексно-сопряженное, при этом фазовые экспоненты скомпенсируются и мы получаем выражение

$$I_f\left(x_f, y_f\right) = \frac{A^2}{\lambda^2 f^2}\left|\iint\limits_{\infty} t_0(x,y)\exp\left[-j\frac{2\pi}{\lambda f}\left(xx_f + yy_f\right)\right]\mathrm{d}x\,\mathrm{d}y\right|^2 \quad (4.63)$$

Предположим, что тонкий объект не прижат к линзе, а находится на некотором расстоянии d_0 от нее (рис.4.15). Подход к анализу прохождения поля через всю систему останется прежним — получаются распределения поля в разных ключевых плоскостях, которые потом связываются дифракционными интегралами. Считаем, что по-прежнему освещающая волна является плоской, предмет и линза тонкими, а лучи, формирующие распределение поля в фокальной плоскости, параксиальными.

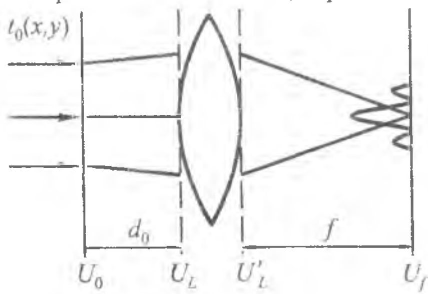

Рис. 4.15. Линза как элемент, осуществляющий преобразование Фурье (транспарант на расстоянии от линзы)

Пусть распределение света после объекта U_0, Фурье-спектр этого распределения F_0; распределение амплитуды света, падающего на линзу U_L, Фурье-спектр этого распределения F_L, а распределение поля в фокальной плоскости U_f.

Считаем, что на расстоянии d_0 справедливо приближение Френеля, а значит, распространение света на это расстояние можно описать с помощью (4.40):

$$F_L\left(f_x, f_y\right) = F_0\left(f_x, f_y\right)\exp\left[-j\pi\lambda d_0\left(f_x^2 + f_y^2\right)\right].$$

Не учитывая конечной апертуры линзы, мы можем теперь записать распределение поля в фокальной плоскости

$$U_f(x_f,y_f)=\frac{\exp\left[j\frac{k}{2f}\left(x_f^2+y_f^2\right)\right]}{j\lambda f}F_L\left(\frac{x_f}{\lambda f},\frac{y_f}{\lambda f}\right).$$

Подставляем сюда предыдущую формулу и получаем распределение амплитуды света в фокальной плоскости

$$U_f(x_f,y_f)=\frac{A\exp\left[j\frac{k}{2f}\left(1-\frac{d_0}{f}\right)\left(x_f^2+y_f^2\right)\right]}{j\lambda f}\times$$
$$\times\iint_{\infty} t_0(x_0,y_0)\exp\left[-j\frac{2\pi}{\lambda f}\left(x_0x_f+y_0y_f\right)\right]\mathrm{d}x_0\,\mathrm{d}y_0. \tag{4.64}$$

Формула (4.64) очень похожа на (4.62), за исключением экспоненты перед интегралом.

Интересен частный случай, когда предмет находится в передней фокальной плоскости линзы ($d_0 = f$). При этом показатель экспоненты, стоящей перед интегралом, обращается в нуль и (4.64) представляет собой просто интеграл преобразования Фурье.

4.5. ФОРМИРОВАНИЕ ИЗОБРАЖЕНИЯ ЛИНЗОЙ В ДИФРАКЦИОННОМ ПРИБЛИЖЕНИИ

Наиболее известное свойство линзы — способность формировать изображение объекта. Этот вопрос достаточно подробно рассмотрен в предыдущей главе, но сейчас необходимо определить, каким образом дифракция влияет на качество этого изображения.

Рассмотрим свободную от аберраций положительную тонкую линзу, освещаемую монохроматическим светом, которая формирует действительное изображение. Это означает, что система (рис. 4.16) линейна относительно комплексной амплитуды поля.

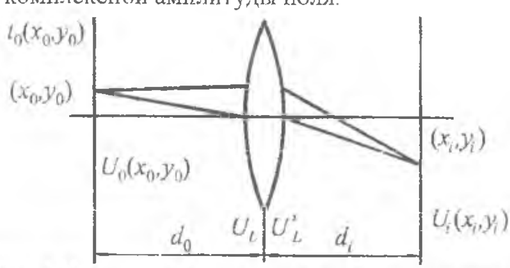

Рис. 4.16. Схема формирования изображения линзой

Комплексное поле сразу за предметом будет иметь вид $U_0(x_0, y_0)$, а на расстоянии d_i за линзой $U_i(x_i, y_i)$. Необходимо определить условия, при которых распределение поля в плоскости за линзой можно с уверенностью назвать изображением предмета.

В силу линейности используемых функций световые распределения можно представить в виде интеграла суперпозиции (4.24):

$$U_i(x_i, y_i) = \iint\limits_{\infty} h(x_i, y_i, x_0, y_0) U_0(x_0, y_0)\, \mathrm{d}x_0\, \mathrm{d}y_0. \tag{4.65}$$

где функция $h(x_i, y_i, x_0, y_0)$ носит название импульсного отклика. Эта функция определяет амплитуду в точке (x_i, y_i), созданную точечным источником единичной амплитуды, расположенным в точке с координатами (x_0, y_0).

Чтобы изображение было высококачественным, необходимо, чтобы поле U_i мало отличалось от U_0. Это значит, что импульсный отклик должен быть очень близок к δ-функции:

$$h(x_i, y_i, x_0, y_0) \approx K\delta(x_i \pm \beta_0 x_0, y_i \pm \beta_0 y_0), \tag{4.66}$$

где K — комплексная постоянная; β_0 — увеличение системы, ± учитывает возможность как прямого, так и перевернутого изображения.

Теперь допустим, что предмет — точечный источник с координатами (x_0, y_0) описываемый δ-функцией. В этом случае на линзу будет падать сферический пучок, расходящийся из этой точки. В параксиальном приближении эту волну можно записать в виде

$$U_L(x, y) = \frac{1}{j\lambda d_0} \exp\left\{ j\frac{k}{2d_0}\left[(x - x_0)^2 + (y - y_0)^2\right]\right\}. \tag{4.67}$$

После прохождения через линзу распределение поля примет вид (с учетом функции зрачка $P(x, y)$)

$$U'_L(x, y) = U_L(x, y) P(x, y) \exp\left[-j\frac{k}{2f}(x^2 + y^2)\right]. \tag{4.68}$$

И, наконец, распределение поля в плоскости за линзой может быть получено с помощью интеграла Френеля (4.39) с учетом того, что это распределение и есть импульсный отклик, так как оно создано точечным источником

$$\begin{aligned} h(x_i, y_i, x_0, y_0) &= \frac{1}{j\lambda d_i} \iint\limits_{\infty} U'_L(x, y) \times \\ &\times \exp\left\{ j\frac{k}{2d_i}\left[(x - x_i)^2 + (y - y_i)^2\right]\right\} \mathrm{d}x\, \mathrm{d}y. \end{aligned} \tag{4.69}$$

Здесь опущен постоянный фазовый множитель.

Объединяя (4.67)—(4.69) в одну формулу, получаем

$$h(x_i, y_i, x_0, y_0) = \frac{1}{\lambda^2 d_0 d_i} \exp\left[j\frac{k}{2d_i}\left(x_i^2 + y_i^2\right)\right] \cdot \exp\left[j\frac{k}{2d_0}\left(x_0^2 + y_0^2\right)\right] \times$$
$$\times \iint_{\infty} \exp\left[j\frac{k}{2}\left(\frac{1}{d_0} + \frac{1}{d_i} - \frac{1}{f}\right)\left(x^2 + y^2\right)\right] \cdot \exp\left\{-jk\left[\left(\frac{x_0}{d_0} + \frac{x_i}{d_i}\right)x + \left(\frac{y_0}{d_0} + \frac{y_i}{d_i}\right)y\right]\right\} \mathrm{d}x\mathrm{d}y \quad (4.70)$$

Полученное выражение (4.70) полностью описывает импульсный отклик в нашей модели, однако оно слишком громоздко и его необходимо упростить. Самыми неприятными компонентами этой формулы являются экспоненты с квадратичными по координатам показателями степени. Заметим, что две из них, стоящие перед интегралом, не зависят от переменных (x,y), они определяют фазовое искривление в плоскостях предмета и изображения.

Первую из этих экспонент, зависящую от (x_i, y_i), можно в расчет не принимать, так как в плоскости изображения, как правило, стоит квадратичный детектор (глаз, фотоприемник и т.д.), который все равно фазу не регистрирует.

К сожалению, от второй экспоненты так просто не избавиться. Вычисляемая функция импульсного отклика входит в подынтегральное выражение (4.65), которое интегрируется по переменным (x_0, y_0). Однако если система достаточно близка к идеальной, а только такая система будет создавать изображение объекта, то ее импульсный отклик должен быть похож на δ-функцию. Это означает, что амплитуда волны в точке (x_i, y_i) будет определяться вкладом только от очень малой области в пространстве предмета с центром в точке, соответствующей идеальному геометрическому изображению. Если внутри этой области аргумент от функции $\exp\left[j\frac{k}{2d_0}\left(x_0^2 + y_0^2\right)\right]$ меняется не более чем на доли радиана, то

$$\exp\left[j\frac{k}{2d_0}\left(x_0^2 + y_0^2\right)\right] \approx \exp\left[j\frac{k}{2d_0} \cdot \frac{\left(x_i^2 + y_i^2\right)}{\beta_0}\right]$$

и теперь его можно опустить, так как здесь нет координат входной плоскости. Еще одно упрощение связано с тем, что система предназначена для формирования изображения. Используя формулу Гаусса с учетом обозначений на рис. 4.16, перепишем (4.70) в следующем виде:

$$h(x_i, y_i, x_0, y_0) = \frac{1}{\lambda^2 d_0 d_i} \iint_{\infty} P(x, y) \cdot \exp\left\{-jk\left[\left(\frac{x_0}{d_0} + \frac{x_i}{d_i}\right)x + \left(\frac{y_0}{d_0} + \frac{y_i}{d_i}\right)y\right]\right\} \mathrm{d}x\mathrm{d}y \quad (4.71)$$

Это соотношение определяет расположенную за линзой точку, в которой пересекаются лучи, выходящие из одной и той же точки предмета.

Для того, чтобы доказать, что полученная функция импульсного отклика соответствует реальности, необходимо, чтобы в предельном переходе $\lambda \to 0$ эта функция превратилась в функцию импульсного отклика геометрической оптики, т.е. в δ-функцию. Покажем это.

Увеличение системы в данном случае равно $\beta_0 = d_i/d_0$. Тогда

$$h(x_i, y_i, x_0, y_0) = \frac{-1}{\lambda^2 d_0 d_i} \iint_\infty P(x, y) \cdot \exp\left\{-j\frac{2\pi}{\lambda d_i}\left[(x_i + \beta_0 x_0)x + (y_i + \beta_0 y_0)y\right]\right\} dx dy \quad (4.72)$$

Фактически это формула дифракции Фраунгофера, что неудивительно, так как при выборе d_i в соответствии с формулой линзы анализируется плоскость, в которой сходится сферическая волна, прошедшая через линзу.

Совершим предельный переход к условиям геометрической оптики. В этом случае дифракционные эффекты должны быть несущественными. Выполним замену переменных

$$\bar{x} = \frac{x}{\lambda d_i}, \quad \bar{y} = \frac{y}{\lambda d_i},$$

тогда

$$h(x_i, y_i, x_0, y_0) = -\beta_0 \iint P(\lambda d_i \bar{x}, \lambda d_i \bar{y}) \times$$
$$\exp\left\{-j2\pi\left[(x_i + \beta_0 x_0)\bar{x} + (y_i + \beta_0 y_0)\bar{y}\right]\right\} d\bar{x}\, d\bar{y}. \quad (4.73)$$

При $\lambda \to 0$ аргументы функции зрачка P стремятся к нулю, а это значит, что $P = 1$, так как по центру апертуры пропускание всегда равно единице:

$$h(x_i, y_i, x_0, y_0) \to -\beta_0 \iint_\infty \exp\left\{-j2\pi\left[(x_i + \beta_0 x_0)\bar{x} + (y_i + \beta_0 y_0)\bar{y}\right]\right\} d\bar{x}\, d\bar{y} =$$
$$= -\beta_0 \delta(x_i + \beta_0 x_0, y_i + \beta_0 y_0) = \frac{-1}{\beta_0}\delta\left(\frac{x_i}{\beta_0} + x_0, \frac{y_i}{\beta_0} + y_0\right) \quad (4.74)$$

Подставляем выражение для импульсного отклика (4.74) в интеграл суперпозиции (4.65) и получаем соотношение между распределениями света в области предмета и изображения

$$U_i(x_i, y_i) = \frac{1}{\beta_0} U_0\left(-\frac{x_i}{\beta_0}, -\frac{y_i}{\beta_0}\right).$$

Таким образом, доказано, что полученная функция импульсного отклика (4.70) правильно описывает процесс формирования изображения.

Анализ формирования изображения линзой показывает, что учет дифракционных эффектов приводит к ухудшению качества изображения за счет сглаживания деталей. Чем сильнее влияние дифракции, тем более крупные детали изображения начинают пропадать. Это связано с отличием функции импульсного отклика от δ-функции. При этом в изображении вместо точек образуются маленькие картины Эйри, которые, накладываясь друг на друга, и определяют разрешение в изображении.

4.6. ПРОСТРАНСТВЕННОЕ РАЗРЕШЕНИЕ В ОПТИКЕ

Только что, при анализе особенностей формирования изображения в оптических системах с учетом дифракции, было показано, что в них происходит снижение разрешающей способности по сравнению с идеальными системами за счет размазывания точек в изображении — превращения их в кружок Эйри. Кроме дифракции на разрешение в изображении влияют аберрации, неточности фокусировки, дополнительные засветки, рассеяние, шумы и др. В этой связи весьма важным как с практической, так и с расчетной точки зрения является вопрос о критерии, которым следует руководствоваться при рассмотрении качества изображения. Очень близкие проблемы возникают и в спектральном анализе, и в голографии, и в системах отображения информации.

Предельное расстояние, на котором глаз еще может обнаружить раздельное существование двух близлежащих точек, зависит во многом от техники наблюдения или регистрации (например, можно увеличить контраст в изображении на фотографии путем дополнительной обработки эмульсии). Однако должен существовать и некоторый объективный критерий, который позволил бы сравнивать различные оптические системы. Таким критерием и является *критерий Релея*. Он основывается на дифракционной модели и определяет, что две близлежащие точки (кружки Эйри) в изображении перестают различаться, когда главный максимум распределения интенсивности одного из них совпадает с первым минимумом другого (рис. 4.17).

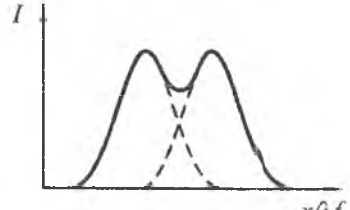

Рис. 4.17. Иллюстрация критерия Релея

Рассмотрим разрешающую способность объектива телескопа для случая бесконечно удаленного предмета, т.е. для формирования изображения в фокальной плоскости объектива. При этом можно считать, что на объектив падает плоская волна, которая фокусируется в картину Эйри (справедливо приближение дифракции Фраунгофера). Первый минимум картины Эйри соответствует случаю $x' = 1{,}22\lambda f/D$, где D — диаметр объектива, f — его фокусное расстояние, а числовой коэффициент определяется преобразованием Фурье—Бесселя, так как рассматривается круглое отверстие. Если перейти от линейного разрешения к угловому, то последнее будет равно $\Theta = 1{,}22\lambda/D$. Очень часто вводят понятие разрешающей способности объектива, понимая под ней величину, обратно пропорциональную Θ.

В последние годы в научной литературе много внимания уделяется вопросу получения в изображении разрешения выше релеевского, т.е. сверхразрешения. Эта возможность носит либо чисто теоретический характер, либо узко конкретный и определяется тем объемом информации, который распространяется через канал связи с помехами и нелинейностями. Поскольку эта величина зафиксирована, то увеличение разрешения, т.е. увеличение числа точек в изображении приведет к уменьшению динамического диапазона в нем (уменьшению количества промежуточных градаций яркости между светлым и темным в картинке) и нелинейному его искажению. Технически это достигается разными методами: и пространственным дифференцированием в изображении и нелинейным изменением контраста, и введением опорного пучка, т.е. переходом к интерференционной картинке.

Теоретическое обоснование сверхразрешения и основанный на нем расчетный механизм связаны с понятием аналитической функции и теоремой Котельникова. Если воспользоваться последней, то ограниченный по спектру сигнал может быть представлен конечным числом отсчетов, расположенных через интервал, обратно пропорциональный ширине спектра. Конечность спектра и ограничивает разрешение в изображении. Однако если сигнал таков, что его спектр является аналитической функцией, то методом аналитического продолжения можно рассчитать его значения в точках (отсчетах), расположенных через такой же интервал и прилегающих к области спектрального представления сигнала. Таким образом, спектр расширяется и соответственно увеличивается разрешение в сигнале. Затем можно рассчитать следующие значения и т.д. Однако полученные новые значения носят вероятностный характер и приводят к искажению представления сигнала.

4.7. ГАУССОВЫ ПУЧКИ

До сих пор, занимаясь скалярной теорией дифракции, мы рассматривали два возможных решения волнового уравнения (для плоской и сферической волн). Это вполне объяснимо, поскольку используя их как элементарные кирпичики, можно построить любое световое распределение, например, через механизм спектрального разложения по плоским волнам. Конечно, это не единственные решения, и в ряде случаев чрезвычайно удобно и наглядно можно использовать другие волновые объекты, например гауссовы пучки.

Понятие подобных пучков возникло при использовании геометрической оптики для анализа сложных объектов, таких как, например, открытые резонаторы. Дело в том, что свет вследствие волновой природы не может распространяться в виде плоской волны (поскольку она должна быть тогда бесконечной) либо в виде бесконечно тонкого луча. Если же рассматривать луч конечного диаметра, то обязательно должна наблюдаться дифракция на той апертуре, которая и сформировала этот луч. Поэтому введение некоторого светового пучка, имеющего в сечении гауссово распределение амплитуды, является весьма хорошим приближением как к понятию луча (если иметь в виду область перетяжки), так и сферического волнового фронта на большом удалении от нее.

Есть и еще одна причина для введения в рассмотрение именно гауссовых пучков. Она связана с устойчивостью световых распределений в сложных оптических системах, в первую очередь в системах с обратной связью. Если обратиться к открытым оптическим резонаторам, то в них световой пучок распространяется от зеркала к зеркалу, формируя некоторое световое распределение, которое получается устойчивым, если через n проходов оно восстанавливает свою конфигурацию. Дифракционные эффекты при этом присутствуют, т.е. фактически приходится описывать распространение волны преобразованием Фурье. Известно, что преобразование Фурье от гауссова распределения есть опять же гауссово распределение, т.е. гауссовы пучки должны являться устойчивыми образованиями для открытых резонаторов.

Определим теперь требования, при которых гауссов пучок удовлетворяет волновому уравнению, а точнее уравнению Гельмгольца

$$\nabla^2 U + k^2 U = 0.$$

Будем искать решение этого уравнения в виде

$$U = U_0 \chi(x, y, z) \exp(-jkz).$$

При этом мы считаем, что функция χ есть чисто амплитудная функция, медленно меняющаяся при распространении вдоль оси z. В связи с этим можно пренебречь производной $\partial^2\chi/\partial x^2$.

Подставим (4.75) в уравнение Гельмгольца и с учетом медленности амплитудной функции по z получим

$$\frac{\partial^2\chi}{\partial x^2} + \frac{\partial^2\chi}{\partial y^2} - 2jk\frac{\partial\chi}{\partial z} = 0. \qquad (4.76)$$

Амплитудная часть решения этого уравнения имеет вид

$$\chi = \exp\left[F_2(z) - \frac{x^2+y^2}{F_1(z)}\right]. \tag{4.77}$$

Функция $F_1(z)$ описывает два эффекта: изменение ширины пучка с увеличением расстояния вдоль оси распространения z и сдвиг фазы в поперечной этому направлению плоскости, связанный с изменением кривизны волнового фронта при изменении его ширины. Функция $F_2(z)$ — изменение амплитуды на оси пучка в связи с изменением его ширины, и дополнительный сдвиг фазы при распространении пучка.

Подставим (4.77) в (4.76) и получим уравнение

$$\left(x^2+y^2\right)\left[\frac{2}{F_1^2(z)} - jk\frac{F_1'(z)}{F_1^2(z)}\right] - \left[\frac{2}{F_1(z)} + jkF_2'(z)\right] = 0, \tag{4.78}$$

где штрих означает дифференцирование по z.

Для того чтобы (4.78) имело решение при всех (x, y), необходимо, чтобы в нем выражения в обеих квадратных скобках были равны нулю, тогда получим систему уравнений для F_1, F_2. Решение этой системы находится в виде

$$\left.\begin{aligned} F_1(z) &= A + \frac{2z}{jk}, \\ F_2(z) &= -\ln\left(z + \frac{jAk}{2}\right) + B, \end{aligned}\right\} \tag{4.79}$$

где A и B — комплексные постоянные, определяемые нулевыми условиями в точке $z = 0$.

Таким образом, показано, что решение в виде гауссова пучка соответствует уравнению Гельмгольца. Соответствующим выбором осей и начала координат, а также выделением действительной и мнимой частей у используемых функций можно преобразовать (4.77) и (4.79) к более понятному виду. В частности, для поперечной компоненты электрического поля распределение амплитуды имеет вид

$$E_x = E_{x0}\frac{w}{w_0}\exp\left[-j(kz+\varphi) - \left(x^2+y^2\right)\frac{jk}{2}\left(\frac{1}{R} - \frac{2j}{kw^2}\right)\right]. \tag{4.80}$$

Здесь w_0 — ширина пучка в области перетяжки (области, где пучок имеет минимальную ширину, рис. 4.18); $w(z)$ — ширина пучка на расстоянии z от перетяжки; $R(z)$ — радиус кривизны волнового фронта на расстоянии z от перетяжки; φ — некоторый фазовый параметр. Эти величины определяются следующими выражениями:

$$
\left.
\begin{aligned}
&\varphi = arctg\left(\frac{kw_0^2}{2z}\right);\\
&w^2(z) = w_0^2\left[1+\left(\frac{2z}{kw_0^2}\right)^2\right];\\
&R(z) = z\left[1+\left(\frac{w_0^2 k}{2z}\right)^2\right].
\end{aligned}
\right\}
\tag{4.81}
$$

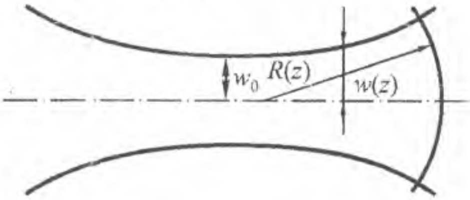

Рис. 4.18. Гауссов пучок

В перетяжке пучка волновой фронт плоский, в то время как на больших расстояниях от нее волновой фронт приближается к фронту сферической волны, исходящей из точечного источника, расположенного в начале координат.

Таким образом, получено решение волнового уравнения в виде гауссова пучка и определены его основные параметры.

4.8. НЕСКОЛЬКО ЗАМЕЧАНИЙ ОТНОСИТЕЛЬНО СТРОГОЙ ТЕОРИИ ДИФРАКЦИИ

Рассматривая методы скалярной теории дифракции, нужно иметь в виду ограниченную ее применимость, связанную с предположениями о диэлектрическом характере препятствий и об одинаковости изменений, происходящими со всеми компонентами электромагнитного поля. Необходимость учета в первую очередь конечной проводимости объектов, на которых происходит дифракция, и их конечной диэлектрической проницаемости приводит к несколько иной постановке задачи.

Для того чтобы продемонстрировать используемые представления и методы решения системы уравнений Максвелла, рассмотрим их вариант для распространения волн в проводнике, относящийся к так называемым задачам металлооптики.

Среда, в которой распространяется электромагнитная волна, является однородной и изотропной и характеризуется величинами $\varepsilon\mu$ и σ — соответственно диэлектрической и магнитной проницаемостью и проводимостью. С учетом того, что $\mathbf{D} = \varepsilon\varepsilon_0\mathbf{E}$, $\mathbf{B} = \mu\mu_0\mathbf{H}$, $j = \sigma\mathbf{E}$, получаем

$$\left.\begin{aligned}&\mathrm{rot}\mathbf{H}-\varepsilon\varepsilon_0\dot{\mathbf{E}}=\sigma\dot{\mathbf{E}};\\&\mathrm{rot}\dot{\mathbf{E}}+\mu\mu_0\mathbf{H}=0;\\&\mathrm{div}\dot{\mathbf{E}}=\frac{1}{\varepsilon\varepsilon_0}\rho;\\&\mathrm{div}\mathbf{H}=0.\end{aligned}\right| \tag{4.82}$$

Можно показать, что для электромагнитного возмущения, падающего на проводник, плотность заряда уменьшается во времени очень быстро, поэтому в третьем уравнении в (4.82) можно правую часть приравнять нулю. Тогда получается видоизмененное волновое уравнение

$$\nabla^2\mathbf{E}=\frac{\mu\varepsilon}{c^2}\ddot{\mathbf{E}}+\mu\mu_0\sigma\dot{\mathbf{E}}. \tag{4.83}$$

Наличие первой производной от напряженности поля означает, как обычно, затухание волны при распространении.

Волновое уравнение (4.83) при учете гармонического характера электромагнитной волны преобразуется в некоторый аналог уравнения Гельмгольца

$$\nabla^2\mathbf{E}+\bar{k}^2\mathbf{E}=0. \tag{4.84}$$

где

$$\bar{k}^2=\frac{\omega^2\mu}{c^2}\left(\varepsilon+j\frac{\mu_0\sigma}{\omega}\right).$$

Если ввести в уравнение (4.84) $\bar{\varepsilon}$, аналог обычной диэлектрической проницаемости,

$$\bar{\varepsilon}=\varepsilon+j\frac{\mu_0\sigma}{\omega},$$

то уравнение (4.83) станет идентичным обычному волновому уравнению для непроводящих сред после замены вещественной диэлектрической проницаемости на комплексную.

Вышеприведенные выкладки являются демонстрацией того, каким образом решаются проблемы теории дифракции, когда она выходит за рамки скалярного приближения. Однако, как правило, методы строгой теории дифракции применяются для решения конкретных задач и не имеют столь общего значения, как скалярная теория дифракции.

ВЫВОДЫ

1. Распределение поля после взаимодействия с препятствием может быть описано с помощью принципа Гюйгенса—Френеля, который представляет это распределение как сумму вторичных волн, испущенных точечными источниками, из которых складывается первичный волновой фронт, с учетом их интерференции.

2. Скалярная теория дифракции позволяет определить поле за экраном исходя из распределения поля в плоскости экрана. При этом должны выполняться условия Кирхгофа—Зоммерфельда о четкости тени, что определяет условия применимости этой теории: нельзя рассматривать область в несколько длин волн за экраном и объекты с размерами, сравнимыми с длиной волны.

3. Интегральные функции, описывающие распределение поля при дифракции, имеют структуру преобразования Фурье либо с точностью до фазового множителя (в области дифракции Фраунгофера), либо с точностью до квадратичного экспоненциального множителя под интегралом в случае дифракции Френеля.

4. Распределение поля в фокальной плоскости линзы представляет собой преобразование Фурье от распределения поля перед линзой с точностью до фазового множителя перед интегралом (аналогично виду дифракционной картины Фраунгофера).

5. Изображение объекта, получаемое с помощью линзы в дифракционном приближении, отличается от идеального более низкой разрешающей способностью, что связано с размыванием каждой его точки в картину Эйри за счет дифракции на ограниченной апертуре линзы. Разрешение в изображении определяется по критерию Релея, который определяет его как расстояние, при котором главный максимум первой картины Эйри совпадает с первым минимумом соседней картины.

6. Использование двумерного преобразования Фурье для разложения в спектр распределения амплитуды на входе оптической системы позволяет сформировать угловой спектр плоских волн, каждая из которых идет под своим углом и со своей амплитудой. При этом анализ прохождения системы одной плоской волной весьма прост, а сложение преобразованных плоских волн на выходе позволяет получить искомое изображение.

7. Гауссовы пучки, являющиеся также одним из решений системы уравнений Максвелла, позволяют рассматривать прохождение света в оптических системах в рамках приближения геометрической оптики с учетом самодифракции ограниченных пучков и их преобразований в пространстве, что приводит к возможности применения расчетных методов геометрической оптики в волновых системах, в том числе с обратной связью.

8. Учет проводимости препятствий в задачах дифракции приводит к необходимости перехода к векторному варианту решений, так как взаимодействие поля с проводящими средами вызывает возникновение в них токов, которые по-разному воздействуют на различные компоненты полевых векторов. Одновременно с этим формально происходит переход к комплексным выражениям для диэлектрической проницаемости, т.е. появляются потери и угловая дисперсия.

ГЛАВА 5
МОДЕЛИРОВАНИЕ СВЕТОТЕХНИЧЕСКИХ И ФОКУСИРУЮЩИХ УСТРОЙСТВ С ДОЭ

5.В. ВВЕДЕНИЕ

В предыдущих разделах основной акцент был сделан на изучении аппарата, предназначенного для моделирования оптических систем с классическими оптическими элементами (линзы, зеркала, призмы и т.п.). Учитывая бурное развитие новых разделов оптики, представляется актуальным изучить методы для моделирования систем, содержащих дифракционные оптические элементы (ДОЭ). Особый интерес представляет применение ДОЭ в фокусирующих и освещающих устройствах. Использование ДОЭ позволяет упростить конструкцию устройств, повысить энергетическую эффективность и оптимальным образом сформировать требуемую диаграмму направленности излучения.

5.1. ФИЗИЧЕСКИЕ ОСНОВЫ МОДЕЛИРОВАНИЯ ОПТИЧЕСКИХ УСТРОЙСТВ С ДОЭ

Рассмотрим сложную оптическую схему, состоящую из точечного источника, отражающих и преломляющих поверхностей и дифракционных оптических элементов. Оптическая установка может быть представлена как набор сред $S_1,\ldots,S_n$.

Каждая среда характеризуется показателем преломления и уравнением границы. ДОЭ вводится как фазовый разрыв вдоль элемента и приводит к существованию поверхности фазовых скачков. Для дальнейшего анализа укажем, что идеальная отражающая поверхность соответствует среде с показателем преломления "-1".

Для моделирования работы светотехнического устройства необходимо определить распределение интенсивности, формируемое исследуемым устройством, на трехмерной поверхности, описываемой следующим выражением

$$\Phi(x_1, x_2, x_3) = 0,$$

где (x_1,x_2,x_3) - координаты на трехмерной поверхности.

Световое поле комплексной амплитуды E в скалярном приближении подчиняется уравнению Гельмгольца

$$\nabla^2 E + k^2 n^2 E = \delta(\mathbf{r} - \mathbf{r}_0), \qquad (5.1)$$

где $k = 2\pi/\lambda$ - волновое число, λ - средняя длина волны, $n(x_1,x_2,x_3)$ - показатель преломления в оптической системе,

$\delta(\mathbf{r}-\mathbf{r}_0)$ - функция Дирака.

Правая часть уравнения Гельмгольца описывает существование точечного источника.

Для решения уравнения Гельмгольца необходимо определить граничные условия каждой среды. Кроме того, необходимо использовать информацию о поведении комплексной амплитуды светового поля вблизи ДОЭ.

Световое поле вдали от источников излучения может быть аппроксимировано более простыми уравнениями. Это упрощение базируется на предположении, что $kL >> 1$, где L - характерная длина, описывающая расстояние наблюдения, размеры препятствий и неоднородностей среды.

Введем приближение Дебая [1]

$$E(\mathbf{x}) = \sqrt{I(\mathbf{x})}\exp\left[ik\Psi(\mathbf{x})\right],$$

$$\mathbf{x} = (x_1, x_2, x_3), \tag{5.2}$$

где I - интенсивность светового поля,

$\Psi(\mathbf{x})$ - эйконал светового поля.

Подстановка уравнения (5.2) в уравнение Гельмгольца показывает, что функции I и Ψ подчиняются системе уравнений

$$\begin{cases} (\nabla\Psi)^2 = n^2(\mathbf{x}) \\ \operatorname{div}\left[I(\mathbf{x})\,\nabla\Psi(\mathbf{x})\right] = 0 \end{cases} \tag{5.3}$$

Граничные условия для уравнения Гельмгольца могут быть преобразованы:

1) для преломляющей поверхности

$$\left[\nabla\Psi_1 \times \mathbf{N}\right] = \left[\nabla\Psi_2 \times \mathbf{N}\right], \tag{5.4}$$

где $\mathbf{N}$ - нормаль к границе между двумя средами с различными показателями преломления;

2) для отражающей поверхности

$$\left[\nabla\Psi_1 \times \mathbf{N}\right] = -\left[\nabla\Psi_2 \times \mathbf{N}\right]; \tag{5.5}$$

3) для дифракционного оптического элемента

$$\left[(\nabla\Psi_2 - \nabla\Psi_1)\times\mathbf{N}\right] = \left[\nabla\varphi \times \mathbf{N}\right]. \tag{5.6}$$

Наличие в оптической схеме источника светового поля описывается с помощью граничных условий на исходной поверхности.

Рассмотрим точечный источник, окруженный исходной поверхностью.

Пусть исходная поверхность задается следующей системой уравнений

$$\begin{cases} x_1 = x_1(t_1,t_2) \\ x_2 = x_2(t_1,t_2), \\ x_3 = x_3(t_1,t_2) \end{cases} \tag{5.7}$$

где t_1, t_2 суть криволинейные координаты на начальной поверхности.

Предположим, что $\Psi(t_1,t_2)$ и $I_0(t_1,t_2)$ - распределения эйконала и интенсивности на исходной поверхности, которые определяются диаграммой направленности и положением точечного источника.

Вместо распределения эйконала на исходной поверхности возможно использование поля направлений лучей. Уравнения (5.3) включают частные производные, однако с помощью лучевых координат эти уравнения могут быть сведены к системе обыкновенных дифференциальных уравнений.

Введем лучевые координаты

$$\begin{cases} x_1 = x_1(t_1,t_2,l) \\ x_2 = x_2(t_1,t_2,l) \\ x_3 = x_3(t_1,t_2,l) \end{cases} \tag{5.8}$$

или в векторной форме

$$\mathbf{r} = \mathbf{r}(t_1,t_2,l). \tag{5.9}$$

Функция $\mathbf{r}(t_1,t_2,l)$ подчиняется уравнению распространения луча

$$\frac{\mathrm{d}}{\mathrm{d}l}\left(n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}l}\right) = \nabla n. \tag{5.10}$$

При $l = 0$ уравнение $\mathbf{r}_0 = \mathbf{r}(t_1,t_2,0)$ описывает исходную поверхность.

Функция $\mathbf{S}_0(t_1,t_2) = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}l}(t_1,t_2,0)$ определяет направление лучей на исходной поверхности.

Решение уравнения переноса в лучевых координатах принимает форму

$$I(t_1,t_2,l) = I_0(t_1,t_2)\frac{\dfrac{\partial(x_1,x_2,x_3)}{\partial(t_1,t_2,0)}}{\dfrac{\partial(x_1,x_2,x_3)}{\partial(t_1,t_2,l)}}, \tag{5.11}$$

где $I_0(t_1,t_2)$ - распределение интенсивности на исходной поверхности, $\dfrac{\partial(x_1,x_2,x_3)}{\partial(t_1,t_2,l)}$ - Якобиан преобразования Декартовых координат к лучевым.

Перепишем последнее уравнение в векторной форме так, как это более удобно для дальнейших рассуждений,

$$I(t_1,t_2,l) = I_0(t_1,t_2)\frac{\left[\mathbf{r}_{t_1} \times \mathbf{r}_{t_2}\right]\mathbf{r}_l\Big|_{l=0}}{\left[\mathbf{r}_{t_1} \times \mathbf{r}_{t_2}\right]\mathbf{r}_l} \tag{5.12}$$

и введем криволинейную систему координат на поверхности регистрации

$$\mathbf{r} = \mathbf{r}\left(t_1,t_2,l\left(t_1,t_2\right)\right) = \mathbf{R}\left(t_1,t_2\right), \tag{5.13}$$

где $l(t_1, t_2)$ - решение уравнения

$$\Phi\left(\mathbf{r}\left(t_1,t_2,l\right)\right) = 0. \tag{5.14}$$

Дифференцируя по t_1 или t_2 можно получить выражения для базисных векторов

$$\mathbf{R}_{t_1} = \mathbf{r}_{t_1} - \mathbf{r}_l \cdot \frac{\nabla\Phi \cdot \mathbf{r}_{t_1}}{\nabla\Phi \cdot \mathbf{r}_l}, \tag{5.15}$$

$$\mathbf{R}_{t_2} = \mathbf{r}_{t_2} - \mathbf{r}_l \cdot \frac{\nabla\Phi \cdot \mathbf{r}_{t_2}}{\nabla\Phi \cdot \mathbf{r}_l}.$$

Принимая во внимание выражения для базисных векторов на исходной поверхности и поверхности регистрации, уравнение (5.12) преобразуется в следующее

$$I(t_1,t_2,l) = I_0(t_1,t_2)\frac{\left[\mathbf{r}_{t_1} \times \mathbf{r}_{t_2}\right]\mathbf{r}_l\Big|_{l=0}}{\left[\mathbf{R}_{t_1} \times \mathbf{R}_{t_2}\right]\mathbf{r}_l}. \tag{5.16}$$

По сравнению с выражением (5.11) уравнение (5.16) содержит распределение интенсивности на исходной поверхности и метрические характеристики исходной поверхности и поверхности регистрации. В качестве поверхности регистрации может использоваться сфера с бесконечно большим радиусом. В этом случае направления лучей, падающих на поверхность регистрации, совпадают с нормалью к ней

$$\left[\mathbf{R}_{t_1} \times \mathbf{R}_{t_2}\right]\mathbf{r}_l = \left|\left[\mathbf{R}_{t_1} \times \mathbf{R}_{t_2}\right]\right|. \tag{5.17}$$

То же самое соотношение возможно на исходной поверхности

$$\left[\mathbf{r}_{t_1} \times \mathbf{r}_{t_2}\right]\mathbf{r}_l = \left|\left[\mathbf{r}_{t_1} \times \mathbf{r}_{t_2}\right]\right|. \tag{5.18}$$

Можно применить эти уравнения для получения более простого выражения для распределения интенсивности на поверхности регистрации.

Однако это выражение не подходит для расчета. Неудобство связано с использованием лучевой системы координат, так как выбор лучевых координат зависит от оптической схемы.

Введем новую криволинейную систему координат (T_1, T_2) на поверхности регистрации

$$\mathbf{r} = \Phi(T_1, T_2). \tag{5.19}$$

Предположим, что криволинейные координаты (T_1, T_2) связаны с лучевыми координатами (t_1, t_2) с помощью следующей системы уравнений

$$\begin{aligned} T_1 &= F_1(t_1, t_2) \\ T_2 &= F_2(t_1, t_2) \end{aligned} \tag{5.20}$$

Подставляя (5.18) и (5.19) в (5.16), получаем

$$I(t_1, t_2, l) = I_0(t_1, t_2) \frac{\left\| \left[\mathbf{r}_{t_1} \times \mathbf{r}_{t_2} \right] \right\|}{\left\| \left[\Phi_{F_1} \times \Phi_{F_2} \right] \right\| \dfrac{\partial(F_1, F_2)}{\partial(t_1, t_2)}}. \tag{5.21}$$

Используя свойства функции Дирака можно переписать последнюю формулу в интегральной форме

$$I = \int I_0(t_1, t_2) \frac{\left\| \left[\mathbf{r}_{t_1} \times \mathbf{r}_{t_2} \right] \right\|}{\left\| \left[\Phi_{F_1} \times \Phi_{F_2} \right] \right\|} \delta\left[T_1 - F_1(t_1, t_2), T_2 - F_2(t_1, t_2) \right] \mathrm{d}t_1\, \mathrm{d}t_2. \tag{5.22}$$

Уравнение (5.22) обеспечивает базис для компьютерного моделирования оптических систем.

Предположим, что исходная поверхность и поверхность регистрации описываются уравнениями

$$\mathbf{r}(t_1, t_2) = \varepsilon\left(\sin t_1 \cos t_2 \cdot \mathbf{e}_1 + \sin t_1 \sin t_2 \cdot \mathbf{e}_2 + \cos t_1 \cdot \mathbf{e}_3 \right),$$

$$\Phi(T_1, T_2) = R_0\left(\sin T_1 \cos T_2 \cdot \mathbf{e}_1 + \sin T_1 \sin T_2 \cdot \mathbf{e}_2 + \cos T_1 \cdot \mathbf{e}_3 \right), \tag{5.23}$$

где ε - радиус исходной поверхности, R_0 - радиус поверхности регистрации, ($\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$) - базисные вектора прямоугольной системы координат.

В этом случае распределение интенсивности на исходной поверхности представляется как

$$I(t_1, t_2) = \frac{I_0}{\varepsilon^2}, \tag{5.24}$$

а распределение интенсивности на поверхности регистрации принимает вид

$$I(T_1, T_2) = \int \frac{I_0}{R_0^2} \frac{\sin t_1}{\sin F_1(t_1, t_2)} \delta\left[T_1 - F_1(t_1, t_2), T_2 - F_2(t_1, t_2) \right] \mathrm{d}t_1\, \mathrm{d}t_2. \tag{5.25}$$

5.2. МОДЕЛИРОВАНИЕ УСТРОЙСТВ С КВАНТОВАННЫМИ ДОЭ

В предыдущем разделе рассмотрен метод моделирования светотехнических устройств с ДОЭ. Однако описанные выше ДОЭ обладают непрерывной фазовой функцией, что возможно только при использовании уникальных технологий изготовления дифракционного микрорельефа, например, на основе эффекта темнового роста в слоях жидких фотополимеризующихся композиций или на основе прямой лазерной записи на фоторезист [3].

Однако большинство разработанных технологий и существующего технологического оборудования не обеспечивает формирование ДОЭ с непрерывной фазовой функцией [3]. В частности, для изготовления ДОЭ успешно используется технология фотолитографического травления, позволяющая сформировать многоуровневый микрорельеф.

В настоящем разделе рассматривается метод моделирования устройств, включающих многоуровневые («квантованные») ДОЭ.

Моделирование аналогичных оптических систем основано на модифицированном методе расчета лучевых траекторий. Для понимания предлагаемого метода рассмотрим прохождение светового пучка через многоуровневый ДОЭ.

Эйконал ДОЭ со ступенчатым профилем микрорельефа может быть представлен как

$$\bar{\varphi} = \Phi\left(\operatorname{mod}_{\lambda} \varphi(t_1, t_2)\right), \tag{5.26}$$

где (t_1, t_2) - Декартовы координаты на дифракционном оптическом элементе, $\Phi(z)$ - функция предыскажения [3].

Функция предыскажения описывает преобразование функции эйконала ДОЭ в результате использования конкретной технологии формирования ДОЭ и возникших при изготовлении технологических погрешностей микрорельефа.

Решение уравнения Гельмгольца может быть записано в форме

$$E(x_1, x_2, x_3) = \frac{i}{\lambda} \int E_0(t_1, t_2) \frac{z}{L} \frac{\exp(ikL)}{L} \exp\left[ik\bar{\varphi}(t_1, t_2)\right] \mathrm{d}t_1 \, \mathrm{d}t_2 . \tag{5.27}$$

Для вычисления интеграла Кирхгофа разложим функцию $\exp(ikf)$ в ряд Фурье

$$\exp(ik\bar{\varphi}) = \sum C_n \exp\left[ikn\varphi(t_1, t_2)\right]. \tag{5.28}$$

Подставляя формулу (5.28) в уравнение (5.27) и вычисляя все интегралы методом стационарной фазы [5] получаем выражение для комплексной амплитуды светового поля

$$E = -\sum C_n E_0\left(t_1^n, t_2^n\right) \frac{\exp\left[ikn\,\varphi\left(t_1^n, t_2^n\right)\right]}{\sqrt{J_n\left(t_1^n, t_2^n, l^n\right)}}, \tag{5.29}$$

где

$$J_n\left(t_1^n, t_2^n, l^n\right) = \frac{\dfrac{\partial\left(x_1, x_2, x_3\right)}{\partial\left(t_1^n, t_2^n, 0\right)}}{\dfrac{\partial\left(x_1^n, x_2^n, x_3^n\right)}{\partial\left(t_1^n, t_2^n, l^n\right)}}$$ - лучевые координаты.

Функция преобразования лучевых координат в Декартовы подчиняется системе дифференциальных уравнений

$$\frac{\mathrm{d}}{\mathrm{d}l^n}\left(\frac{\mathrm{d}x_i}{\mathrm{d}l^n}\left(t_1^n, t_2^n, l^n\right)\right) = 0 \tag{5.30}$$

с начальными условиями

$$x_i\left(t_1^n, t_2^n, 0\right) = t_i,\ i=1,2;$$

$$x_3\left(t_1^n, t_2^n, 0\right) = 0;$$

$$\frac{\mathrm{d}x_i}{\mathrm{d}l}\left(t_1^n, t_2^n, 0\right) = \frac{\partial\left[n\varphi\left(t_1^n, t_2^n\right)\right]}{\partial t_i^n},\ i=1,2; \tag{5.31}$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}l}\left(t_1, t_2, 0\right) = \sqrt{1 - \left(\frac{\mathrm{d}x_1}{\mathrm{d}l}\right)^2 - \left(\frac{\mathrm{d}x_2}{\mathrm{d}l}\right)^2}.$$

На основе формулы (5.29) можно получить распределение интенсивности

$$I\left(x_1, x_2, x_3\right) = \sum \frac{\left(C_n\right(^2 I_0\left(t_1^n, t_2^n\right)}{J_n\left(t_1^n, t_2^n, l\right)} +$$

$$+\sum_{m=n} \frac{C_n C_m E_0\left(t_1^n, t_2^n\right) E_0^*\left(t_1^m, t_2^m\right) \exp\left\{ik\left[n\varphi\left(t_1^n, t_2^n\right) - m\varphi\left(t_1^m, t_2^m\right)\right]\right\}}{\sqrt{J_n\left(t_1^n, t_2^n, l^n\right)}\sqrt{J_m\left(t_1^m, t_2^m, l^m\right)}}. \tag{5.32}$$

Попробуем проанализировать полученное выражение. Первый член в уравнении (5.32) представляет собой интенсивность светового поля, сформированного дифракционным оптическим элементом с функцией эйконала $n\varphi$. Второй член этого выражения описывает интерференцию между световыми полями. Обычно источник света в исследуемых светотехнических устройствах является некогерентным, поэтому интерференцией между различными световыми пучками можно пренебречь.

5.3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ СВЕТОТЕХНИЧЕСКОГО УСТРОЙСТВА С ДОЭ

В качестве примера моделирования оптической установки рассмотрим простейшее светотехническое устройство, схематически изображенное на рисунке 5.1. Устройство состоит из параболического отражателя, в фокусе которого расположен центр осевой нити накаливания, и дифракционного оптического элемента. Целью работы исследуемого светотехнического устройства будем считать формирование диаграммы направленности излучения в форме латинской буквы F. Результат, полученный для буквы F, может быть легко обобщен на любую требуемую диаграмму направленности.

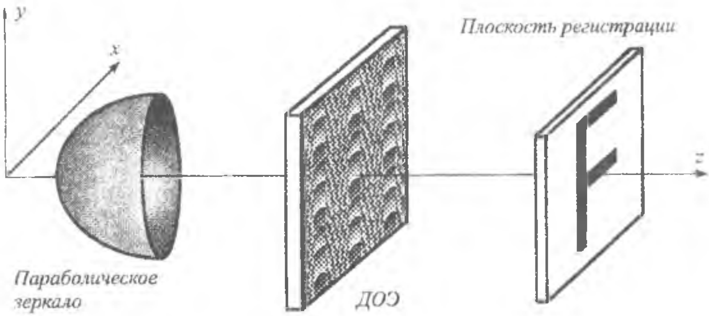

Рис.5.1. Оптическая схема моделируемого светотехнического устройства

Для оценки качества работы светотехнического устройства на стадии проектирования необходимо проанализировать световое поле, формируемое исследуемой оптической схемой в заданной области пространства. В рамках численного моделирования исследуется диаграмма направленности и анализируется связь между параметрами оптической схемы и качеством формируемой диаграммы направленности.

Для характеристики качества диаграммы направленности используются следующие величины

$$\delta = \frac{\sqrt{\int \left(I(\bar{x}) - I_0(\bar{x})\right)^2 d\bar{x}}}{\int I_0(\bar{x}) d\bar{x}}, \tag{3.1}$$

$$\varepsilon = \frac{E_1}{E_0}, \tag{3.2}$$

где ε - значение энергетической эффективности устройства, а δ - значение среднеквадратичного отклонения $I_0(\overline{x})$ от $I(\overline{x})$; $I_0(\overline{x})$ - идеальное (требуемое) распределение интенсивности, $I(\overline{x})$ - результат моделирования; E_1 - энергия, попадающая в требуемую область, E_0 - энергия источника излучения. Энергетическая эффективность характеризует долю энергии источника, попадающую в заданную область пространства.

В рассматриваемой оптической схеме (рис. 5.1) ДОЭ вычислялся в приближении геометрической оптики на основе метода согласованных прямоугольников [3] с последующим повторением полученной фазовой функции по вертикали и горизонтали. Сложность расчета ДОЭ в подобном устройстве заключается в том, что в каждую точку оптического элемента приходит конус лучей от протяженного источника и отражателя. В результате практически невозможно рассчитать ДОЭ, оптимально управляющий приходящими разнородными пучками, что подтверждается результатами моделирования (рис. 5.2 и табл. 5.1).

Численное моделирование показывает, что качество формируемой диаграммы направленности может быть улучшено при уменьшении отношения d/f, где d - размер (длина нити накаливания) источника, а f - фокусное расстояние параболического отражателя (см. табл. 5.1 и рис. 5.2). Этот факт может быть подтвержден и с помощью аналитического исследования.

Таблица 5.1.

Результаты моделирования светотехнического устройства

d/f	ε (%)	δ	*Номер рисунка*
2/3	69,00	1,485	5.2.a
2/5	70,89	1,435	5.2.b
4/15	59,63	1,357	5.2.c
1/5	47,88	1,243	5.2.d
2/15	30,44	1,009	5.2.e
4/35	24,60	0,907	5.2.f
1/10	20,22	0,837	5.2.g

При моделировании были зафиксированы глубина отражателя и размер источника излучения (т.е. длина нити накаливания), а величина параметра d/f регулировалась за счет увеличения фокусного расстояния отражателя. При фиксированной глубине это приводило к увеличению раскрыва отражателя, и, как следствие, уменьшению энергетической эффективности светотехнического устройства (см. табл.5.1).

Другое направление проведенных исследований - это необходимая точность юстировки светотехнического устройства для формирования требуемой диаграммы направленности.

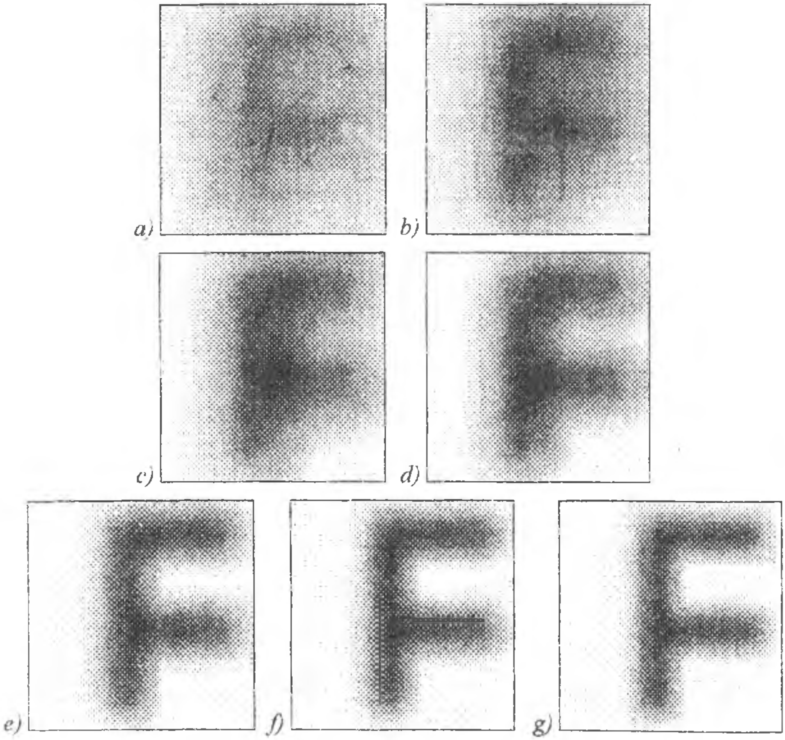

Рис. 5.2. Распределение интенсивности в плоскости регистрации для параметров табл. 5.1

На рисунке 5.3 и в таблице 5.2 показано изменение формы диаграммы направленности при перемещении центра нити накаливания вдоль оси симметрии отражателя в зависимости от значения относительной величины сдвига $\Delta z/f$, где f - фокусное расстояние отражателя. В результате такого сдвига при непринципиальном изменении энергетической эффективности падающего на область регистрации светового потока происходит существенное искажение формируемой диаграммы направленности.

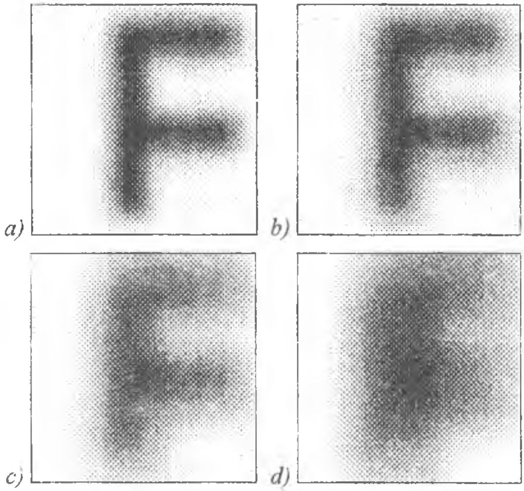

Рис. 5.3. Изменение диаграммы направленности для параметров таблицы 5.2.

Таблица 5.2. Характеристики изменения формы диаграммы направленности при перемещении нити накаливания вдоль оптической оси

d/f	Сдвиг $\Delta z/f$	δ	*Номер рисунка*
1/10	1/60	0.876	5.3.a
1/10	1/30	1.027	5.3.b
1/10	1/20	1,273	5.3.c
1/10	1/15	1.495	5.3.d

ВЫВОДЫ

1. Для моделирования освещающих и фокусирующих устройств, содержащих дифракционные оптические элементы, возможно использование специально адаптированных для этих целей методов геометрической оптики.

2. При выборе аппарата для моделирования оптических устройств, содержащих ДОЭ, необходимо учитывать вид дифракционного микрорельефа ДОЭ (многоуровневый ступенчатый или непрерывно-гладкий внутри каждой зоны).

3. Сложность расчета светотехнических устройств, формирующих требуемую диаграмму направленности, требует не только использования уникальных возможностей дифракционной оптики, но и обязательного моделирования и оптимизации работы такого класса оптических систем на этапе их проектирования. Проведенные исследования светотехнического устройства, состоящего из ДОЭ, параболического отражателя и протяженного источника, демонстрируют возможности моделирования и дифракционных оптических элементов.

ГЛАВА 6
МОДЕЛИРОВАНИЕ ФОКУСИРУЮЩИХ ДОЭ

6.В. ВВЕДЕНИЕ

На основе анализа, проведенного в первой главе (см. § 1.8), можно сделать вывод, что в большинстве случаев для моделирования дифракционных оптических элементов, предназначенных для фокусировки лазерного излучения, достаточно использования скалярной теории дифракции. ДОЭ, предназначенный для фокусировки лазерного излучения в заданную область пространства с требуемым распределением интенсивности внутри нее, по предложению лауреата нобелевской премии академика А.М. Прохорова получил название «фокусатор» [3].

Существующие коммерческие программы создавались для моделирования общего случая скалярной теории дифракции, поэтому большинство проектировщиков вынуждены использовать свои собственные численные приложения для вычисления соответствующего скалярного дифракционного интеграла. Таким образом, для такого класса элементов дифракционной микрооптики, который образуют фокусаторы лазерного излучения, необходима разработка своих методов расчета скалярного дифракционного интеграла, учитывающих особенности этого класса.

6.1. ДИФРАКЦИЯ ЛАЗЕРНОГО ИЗЛУЧЕНИЯ НА СИНТЕЗИРОВАННЫХ ОПТИЧЕСКИХ ЭЛЕМЕНТАХ

6.1.1. Постановка задачи фокусировки лазерного излучения с помощью ДОЭ

Если задана фазовая функция оптического элемента, то, решив задачу дифракции волн на таком оптическом элементе, можно получить распределение поля в интересующей нас области. Это - прямая задача теории дифракции. Для получения фазовой функции дифракционного оптического элемента, фокусирующего излучение в требуемую область пространства с заданным распределением интенсивности, необходимо решить обратную задачу дифракции. Для уяснения физической сущности обратной задачи дифракции рассмотрим Рис. 6.1. Плоский оптический элемент Φ, расположенный в области G плоскости $\mathbf{u} = (u, v)$, освещается пучком E монохроматического излучения длины волны λ. Требуется сформировать в области D плоскости $\mathbf{x} = (x, y)$ волновое поле $w(\mathbf{x}, z)$. Фазовая функция оптического элемента полностью определяет поведение пучка за плоскостью **u** и, в частности, в интересующей нас области D. Задача состоит в отыскании фазовой функции оптического элемента $\varphi(u,v)$, обеспечивающей формирование требуемого волнового поля.

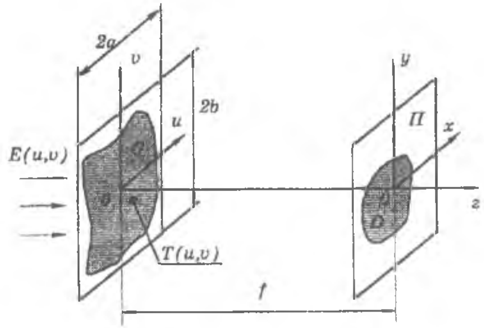

Рис. 6.1. Геометрия задачи фокусировки

С математической точки зрения обратная задача является некорректной: во-первых, решение может вообще не существовать; во-вторых, оно может быть неустойчивым.

Типичной обратной задачей дифракции служит задача создания ДОЭ, фокусирующего излучение в заданную область пространства. Пусть необходимо сфокусировать излучение в какую-либо фигуру или линию в фокальной плоскости. Поскольку в области фокусировки задается лишь распределение интенсивности, а фаза произвольна, то указанная степень свободы позволяет построить фазовый оптический элемент, получивший название фокусатор. Впервые задача синтеза фокусатора поставлена и решена в 1981 году в работе М.А. Голуба, А.М. Прохорова, В.А. Сойфера и др. Учитывая сложность обратной задачи фокусировки, решение, как правило, удается получить лишь в приближении геометрической оптики.

Геометрооптические фокусаторы рассчитаны на получение равномерного распределения интенсивности в области фокусировки. Однако для достижения оптимальных условий ряда технологических процессов лазерной обработки материалов бывает нужно не просто сфокусировать излучение в заданную область фокальной плоскости, а сформировать в ней заданное распределение интенсивности. При этом размеры области фокусировки обычно сравнимы с шириной дифракционно-ограниченного пятна, так что определяющую роль в формировании требуемого профиля интенсивности играет дифракция. Некоторые задачи требуют фокусировки излучения в области, продольные размеры которых значительно больше поперечных, обычно совпадающих с шириной дифракционно-ограниченного пятна. В таких случаях говорят о фокусировке излучения в линию (отрезок, набор отрезков) или кривую.

К настоящему времени разработан ряд дифракционных методов расчета фазовых функций фокусирующих ДОЭ. Эти методы учитывают и частично компенсируют дифракционные эффекты фокусировки. К таким методам относятся: метод дифракционных поправок геометрооптической фазовой функции, метод нелинейных предыскажений фазовой функции, различные итерационные методы расчета. Кроме дифракционных разработаны также различные менее эффективные методы решения задачи фокусировки: численные, эмпирические, квазипериодические, голографические и другие. В связи с наличием большого количества разработанных методов решения обратной задачи дифракции представляется актуальным проведение сравнительного анализа существующих методов на эталонных задачах фокусировки.

6.1.2. Методы автоматизированной записи фазовой функции на физическую среду

Обычно фокусатор представляет собой стеклянную или медную подложку, лицевая поверхность которой имеет дифракционный микрорельеф. Рельеф повторяет профиль формируемой фазы, взятой по модулю 2π. Такие элементы могут изготовляться как для работы на отражение, так и на пропускание. Изготовление фокусаторов осуществляется методами компьютерной оптики. В истоках этих методов лежит метод киноформа.

Основные этапы получения ДОЭ показаны на Рис. 6.2. Начальным этапом является теоретическое исследование, результаты которого в виде геометрооптического (или другого) решения, начального приближения или (и) метода расчета используются на этапе 1. Результатом этапа 1 является матрица (массив) отсчетов фазовой функции ДОЭ или две матрицы (два массива) отсчетов амплитудного и фазового пропускания ДОЭ, т.е. уже осуществленной оказывается операция дискретизации. Затем полученная матрица подвергается операции кодирования, в частности, в случае фокусатора - приведения фазы к интервалу $[0, 2\pi)$. Затем выполняется квантование фазовой функции, и моделирование работы полученного элемента. Этапы 4-5 связаны с физической реализацией ДОЭ. Все этапы в настоящее время осуществляются на компьютере (этапы 1-3) или под его управлением (этапы 4-6).

Особую проблему представляет отображение фазового пропускания ДОЭ (этапы 4-5). Первые фазовые элементы были получены по технологии цифровой голографии: фаза приводится к интервалу $[0, 2\pi)$ и отображается с помощью многоградационного фотопостроителя в виде вариаций плотности почернения фотоматериала; полученная амплитудная маска фазового элемента подвергается отбеливанию, в результате чего формируется фазовый рельеф. В дальнейшем эта технология была дополнена операцией фотоуменьшения амплитудной маски для получения более высокого пространственного разрешения.

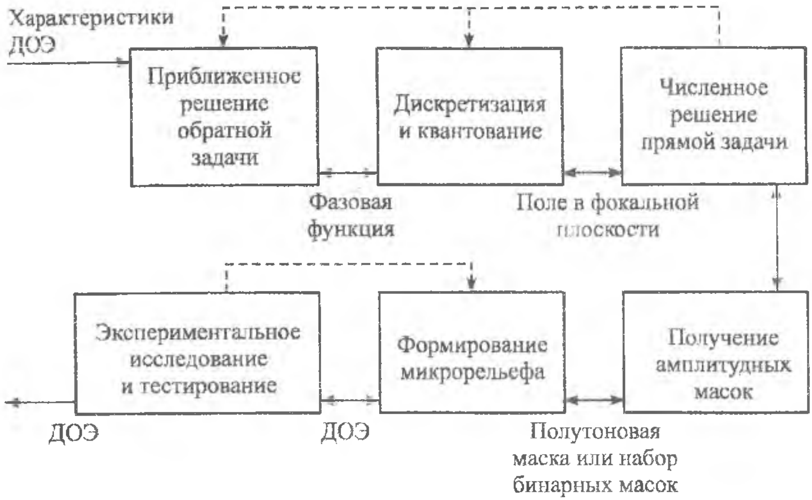

Рис. 6.2. Основные этапы синтеза ДОЭ

Первые отражательные фокусаторы были получены путем нанесения металлического покрытия на стеклянную подложку с фазовым рельефом. В дальнейшем был применен процесс гальванопластики, в результате которого сложный фазовый рельеф переносится с металлизированного стекла на медь и получается силовой оптический элемент.

При использовании бинарных фотопостроителей получают наборы масок и, последовательно применяя их в фотолитографической установке, получают многоградационный фазовый рельеф.

При сканировании плоскости среды, регистрирующей ДОЭ, наибольшее распространение получили прецизионные сканирующие электромеханические и лазерные устройства вывода изображений, а также электронные литографы. При этом бинарные генераторы изображений и электронные литографы позволяют добиться значительного разрешения (0,02 - 4 мкм), в то время как многоградационные фотопостроители при ограниченном разрешении (1 - 50 мкм) позволяют получать до 256 градаций яркости.

Различные типы устройств регистрации ДОЭ тем не менее позволяют описывать ДОЭ общей моделью, вполне достаточной для изучения влияния дискретизации и квантования фазовой функции на эффективность фокусировки.

В рамках этой модели считается, что на поверхности подложки ДОЭ имеются координаты $\mathbf{u} = (u, v)$. Область G, занимаемая ДОЭ, разбивается на N непересекающихся ячеек (модулей) G_n, нумеруемых индексами $n \in I_N$:

$$\bigcup_{n \in I_N} G_n = G, \qquad G_n \bigcap G_{n'} = 0 \quad (n \neq n'). \tag{6.1}$$

Каждая ячейка соответствует одному положению сканирующего устройства. Значение t_n функции амплитудно-фазового пропускания синтезируемого ДОЭ в пределах ячейки полагается постоянным и называется отсчетом этой функции.

Полученная функция комплексного пропускания синтезированного ДОЭ является кусочно-постоянной и описывается выражением

$$T(\mathbf{u}) = \sum_{n \in I_N} t_n x_n(\mathbf{u}), \qquad \mathbf{u} \in G, \tag{6.2}$$

где

$$x_n(\mathbf{u}) = \begin{cases} 1, & \mathbf{u} \in G_n \\ 0, & \mathbf{u} \overline{\in} G_n \end{cases}. \tag{6.3}$$

При переходе от ячейки к ячейке значение отсчета изменяется в соответствии с сигналом, поступающим от компьютера. В памяти ЭВМ n-ному отсчету соответствует последовательность из m двоичных разрядов $(d_{n_1}, d_{n_2}, \dots, d_{n_m})$ - двоичный код. Соответствие

$$t_n = f\left(d_{n_1}, d_{n_2}, \dots, d_{n_m}\right) \tag{6.4}$$

определяется выбранным способом кодирования, а также нелинейными характеристиками устройства регистрации ДОЭ.

Введение интерполирующей функции (6.3) соответствует "равномерной засветке" в пределах одного элемента разрешения амплитудной маски, что имеет место в известных устройствах. Наибольшее распространение при синтезе фокусирующих ДОЭ получили два типа устройств.

1. Устройство с построчной разверткой. Для него имеют место соотношения

$$G = \left\{\mathbf{u} = (u, v): \quad |u| < U, \quad |v| < V\right\}, \tag{6.5}$$

$n = (p, l)$ - двойной индекс, $N = N_1 \cdot N_2$, $\left(p = \overline{1, N_1}; \quad l = \overline{1, N_2}\right)$,

$$G_{pl} = \left\{\mathbf{u} = (u, v): u_{p-1} \le u < u_p, v_{l-1} \le v < v_l\right\}, \tag{6.6}$$

$$u_p = \left(p - \frac{N_1}{2} \right) \delta u, \qquad v_l = \left(l - \frac{N_2}{2} \right) \delta v, \tag{6.7}$$

$$\delta u = \frac{2U}{N_1}, \qquad \delta v = \frac{2V}{N_2}. \tag{6.8}$$

Величины δu, δv естественно называть разрешением устройства по осям u и v соответственно.

ДОЭ представляется в виде двумерной системы прямоугольников $\delta u \cdot \delta v$ с различными коэффициентами комплексного пропускания.

2. Для устройств с круговой разверткой

$$G = \left\{ \mathbf{u} = (u, v): \ \sqrt{u^2 + v^2} \le a \right\}, \tag{6.9}$$

n - одномерный индекс, $n = \overline{1, N}$,

$$G_n = \left\{ \mathbf{u} = (u, v): \ r_{n-1} \le \sqrt{u^2 + v^2} < r_n \right\}, \tag{6.10}$$

где a - радиус ДОЭ,

N - число колец разрешения,

r_k - внешний радиус k-го кольца,

$r_0 = 0$, $r_N = a$.

ДОЭ радиусом a представляется в виде системы концентрических колец с различными значениями коэффициента амплитудно-фазового пропускания.

При этом следует подчеркнуть, что для ДОЭ с прямоугольным растром дискретизации возникает естественное разбиение на одинаковые элементы разрешения размером $\delta u \cdot \delta v$ каждый. С другой стороны, для цилиндрических (линейных) ДОЭ или ДОЭ с вращательной симметрией современная технология расчета и изготовления позволяет получать оптические элементы как с равномерным, так и с неравномерным (адаптируемым к функции пропускания) шагом дискретизации.

В этом случае функция комплексного пропускания радиально-симметричного ДОЭ будет иметь вид:

$$T(r) = \sum_{k=1}^{N} t_k \cdot \mathrm{rect}\left(\frac{r - r_k}{\delta_k} + \frac{1}{2} \right), \qquad 0 \le r \le a, \tag{6.11}$$

где

$$\mathrm{rect}(x) = \begin{cases} 1, & |x| \le 0{,}5; \\ 0, & |x| > 0{,}5; \end{cases} \tag{6.12}$$

$$r=\sqrt{u^2+v^2},$$

N - число кольцевых модулей;

δ_k - ширина, а r_k - внешний радиус k-го кольца с комплексным пропусканием t_k, $r_N = a$.

Функция комплексного пропускания цилиндрического ДОЭ имеет вид

$$T(u)=\sum_{k=1}^{N} t_k \cdot \mathrm{rect}\left(\frac{u-u_k}{\delta_k}+\frac{1}{2}\right), \qquad |u|\le U, \tag{6.13}$$

где δ_k- ширина, а u_k - большая граница k-го элемента разрешения с комплексным пропусканием t_k; $u_0 = -U$; $u_N = U$.

Для ДОЭ с прямоугольным растром дискретизации функция комплексного пропускания имеет вид

$$T(u,v)=\sum_{p=1}^{N_1}\sum_{l=1}^{N_2} t_{pl} \cdot \mathrm{rect}\left(\frac{u-u_p}{\delta u}\right)\cdot \mathrm{rect}\left(\frac{v-v_l}{\delta v}\right), \tag{6.14}$$

где $|u| \le U$, $|v| \le V$,

(u_p, v_l) - центр (p,l)-го модуля ДОЭ с функцией комплексного пропускания t_{pl}.

$$u_p=-U+\left(p-\frac{1}{2}\right)\cdot\delta u, \qquad p=\overline{1,N_1},$$

$$v_l=-V+\left(l-\frac{1}{2}\right)\cdot\delta v, \qquad l=\overline{1,N_2}, \tag{6.15}$$

$$\delta u=\frac{2U}{N_1}, \qquad \delta v=\frac{2V}{N_2}.$$

При этом для ДОЭ функция комплексного пропускания полностью определяется фазой (т.е. амплитудная составляющая тождественно равна единице для всего элемента и, соответственно, для всех модулей).

Таким образом, на основе анализа методов расчета и изготовления предлагается растровая модель, основанная на представлении ДОЭ в виде совокупности элементарных модулей (элементов растра). Для фокусаторов с учетом дискретизации и квантования фазовой функции указанные модули могут быть кольцевой, линейной или прямоугольной формы.

6.1.3. Характеристики фокусирующих ДОЭ. Фокусатор как объект исследования

Учитывая принципиальную новизну и широкие прикладные возможности фокусирующих ДОЭ, представляется актуальным теоретическое и экспериментальное исследование характеристик фокусаторов.

Среди характеристик, описывающих процессы создания и функционирования фокусатора, можно выделить три вида параметров.

К первому виду относятся физические параметры, положенные в основу расчета фазовой функции ДОЭ, - фокусное расстояние; рабочая длина волны; размеры фокусатора и области фокусировки, а также характеристики, описывающие ее форму и форму фокусируемого пучка; угол падения излучения на оптический элемент и т.п.

Ко второму виду относятся параметры дискретизации и квантования фазовой функции фокусатора, размеры и форма элементов разрешения (модулей) фокусатора. Эти параметры связаны с выбором устройства регистрации ДОЭ.

К третьему виду относятся дифракционные характеристики фокусатора - энергетическая эффективность, ширина фокальной линии, среднеквадратическое отклонение полученного распределения интенсивности в фокальной области от требуемого и т.п.

Для проектирования фокусирующих ДОЭ первые два вида параметров являются внутренними, а дифракционные параметры - внешними, получающимися в результате функционирования фокусатора с выбранными внутренними параметрами. Для исследования фокусатора важно выявить связь внешних и внутренних параметров проектирования. Причем, учитывая трудоемкость, многовариантность процедуры изготовления ДОЭ, исследовать характеристики фокусатора необходимо уже на стадии проектирования.

Охарактеризуем важнейшие внешние параметры фокусаторов. Пусть линия фокусировки задается параметрически $\mathbf{x} = \mathbf{x}_0(\xi)$, $\xi \in [0, \kappa]$.

Здесь $\mathbf{x} = (x, y)$ - декартовы координаты в плоскости фокусировки, ξ - натуральный параметр кривой. Далее введем в плоскости фокусировки криволинейные координаты (ξ, η), связанные с (x, y) соотношениями

$$\begin{cases} y = y_0(\xi) - \dfrac{dx_0(\xi)}{d\xi}\eta \\ x = x_0(\xi) - \dfrac{dy_0(\xi)}{d\xi}\eta \end{cases} \tag{6.16}$$

Координата η, выбранная таким образом, характеризует отклонение точки $\mathbf{x}$ от линии фокусировки, отсчитанное по нормали. В криволинейных координатах (ξ, η) ширину фокальной линии $\Delta\eta$ можно определить из системы уравнений

$$\begin{cases} I(\xi, \eta_+) = \Theta \cdot I(\xi, 0) \\ I(\xi, \eta_-) = \Theta \cdot I(\xi, 0) \end{cases} \tag{6.17}$$

где $\eta_+ > 0, \eta_- < 0$,

$I(\xi, \eta)$ - интенсивность светового поля в точке (ξ, η)

Θ - уровень интенсивности относительно интенсивности на линии фокусировки,

$\Delta\eta_\Theta = \eta_+ - \eta_-$ - ширина фокальной линии, определяемая по уровню Θ.

Энергетическая эффективность фокусировки ε_Θ может характеризоваться долей падающей на фокусатор энергии, попадающей в фокальную линию:

$$\varepsilon_\Theta = \frac{\int\limits_0^\kappa \left(\int\limits_{\eta_-}^{\eta_+} I(\xi, \eta) \mathrm{d}\eta \right) \mathrm{d}\xi}{\iint\limits_G |E(u, v)|^2 \mathrm{d}u \, \mathrm{d}v}, \tag{6.18}$$

где $E(u,v)$ - комплексная амплитуда, падающего на фокусатор пучка света. В случае фокусировки в плоскую область энергетическая эффективность ДОЭ оценивается также формулой (6.18), только вместо поперечных границ фокальной линии (η_+, η_-) необходимо использовать требуемые поперечные границы фокальной области.

При исследовании фокусатора необходимо оценить световое поле в фокальной области, по которому легко определяются дифракционные характеристики ДОЭ.

Рассмотрим оптическую схему рисунка 6.1, в которой ДОЭ с функцией амплитудно-фазового пропускания $T(u,v)=A(u,v)\exp[i\varphi(u,v,)]$ освещается световым пучком с комплексной амплитудой $E(u,v)=B(u,v)\exp[i\psi(u,v,)]$, а световое поле w наблюдается вблизи плоскости Π, отстоящей на расстоянии f от плоскости элемента. Точку наблюдения будем характеризовать координатами (x, y, z), где z - расстояние до плоскости Π. При условии, что элементы разрешения ДОЭ много больше длины волны освещающего пучка, для оценки светового поля $w(x,y,z)$ можно воспользоваться интегралом Кирхгофа

$$w(x,y,z)=\frac{f+z}{i\lambda}\times$$

$$\times\iint\limits_G \frac{E(u,v)\cdot T(u,v)\cdot \exp\left\{ik\sqrt{(f+z)^2+(u-x)^2+(v-y)^2}\right\}}{\left[(f+z)^2+(u-x)^2+(v-y)^2\right]}\mathrm{d}u\,\mathrm{d}v, \tag{6.19}$$

где G - область, занимаемая ДОЭ,

$k=2\pi/\lambda$, λ - длина волны освещающего пучка.

Возможности вычисления на современных персональных компьютерах этого интеграла с быстроосциллирующим (ввиду малости λ) подынтегральным выражением представляются весьма ограниченными. Имеется возможность существенно уменьшить вычислительные трудности, используя следующее разложение в ряд:

$$\sqrt{(f+z)^2+(u-x)^2+(v-y)^2}\cong(f+z)+$$

$$+\frac{(u-x)^2+(v-y)^2}{2(f+z)}-\frac{\left[(u-x)^2+(v-y)^2\right]^2}{8(f+z)^3}+\ldots \tag{6.20}$$

Оставляя только первые два члена разложения (6.20) при аппроксимации фазовой составляющей подынтегрального выражения (6.19) и только первый член разложения (6.20) при аппроксимации амплитудной составляющей, приходим к приближению, которое называется приближением Френеля. Фазовая составляющая подынтегрального выражения (6.19) более чувствительна к аппроксимации, чем амплитудная. Так, использование двух первых членов при аппроксимации фазовой составляющей (6.19) является более строгим ограничением, чем использование одного члена при аппроксимации амплитудной составляющей:

$$w(x,y,z)=\frac{\exp\left[ik(f+z)\right]}{i\lambda(f+z)}\iint\limits_G E(u,v)\cdot T(u,v)\times$$

$$\times\exp\left\{\frac{ik(u-x)^2+(v-y)^2}{2(f+z)}\right\}\mathrm{d}u\,\mathrm{d}v. \tag{6.21}$$

Исследованию точности приближения Френеля посвящены многие работы. В них сравнением результатов прямого численного интегрирования по формулам (линейный случай)

$$w_1(x,z)=\sqrt{\frac{i}{\lambda}}\int\limits_{-a}^{a}\frac{E(u)\cdot\exp(iks)}{\sqrt{s}}\cdot\frac{(f+z)}{s}\mathrm{d}u \tag{6.22}$$

и

$$w_2(x,z)=\sqrt{\frac{i}{\lambda(f+z)}}\cdot\exp\left[ik(f+z)\right]\times$$
$$\times\int_{-a}^{a}E(u)\cdot\exp\left[\frac{ik(x-u)^2}{2(f+z)}\right]du\,, \qquad (6.23)$$

где

$$s=\sqrt{(f+z)^2+(x-u)^2}, \qquad (6.24)$$

показано, что при $ks >> 1$ и достаточно больших F-числах ($F = f/2a \geq 12$) результаты для модуля амплитуды $|w_j(x,z)|$, j=1,2 отличаются не более чем на 2%.

Учитывая предложенную модель ДОЭ (6.2), интеграл (6.19) можно представить следующим образом

$$w(x,y,z)=\frac{f+z}{i\lambda}\sum_{n\in I_N}t_n\iint_{G_n}\frac{E(u,v)\cdot\exp\{iks\}}{s^2}\,\mathrm{d}u\,\mathrm{d}v\,, \qquad (6.25)$$

где

$$s=\sqrt{(f+z)^2+(x-u)^2+(v-y)^2}, \qquad (6.26)$$

В приближении Френеля выражение (6.25) можно представить следующим образом:

$$w(x,y,z)=\frac{f+z}{i\lambda}\sum_{n\in I_N}\frac{t_n\exp(iks)}{s_n^2}\iint_{G_n}E(u,v)\times$$
$$\times\exp\left\{\frac{ik\left[(u_n-u)(2x-u-u_n)+(v_n-v)(2y-v-v_n)\right]}{2s_n}\right\}\mathrm{d}u\,\mathrm{d}v\,, \qquad (6.27)$$

где

$$s_n=\sqrt{(f+z)^2+(x-u_n)^2+(y-v_n)^2}, \qquad (6.28)$$

(u_n, v_n) - "центральная" точка модуля G_n.

Интенсивность поля оценивается как $I(x,y,z) = |w(x,y,z)|^2$, где $w(x,y,z)$ определяется в соответствии с выражением (6.27), интеграл в котором вычисляется на основе базовых аналитических решений задачи дифракции на отдельных модулях фокусатора. При исследовании принципиально новых возможностей фокусирующих ДОЭ важно выявить ограничения (на физические параметры) геометрооптического подхода, положенного в основу расчета фокусатора, наметить пути оптимизации фазовой функции. В ходе проектирования, когда уже заданы физические параметры оптической схемы и известны принципиальные возможности проектируемого ДОЭ, важно выяснить влияние каждой произведенной операции (см. рис. 6.2) - дискретизации, кодирования, квантования - на дифракционные характеристики синтезируемого фокусатора. После каждого расчетного этапа синтеза ДОЭ (1-3) в ходе диалога с программой проектировщику важно получить наглядное изображение распределения интенсивности в фокальной области и оценки дифракционных параметров фокусатора. После анализа полученных данных проектировщик решает, продолжить ли процесс синтеза или, изменив параметры дискретизации и квантования, вернуться на несколько этапов назад.

6.1.4. Базовые аналитические решения прямой задачи фокусировки - задачи дифракции на фокусаторе

Используемые при расчете поля в фокальной области ДОЭ базовые аналитические решения определяются структурой и формой модулей исследуемого фокусатора.

При использовании для изготовления ДОЭ с вращательной симметрией (6.11) фотопостроителя с круговым сканированием и при аксиальной симметрии освещающего пучка $E(u,v) = E(r) = B(r) \cdot exp[i\psi(r)]$ выражение (6.27) будет выглядеть так

$$w(\rho, z) = \frac{f+z}{i\lambda} \cdot \sum_{n=1}^{N} \frac{t_n \cdot \exp\left[ik\left(L_n + \frac{-\bar{r}_n^2 + 2\rho\bar{r}_n}{2L_n}\right)\right]}{L_n^2} \times$$
$$\times \int_{r_{n-1}}^{r_n} E(r) \cdot \int_0^{2\pi} \exp\left\{ik\frac{r^2 - 2\rho r\cos(\Theta - \varphi)}{2L_n}\right\} r\,\mathrm{d}r\,\mathrm{d}\Theta. \tag{6.29}$$

Здесь

$$L_n = \sqrt{(f+z)^2 + (\rho - \bar{r}_n)^2}, \qquad \bar{r}_n = \frac{r_n + r_{n-1}}{2},$$
$$\rho = \sqrt{x^2 + y^2}, \qquad r = \sqrt{u^2 + v^2},$$

N - число кольцевых модулей ДОЭ;
r_n - внешний радиус n-го кольца с комплексным пропусканием t_n;
$r_0 = 0$, $x = \rho \cdot \cos\varphi$, $y = \rho \cdot \sin\varphi$, $u = r \cdot \cos\Theta$, $v = r \cdot \sin\Theta$.

Проинтегрировав (6.29) по Θ, получаем

$$w(\rho, z) = -ik(f+z) \cdot \sum_{n=1}^{N} \frac{t_n \cdot \exp\left[ik\left(L_n - \frac{\bar{r}_n^2 - 2\rho\bar{r}_n}{2L_n}\right)\right]}{L_n^2} \times$$
$$\times \int_{r_{n-1}}^{r_n} E(r) \cdot \exp\left\{\frac{ikr^2}{2L_n}\right\} \cdot J_0\left(\frac{k\rho r}{L_n}\right) r\,\mathrm{d}r\,. \tag{6.30}$$

Здесь $J_0(\xi)$ - функция Бесселя 1-го рода нулевого порядка. Аппроксимируя амплитудную составляющую фокусируемого пучка, получаем вместо (6.30)

$$w(\rho, z) = -ik(f+z) \cdot \sum_{n=1}^{N} \frac{t_n \cdot \exp\left[ik\left(L_n - \frac{\bar{r}_n^2 - 2\rho\bar{r}_n}{2L_n}\right)\right] \cdot B(\bar{r}_n)}{L_n^2} \times$$
$$\times \int_{r_{n-1}}^{r_n} \exp\left\{i\left[\psi(r) + \frac{kr^2}{2L_n}\right]\right\} \cdot J_0\left(\frac{k\rho r}{L_n}\right) r\,\mathrm{d}r\,. \tag{6.31}$$

Здесь $B(\bar{r}_n)$ - амплитуда в центре n-го кольца.

В зависимости от числа колец разрешения N и физических параметров оптической схемы возможно несколько способов вычисления интеграла в уравнении (6.31).

В силу линейности интеграл в (6.31) можно представить как

$$\int_{r_{n-1}}^{r_n} \exp\left\{i\left[\psi(r) + \frac{kr^2}{2L_n}\right]\right\} \cdot J_0\left(\frac{k\rho r}{L_n}\right) r\,\mathrm{d}r = I_n - I_{n-1}\,, \tag{6.32}$$

где

$$I_k = \int_{0}^{r_k} \exp\left\{i\left[\psi(r) + \frac{kr^2}{2L_n}\right]\right\} \cdot J_0\left(\frac{k\rho r}{L_n}\right) r\,\mathrm{d}r\,, \quad k = n-1,\ n. \tag{6.33}$$

Для плоского или сходящегося сферического пучков $E(r)$ интеграл (6.33) можно вычислить через функции Ломмеля

$$I_k = \frac{r_k^2}{q_k} \cdot \begin{cases} \exp\left(-\frac{iq_k}{2}\right) \cdot U_1(q_k, p_k) + iU_2(q_k, p_k), & при \quad \left|\frac{q_k}{p_k}\right| \le 1; \\ -i\exp\left(\frac{ip_k^2}{2q_k}\right) - \exp\left(-\frac{iq_k}{2}\right) \cdot \left[V_1(q_k, p_k) - iV_0(q_k, p_k)\right] \\ \qquad при \left|\frac{q_k}{p_k}\right| > 1; \end{cases} \qquad (6.34)$$

где $k = n-1, n$; $p_k = (kr_k \rho)/L_n$; $q_k = k(r_k)^2(1/L_0-1/L_n)$;
функции Ломмеля -

$$\begin{cases} U_m = \sum_{l=0}^{\infty} (-1)^l \left(\frac{\xi}{\eta}\right)^{2l+m} \cdot J_{m+2l}(\eta) \\ V_m = \sum_{l=0}^{\infty} (-1)^l \left(\frac{\eta}{\xi}\right)^{2l+m} \cdot J_{m+2l}(\eta) \end{cases} \qquad (6.35)$$

f - фокусное расстояние ДОЭ;
L_0 - фокусное расстояние: точка $z = L_0 - f$ - точка фокусировки ($2a/L_0 \ll 1$) падающего на ДОЭ пучка. Для плоского пучка $L_0 = \infty$.
$J_l(\eta)$ - функция Бесселя первого рода l-го порядка.

Описанный метод расчета интеграла (6.33) обобщает и уточняет известные формулы. Основанный на нем способ вычисления поля от ДОЭ (6.31)-(6.35) имеет существенное преимущество по сравнению с методом, используемым в работе для вычисления поля от плоской линзы. В ряде работ для вычисления поля от круглого отверстия используется метод граничной дифрагирующей волны (ГДВ), непригодный для вычисления поля в каустике. Учитывая протяженность каустики фокусатора, ГДВ-метод является неприсмлемым для расчета поля от фокусирующих ДОЭ.

Однако методы, основанные на вычислении поля от круглого отверстия, удобно применять при малом числе кольцевых модулей N. При большом N целесообразно использование других способов, например, квадратурных методов расчета интеграла (6.32).

Для плоского освещающего пучка расчет поля вблизи оптической оси целесообразно основывать на модификации схемы Филона с целью вычисления интеграла (6.32). При этих условиях амплитуда осцилляции функции Бесселя $J_0(k\rho r/L_n)$ много меньше амплитуды осцилляции действительной и мнимой частей экспоненты $exp(ikr^2/2L_n)$, а период осцилляции функции Бесселя во столько раз больше соответствующего периода слагаемых экспоненты (по r), во сколько раз ρ меньше, чем $r/2$. В этом случае $J_0(k\rho r/L_n)$ на выбранном интервале интегрирования (r_{n-1}, r_n) можно заменить константой:

$$\int\limits_{r_{n-1}}^{r_n} \exp\left\{\frac{ikr^2}{2L_n}\right\}\cdot J_0\left(\frac{k\rho r}{L_n}\right) r\,\mathrm{d}r \cong J_0\left(\frac{k\rho \bar{r}_n}{L_n}\right)\cdot \int\limits_{r_{n-1}}^{r_n} \exp\left\{\frac{ikr^2}{2L_n}\right\} r\,\mathrm{d}r. \quad (6.36)$$

Последний интеграл в (6.36) легко вычисляется, и даже устраняются неопределенности вычитания:

$$\int\limits_{r_{n-1}}^{r_n} \exp\left\{\frac{ikr^2}{2L_n}\right\} r\,\mathrm{d}r = \frac{2L_n}{k}\cdot \exp\left[\frac{ik\left(r_{n-1}^2 + r_n^2\right)}{4L_n}\right]\cdot \sin\left[\frac{k\left(r_{n-1}^2 - r_n^2\right)}{4L_n}\right]. \quad (6.37)$$

Предлагаемый метод (6.36)-(6.37) точнее обычного квадратурного метода вычисления интеграла (6.32).

Простейший, но менее точный, способ вычисления интеграла (6.32) использует разложение интеграла в ряд по степеням малого параметра $\delta = r_n - r_{n-1}$ согласно соотношению

$$\begin{aligned}
&\int\limits_{c}^{c+\delta} \exp\left\{i\left[\psi(r) + \frac{kr^2}{2L_n}\right]\right\}\cdot J_0\left(\frac{k\rho r}{L_n}\right) r dr = \exp\left\{i\left[\psi(c) + \frac{kc^2}{2L_n}\right]\right\}\times \\
&\times J_0\left(\frac{k\rho c}{L_n}\right) c\cdot\delta + \frac{1}{2}\exp\left\{i\left[\psi(c) + \frac{kc^2}{2L_n}\right]\right\}\left\{J_0\left(\frac{k\rho c}{L_n}\right) - J_1\left(\frac{k\rho c}{L_n}\right) c + \right. \qquad (6.38) \\
&\left. + iJ_0\left(\frac{k\rho c}{L_n}\right)\left[\left.\frac{\partial\psi(r)}{\partial(r)}\right|_{r=c} + \frac{kc}{L_n}\right] c\right\}\cdot\delta^2 + \ldots
\end{aligned}$$

Последний способ целесообразно использовать при больших N в случае, когда неприемлемо вычисление по модифицированной схеме Филона (6.36)-(6.37).

При исследовании средствами вычислительного эксперимента цилиндрических фокусаторов (6.13) рассматривается двумерная задача, описываемая интегралом типа (6.22):

$$w(x, z) = \sqrt{\frac{i}{\lambda}}\cdot \int\limits_{-a}^{a} \frac{E(u)\cdot T(u)\cdot \exp(iks)}{\sqrt{s}}\cdot \frac{f+z}{s}\,\mathrm{d}u\,, \qquad (6.39)$$

где $s = \sqrt{(f+z)^2 + (x-u)^2}$,

$E(u) = A(u)\cdot\exp[i\psi(u)]$ - фокусируемый пучок.

В приближении Френеля для цилиндрических ДОЭ (6.13) имеем

$$w(x,z)=\sqrt{\frac{i}{\lambda}}\cdot(f+z)\cdot\sum_{n=1}^{N}\frac{\exp\left\{ik\left[s_n-\frac{(x-\bar{u}_n)^2}{2s_n}\right]\right\}}{\sqrt[3]{s_n}}\times$$
$$\times t_n\cdot A(\bar{u}_n)\cdot\int_{u_{n-1}}^{u_n}\exp\left\{i\left[\psi(u)+\frac{k(x-u)^2}{2s_n}\right]\right\}\mathrm{d}u\,, \qquad (6.40)$$

где

$$s_n=\sqrt{(f+z)^2+(x-\bar{u}_n)^2},\qquad \bar{u}_n=\frac{u_n+u_{n-1}}{2}.$$

Существует целый ряд методов вычисления интеграла в (6.40) для плоского и сходящегося цилиндрического освещающих пучков.

Наиболее простой способ заключается в сведении интеграла в (6.40) к разности двух значений интегралов Френеля, программа вычисления которых имеется в Библиотеке стандартных программ. Например, для плоского освещающего пучка

$$\int_{u_{n-1}}^{u_n}\exp\left\{\frac{ik(x-u)^2}{2s_n}\right\}\mathrm{d}u=\sqrt{\frac{\lambda s_n}{2}}\cdot\Bigg\{\mathrm{sign}(u_n-x)\times$$
$$\times\left[C\left[\frac{k(u_n-x)^2}{2s_n}\right]+iS\left[\frac{k(u_n-x)^2}{2s_n}\right]\right]- \qquad (6.41)$$
$$-\mathrm{sign}(u_n-x)\cdot\left[C\left[\frac{k(u_{n-1}-x)^2}{2s_n}\right]+iS\left[\frac{k(u_{n-1}-x)^2}{2s_n}\right]\right]\Bigg\},$$

где

$$\mathrm{sign}(y)=\begin{cases}-1, & y<0\\ 0, & y=0\\ 1, & y>0\end{cases}, \qquad (6.42)$$

а

$$\begin{cases}C(y)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^y\frac{\cos t}{\sqrt{t}}\,\mathrm{d}t\\ S(y)=\dfrac{1}{\sqrt{2\pi}}\cdot\displaystyle\int_0^y\frac{\sin t}{\sqrt{t}}\,\mathrm{d}t\end{cases} \qquad (6.43)$$

- интегралы Френеля.

Данный способ приемлем в случае малого числа элементов дискретизации N. Для больших N при вычислении интеграла в (6.40) возможно применение обычных квадратурных методов (например, метода Симпсона), но наиболее эффективно использование метода локальной линейной аппроксимации фазовой функции подынтегрального выражения. Для плоского освещающего пучка в приближении Френеля реализация этого метода выглядит следующим образом

$$\int_{u_{n-1}}^{u_n} \exp\left\{\frac{ik(x-u)^2}{2s_n}\right\} \mathrm{d}u = \sum_{j=1}^{k} \int_{u_{n_{j-1}}}^{u_{n_j}} \exp\left\{\frac{ik(x-u)^2}{2s_n}\right\} \mathrm{d}u \approx$$

$$\approx \sum_{j=1}^{k}\left(u_{n_j} - u_{n_{j-1}}\right) \times \sin c\left[\frac{\pi\left(u_{n_{j-1}} - x\right)\left(u_{n_j} - u_{n_{j-1}}\right)}{\lambda s_n}\right] \times \tag{6.44}$$

$$\times \exp\left[\frac{i\pi\left(u_{n_{j-1}} - x\right)\left(u_{n_j} - x\right)}{\lambda s_n}\right],$$

$$u_{n_0} = u_{n-1}, \qquad u_{n_k} = u_n.$$

Результаты, аналогичные (6.41), (6.44), легко получаются и для сходящегося цилиндрического пучка $E(u)$, для которого

$$\psi(u) = -\frac{ku^2}{2f}. \tag{6.45}$$

В общем случае поле $w(x,y,z)$ в фокальной области фокусатора с прямоугольным растром дискретизации в соответствии с (6.14) вычисляется следующим образом

$$w(x,y,z) = \frac{(f+z)}{i\lambda} \times$$

$$\times \sum_{n=1}^{N_1} \sum_{l=1}^{N_2} \frac{t_{nl} \exp\left\{ik\left[s_{nl} - \frac{(u_n - x)^2 + (v_l - y)^2}{2s_{nl}}\right]\right\}}{s_{nl}^{\,2}} \times \tag{6.46}$$

$$\times \int_{u_n - \frac{\delta u}{2}}^{u_n + \frac{\delta u}{2}} \int_{v_l - \frac{\delta v}{2}}^{v_l + \frac{\delta v}{2}} E(u,v) \cdot \exp\left\{\frac{ik\left[(x-u)^2 + (y-v)^2\right]}{2s_{nl}}\right\} \mathrm{d}u\, \mathrm{d}v,$$

где

$$s_{nl}^2 = (f+z)^2 + (x-u_n)^2 + (y-v_l)^2. \tag{6.47}$$

Для плоского или сферического освещающих пучков в приближении Френеля

$$E(u,v) = A(u,v)\cdot\exp\{i[\Psi_1(u) + \Psi_2(v)]\} \tag{6.48}$$

и интеграл в (6.46) факторизуется, т.е.

$$w(x,y,z) = \frac{f+z}{i\lambda}\sum_{n=1}^{N_1}\int\limits_{u_n-\frac{\delta u}{2}}^{u_n+\frac{\delta u}{2}} \exp\left\{i\left[\Psi_1(u)+\frac{k(x-u)^2}{2s_n}\right]\right\}\mathrm{d}u \times$$

$$\times\sum_{l=1}^{N_2}\frac{A(u_n,v_l)\cdot t_{nl}}{s_{nl}^2}\cdot\exp\{ik(f+z)\}\times \tag{6.49}$$

$$\times\int\limits_{v_l-\frac{\delta v}{2}}^{v_l+\frac{\delta v}{2}} \exp\left\{i\left[\Psi_2(v)+\frac{k(y-v)^2}{2s_l}\right]\right\}\mathrm{d}v,$$

$$s_n = \sqrt{(f+z)^2+(x-u_n)^2}, \qquad s_l = \sqrt{(f+z)^2+(y-v_l)^2}.$$

Из формулы (6.49) видно, что для вычисления поля в точке (x, y, z) необходимо вычислить N_1 интегралов

$$A_n(x) = \int\limits_{u_n-\frac{\delta u}{2}}^{u_n+\frac{\delta u}{2}} \exp\left\{i\left[\Psi_1(u)+\frac{k(x-u)^2}{2s_n}\right]\right\}\mathrm{d}u \tag{6.50}$$

и N_2 интегралов

$$B_l(y) = \int\limits_{v_l-\frac{\delta v}{2}}^{v_l+\frac{\delta v}{2}} \exp\left\{i\left[\Psi_2(v)+\frac{k(y-v)^2}{2s_l}\right]\right\}\mathrm{d}v. \tag{6.51}$$

Таким образом, для вычисления поля в одной точке требуется найти $N_1 + N_2$ интегралов типа интеграла Френеля, а не $N_1 \cdot N_2$, хотя полной факторизации сумм добиться невозможно, так как коэффициент амплитудно-фазового пропускания прямоугольных модулей t_{nl} свойствами факторизации обладать не обязан. В случае, когда необходимо найти матрицу отсчетов поля $\{w(x_k, y_m, z_o), k = 1,..., K; m = 1,..., M\}$ в сечении фокальной области $z = z_o$, достаточно вычислить не $K{\cdot}M{\cdot}N_1{\cdot}N_2$ интегралов, а только $K{\cdot}N_1$ интегралов $A_n(x_k)$ по формуле (6.50) и $M{\cdot}N_2$ интегралов $B_l(y_m)$ по формуле (6.51). Интегралы (6.50), (6.51) вычисляются в соответствии с ранее описанными методами (6.41)-(6.44).

Следует отметить, что предлагаемый подход (6.49) точнее, чем широко распространенные методы дифракционного расчета, использующие быстрое преобразование Фурье (БПФ). Основой дифракционного расчета с использованием БПФ являются формулы (4.39), (4.40), (4.62)-(4.64). Подход (6.49) точнее за счет более точного вычисления амплитудного знаменателя ($(f+z)/s_{nl}^2$ вместо $1/(f+z)$). Но основным недостатком метода, использующего БПФ, является пропорциональность размеров (и одинаковая размерность матриц) фокусатора и исследуемого сечения фокальной области. Следовательно, на основе БПФ затруднено исследование энергетических характеристик большинства фокусаторов.

Как видно из анализа предлагаемых методов расчета поля от фокусаторов, решение задачи дифракции на фокусирующем ДОЭ сводится к суперпозиции множества базовых аналитических решений задачи дифракции на элементарных апертурах - круглом отверстии или щели.

6.1.5. Анализ ограничений физического и математического характера и источников энергетических потерь при фокусировке лазерного излучения дифракционными оптическими элементами

В рамках данного пособия предполагается, что фокусатор относится к объектам дифракционной микрооптики, и, следовательно, для него выполняются условия приближения Френеля скалярной теории дифракции. Это означает, что рассматриваемые методы применимы для исследования дифракционных оптических элементов, имеющих небольшие размеры относительно фокусного расстояния и шаг дискретизации, который превосходит рабочую длину волны.

Аналитические дифракционные исследования фокальных областей для простейших геометрооптических фазовых функций фокусирующих оптических элементов показывают неизбежность потери 10-12% энергии фокусируемого пучка, не попадающей в фокальную область дифракционной ширины. Потери обусловлены несовершенством геометрооптических фазовых функций, рассеянием за счет эффектов дифракции на апертуре оптического элемента. Сравнительный анализ проведенных исследований различных оптических элементов показывает, что дифракционные потери тем больше, чем меньше область фокусировки. В частности, для дифракционной ширины области фокусировки, определяемой по уровню спада 0,1 относительно максимальной интенсивности, энергетическая эффективность составляет: для безаберрационной линзы - 81-82%, для фокусатора в кольцо - 87-88%, для фокусатора в отрезок - 89-90%. С другой стороны, чем больше область фокусировки, тем сложнее добиться небольших значений среднеквадратичного отклонения получающегося распределения интенсивности от требуемого.

В связи с технологическими погрешностями изготовления дифракционного микрорельефа, а также неизбежными при расчете и создании фотошаблонов операциями дискретизации и квантования фазовой функции такие значения энергетической эффективности практически недостижимы. Кроме представляемых в данной главе средств моделирования для оценки дополнительных энергетических потерь разработаны и применяются различные методы, в основе которых лежит анализ Фурье.

Из теории сигналов известно: для того, чтобы дискретизация была обратимой, т.е. чтобы имелась возможность по функции $f(n)$ восстановить функцию $f(u)$, следует выполнить условие

$$\Delta < \frac{1}{2F} \tag{6.52}$$

где F - ширина спектра функции $f(u)$, а Δ - шаг дискретизации.

Неравенство (6.52) является отправным в выборе шага дискретизации, однако оно не указывает количественно, каким же должен быть он выбран. Имеются [2] некоторые дополнительные количественные оценки. Для этого на фазовую функцию $\varphi(u,v)$ накладывается достаточно очевидное условие допустимости дискретизации в виде

$$\left|\delta_1 \frac{\partial\varphi(u,v)}{\partial u} + \delta_2 \frac{\partial\varphi(u,v)}{\partial v}\right| << 2\pi \tag{6.53}$$

или сокращенно

$$\left|\nabla_\delta \varphi\right| << 2\pi .$$

Влияние дискретизации может быть охарактеризовано максимальной или среднеквадратичной погрешностью (флуктуацией) фазы, выражения для которых соответственно имеют вид

$$\Delta\varphi_{\max} = \max_{\mathbf{u}\in G}\left|\Delta\varphi(\mathbf{u})\right|, \tag{6.54}$$

$$\Delta\varphi_c = \left[\frac{1}{|G|}\int\limits_G \left|\Delta\varphi(\mathbf{u})\right|^2 \mathbf{d}^2 u\right]^{\frac{1}{2}}, \tag{6.55}$$

где $|G|$ - площадь области G

В [2] приведены оценки погрешности кусочно-постоянной аппроксимации функции $\varphi(\mathbf{u})$ для прямоугольного сканирующего пятна (6.3) и гладкой функции $\varphi(u,v)$. Для ДОЭ с прямоугольным растром дискретизации (6.14) в [2] приведены следующие оценки для максимальной погрешности

$$\Delta\varphi_{\max} \le \frac{1}{2}\max_{\mathbf{u}\in G}\left|\nabla_\delta \varphi(\mathbf{u})\right|_+, \tag{6.56}$$

где символ $|\cdot|_+$ означает сумму модулей компонент вектора, и для среднеквадратичной погрешности дискретизации фазы $\Delta\varphi_s$

$$\Delta\varphi_{\mathbf{s}} = \left\{\frac{1}{12|G|}\int_G \left[\nabla_\delta \varphi(\mathbf{u})\right]^2 \mathbf{d}^2\mathbf{u}\right\}^{\frac{1}{2}}. \tag{6.57}$$

Оценки (6.57), (6.56) позволяют подсчитать погрешности дискретизации непосредственно по параметрам дискретизации δ_1,δ_2 и первым производным фазы.

Для цилиндрического ДОЭ (6.13) с $\varphi(\mathbf{u}) \equiv \varphi(u)$, $|u| \le a$,

$$\Delta\varphi_{\max} \le \frac{\delta}{2}\max_{|u|\le a}\left|\varphi'(\mathrm{u})\right|, \quad \varphi' = \frac{\mathrm{d}\varphi}{\mathrm{d}u}, \tag{6.58}$$

$$\Delta\varphi_{\mathrm{s}} = \left\{\frac{\delta^2}{24a}\int_{-a}^{a}\left[\varphi'(\mathrm{u})\right]^2 \mathrm{d}u\right\}^{\frac{1}{2}}. \tag{6.59}$$

Для радиально-симметричных ДОЭ (6.11) $\varphi(\mathbf{u}) \equiv \varphi(r)$, $r = |\mathbf{u}| \le a$,

$$\Delta\varphi_{\max} \le \frac{\delta}{2}\max_{0\le r\le a}\left|\varphi'(r)\right|, \quad \varphi' = \frac{\mathrm{d}\varphi}{\mathrm{d}r}, \tag{6.60}$$

$$\Delta\varphi_{\mathrm{s}} = \left\{\frac{\delta^2}{6a^2}\int_0^a\left[\varphi'(r)\right]^2 r\,\mathrm{d}r\right\}^{\frac{1}{2}}. \tag{6.61}$$

Оценки [2] дают непосредственную связь возмущений дискретизации с производными гладкой фазовой функции.

Если допустить, что погрешность квантования $\Delta\varphi_k$ принимает равновероятные значения в диапазоне $[-\pi m/M, \pi m/M]$, где $2\pi m$ - максимальная фаза на оптическом элементе (m - целое число), а M - число уровней квантования фазы, то нетрудно вычислить среднеквадратичное значение погрешности квантования фазы

$$\Delta\varphi_s^k = \frac{1}{2\pi m/M} \int_{-\pi m/M}^{\pi m/M} \left(\Delta\varphi^k\right)^2 d\left(\Delta\varphi^k\right) = \frac{1}{3}\left(\frac{\pi m}{M}\right)^2. \qquad (6.62)$$

При больших значениях M величина среднеквадратичной погрешности квантования фазы составляет доли процента. В этом случае эффект квантования фазы может быть сведен к появлению в плоскости ДОЭ дополнительного фазового транспаранта со случайной по полю апертуры фазой, варьирующейся в узком диапазоне $[-\pi m/M, \pi m/M]$, и описывается моделью случайного фазового диффузора.

Для оценки влияния дискретизации фазовой функции на характеристики фокусирующих ДОЭ можно рассмотреть оценки поля от ДОЭ в приближении Фраунгофера, т.е. на больших расстояниях от плоскости ДОЭ.

Пусть r_0 - расстояние от начала координат в плоскости ДОЭ до точки наблюдения (x,y,z), а $\alpha = x/r_0$ и $\beta = y/r_0$ - направляющие косинусы линии. Тогда, если r_0 велико по сравнению с размером области G, то поле в точке (x,y,z) выражается через преобразование Фурье $F(\omega_x, \omega_y)$ функции пропускания $T(u,v)$ ДОЭ.

В случае дифракционного оптического элемента, функция комплексного пропускания которого подверглась дискретизации с прямоугольным растром (6.14), выражение для поля в выходной плоскости имеет вид

$$g(\omega_x,\omega_y) = A\cdot\frac{\exp(ikR_0)}{R_0}\cdot\exp\left[-i\omega_x\left(-A_1+\frac{A_1}{N_1}\right)-i\omega_y\left(-A_2+\frac{A_2}{N_2}\right)\right]\times \qquad (6.63)$$
$$\times P(\omega_x,\omega_y)\cdot Q(\omega_x,\omega_y),$$

где

$$\omega_x = k\alpha, \quad \omega_y = k\beta;$$

A_1. A_2 - размеры оптического элемента;
N_1, N_2 - число отсчетов по осям координат;

$$Q(\omega_x,\omega_y) = \sum_{j_1=0}^{N_1-1}\sum_{j_2=0}^{N_2-1} \tilde{A}(u_{j_1},v_{j_2})\cdot\exp\left(-ij_1\omega_x 2A_1/N_1\right)\times \qquad (6.64)$$
$$\times\exp\left(-ij_2\omega_y 2A_2/N_2\right),$$

$$P(\omega_x,\omega_y) = 4\frac{A_1A_2}{N_1N_2}\cdot\mathrm{sinc}\left[\frac{\omega_x A_1}{N_1}\right]\mathrm{sinc}\left[\frac{\omega_x A_2}{N_2}\right]. \qquad (6.65)$$

Функция $Q(\omega_x,\omega_y)$ (6.64) представляет собой двумерную периодическую функцию с периодом по осям $\pi N_1/A_1$ и $\pi N_2/A_2$ соответственно.

Эта функция описывает хорошо известный эффект размножения спектров, возникающий при дискретизации исходного сигнала. При увеличении N_1, N_2 - числа точек дискретизации исходного дифракционного оптического элемента, расстояние между различными порядками возрастает. На практике число отсчетов необходимо выбирать таким образом, чтобы различные порядки в спектре были пространственно разнесены и не перекрывали друг друга (см. (6.52)). Функция $P(\omega_x, \omega_y)$ по своей форме описывает дифракцию света на прямоугольном растре при дискретизации дифракционного оптического элемента. Наличие этой функции уменьшает долю энергии, приходящуюся на дополнительные спектры. При увеличении числа точек дискретизации на дифракционном оптическом элементе эта доля энергии убывает. Характер убывания величины P с ростом N_1, N_2 определяется зависимостью $P \sim 1/N_1 N_2$. В то же время указанная оценка может служить лишь ориентиром, так как фокусаторы (в отличие, например, от оптических антенн, формирующих требуемую диаграмму направленности) рассчитаны на работу в ближней зоне, для оценки поля в которой необходимо использование приближения Френеля-Кирхгофа.

Более интересные оценки можно получить для влияния числа уровней квантования фазы на дифракционные характеристики фокусирующих ДОЭ. В частности, оценить энергетическую эффективность фокусирующего ДОЭ (6.18) для квантованной фазовой функции, приведенной к интервалу $[0, 2\pi)$.

Квантование фазовой функции дифракционного оптического элемента может быть описано как результат применения нелинейного преобразования $\hat{\Phi}(\varphi)$ к фазовой функции оптического элемента $\varphi(u,v)$. Нелинейное предыскажение в случае квантования по M уровням фазовой функции $\varphi(u,v)$, приведенной к интервалу 2π, имеет вид

$$\hat{\Phi}(\varphi) = \mathrm{int}\left[\frac{\varphi}{\Delta}\right]\Delta; \quad \Delta = 2\pi/M, \tag{6.66}$$

где $\mathrm{int}[x]$ - операция взятия целой части числа x; M - число уровней квантования фазы.

В случае квантования функция комплексного пропускания ДОЭ принимает вид

$$\tilde{A}(u,v) = \exp\left\{i\hat{\Phi}\left[\varphi(u,v)\right]\right\}. \tag{6.67}$$

Для описания работы дифракционного оптического элемента разложим функцию комплексного пропускания в ряд Фурье по переменной $\varphi(u,v)$

$$\bar{A}(u,v)=\sum_{n=-\infty}^{n=\infty} c_n \exp\left[i\frac{n}{m}\varphi(u,v)\right], \tag{6.68}$$

где $2\pi m$ - максимальная фаза на оптическом элементе,

$$c_n=\frac{1}{2\pi m}\int_0^{2\pi m}\exp\left(i\hat{\Phi}(\xi)-i\frac{n}{m}\xi\right)d\xi. \tag{6.69}$$

С учетом (6.68) комплексная амплитуда поля в фокальной плоскости элемента представима в виде

$$W(x,y)=\sum_{n=-\infty}^{n=\infty} c_n W_n(x,y). \tag{6.70}$$

В приближении Френеля-Кирхгофа W_n имеет вид

$$W_n(\mathrm{x})=\frac{1}{i\lambda f}\cdot\exp\left(i\frac{k}{2f}\mathrm{x}^2\right)\cdot\iint_D E_0(\mathbf{u})\cdot\exp\left(i\frac{n}{m}\varphi(\mathbf{u})-i\frac{k\mathbf{xu}}{f}\right)d^2\mathbf{u}, \tag{6.71}$$

где (u,v) - плоскость, непосредственно прилегающая к оптическому элементу, (x,y) - плоскость фокусировки.

Согласно выше приведенным формулам дифракционный оптический элемент формирует дифракционные порядки $W_n(x,y)$, $n=-\infty,\infty$. Требуемый процесс фокусировки в m-ом рабочем (полезном) порядке дифракции описывается величиной $W_m(x,y)$. Из общего вида фазовой функции следует, что слагаемые с другими n соответствуют расфокусированным изображениям с геометрическими размерами, в n/m раз большими.

В общем случае расфокусированные изображения $W_n(x,y)$ влияют на $W_m(x,y)$, при этом согласно (6.70) получается следующая оценка интенсивности

$$I(x,y)\geq\left|c_m\right|^2 I_m(x,y), \tag{6.72}$$

где

$$I(x,y)=\left|W(x,y)\right|^2, \qquad I_m(x,y)=\left|W_m(x,y)\right|^2. \tag{6.73}$$

Для энергетической эффективности ДОЭ имеет место следующая оценка:

$$\varepsilon\geq\left|c_m\right|^2\cdot\varepsilon_0, \tag{6.74}$$

где ε_0 - энергетическая эффективность дифракционного оптического элемента с непрерывной (неквантованной) фазовой функцией $\varphi(u,v)$.

Рассмотрим далее квантование фазовой функции оптического элемента с максимальной фазой, равной 2π (т.е. $m=1$). Пусть число уровней квантования равно M. В этом случае выражение для c_n принимает следующий вид

$$c_n = \frac{1}{2\pi}\sum_{k=0}^{M-1}\int\limits_{\frac{2\pi k}{M}}^{\frac{2\pi(k+1)}{M}} \exp\left(i\frac{2\pi k}{M} - in\xi\right)\mathrm{d}\xi. \qquad (6.75)$$

Интересно отметить, что выражение для мощности дифракционных порядков не содержит фазовой функции исходного дифракционного оптического элемента.

После вычисления интеграла данное выражение принимает вид:

$$c_n = \begin{cases} \frac{iM}{2\pi n}\left[\exp\left(-\frac{2\pi i n}{M}\right) - 1\right], & (n-1) = s\cdot M, \\ 0, \quad (n-1) \neq s\cdot M \end{cases}. \qquad (6.76)$$

где s - целое число.

Рассмотрим случай $n = sM+1$ (при других n согласно (6.76) $c_n = 0$), где s - произвольное целое число, определяющее ненулевые дифракционные порядки:

$$c_{sM+1} = \frac{i}{2\pi\left(s + \frac{1}{M}\right)}\cdot\left[\exp\left(-\frac{2\pi i}{M}\right) - 1\right]. \qquad (6.77)$$

Анализ (6.79) показывает, что при малом числе уровней квантования (M = 2 - 6) существенный вес имеют первые 2 - 3 порядка дифракции (абсолютная величина соответствующих коэффициентов сравнима с $|c_1|^2$). При увеличении числа уровней квантования фазы M дифракционного оптического элемента все коэффициенты, кроме первого, стремятся к нулю. Это означает, что вклад побочных порядков становится очень малым. Результаты исследований [3] плоских осевой и цилиндрической линз, фокусатора в кольцо подтвердили высокую точность оценок (6.74), (6.77) энергетической эффективности в случае неравномерной (оптимальной) дискретизации ДОЭ для исследуемого числа уровней квантования фазы.

Общим недостатком даже известных аналитических оценок является их достаточная громоздкость, которая приводит к необходимости программирования (и отладки программ) расчета по аналитическим формулам эффективности ДОЭ с непрерывной фазовой функцией для каждого типа фокусирующих областей и элементов. Только после получения таких оценок можно воспользоваться формулами (6.74), (6.77) для окончательного расчета эффективности ДОЭ. Поэтому разработка универсальных средств моделирования ДОЭ является весьма актуальной.

Приведенные в настоящем разделе оценки не позволяют рассчитать совокупное влияние дискретизации и квантования фазовой функции фокусирующего ДОЭ на его дифракционные характеристики. Для большинства случаев фокусировки (за исключением простейших) невозможно оценить энергетическую эффективность ε_0 ДОЭ с непрерывной фазовой функцией и, соответственно, воспользоваться формулами (6.74), (6.77) для оценки эффективности квантованного фокусатора. Получить совокупные количественные оценки влияния дискретизации и квантования фазовой функции фокусирующих ДОЭ на качество фокусировки, визуально оценить получаемое распределение интенсивности в фокальной области ДОЭ можно только в рамках вычислительного эксперимента.

Полученные в данном разделе оценки и сформулированные ограничения очерчивают рамки исследовательских задач, решаемых средствами вычислительного эксперимента.

6.1.6. Методика проведения вычислительного эксперимента с ДОЭ на основе растровой модели

Проведенный анализ технологии изготовления фокусаторов и методов дифракционного расчета структуры сфокусированного излучения позволяет построить растровую математическую модель фокусировки лазерного излучения дифракционными оптическими элементами.

В рамках данной модели фокусатор представляется совокупностью модулей (т.е. в основе описания лежит растровая структура ДОЭ), а поле от фокусатора описывается суммой базовых решений задачи дифракции света на отдельных модулях (т.е. используется результат решения задачи дифракции на растре), и это описание применяется для получения дифракционных характеристик всего фокусатора, употребление термина “растровая” представляется обоснованным.

Данная математическая модель учитывает (и включает в себя) следующие важнейшие особенности решаемой задачи:

1) наличие базовых аналитических решений задачи дифракции света на простейших апертурах;

2) модульный характер ДОЭ, позволяющий представить фокусатор в виде совокупности кольцевых, щелевых или прямоугольных элементов разрешения;

3) дискретизацию и квантование фазовой функции, обусловленных современной технологией изготовления ДОЭ;

4) специфику фокусаторов - сложность и протяженность области фокусировки ДОЭ, сложную форму геометрооптической функции фазового пропускания ДОЭ;

5) линейность, приводящую к представлению общего (растрового) решения в виде суперпозиции множества базовых решений задачи дифракции на отдельных модулях фокусатора.

Функциональной реализацией предлагаемой математической модели является численный метод дифракционного расчета светового поля в фокальной области дифракционного оптического элемента, которая может иметь довольно сложную форму. Метод заключается в представлении поля от фокусатора в виде суммы полей, созданных его модулями, и в применении приближения Френеля-Кирхгофа к расчету поля от каждого модуля фокусатора, что позволяет получить решение задачи дифракции на ДОЭ в виде суперпозиции множества аналитических решений задачи дифракции на круглом отверстии или щели.

Разработанная математическая модель определяет методику проведения вычислительного эксперимента с фокусирующими ДОЭ.

На начальном этапе исследований важно провести аналитический дифракционный расчет структуры сфокусированного излучения с учетом конечных размеров фокусатора. Анализ полученных дифракционных соотношений позволяет исследовать ограничения положенного в основу расчета фазовой функции фокусатора геометрооптического подхода, выявив начальные значения физических параметров, при которых происходит разрушение требуемой формы области фокусировки. Однако аналитические исследования можно провести только для простейших фазовых функций. К тому же в рамках аналитического исследования невозможно учесть влияние дискретизации и квантования фазовой функции фокусатора, возникающих в ходе изготовления ДОЭ.

Поэтому на первом этапе вычислительного эксперимента необходимо показать принципиальную работоспособность исследуемого геометрооптического фокусатора с учетом дифракционных эффектов на элементах разрешения ДОЭ, правильность алгоритмической и программной реализации методов расчета фазовой функции фокусирующего ДОЭ. Такое исследование целесообразно проводить для реальных физических параметров и при максимально возможном числе M уровней квантования фазовой функции фокусатора ($M > 15$). При равномерной дискретизации M полагается равным бесконечности или 256.

Следует подчеркнуть, что наиболее качественный фазовый рельеф получается с помощью фотолитографических методов травления. Однако из-за высочайших требований к точности совмещения бинарных фотошаблонов, используемых для формирования фазового рельефа по данной технологии, затруднительно получение большого числа градаций фазовой функции. Поэтому следующий важный этап вычислительного эксперимента - это исследование структуры сфокусированного излучения для малого числа уровней квантования фазы фокусатора ($M = 2; 4$).

Перечисленные важнейшие направления исследования не исчерпывают возможностей разработанного математического обеспечения. В рамках вычислительного эксперимента возможны исследования хроматических аберраций фокусаторов, влияние погрешностей технологии изготовления ДОЭ и изменения физических параметров на структуру сфокусированного излучения и дифракционные характеристики фокусаторов.

Результаты проведенных исследований важно представить в виде привычном для оптика-экспериментатора: полутоновые распределения яркости на экране монитора, графики кривых, изофоты (линии равной интенсивности [1]), изометрии пространственных распределений интенсивности. В результате автоматизированного анализа структуры сфокусированного излучения исследователь должен получить оценки дифракционных параметров фокусатора: энергетической эффективности, ширины фокальной линии, среднеквадратического отклонения полученного распределения интенсивности от требуемого и т.п.

Количество отсчетов в фокальной области ДОЭ должно быть выбрано достаточно большим, чтобы отразить все изменения быстроосциллирующего поля. При этом отображается не только линия (область) фокусировки, но и прилегающие к ней участки, на которых возможны заметные всплески интенсивности. Дифракционные характеристики фокусаторов рассчитываются при более тщательном исследовании поля (с мелким шагом) в пределах фокальной линии, ширина которой определяется по уровню Θ спада интенсивности относительно максимальной (обычно $\Theta = 0{,}1$).

6.2. ПРИМЕРЫ МОДЕЛИРОВАНИЯ ФОКУСАТОРОВ

6.2.1. Моделирование дифракционной цилиндрической линзы

Фазовая функция цилиндрической линзы в параксиальном приближении имеет вид:

$$\varphi(u) = -\frac{ku^2}{2f}, \qquad |u| \le a \tag{6.78}$$

Координаты границ линейных зон при M уровнях квантования фазы дифракционной цилиндрической линзы имеют вид:

$$u_n = \pm\sqrt{2n\lambda f / M}\,. \tag{6.79}$$

При этом общее количество линейных модулей

$$N = \mathrm{int}\left[Ma^2 / \lambda f \right]. \qquad (6.80)$$

На рис. 6.3 представлены распределения нормированной интенсивности $I(\bar{x},0)/I_{ид}$ в фокальной плоскости дифракционной цилиндрической линзы с неравномерной дискретизацией (6.79) для M = 2, 4, 16 уровней градаций фазы.

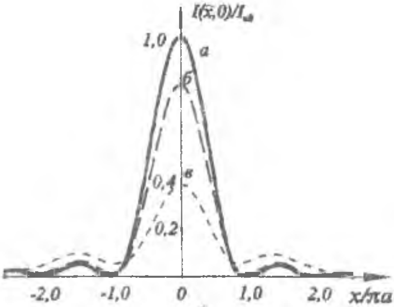

Рис. 6.3. Распределения нормированной интенсивности в фокальной плоскости дифракционной цилиндрической линзы для различного числа уровней квантования M: (а) M = 16, (б) M = 4, (в) M = 2.

Здесь $\bar{x} = x/\pi a$ - безразмерная координата в фокальной плоскости, а

$$I_{ид} = \frac{4a^2}{\lambda f} \cdot I_0 \qquad (6.81)$$

интенсивность в фокусе неквантованной цилиндрической линзы.

Рис. 6.3 показывает, что при уменьшении числа уровней квантования фазы цилиндрической линзы происходит рост «фоновой» интенсивности. Фоновая интенсивности создается паразитными дифракционными порядками и приводит к подъему уровня боковых лепестков.

На рис.6.4 представлены распределения нормированной интенсивности $I(0,z)/I_{ид}$ вдоль "оптической оси" дифракционной цилиндрической линзы с неравномерной дискретизацией (6.79) для M = 2, 4, 16 уровней квантования фазы при параметрах: a = 1 мм; f = 25 мм; λ= 10,6 мкм. Анализ рис. 6.4 показывает, что величина фокального сдвига (смещение точки максимума интенсивности из фокальной плоскости) зависит не только от физических параметров, но и от числа уровней квантования фазы и составляет для исследуемой линзы: $\Delta z = -0{,}05f = -1{,}25$ мм при $M = 16$; $\Delta z = 0$ при $M = 4$; и $\Delta z = 0{,}09f = 2{,}25$ мм при $M = 2$. При этом общий характер распределения и величина фокального сдвига с увеличением M асимптотически приближаются к распределению света от неквантованной цилиндрической линзы. Соответствие полученных результатов с данными известных работ по исследованию рефракционной цилиндрической линзы говорит о корректности использованного метода расчета интеграла Френеля-Кирхгофа.

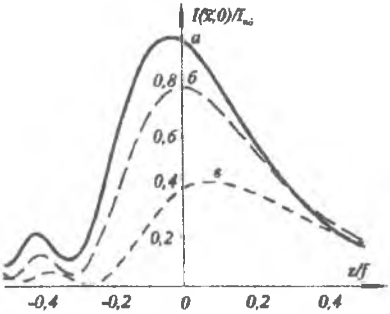

Рис. 6.4. Распределения нормированной интенсивности вдоль "оптической оси" дифракционной цилиндрической линзы для различного числа уровней квантования M: (а) M = 16, (б) M = 4, (в) M = 2.

6.2.1. Вычислительный эксперимент с фокусатором в полукольцо

Фазовая функция геометрооптического фокусатора плоского пучка в полукольцо с радиусом r_0, имеет вид:

$$\varphi(u,v)=\begin{cases}-k\sqrt{f_0^{\,2}+(r-r_0)^2}, & u\ge 0,\\ -k\sqrt{f_0^{\,2}+(r+r_0)^2}, & u<0,\end{cases}\qquad r=\sqrt{u^2+v^2}\le a\,. \tag{6.82}$$

В параксиальном приближении фаза фокусатора сходящегося сферического пучка принимает вид:

$$\varphi(u,v)=\frac{krr_0}{\sqrt{f_0^{\,2}+r_0^2}}\,\mathrm{Sign}(u),\qquad r=\sqrt{u^2+v^2}\le a\,. \tag{3.83}$$

На рис. 6.5, 6.6 представлены результаты исследования фокусатора (3.83) при f_0 = 750 мм; λ = 0,6328 мкм; r_0 = 1 мм; $2a$ = 25,6 мм и при растровой дискретизации фазы на сетке из 128 × 128 отсчетов. Распределение интенсивности в фокальной плоскости фокусатора на рис. 6.5 показывает сохранение всплеска интенсивности на оптической оси, характерного для фокусатора в кольцо.

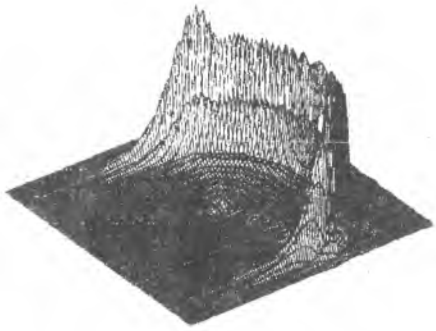

Рис. 6.5. Распределение интенсивности в фокальной плоскости фокусатора в полукольцо.

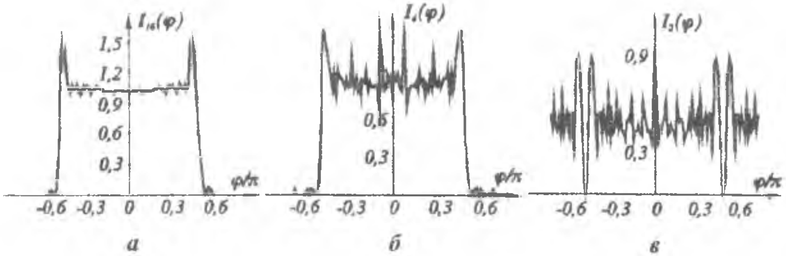

Рис. 6.6. Распределения нормированной интенсивности $I_M(\varphi)=I_M(r_0,\varphi)/I_{16}(r_0,0)$ *по углу вдоль полукольца фокусировки при (а) M = 16; (б) M = 4; (в) M = 2 и параметрах:* f_0 *= 750 мм; λ = 0,6328 мкм;* r_0 *= 1 мм; 2a = 25,6 мм; N = 128.*

Согласно принятой модели фокусатор состоит из $N_1\times N_2$ квадратных модулей, фазовая функция на каждом из которых определяется в соответствии с (3.83) своим значением в центре ячейки. При этом в ходе моделирования для модулей, центры которых лежат за пределами радиуса a, амплитудное пропускание считается равным нулю. Поскольку круг $r\le a$ аппроксимируется малыми квадратными модулями, вдоль полукольца фокусировки возникают всплески интенсивности, особенно заметные на рис. 6.6. На рис. 6.6 в полярных координатах представлены распределения интенсивности в полярных координатах вдоль полукольца фокусировки (при $\rho=r_0$, $-2\pi/3\le\varphi\le 2\pi/3$) при различном числе уровней квантования фазы. Всплески интенсивности по краям полукольца могут быть обусловлены разрезом фазовой функции фокусатора (3.83) при $u=0$. При 4-х уровнях градации фазы (рис. 6.6б) происходит уменьшение среднего уровня интенсивности вдоль линии фокусировки с одновременным увеличением неравномерности распределения интенсивности. Поскольку бинарный ДОЭ позволяет сформировать только центрально-симметричное распределение интенсивности, то при двух уровнях квантования фазы (рис. 6.6в) происходит разрушение области фокусировки и вместо фокусировки в полукольцо происходит фокусировка в кольцо с перепадами интенсивности вдоль него.

ВЫВОДЫ

1. Рассматриваемые в настоящей работе фокусирующие ДОЭ относятся к объектам дифракционной микрооптики, работа которых с высокой точностью моделируется в приближении Френеля-Кирхгофа.

2. Процесс фокусировки лазерного излучения дифракционными оптическими элементами целесообразно описывать растровой математической моделью, учитывающей следующие важнейшие факторы:

а) модульный характер ДОЭ;

б) дискретизацию и квантование фазовой функции ДОЭ;

в) малые по сравнению с фокусным расстоянием размеры модулей ДОЭ;

г) наличие базовых аналитических решений задачи дифракции света на элементарных отверстиях.

3. В рамках растровой модели решение задачи дифракции на фокусаторе представляется в виде суперпозиции множества аналитических решений задачи дифракции на круглом отверстии или щели, что позволяет провести численные исследования структуры светового поля вдоль сложной линии фокусировки ДОЭ.

4. Алгоритмические средства исследования фокусирующих ДОЭ обязаны обеспечивать возможность управления физическими параметрами и параметрами дискретизации фокусатора в ходе вычислительного эксперимента, а также получение, хранение и такую обработку данных большого объема, которая учитывает значительное число вариантов параметров исследуемого поля, наличие набора критериев качества, требования естественности восприятия результатов расчета, организацию диалога с исследователем и служит для установления связи между внешними и внутренними параметрами проектирования ДОЭ.

5. Актуальной является задача исследования фокусатора с малым (2-4) числом градаций фазового микрорельефа, порождаемого технологией изготовления ДОЭ.

6. Исследование методом вычислительного эксперимента фокусатора в полукольцо позволило продемонстрировать влияние параметров дискретизации ДОЭ на структуру сфокусированного излучения. Установлено полное разрушение области фокусировки при двух уровнях градации фазы.

ГЛАВА 7.
ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ МОДЕЛИРОВАНИЯ ОПТИЧЕСКИХ СИСТЕМ

7.В. ВВЕДЕНИЕ

Имеется достаточно большое количество программных продуктов, предназначенных для моделирования оптических систем. Все программные средства, используемые для проектирования оптических устройств, в той или иной степени обладают возможностями для моделирования конструируемых оптических схем. В зависимости от своих расчетных возможностей и наполнения баз данных, содержащих программное описание выпускаемых мировыми производителями оптических элементов и источников света, существующие коммерческие программы сильно отличаются по цене. В частности, ряд программных продуктов (CODE V, OSLO), позволяют проектировать только изображающие оптические системы, например, объективы фотоаппаратов. Наиболее мощные программные комплексы (TracePro Expert, ASAP) позволяют проектировать не только изображающие, но и освещающие, фокусирующие и другие оптические системы, однако стоимость таких программных продуктов доходит до 40 тысяч долларов США (ASAP).

7.1. КРАТКИЙ ОБЗОР ПРОГРАММНЫХ ПРОДУКТОВ

В данном параграфе (Табл. 7.1) приведены краткие характеристики ряда известных программных продуктов, а в последующих параграфах дано более подробное описание программного обеспечения, имеющегося в распоряжении Института компьютерных исследований Самарского государственного аэрокосмического университета (СГАУ).

Таблица 7.1. Список некоторых программных продуктов

Название	Описание	Производитель
OPAL	Одна из самых известных российских программ для расчета и оптимизации оптических систем.	Санкт-Петербургский государственный институт информационных технологий, точной механики и оптики (технический университет) **http://aco.ifmo.ru/**
SARO	Программа для расчета и оптимизации оптических систем.	Всероссийский научный центр «Государственный оптический институт им. С.Н.Вавилова» (Санкт-Петербург)

ZEMAX, LensVIEW (ZELUM, ZEBASE)	ZEMAX - самая известная и популярная на сегодняшний день программа для расчета оптики на персональных компьютеров. Эта программа позволяет анализировать оптические системы на основе последовательного или непоследовательного расчета лучей, выполнять глобальную и локальную оптимизацию параметров оптической системы. Программа обладает всеми необходимыми возможностями, позволяющими проектировать современные оптические системы.	Компания Focus Software, Inc. Доступны демонстрационные версии: ZEMAX, ZELUM и LensVIEW. **http://www.focus-software.com/** **http://www.optima.co.uk/** Модули ZEBASE и LensVIEW позволяют работать с самой обширной библиотекой оптических систем, которая создана и постоянно обновляется специалистами компании. Дополнительный программный модуль ZELUM помогает разработать осветительную оптическую систему.
CODE V, LightTools	Программа CODE V, разрабатываемая уже на протяжении нескольких десятилетий, обладает самым обширным арсеналом средств для синтеза, анализа и оптимизации оптических систем. Представляет интерес и относительно новый программный продукт LightTools, созданный специалистами этой компании.	Optical Research Associates (ORA) является лидером в области компьютерного проектирования оптических систем. **http://www.opticalres.com/**
OSLO	OSLO - одна из старейших программ для проектирования оптических систем. Кроме информации о возмож-	Фирма Sinclair Optics. **http://www.sinopt.com/** Права на распространение и дальнейшее развитие программы переданы Lambda Research Corporation.

	ностях программы на сайте можно найти учебные материалы по расчету оптики с использованием OSLO. Доступна бесплатная версия OSLO LT.	
TracePro	TracePro - программа расчета лучей с учетом поглощения, отражения, преломления, рассеяния и дифракции света при распространении через оптическую систему.	Компания Lambda Research Corporation предлагает программное обеспечение для проектирования и анализа работы оптических систем. **http://www.lambdares.com/**
OSD Synosys		**http://www.gwi.net/osd**
ASAP, ReflectorCAD, APART	Программа ASAP (Advanced Systems Analysis Program) широко используется для проектирования сложных изображающих и осветительных систем специального назначения. Программа APART предназначена для анализа паразитных засветок. Компания также предлагает специализированный программный продукт ReflectorCAD для конструирования отражателей.	Компания Breault Research Organization, известная также как BRO, является авторитетным разработчиком оптического программного обеспечения. **http://www.breaut.com** На сайте доступна демонстрационная версия программы ReflectorCAD. На сайте компании работает электронная библиотека источников освещения (BRO Light Source Library), которая содержит конструктивные параметры и математические модели всевозможных источников оптического излучения.
PARAXIA-Plus, OPTEC, SIGRAPH-OPTIK/CAOS	PARAXIA-Plus - программа для проектирования лазерных систем и моделирования распространения пучков в лазерных резонаторах, OPTEC - программа для проектирования оптиче-	Компания SCIOPT Enterprises. **http://www.sciopt.com** Компания предлагает широкий спектр оптических программных продуктов, в том числе: PARAXIA-Plus, OPTEC, SIGRAPH-OPTIK/CAOS

	ских систем и расчета лучей, SIGRAPH-OPTIK/CAOS - программа для проектирования интегральной оптики.	
OptiCAD	Трассировка лучей через стандартные или определяемые пользователем компоненты. Позволяет импортировать IGES, STL объекты, существует механизм перевода объектов из ZEMAX в Opticad.	Opticad Corp., США. **www.landform.com/opticad** На сайте производителя доступны для скачивания Демо и полная (!) рабочая версия. Однако без HASP она не работает.
OptisWorks	Программа для расчета и оптимизации оптических систем, интегрированная в SolidWorks.	**http://www.optics.fr/software**
QuickDOE, IterDOE, IterMODE	Программное обеспечение по компьютерной оптике	Институт систем обработки изображений РАН (Самара). **www.ipsi.smr.ru**
SimuLight	Программное обеспечение для моделирования дифракционных оптических элементов	Институт систем обработки изображений РАН (Самара). **www.ipsi.smr.ru**

По функциональности оптические программы можно разделить на несколько групп:

- Инструменты для проектирования и расчета детерминированных оптических систем (обычно используют «Straight ray analysis» - анализ методом прямых лучей). К ним относятся объективы фотоаппаратов, телескопы, бинокли и т.д. Здесь может быть предсказана траектория каждого светового луча. Одновременно осуществляется управление любым световым лучом.
- Программы для расчета оптических систем, где предсказать ход луча невозможно. В английской терминологии этот процесс называется «Nonsequential ray tracing» - непоследовательная трассировка лучей.
- Программы для проектирования светотехники, понимаемой как недетерминированная оптическая система (светильники, автомобильные фары, и фонари).

- Средства для обработки результатов оптических измерений, подготовки математических моделей источников света.

Первый и последний пункты интересуют весьма узкий круг специалистов, требуя сугубо специальных знаний. Поэтому рассмотрим следующие программы: «TracePro» - разработку фирмы «Lambda Research Corporation» (США), и программное обеспечение по компьютерной оптике.

7.2. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ПО КОМПЬЮТЕРНОЙ ОПТИКЕ

Программное обеспечение «QUICK-DOE». Для моделирования работы дифракционных оптических элементов в «QUICK-DOE» в приближении Френеля реализован дифракционный метод расчета, основанный на быстром преобразовании Фурье (БПФ) [3], а для более сложных оптических схем – метод расчета трассировки лучей. Ниже дано краткое описание других возможностей программного обеспечения (рис. 7.1)[3].

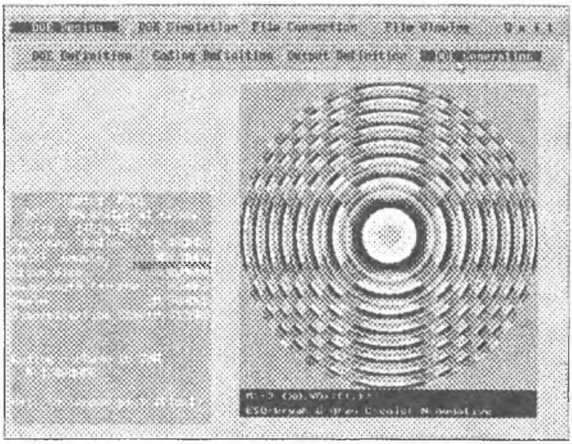

Рис. 7.1. Экран программного обеспечения «QUICK-DOE».

Это программное обеспечение включает реализацию аналитических методов расчета матриц фотошаблонов, описывающих фазовые функции ДОЭ, и методов кодирования функций пропускания оптических элементов. Программное обеспечение также оснащено средствами формирования файлов для вывода на растровые фотопостроители, расчета цифровых голограмм, выполнения ряда вспомогательных функций. В частности, к вспомогательным операциям относятся преобразование форматов файлов, визуализация маски файла, моделирование работы ДОЭ на основе БПФ, описание и подготовка составного изображения для вывода на фотоплоттер. Программное обеспечение предназначено для использования физиками-оптиками, проектировщиками дифракционной оптики и программистами, разрабатывающими программы расчета ДОЭ.

«QUICK-DOE» содержит программы расчета следующих аналитически описанных ДОЭ:

- различных вариантов дифракционных линз (радиальная, бифокальная, многофокусная, цилиндрическая, скрещенные цилиндрические, с повышенной глубиной фокуса [3]);
- фокусаторов (в кольцо, в крест, в контур квадрата, в отрезок в фокальной плоскости, в прямоугольную область – см. [3]);
- компенсаторов для преобразования сферического волнового фронта в волновой фронт с осесимметричной поверхностью вращения второго порядка (включая параболоид, эллипсоид и гиперболоид – см. [3]);
- оптических фильтров (Карунена-Лоева базисные функции, фильтр Винера, регуляризующий фильтр Тихонова – см. [3]);
- ДОЭ, согласованных с модами (моды Гаусса-Эрмита и Гаусса-Лагерра – см. [3]) когерентного излучения (моданов).

Список рассчитываемых ДОЭ может быть дополнен пользовательскими ДОЭ. Разработан и реализован несложный метод включения новых оптических элементов. Для введения нового элемента необходимо отредактировать программу проведения диалога для ввода требуемых параметров и написать программу расчета поля (фазы или амплитуды) в точке.

Программное обеспечение содержит различные варианты кодирования фазы и амплитуды:

- кодирование фазы по модулю 2π,
- кодирование из фазы в амплитуду с наложением несущей частоты,
- кодирование из амплитуды в фазу со значениями $\{0,\pi\}$,
- кодирование из амплитуды в фазу с наложением несущей частоты методом Кирка-Джонса [2, 3],
- кодирование методами Ломана и Ли [3],
- кодирование из амплитуды в амплитуду с наложением несущей частоты,
- преобразование амплитуды по абсолютной величине (амплитуда),
- преобразование амплитуды в интенсивность.

Все преобразования могут быть выполнены не только для существующих элементов, но и для элементов, включенных позже пользователем.

При расчете элемента могут быть получены полутоновые или бинарные маски. Маска может быть записана в файл в следующих вариантах:

- каждый отсчет как действительное число,
- байт на отсчет,
- бит на отсчет (только для бинарных масок).

Если нужен набор из нескольких бинарных масок (2 - 4), то можно рассчитать их последовательно или записать полутоновую маску в файл и затем получить все требуемые бинарные маски, используя преобразование.При расчете маски ДОЭ и изменении формата файла маски ДОЭ выполняется синхронная визуализация.

При работе «QUICK-DOE" поддерживается внутренний (растрового типа) формат файлов, записанный построчно с комментариями, содержащимися в заголовке и в окончании файла. Для стыковки с другим программным обеспечением предусмотрены перекодировщики в TIFF-формат байт/отсчет и обратно, и в поле байт/отсчет с построчной организацией. Для облегчения работы пользователя с программным обеспечением имеется удобная оболочка.

Программное обеспечение «Iter-DOE». В случае необходимости, рассчитанные с помощью «QUICK-DOE" фазовые функции оптических элементов могут быть оптимизированы с помощью программного обеспечения, реализующего итерационные методы расчета на основе БПФ. Такое программное обеспечение используется также, когда нельзя аналитически получить и запрограммировать фазовую функцию ДОЭ. На рис.7.2 представлен экран программного обеспечения «Iter-DOE», предназначенного для расчета фазовых оптических элементов (киноформов, фокусирующих ДОЭ, многофокусных линз, аксиконов, компенсаторов) на основе быстрых интегральных преобразований.

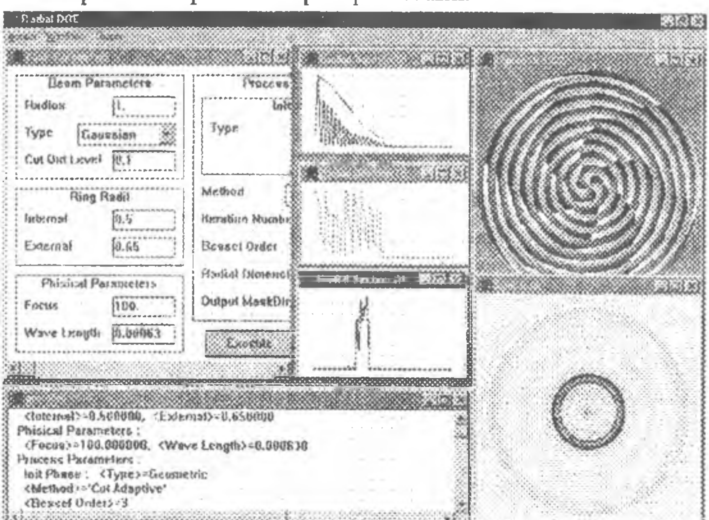

Рис. 7.2. Экран программного обеспечения «Iter-DOE (на примере расчета фазового ДОЭ, фокусирующего в кольцо).

Многооконный интерфейс позволяет наглядно представить все изменения фазовой функции и качества работы ДОЭ в процессе итерационного расчета оптического элемента.

Программное обеспечение «Iter-MODE». На рис.7.3 представлен экран другого программного обеспечения, основанного на итерационных методах расчета ДОЭ. Это - программное обеспечение «Iter-MODE», предназначенное для расчета дифракционных оптических элементов, формирующих моды Гаусса-Лагерра, Гаусса-Эрмита и Бесселя на основе алгоритмов аппроксимации комплексной функции конечным числом базисных функций. Многооконный интерфейс позволяет наглядно представить вид фазовой маски рассчитываемого 9-модового пространственного фильтра и результат моделирования воздействия этого фильтра на освещающий пучок с соответствующим модовым составом.

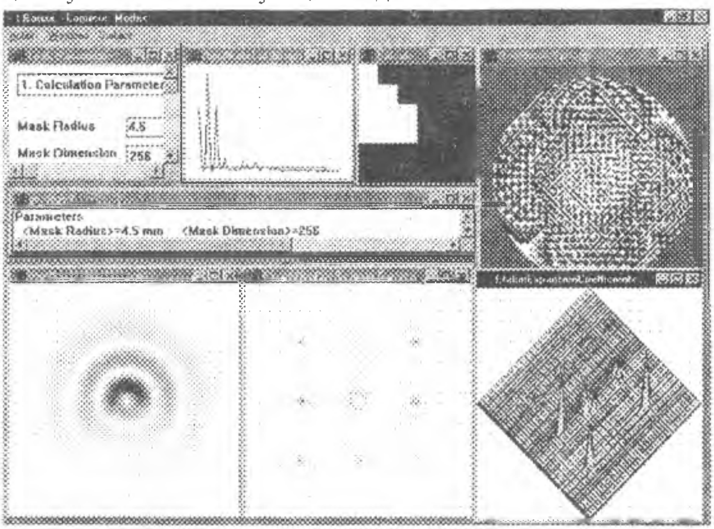

Рис. 7.3. Экран программного обеспечения «Iter-MODE» (на примере расчета ДОЭ, формирующего многомодовый пучок Гаусса-Лагерра).

Программное обеспечение «GRATING SOLVER». Если в процессе итерационного расчета выясняется, что итерационными методами в приближении Френеля-Кирхгофа невозможно добиться решения поставленной задачи, то приходится использовать более точные методы электромагнитного подхода [3]. Эти методы приходится применять также в случаях, когда нарушаются условия приближения Френеля. Программное обеспечение «GRATING SOLVER» (Рис. 7.4) предназначено для расчета и моделирования дифракционных решеток в рамках электромагнитной теории.

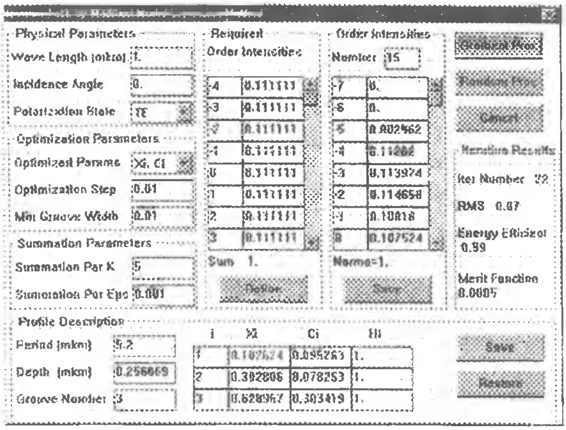

Рис. 7.4. Экран программного обеспечения «GRATING SOLVER» (на примере расчета 9-порядковой решетки с 3 штрихами).

Это программное обеспечение охватывает:

- моделирование отражающих и пропускающих дифракционных решеток с непрерывным и бинарным профилями рельефа;
- градиентные методы и стохастические алгоритмы решения обратной задачи расчета бинарного профиля из условия формирования заданных порядков дифракции;
- градиентные методы решения обратной задачи расчета непрерывного отражающего профиля в приближении Релея.

На Рис. 7.4 представлен экран программного обеспечения «GRATING SOLVER» (на примере расчета 9-порядковой решетки с 3 штрихами). Метод расчета основан на оптимизации координат (x_i, c_i, h_i) штрихов профиля из условия минимизации функции невязки расчетных и требуемых интенсивностей порядков.

7.3. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ МОДЕЛИРОВАНИЯ ДИФРАКЦИОННЫХ ОПТИЧЕСКИХ ЭЛЕМЕНТОВ

Программное обеспечение «*SimuLight*» предназначено для моделирования дифракционных оптических элементов (ДОЭ) и оптических систем, состоящих из нескольких ДОЭ. Программное обеспечение также обеспечивает средства для преобразования файлов форматов микроэлектроники: из векторного многоуровневого формата GDS в набор бинарных GDS файлов. Программное обеспечение предусматривает работу на следующих программных платформах (операционных системах): Win-XP/Millenium/98/NT. Рабочий язык оболочки – английский.

Моделирование работы ДОЭ в приближении скалярной теории дифракции реализовано на основе специальных методов расчета, использующих последние достижения в области численных методов. Программное обеспечение «*SimuLight*» включает приложение «*SimuLight*», предназначенное для моделирования оптических систем, а утилита *Gds2Mask* предназначена для преобразования файлов из векторного многоуровневого формата GDS в набор бинарных GDS файлов.

7.3.1 Возможности «SimuLight» для моделирования оптических систем

Цель: моделирование оптических систем с последующей визуализацией результата моделирования. Оптическая система включает источник когерентного излучения и один или несколько ДОЭ. Предполагается, что источник когерентного излучения и дифракционные оптические элементы расположены на одной оптической оси *OZ*. Для моделирования процесса распространения света используется скалярная теория дифракции (интегральное преобразование Кирхгофа). При этом предполагается, что минимальный характеристический размер микрорельефа ДОЭ [3] больше, чем длина волны источника света.

Визуализируются распределения интенсивности и фазы:

- непосредственно от источника света,
- освещающего пучка в плоскости непосредственно перед ДОЭ,
- прошедшего через ДОЭ пучка сразу после ДОЭ,
- в выбранной области наблюдения,

а также - фазовая функция ДОЭ.

Выбранная область наблюдения определяется как прямоугольник в плоскости, параллельной одной из базовых плоскостей (*XY*, *ZY*, *ZX*). При этом стороны прямоугольника параллельны базовым осям координат (*OX*, *OY*, *OZ*).

Описание:

Данное приложение реализовано как *MDI*-приложение. Окно задачи состоит из двух частей. Верхняя часть содержит таблицу с описаниями объектов. Нижняя часть содержит изображения описанных в верхней части таблицы объектов. Окно задачи также включает в себя строку состояний (в нижней части окна).

В данном приложении доступны следующие типы объектов:

- когерентный источник света (освещающий пучок - «*Beam*»);
- дифракционный оптический элемент («*DOE*»);
- область визуализации в плоскости («*Rectangle across OZ*»);
- область визуализации в плоскости *ZX* («*Rectangle across OY*»);
- область визуализации в плоскости *ZY* («*Rectangle across OX*»)

Замечание: *область визуализации в плоскости XY имеет общую оптическую ось (ось OZ) с источником света и ДОЭ.*

Замечание: *В описываемом программном обеспечении фаза ДОЭ описывается с помощью bmp-файла. При этом предполагается, что горизонтальный и вертикальный размеры файла в пикселах одинаковы.*

Таблица объектов содержит следующие столбцы:

- тип объекта, название столбца «Type»;
- плоскость визуализации (*XY*, *ZX*, *ZY*), название столбца «Plane»;
- отклонение плоскости наблюдения от точки с координатами (0,0,0), название столбца «Offset», название координаты (x, y или z) соответствует направлению сдвига;
- физические размеры области визуализации, название столбца «Physical Sizes», первое значение соответствует оси ординат выбранной плоскости (x – для плоскости *XY*, z – для *ZX* или *ZY*), второе значение соответствует оси абсцисс;
- размеры отсчетов в области визуализации;
- другие параметры.

Строка состояний включает актуальную текущую информацию о задаче:

- значение длины волны;
- число обрабатываемых объектов;
- текущий процент обработки объекта;
- обозначение обрабатываемого объекта.

Обработка объектов реализована в фоновом режиме, чтобы предотвратить зависание компьютера. Имеются отдельные окна для каждого объекта. Эти окна содержат визуализируемые изображения. Обозначение объекта приводится в заголовке окна.

Параметры объекта:

В настоящее время реализованы следующие типы освещающих пучков:

- Гауссов пучок;
- плоский пучок;
- круглый равномерный пучок;
- пучок, интенсивность и фаза которого определяются из файла.

Замечание: список параметров пучка зависит от выбранного типа пучка.

Замечание: задаваемое сечение пучка всегда имеет нулевую координату ($z = 0$).

Параметры Гауссова пучка:

- размеры апертуры (в *мм*), название параметра «Aperture Size»;

- радиус перетяжки Гауссова пучка, название параметра «Gaussian Radius»;

- размер апертуры в пискелах, название параметра «Pixel Number per Side»;

Параметры равномерного пучка:

- размеры апертуры (в мм), название параметра «Aperture Size»;

- размер апертуры в пикселах, название параметра «Pixel Number per Side»;

Параметры равномерного кольцевого пучка:

- размеры апертуры (в мм), название параметра «Aperture Size»;

- размер апертуры в пикселах, название параметра «Pixel Number per Side»;

Параметры пучка, значения интенсивности и фазы которого задаются из файлов:

- размеры апертуры (в мм), название параметра «Aperture Size»;

- имя файла, содержащего значения интенсивности;

- имя файла, содержащего значения фазы.

Замечание: *файлы, содержащие значения интенсивности и фазы, должны быть представлены в формате bmp и иметь одинаковые размеры в пикселах. Предполагается, что вертикальный и горизонтальный размеры апертуры в пикселах одинаковы.*

Параметры ДОЭ:

- положение вдоль оси OZ (мм);

- размеры апертуры ДОЭ (в мм), название параметра «Aperture Size»;

- имя файла, содержащего значения фазовой функции ДОЭ.

Параметры области визуализации (прямоугольная область, перпендикулярная оси OZ - «Rectangle across OZ»):

- положение вдоль оси OZ (мм);

- размеры области (в мм), название параметра «Aperture Size»;

- размеры области визуализации в пикселах, название параметра «Pixel Number per Side»;

Параметры области визуализации (прямоугольная область, перпендикулярная оси OY - «Rectangle across OY»):

- положение вдоль оси OY (мм);

- размеры области вдоль оси OX (в мм),

- количество отсчетов в области визуализации вдоль оси OX,

- начальное (минимальное) значение координаты z (в мм),

- конечное (максимальное) значение координаты z (в мм),

- количество отсчетов в области визуализации вдоль оси OZ.

Параметры области визуализации (прямоугольная область, перпендикулярная оси *OX* - «Rectangle across *OX*»):

- положение вдоль оси *OX* (мм);
- размеры области вдоль оси *OY* (в мм),
- количество отсчетов в области визуализации вдоль оси *OY*,
- начальное (минимальное) значение координаты *z* (в мм),
- конечное (максимальное) значение координаты *z* (в мм),
- количество отсчетов в области визуализации вдоль оси *OZ*.

Основные операции, выполняемые приложением:

Название операции	*Элемент меню*	*Клавиатура, действия*	*Результат*
Изменение параметров системы	Action-> Change System Parameters	Ctrl-S	Диалоговое окно для ввода длины волны
Добавление пучка	Action-> Add Beam	Ctrl-B	Диалоговое окно для ввода параметров пучка
Добавление фазы ДОЭ	Action-> Add Phase DOE	Ctrl-D	Диалоговое окно для ввода параметров ДОЭ
Добавление прямоугольной области вдоль оси OZ	Action-> Add Rectangle across OZ	Ctrl-Z	Диалоговое окно для ввода параметров области
Добавление прямоугольной области вдоль оси OY	Action-> Add Rectangle across OY	Ctrl-Y	Диалоговое окно для ввода параметров области
Добавление прямоугольной области вдоль оси OX	Action-> Add Rectangle across OX	Ctrl-X	Диалоговое окно для ввода параметров области
Выбор существующего объекта	-	Использование стрелок вверх / вниз; выбор мышкой в таблице; выбор мышкой окна	Описание выбранного объекта помечается; активизируется окно объекта.

Изменение текущего объекта	Action->Modify Current Item	Ctrl-M, Enter	Соответствующее диалоговое окно
Удаление текущего объекта	Action->Remove Current Item	Ctrl-R, Del, Ctrl-Del	Удаление объекта
Конец сеанса	Action-> Exit	Ctrl-E	Окно задачи закрывается без сохранения результата
Выбор изображения для визуализации из отмеченного файла	Pictures->Show*, *={Intensity, Phase, Phase before, DOE Phase, Phase After}	Alt – стрелка налево, Alt – стрелка направо, выбор мышкой в списке окон	Изображение меняется на новое

7.3.2. Утилита Gds2Mask для преобразования из векторного многоуровневого формата GDS в набор бинарных GDS файлов

Утилита Gds2Mask предназначена для преобразования из векторного многоуровневого формата GDS (формат версий 0, 3, 4, 5 или 600) в набор бинарных GDS файлов, соответствующих различным уровням исходного GDS-файла. Утилита реализована в среде компилятора Borland C++. Имеется проверка на корректность исходных данных.

Формат командной строки:

Gds2Mask.exe <source GDS-file name> <wavelength> <refractive index>

<source GDS-file name> - имя исходного многоуровневого gds-файла. Имя файла должно быть введено с расширением .GDS или без него;

<wavelength> - длина волны освещающего пучка для ДОЭ (в микронах);

<refractive index> - показатель преломления подложки ДОЭ.

Исходный GDS-файл может содержать следующие GDS-коды:

HEADER;
ENDLIB;
ENDSTR;
ENDEL;
BOUNDARY;
PATH;
SREF;

AREF;
TEXT;
LAYER;
DATATYPE;
PATHTYPE;
GENERATIONS;
UNITS;
LIBNAME;
XY;
BGNLIB;
STRNAME;
SNAME.

Выходные файлы: DEPTH.TXT, MASK0.GDS, MASK1.GDS, ..., MASKN.GDS, где MASKM.GDS – бинарный файл формата GDS, соответствующий M-тому слою исходного многоуровневого файла GDS-формата.

DEPTH.TXT – файл типа «txt», содержащий значения глубин травления для каждого уровня микрорельефа.

Пример:

Многоуровневый файл relief.gds, содержащий рассчитанные значения микрорельефа ДОЭ, спроектированного для работы с гелий-неоновым (He-Ne) лазером (длина волны 0,633 мкм), необходимо преобразовать в набор бинарных файлов gds-формата, описывающих отдельные уровни исходного файла. Показатель преломления материала – 1,5. Для запуска преобразования необходимо ввести следующую командную строку:

C:\UTILITY\ gds2mask 0.633 1.5

В результате мы получим сгенерированные файлы DEPTH.TXT, MASK0.GDS, MASK1.GDS, MASK2.GDS, MASK3.GDS, MASK4.GDS, MASK5.GDS, MASK6.GDS, MASK7.GDS в той же самой директории. На рис. 7.5 показан экран программного обеспечения CleWin во время визуализации исходного GDS-файла. На рис. 7.6 показан экран программного обеспечения CleWin во время визуализации бинарного GDS-файла MASK1.GDS.

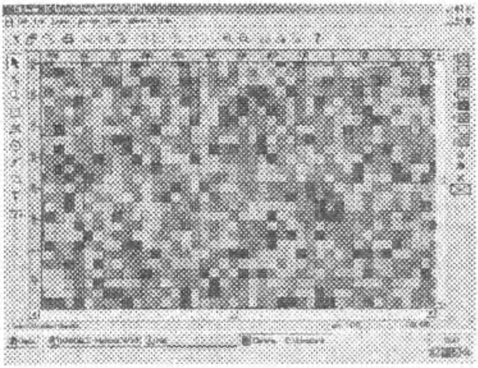

Рис. 7.5. Экран специализированного программного обеспечения во время визуализации исходного многоуровневого GDS-файла

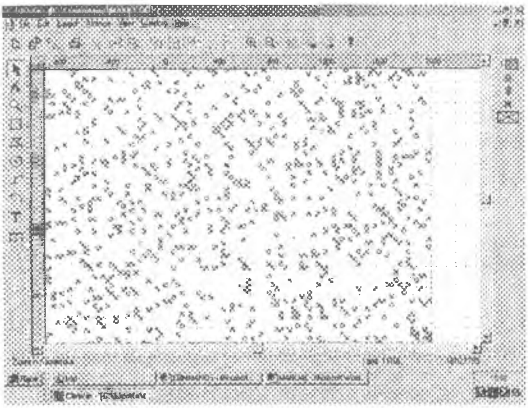

Рис. 7.6. Экран специализированного программного обеспечения во время визуализации бинарного GDS-файла MASK1.GDS

Файл типа «txt» Depth.txt, сгенерированный в той же директории, содержит следующие значения:

```
    Etching Depths for Masks (глубина травления для
масок):
    mask (маска) N 0 (глубина) depth=0.000000
    mask (маска) N 1 (глубина) depth=0.157500
    mask (маска) N 2 (глубина) depth=0.315000
    mask (маска) N 3 (глубина) depth=0.472500
    mask (маска) N 4 (глубина) depth=0.630000
    mask (маска) N 5 (глубина) depth=0.787500
    mask (маска) N 6 (глубина) depth=0.945000
    mask (маска) N 7 (глубина) depth=1.102500
```

7.4 ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ «TracePro»

7.4.1 Методы и алгоритмы, используемые «TracePro»

В данном разделе (7.4) будет рассмотрен пример программного обеспечения, предназначенного для моделирования оптических процессов - «TracePro». Акцент сделан на анализе возможностей, наиболее востребованных в инженерной практике.

Кратко рассмотрим историю развития программ, связанных с моделированием светотехнических устройств. Первоначальная цель такого моделирования состояла в расчете освещенности сцен. При этом главным критерием была визуальная достоверность полученной сцены. Реальные оптические процессы тогда мало интересовали разработчиков. Одной из первых стала локальная модель освещения, в которой учитывались эффекты зеркального и — на примитивном уровне — диффузного отражения. Кроме того, вводилось так называемое фоновое освещение, учитывающее наличие излучения, порождаемого «неформализуемыми» источниками. Следующим шагом было появление модели локального затенения, одной из особенностей которого является интерполяция освещенности по поверхности объектов, а также наличие ряда моделей характеристик поверхности.

Важным этапом стало появления метода «Ray tracing» (Трассировка лучей), который заключается в отслеживании траекторий лучей и моделировании того, как каждый луч взаимодействует со встречающимися на его пути телами и поверхностями. Этот подход позволяет ввести в рассмотрение эффекты преломления, затенения, многократного зеркального отражения. Исторически первым был метод Forward ray tracing (Прямая трассировка), когда лучи, как это и происходит в реальности, испускаются источником света, а затем последовательно взаимодействуют с объектами сцены. Именно этот алгоритм стал базой для развития светотехнических программ, и именно тогда произошло их выделение в самостоятельную ветвь программного обеспечения. Программные продукты для «эстетической» визуализации (рендеринга) стали развиваться отдельно. Взаимодействие между этими направлениями происходило практически только на уровне идей, но предназначение программ и, что немаловажно, клиентская база, были различны.

Круг алгоритмов, которые используются в обеих группах программ, включает в себя метод «Distribute raytracing» (Распределенная трассировка лучей) или, в терминах «TracePro» — «Ray splitting» (Расщепление луча). Смысл его в том, что при взаимодействии с поверхностью или средой луч может расщепляться на несколько, причем отдельные лучи несут зеркально отраженную, рассеянную, преломленную и т.д. компоненты светового потока. Это является базой для корректного описания эффектов диффузного отражения и рассеяния в материале.

Еще одной общей идеей стал алгоритм «Backward ray tracing» (Обратная трассировка лучей) или, как он называется в «TracePro» – «Reverse raytracing». Он базируется на том, что лучи испускаются объектами, освещенность которых интересует наблюдателя, после чего прослеживаются их пути к источнику. В «TracePro» данный алгоритм появился только в 2004 году.

Число трассируемых лучей является ключевым параметром, влияющим на точность анализа. В методе прямой трассировки лучи испускаются источниками света, которые в рассматриваемом программном обеспечении могут быть поверхностными, или же назначаться на сетке. В первом случае, независимо от того, какой тип распределения лучей в пространстве выбирается пользователем, координаты точек, из которых испускается луч, выбираются программой случайным образом. Если используется пространственное распределение, которое не является равномерным, то направления также случайны (хотя, разумеется, подчиняются некоторой функции распределения). Этими процессами управляет метод Монте-Карло, функция которого в данной ситуации генерировать точки старта и направления. Если же алгоритм трассировки имитирует рассеяние света при взаимодействии с поверхностью или в объеме, то здесь также используется генератор случайных чисел. В первой ситуации, в зависимости от соотношения между долями отраженного, прошедшего и рассеянного света формируется луч, который, может имитировать как отраженный, преломленный, так и рассеянный свет. В последнем случае направление луча назначается программой исходя из заданного закона распределения рассеянной энергии, сформулированного с позиции теории вероятности. При описании объемного рассеяния используемые модели подразумевают изменение направления луча (направление выбирается статистически исходя из назначенного закона распределения) при прохождении им некоторого расстояния. Здесь также может имитироваться расщепление: в зависимости от соотношения между долями отраженного, прошедшего и рассеянного света формируется луч, который, может имитировать как отраженный, преломленный, так и рассеянный свет.

В качестве источников света может использоваться совокупность лучей, исходящих из вершин некоторой сетки. Она может быть равномерной (прямоугольной или круговой), а может быть назначена статистически или с учетом статистического отклонения от «правильной» формы. Возникает вопрос: для чего введены эти усложнения. Во-первых, взаимодействие света со средой имеет статистическую природу, поскольку оптические характеристики среды также имеют вероятностный характер: параметры шероховатости поверхности, а также отклонения ее формы, оптическая неоднородность прозрачной среды, например, локальная анизотропия пластмасс, наличие включений и т.д. Реальная атмосфера в силу различия температуры, запыленности, движения потоков с отличающейся плотностью, не является оптически-однородной. Поэтому идеализацию оптических характеристик объектов необходимо компенсировать вводом статистической компоненты в законы излучения и распространения света. Кроме того, вывод изображения на экран монитора, представляющий собой совокупность пикселей, также порождает «Aliasing» (Алиасинг) — зубчатый характер контуров, а также сопровождается появлением повторяющихся структур на однородном, или плавно изменяющемся в действительности изображении.

Алгоритм рекурсивной вокселизации, предназначенный для ускоренной идентификации объектов, встречающихся на пути луча, в «TracePro» был введен в 2003 году. В то же время простая вокселизация — разбиение пространства на виртуальные параллелепипеды с установлением связи между объектами сцены и параллелепипедами, в которые они могут входить, в этой программе существовал изначально.

Алгоритм, который не используется в программах светотехники, но является вполне традиционным для процедур рендеринга — метод «Radiosity» (Метод излучательности), предполагающий разбиение поверхностей на участки с последующим анализом взаимодействия участков, описывающих все объекты сцены, между собой. Многие идеи данного алгоритма перекликаются с методом конечных элементов (МКЭ). Надо сказать, что теоретическая база для моделирования теплопереноса излучением, использованная в МКЭ, имеет много общего с данным алгоритмом. Последние реализации программ рендеринга содержат опции, позволяющие, в какой-то степени, объединять два алгоритма: радиосити и трассировки. Эти гибридные алгоритмы носят название «Photon Maps» (Метод фотонных карт).

Для увеличения визуального правдоподобия в алгоритмах рендеринга, основанных на трассировке лучей, в дополнение к используемым BRDF и BTDF— функциям, описывающим процесс рассеяния при отражении и прохождении через поверхность, разработан метод BSSRDF («Bidirectional Surface Scattering Reflection Distribution Function» — Двунаправленная функция распределения отражения рассеянного света). Эта модель описывает поведение луча, попавшего на поверхность тела, обладающего объемным рассеянием. При этом часть светового потока, в силу объемного рассеяния, возвращается обратно к поверхности, а затем — в среду, из которой этот луч пришел. То есть, по сути, объединяются эффекты, которые в «TracePro» обрабатываются принципиально различными инструментами. Для имитации сцены, когда преследуется цель повысить визуальное правдоподобие, объединение эффектов объемного и поверхностного рассеяния более результативно, по крайней мере, с точки зрения производительности вычислений (возможность прохождения света через тело с такой поверхностью отсутствует в принципе).

Как видно из краткого обзора, имеется достаточно число точек соприкосновения двух направлений компьютерного моделирования при визуализации. Собственно, последние релизы «TracePro», а также других конкурирующих продуктов, включают в себя модули рендеринга (они поставляются опционно). Суть соответствующего приложения «TracePro» в том, что в качестве исходной информации оно используют результаты расчета базовой программы, после чего изображение формируется с учетом условностей, которые порождают ощущение визуального правдоподобия.

Самое важное отличие программ светотехнического анализа от программ визуализации: «TracePro» и родственные продукты — полноценные члены семейства CAD/CAE. В качестве геометрического ядра они имеют одну из известных систем: ACIS («TracePro»), Parasolid, математику Catia V5 и т.д. Соответственно доступны (иногда за дополнительную плату) интерфейсы через форматы STEP и IGES, в то время как организация взаимодействия программ "эстетического" рендеринга и CAD-систем весьма щекотливый вопрос.

Упомянем еще два типа программ, оперирующих с оптическими явлениями. Первые — программы детерминированной оптики, то есть не предполагающие статистического характера распределения света. Эти продукты предназначены для проектирования оптических приборов: объективов, телескопов и т.д. Примером данной программы является «OSLO», которая, так же, как и «TracePro» разрабатывается фирмой «Lambda Research». Вторая группа программ — инструменты проектирования, но, на сей раз, в светотехнике. Одна из них — «Reflector-CAD».

7.4.2 *Краткое изложение возможностей «TracePro»*

«TracePro» представляет собой уникальный по функциональности инструмент, который можно использовать в широком спектре приложений и характеризуется разработчиком как «интегрированная система оптического анализа и твёрдотельного моделирования» (Рис. 7.7 и 7.8).

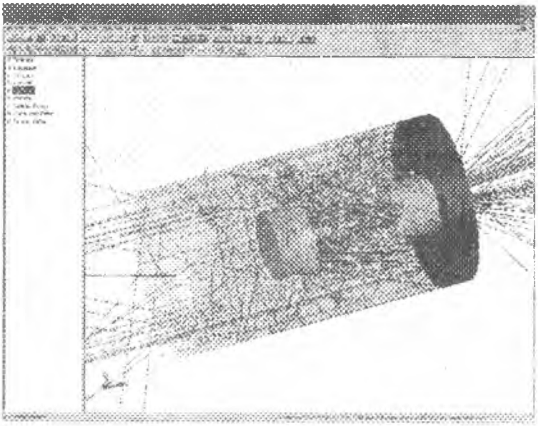

Рис. 7.7. «TracePro» с успехом решает задачи трассировки лучей в произвольной оптической системе. Анализ распространения света с учетом отражения, рассеяния и поглощения выполняется с максимальной наглядностью

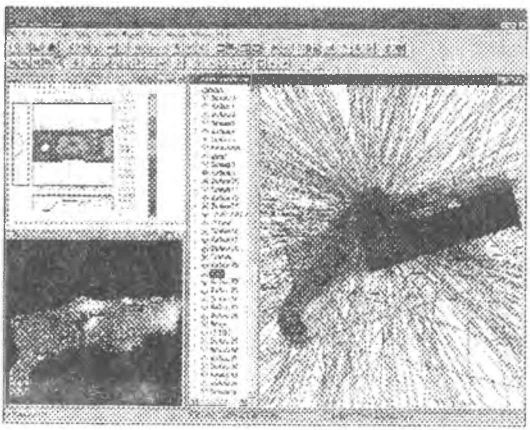

Рис. 7.8. Виртуальные прототипы световодов могут быть рассчитаны с получением фактического распределения света, плоского и пространственного отображения освещенности и яркости, а также качественного рендеринга сцены

«TracePro» имеет следующие возможности для анализа освещения, расчёта осветительных приборов и источников света:

• программа имитирует работу ламп накаливания, светодиодов, флуоресцентных ламп, разнообразных рефлекторов и линзовых систем;

• каталог ламп содержит более 200 стандартных объектов, включая каталоги «OSRAM» и «Phillips»;

• «TracePro» обеспечивает импорт источников, представленных распределением интенсивности света;

• возможен экспорт результатов в стандартах «IES» и «Eulumdat».

«TracePro» предоставляет следующие возможности для моделирования дисплейных систем, включая жидкокристаллические мониторы (LCD):

• программа позволяет рассчитывать дисплеи с задней и передней подсветкой, например, дисплеи компьютеров и мобильных телефонов;

• имеются интегрированные в программу средства для моделирования работы LCD систем и световодов;

• алгоритм «RepTile™» является уникальным средством моделирования устройств, содержащих миллионы повторяющихся объектов, с учетом яркости и рассеяния (см. Рис.7.9);

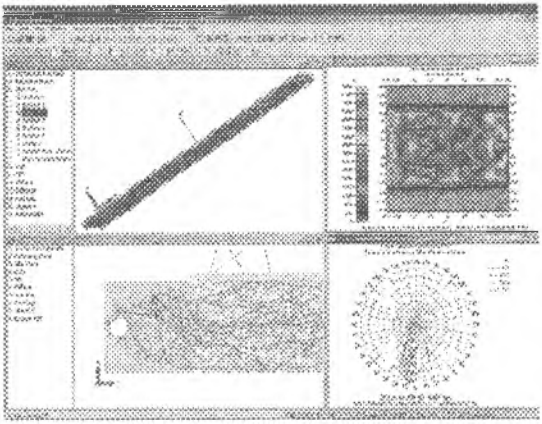

Рис. 7.9. Алгоритм «RepTile» применительно к задним фонарям даёт возможность полноценного анализа сложной конструкции, содержащей LED панель. Приведённая модель содержит семь миллионов рассеивающих точек. Для её расчета потребовалось не более десяти минут

• моделирование поляризации с учетом векторов Стокса и матриц Мюллера в зависимости от угла падения, длины волны и температуры;

• используя «Bitmap», можно создать источник с полной информаций о распределении энергии по длинам волн - этот источник можно использовать в расчетах для анализа искажений и качества цветопередачи;

• диаграммы яркости и освещенности обеспечивают наглядное отображение результатов.

Возможности «TracePro» для систем машинного зрения:

• посредством «TracePro» можно создавать изделия, предназначенные для дефектоскопии, робототехники, сканирования и т.д.;

• полноценное моделирование эффектов рассеяния, с имитацией анизотропии свойств, а также асимметричной BSDF.

Возможности и средства «TracePro» для решения задач автомобилестроения:

• анализ паразитной засветки, в частности, на дисплеях панелей приборов и на лобовых стеклах, используя алгоритм обратной трассировки;

• обширный каталог автомобильных ламп, а также специальные возможности для создания фасетных отражателей;

• инструменты для генерации массивов линзовых объектов в совокупности с фасетными отражателями, используемые применительно к задним фонарям;

• интегрированные IGES и STEP трансляторы для обмена с универсальными CAD-системами;

• инструменты для моделирования линз Френеля;

• инструменты для создания моделей практически любых типов источников, а также импорта источников из каталогов фирмы «Lambda Research» или библиотек сторонних производителей.

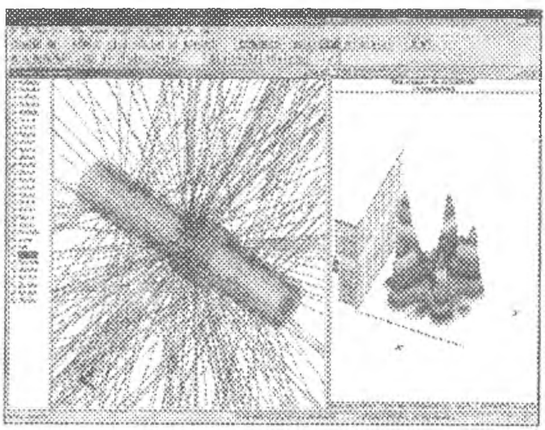

Рис. 7.10. Анализ прохождения светового потока в объёме тела. Программа позволяет рассчитывать объёмное распределение потока энергии в лазерных резонаторах, выделять величину попавшей, отраженной и поглощенной энергии

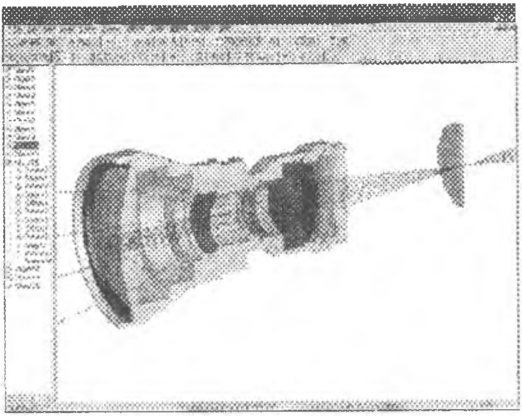

Рис. 7.11. После импорта твердотельной и поверхностной геометрии из CAD систем соответствующим объектам могут быть назначены оптические характеристики и выполнен расчет

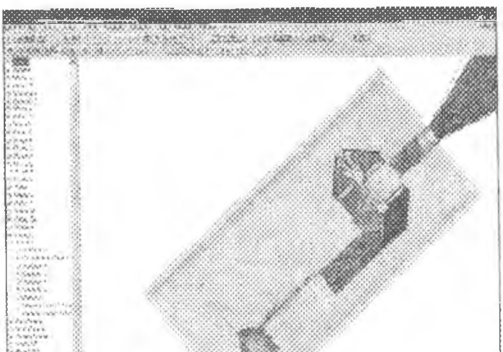

Рис. 7.12. Результаты трассировки могут выводиться на фоне прозрачной модели с отображением лучей линями различного цвета. Эти цвета могут соответствовать как величине потерянной энергии, так и длине световой волн

Возможности и средства «TracePro» для решения задач аэрокосмической промышленности и оборонной техники:

• эффективный анализ CHMSL систем, которые содержат массивы элементов, состоящих из линз Френеля и жидкокристаллических источников, посредством алгоритма RepTile;

• использование выборки по значимости для оптимальной трассировки рассеянного света по направлению к заданным датчикам;

• наличие моделей асимметричного рассеяния, соответствующего оптически-анизотропным поверхностям; учет зависимости параметров рассеяния от направления падения, длины волны и температуры;

• моделирование расщепления лучей при взаимодействии с поверхностью на пять составляющих: пропущенную, поглощенную, отраженную, пропущенную и отраженную с рассеянием;

• эффекты дифракции на апертурах и в коронографах, учет краевых эффектов на экранах;

• получение результатов для произвольных объектов: тел и поверхностей с учетом распределения по длинам волн и с анализом долей попавшей, поглощенной и потерянной энергии;

• доступ к траектории произвольного луча, включая длину траектории, величину переносимого светового потока, информацию о поляризации; выделение совокупности лучей, попавших на поверхность;

• получение информации о распределении энергии в потоке с произвольным спектром излучения.

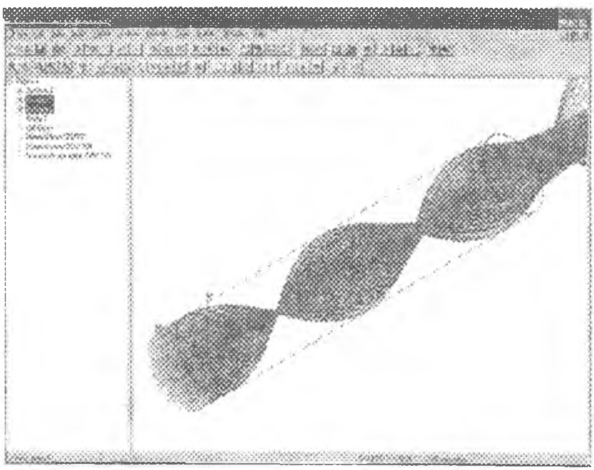

Рис.7.13. Пример расчета волокна, имеющего переменный коэффициент преломления. Такие характеристики содержатся в каталоге программы или могут быть вставлены на основе информации пользователя

«TracePro» - для медицинской техники:

• моделирование взаимодействия излучения с кожей;

• наличие библиотеки характеристик компонентов биологических тканей.

Лазеры:

• трассировка потока в резонаторах и отображение этого на экране;

• экспорт информации о распределении потока в специализированные приложения для расчёта лазеров;

• идентификация рассеянной компоненты энергии, включая процессы в кристаллах АИГ, в частицах воды и в лазерных диодах;

• назначение множественных источников, включая анизотропно-излучающие поверхности;

• учет поляризации посредством векторов Стокса.

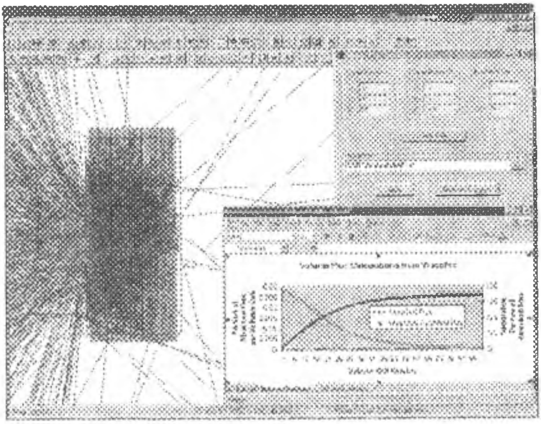

Рис.7.14. Возможности программы по моделированию эффектов объемного рассеяние демонстрируются на примере взаимодействия кожи со светом. Программа содержит разнообразные модели объемного рассеяния

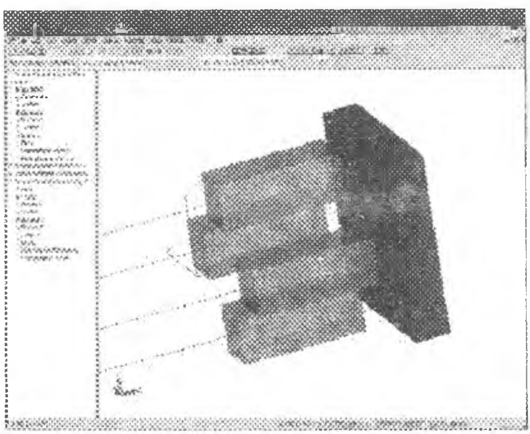

Рис.7.15. Моделирование двоякопреломляющих кристаллов является новой возможностью программы

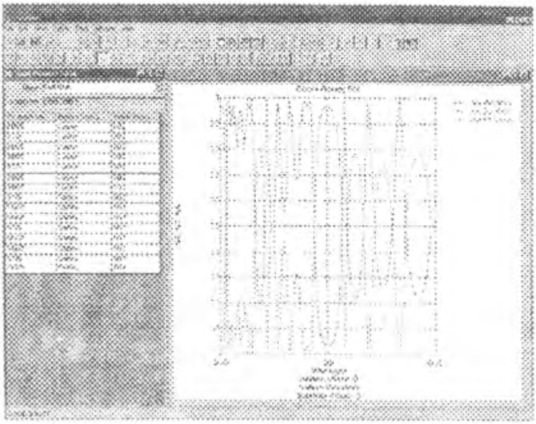

Рис.7.16. Специальные инструменты предназначены для визуализации параметров тонкопленочных объектов. На иллюстрации показаны оптические характеристики 15-слойного покрытия в зависимости от длины волны

ЗАКЛЮЧЕНИЕ К ГЛАВЕ 7

В данной главе дан краткий обзор существующих программных продуктов, предназначенных для моделирования оптических систем. Описано доступное студентам СГАУ (имеющееся в распоряжении Института компьютерных исследований СГАУ) программное обеспечение по компьютерной оптике и программное обеспечение «TracePro», разрабатываемое фирмой «Lambda Research Corporation» (находится в одном из центров мировой оптической индустрии - городе Tucson, США). Описанные продукты дают достаточное представление о современных тенденциях в создании оптического программного обеспечения.

ВЫВОДЫ

1. В настоящее время на рынке имеется большой выбор программных продуктов, предназначенных для моделирования и проектирования оптических систем.

2. Стоимость больших программных комплексов, включающих разнообразные методы расчета и гигантские базы данных с информацией о выпускаемых во всем мире источниках света и оптических элементах, достигает 40 тысяч долларов США (ASAP). Несмотря на такую высокую стоимость время работы программного обеспечения (срок действия лицензии) составляет один год, а стоимость продления лицензии на следующий год (вместе с техническим обслуживание) доходит до 10 тысяч долларов США.

3. В Институте компьютерных исследований СГАУ имеется ряд дополняющих друг друга программных комплексов по компьютерной оптики, позволяющих решать широкий круг задач расчета, оптимизации и моделирования работы дифракционных оптических элементов.

4. В Институте компьютерных исследований СГАУ имеется сетевая учебная версия программного обеспечения «TracePro» - мощного программного продукта, обладающего широким спектром возможностей для моделирования и проектирования разнообразных оптических систем (освещающих устройств, дисплеев, систем машинного зрения, светотехнических устройств для автомобилестроения, для медицинских приборов, лазерной техники, решения задач аэрокосмической промышленности и оборонной техники).

ЛИТЕРАТУРА

1. Борн М., Вольф Э. Основы оптики. - М.: Наука, 1973. - 720 с.
2. Сойфер В.А. Введение в дифракционную микрооптику. - Самара: СГАУ, 1996. - 95с.
3. Методы компьютерной оптики / Под редакцией В.А. Сойфера. - М.: Физматлит, 2003. – 688с.
4. Самарский А.А. Математическое моделирование и вычислительный эксперимент // Вестник АН СССР. - 1979. - № 5. - С.38-41.
5. Самарский А.А. Математическое моделирование и вычислительный эксперимент // Сборник лекций «Современные методы математического моделирования» по материалам Международной конференции «Математическое моделирование – 2001», Самара: СГАУ, 2001, с.4-12.
6. Математическое моделирование / Под редакцией Дж. Эндрюса и Р. Мак-Лоуна. – М.: Мир, 1979. – 277с.
7. Неймарк Ю.И. Математические модели в естествознании и технике. – Н.Новгород: ННГУ, 2004. – 401с.
8. Информационная оптика / Под редакцией Н.Н. Евтихиева. – М.: МЭИ, 2000. – 612с.
9. Farn M.W. Modeling of diffractive optics // OSA Proceedings of the International Optical Design Conference. - 1994. - Vol.22. - P.246-250.
10. Ландау Л.Д., Лифшиц Е.М. Теория поля, изд. 5-е. М.: Наука, Физматгиз, 1967.
11. Федорюк М.В. Асимптотика: Интегралы и ряды. - М.: Наука, 1987. - 544с.
12. Янке Е., Эмде Ф., Лет Ф. Специальные функции (формулы, графики, таблицы). - М.: Наука, 1968. - 344с.
13. Алямовский А.А. Компьютерное моделирование в инженерной практике. – М.: ЗАО «Сикор». 2004. – 186 с.
14. Гудмен Дж. Введение в Фурье-оптику. – М.: Мир, 1970.
15. Применение методов Фурье-оптики / Под ред. Г. Старка: М.: Радио и связь, 1988.
16. Сороко Л.М. Основы голографии и когерентной оптики. – М.: Наука, 1971.
17. Родионов С.А. Автоматизация проектирования оптических систем. – Л.: Машиностроение, 1982.

СОДЕРЖАНИЕ

Учебное издание

Казанский Николай Львович

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ОПТИЧЕСКИХ СИСТЕМ

Учебное пособие

Компьютерная верстка
М.А. Вахе, С.В.Смагин, Я.Е. Тахтаров
Подготовка иллюстраций - С.В.Смагин

Подписано в печать 01.12.2005 г. Формат 60×84 1/16.
Бумага офсетная. Печать офсетная.
Усл. печ. л. 14,1. Усл. кр.-отт. 14,4. Уч.-изд. л. 15,0.
Тираж 100 экз. Заказ 5 .

Самарский государственный аэрокосмический
университет им. академика С.П. Королева
443086 Самара, Московское шоссе, 34

РИО Самарского государственного
аэрокосмического университета.
443086 Самара, Московское шоссе, 34