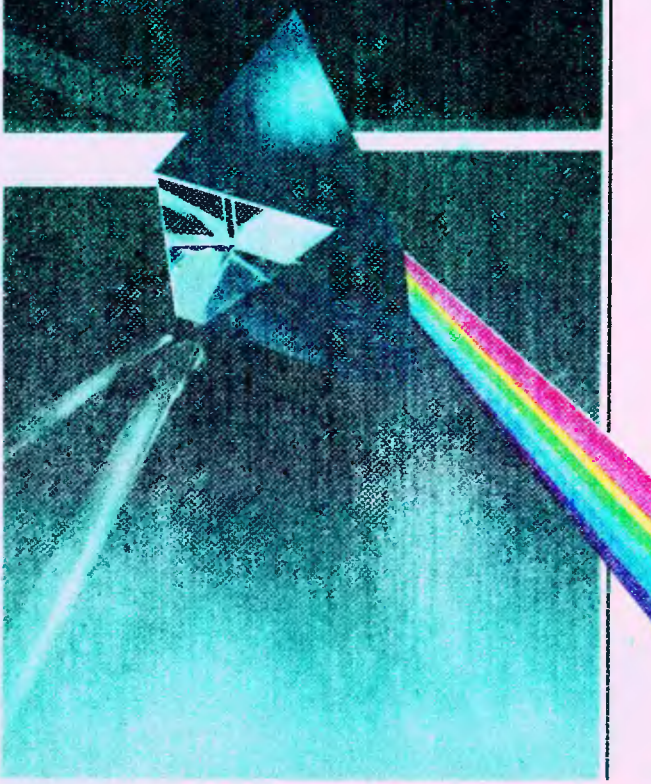
535
Б-93
S7097

Е. И. БУТИКОВ

ОПТИКА



Е. И. БУТИКОВ

ОПТИКА

Под редакцией проф. Н. И. Калитеевского

Допущено
Министерством высшего и среднего
специального образования СССР
в качестве учебного пособия
для студентов физических специальностей
вузов

МОСКВА «ВЫСШАЯ ШКОЛА» 1986

ББК 22.34
Б 93
УДК 535

Рецензенты:

кафедра общей физики Московского физико-технического института (зав. кафедрой — проф. С. П. Капица); проф. В. А. Фабрикант

Бутиков Е. И.

Б93 Оптика: Учеб. пособие для вузов / Под ред. Н. И. Калитеевского. — М.: Высш. шк., 1986. — 512 с.: ил.

Изложение современных фундаментальных понятнй оптнки построено в учебнике на основе требования единства экспернмента н теорнн. Наряду с традиционными вопросами рассмотрены статистическне н когереитные свойства квазнмонохроматического нзлучения, спектральное разложение, электронная теорня дисперсии, основы нелинейной оптики. Большое внимание уделено свойствам лазерного излучення и применению лазеров в фнзическом эксперименте. К каждому параграфу даны контрольные вопросы и задачн.

Б $\frac{1704050000-541}{001(01)-86}$ 79—86

ББК 22.34
535

© Издательство «Высшая школа», 1986

ОГЛАВЛЕНИЕ

Предисловие

Введение

Электромагнитные волны в вакууме.
Испускание волн. Квазимонохроматический свет

1.1	Плоские монохроматические электромагнитные волны в вакууме	12
1.2	Поляризация плоских монохроматических волн	18
1.3	Стоячие электромагнитные волны	25
1.4	Энергия электромагнитных волн	30
1.5	Испускание электромагнитных волн. Сферические волны	35
1.6	Спектральное разложение излучения	45
1.7	Квазимонохроматический свет	50
1.8	Спектральные линии. Поляризация квазимонохроматического света	54
1.9	Эффект Зеемана	62
1.10	Основы фотометрии	67

Распространение света в изотропных средах

2.1	Уравнения Максвелла для волн в веществе	72
2.2	Материальные уравнения. Плоские монохроматические волны в изотропной среде	75
2.3	Классическая электронная теория дисперсии	82
2.4	Дисперсия вдали от линии поглощения	86
2.5	Аномальная дисперсия	
2.6	Дисперсия в металлах и плазме. Показатель преломления рентгеновских лучей	94
2.7	Дисперсия в ионных кристаллах. Ориентационная дисперсия	98
2.8	Поворот направления линейной поляризации в магнитном поле (эффект Фарадея)	102
2.9	Естественное вращение направления поляризации	108
2.10	Рассеяние света	115
2.11	Скорость света. Фазовая и групповая скорости	126
2.12	Излучение Вавилова — Черенкова	136

3

Отражение и преломление света на границе

3.1	Законы отражения и преломления света	142
3.2	Формулы Френеля	145
3.3	Полное отражение	153
3.4	Отражение света от поверхности металлов	161
3.5	Световое давление. Импульс электромагнитной волны	167

4

Распространение света в анизотропной среде

4.1	Двойное лучепреломление	175
4.2	Плоские монохроматические волны в анизотропной среде. Одноосные кристаллы	180
4.3	Преломление на границе анизотропной среды. Построение Гюйгенса	186
4.4	Поляризационные призмы и поляроиды	191
4.5	Искусственная анизотропия. Эффект Керра	194

5

Интерференция света

5.1	Интерференция монохроматического света	202
5.2	Интерференционные опыты по методу деления волнового фронта	206
5.3	Деление амплитуды. Локализация интерференционных полос	211
5.4	Интерференция квазимонохроматического света. Временна́я когерентность	218
5.5	Роль конечных размеров источника света. Пространственная когерентность	234
5.6	Двухлучевые интерферометры	247
5.7	Многолучевая интерференция	255

6

Дифракция света

6.1	Принцип Гюйгенса — Френеля. Зоны Френеля	268
6.2	Дифракция Френеля на прямолинейном крае экрана	278
6.3	Дифракция Фраунгофера	284
6.4	Гауссовы пучки. Оптические резонаторы	297
6.5	Дифракционные решетки	305
6.6	Спектральные приборы	314

Геометрическая оптика и роль дифракции в оптических приборах

7.1	Основные положения геометрической оптики	329
7.2	Центрированные оптические системы	337
7.3	Ограничение световых пучков в оптических системах	347
7.4	Аберрации оптических систем	352
7.5	Яркость и освещенность оптических изображений	361
7.6	Разрешающая способность оптических инструментов	365
7.7	Физические принципы голографии	377

Оптика движущихся тел

8.1	Безуспешные поиски «светоносной среды»	392
8.2	Основные положения частной теории относительности	400
8.3	Теория относительности и оптика движущихся тел	406
8.4	Эффект Саньяка. Лазерный гироскоп	413

Термодинамика излучения. Световые кванты

9.1	Тепловое излучение в замкнутой полости. Черное тело	418
9.2	Спектральная плотность равновесного излучения. Формула Планка	425
9.3	Световые кванты. Спонтанное и вынужденное излучения	434
9.4	Лазеры	444
9.5	Фотоэлектрический эффект	456
9.6	Энергия и импульс фотона. Дуализм света	466

Основы нелинейной оптики

10.1	Некогерентные нелинейные эффекты	479
10.2	Материальные уравнения для нелинейных сред	481
10.3	Генерация второй гармоники	488
10.4	Параметрическое преобразование частоты	494
10.5	Вынужденное рассеяние Мандельштама — Бриллюэна	497
10.6	Вынужденное комбинационное рассеяние	503
	Заключение	507
	Предметный указатель	508

Предисловие

Последние десятилетия ознаменовались важными научными результатами в физической и прикладной оптике. Оптические методы и аппаратура все шире проникают в самые различные области науки и техники и играют значительную роль в ускорении научно-технического прогресса. В современных условиях, когда наука становится непосредственной производительной силой, особое значение приобретают вопросы совершенствования подготовки кадров. Курс оптики представляет существенную часть программы обучения будущих физиков и инженеров независимо от того, по какой специальности они будут работать.

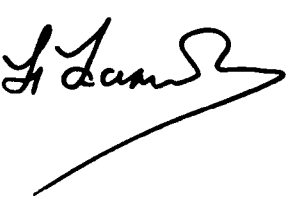
Данное учебное пособие расширяет и углубляет учебный материал, изложенный в «Волновой оптике» Н. И. Калитеевского (М.: Высш. шк., 1978). Оно предназначено для студентов физических специальностей университетов и вузов (при изучении соответствующего раздела общего курса физики), но может быть полезно и для специалистов, работающих в области физической оптики. Потребность в новом пособии вызвана прежде всего происходящим на наших глазах бурным развитием оптики, которое связано с созданием лазеров — принципиально новых источников когерентного излучения. Применение лазерного излучения стремительно расширяется и охватывает практически всю экспериментальную физику, как и многие области техники и технологии. Лазеры не только привели к возникновению новых современных направлений, таких как нелинейная оптика и голография, но и оказали большое влияние на многие классические разделы оптики. Естественно, что все это не могло не отразиться на содержании и методах преподавания оптики, как и на общем увеличении роли этого раздела в общем курсе физики.

Содержание пособия соответствует действующей программе курса общей физики для физических специальностей вузов. От существующих учебных пособий оно отличается тем, что в нем в сравнительно небольшом объеме наряду с традиционными вопросами строже и подробнее, чем это обычно принято, рассматриваются статистические и когерентные свойства квазимонохроматического излучения, спектральное разложение, электронная теория дисперсии, оптические резонаторы, разрешающая сила оптических и спектральных приборов, фотоэлектрические измерения, основы нелинейной оптики. Большое внимание уделяется объяснению свойств лазерного излучения и применению лазеров в оптическом эксперименте. Изложение учебного материала проводится на основе электромагнитной теории света, с соблюдением требования единства теории и эксперимента, обязательного при изучении курса общей физики.

Оптические явления в движущихся телах трактуются в рамках частной теории относительности, основы которой здесь лишь кратко напомииаются, так как подробное их изучение в соответствии с программой курса общей физики предусматривается в разделе «Механика».

Каждый параграф снабжен контрольными вопросами, которые помогут студенту обратить внимание на главные положения изучаемой темы. Задачи, к которым даны решения, указания или ответы, составляют органическую часть учебного материала пособия, необходимую для исчерпывающего понимания изучаемого вопроса. Однако приведенных в пособии задач совершенно недостаточно для активного, творческого овладения материалом. Изучение оптики, как и других разделов общей физики, немыслимо без практических занятий с решением задач и выполнением лабораторных работ, равно как и без лекций с демонстрациями физических явлений на опыте.

Несомненно, что ознакомление с волновыми свойствами оптического излучения, которым в основном посвящена данная книга, представляет собой необходимый этап при изучении современной физики.



Введение

Оптикой называют учение о физических явлениях, связанных с распространением и взаимодействием с веществом коротких электромагнитных волн, длина которых лежит в интервале 10^{-4}—10^{-9} м. Большое значение этой области спектра электромагнитных волн для практической деятельности человека обусловлено прежде всего тем, что внутри нее в узком интервале длин волн от 0,4 до 0,7 мкм лежит участок видимого света, непосредственно воспринимаемого человеческим глазом.

С точки зрения физики происходящих процессов выделение столь узкой спектральной области видимого света не имеет особого смысла, поэтому в понятие оптического диапазона включают обычно еще инфракрасное и ультрафиолетовое излучения. Но и для них принятые границы спектра в значительной степени условны. По существу, они определяются используемыми способами получения и регистрации электромагнитных волн.

Электромагнитная теория света возникла в итоге длительного развития взглядов на природу света. Ей предшествовала волновая теория, в которой свет рассматривался как упругое возмущение, распространяющееся в гипотетической среде — эфире. В трудах Френеля и других выдающихся физиков прошлого столетия эта теория была доведена до высокой степени совершенства, но в то же время в ней выявились трудности принципиального характера. Неудовлетворительность старой волновой теории проявлялась прежде всего в том, что для объяснения наблюдаемых оптических явлений эфир приходилось наделять весьма экзотическими и противоречивыми свойствами, несовместимыми с законами механики.

В середине прошлого столетия на основе связанных главным образом с исследованиями Фарадея экспериментальных открытий в области электрических и магнитных явлений Максвелл сформулировал систему уравнений электродинамики, подытожив все имеющиеся в этой области знания. Наиболее важным следствием уравнений Максвелла оказалась возможность существования электромагнитных волн, распространяющихся в вакууме со скоростью, значение которой равно электродинамической постоянной c, входящей в эти уравнения. Значение c впервые было получено Кольраушем и Вебером в 1856 г. на основе чисто электрических измерений. Найденная таким образом скорость электромагнитных волн совпала со скоростью света в вакууме, измеренной к тому времени достаточно точно. Это совпадение и навело Максвелла на мысль, что свет представляет собой электромагнитные волны.

Интересно, что решение центральной для оптики проблемы природы света пришло со стороны, как это часто бывает в истории науки. Электрические и магнитные явления изучались параллельно и независимо от оптических. Наиболее важные этапы развития физики — это периоды великих обобщений, когда явления, казавшиеся разобщенными и не связанными между собой, неожиданно становятся всего лишь разными аспектами одного и того же процесса.

Таким образом, несмотря на очевидные различия в способах возбуждения и регистрации электромагнитных волн разных диапазонов, все эти волны имеют единую природу и законы их распространения описываются одними и теми же дифференциальными уравнениями — уравнениями Максвелла.

Дадим краткую качественную характеристику электромагнитных волн разных частотных диапазонов и способов их возбуждения и регистрации.

Электромагнитное излучение возникает при ускоренном движении электрических зарядов. Электромагнитные волны (за исключением света) не наблюдались до 1887 г., когда Герцу удалось генерировать волны длиной от 10 до 100 м с помощью искрового разряда между заряженным и заземленным металлическими шарами. Основной недостаток такого излучателя — затухание колебаний и большая ширина спектра частот излучаемых волн. С помощью современных методов, основанных на использовании электронных ламп и транзисторов, можно генерировать монохроматические электромагнитные волны с частотами до 10^{13} Гц. Эта область частот простирается от радиоволн до микроволн. Диапазон радиоволн используют для радиовещания (длинные, средние и короткие волны), телевидения и космической связи (ультракороткие волны). Радиолокация и радиорелейные линии используют микроволновый диапазон.

Для возбуждения излучения с частотами выше границы микроволнового диапазона традиционные электронные методы непригодны. В лабораторном эксперименте излучение оптического диапазона получают с помощью методов, основанных на следующих физических явлениях.

1. Свечение раскаленных твердых или жидких тел (тепловое излучение), обеспечивающее испусканне сплошного спектра. Тепловое излучение особенно эффективио в ближней инфракрасной области. Электрическая дуга между угольными электродами и электрическнй разряд в лампах высокого давления являются хорошими источниками сплошного спектра в видимой и ближней ультрафиолетовой областях.

2. Свечение возбужденных разреженных газов, вызванное квантовыми переходами внешних электронов в атомах и молекулах с высоких энергетических уровней на низкие. Это излучение имеет дискретный спектр, т. е. состоит из узких спектральных линий. Ртутная дуга низкого давления дает пример источника линейчатого спектра, отдельные линии которого можно выделить с помощью фильтров. Широкое распространеиие в лабораторной практике получили безэлектродные лампы. возбуждаемые СВЧ-разрядом.

3. Свечение различных тел под действием излучения (фотолюминесценция) или бомбардировки электронным пучком (катодолюминесценция).

Все перечисленные источники оптического излучения принципиально отличаются от источников радио- и СВЧ-диапазонов. Излучение электромагнитных волн радиодиапазона происходит при ускоренном движении электронов в антенне радиопередатчика. Все электроны в антенне движутся согласованно: они совершают вынужденные колебания в одинаковой фазе. Так как эти колебания могут поддерживаться очень долго и с высоким постоянством частоты, то излучаемые при этом волны с большой степенью точности можно считать монохроматическими (когерентными). Но любой из упомянутых источников света — это скопление множества возбужденных или все время возбуждаемых атомов, излучающих волновые цуги конечной протяженности. Даже в том случае, когда эти цуги можно характеризовать одной и той же длиной волны. из-за независимого характера актов спонтанного испускания света отдельными атомами соотношения фаз между цугами волн имеют совершенно случайный характер и непрерывно изменяются. Излучение обычных источников света, таких, как раскаленные тела, возбуждаемые электрическим разрядом газы и т. п., представляет собой наложение огромного числа не согласованных между собой цугов волн, т. е. фактически «световой шум» — беспорядочные, некогерентные колебания электромагнитного поля.

Чрезвычайно большие возможности открывает применение в оптических экспериментах принципиально новых источников света - лазеров (оптических квантовых генераторов). Благодаря использованию вынужденного излучения в таком источнике все возбужденные атомы испускают электромагнитные волны согласованно, подобно тому, как это происходит в антенне радиопередатчика. В результате образуется световая волна. близкая по своим свойствам к монохроматической, — когерентная электромагнитная волна. Особые свойства таких источников света. заключающиеся в способности концентрировать энергию в спектре, во времени, в пространстве, связаны с высокой когерентностью их излучения. Сам факт их существования заставляет по-иному подходить к изучению многих оптических явлений.

Наиболее важные способы регистрации электромагнитных волн оптического диапазона основаны на измерении переносимого волной потока энергии. Для этой цели используются фотоэлектрические явления (фотоэлементы. фотоумножители, электронно-оптические преобразователи, фоторезисторы и фотодиоды), фотохимические явления (фотоэмульсии), фотолюминесценция (различные люминесцирующие экраны), термоэлектрические явления (термостолбики, болометры).

Визуальный метод регистрации излучения основан на чувствительности человеческого глаза к свету с длинами волн от 0,4 до 0,7 мкм. В пределах этого интервала чувствительность глаза неодинакова. Закон ее изменения иллюстрирует кривая видности, приведенная на переднем форзаце (по оси ординат отложена относительная чувствительность глаза, т. е. величина. обратная мощности монохроматического излучения, вызывающей одинаковое зрительное ощущение). Несмотря на субъективный характер таких оценок, кривая видности мало изменяется при переходе от одного наблюдателя к другому. Лишь у некоторых людей спектральная чувствительность глаза заметно отличается от иормы. Чувствительность нормального глаза достигает максимального значения в зеленой области спектра при $\lambda = 555$ нм.

На видимую область спектра приходится значительная доля энергии излучения Солнца. В этом иет ничего удивительного, ибо особенности человеческого зрения обусловлены длительным приспособлением органов чувств человека к условиям,

сложившимся на нашей планете, где самым важным источником света всегда было Солнце. А вот совпадение видимой глазом области спектра с одним из «окон прозрачности» земной атмосферы, также играющее исключительно важную роль в повседневной жизни, следует уже рассматривать как подарок природы.

Отметим, что вопрос о восприятии света глазом относится, строго говоря, к физиологической оптике, которая не рассматривается в этой книге, содержащей изложение физической оптики. Впрочем, с восприятием света связано множество явлений, в которых тесно переплетаются и физические, и физиологические процессы. Например, в формировании изображений предметов, воспринимаемых глазом, важную роль играют физические процессы преломления света в хрусталике и стекловидном теле. А особенности восприятия цветов (например, исчезновение окраски предметов при слабом освещении, связанное с наличием двух сортов рецепторов в сетчатке глаза, и различие спектральной чувствительности при ярком и сумеречном освещении) и особенности восприятия движущихся объектов обусловлены характером переработки информации на пути к зрительным центрам головного мозга и относятся к области физиологии.

Всем приемникам света присуща инерционность, характеризуемая временем разрешения. При визуальном наблюдении $\tau \approx 0{,}1$ с. Колебания интенсивности, происходящие за меньшие времена, глаз различить не в состоянии, и зрительное ощущение определяется средней за время τ интенсивностью. Поэтому мы не замечаем быстрых (по сравнению с τ) мельканий света от люминесцентных ламп (эти мелькания характеризуются периодом $T = 0{,}01$ с), от экрана телевизора или киноэкрана ($T = 0{,}02$ с). Для фотоматериалов (т. е. при фотографической регистрации) время экспозиции обычно составляет 10^{-2}—10^{-6} с. В приемниках света, использующих ячейки Керра в качестве затворов, время τ может быть доведено до 10^{-9} с. Наименее инерционные фотоэлектрические приемники характеризуются временем разрешения порядка 10^{-10} с.

За пределами примыкающей к видимому свету со стороны высоких частот ультрафиолетовой области начинается рентгеновское излучение, обладающее большой проникающей способностью.

Различают непрерывный и линейчатый спектры рентгеновского излучения. Непрерывный спектр связан с излучением быстрого электрона при его торможении в теле антикатода. При увеличении ускоряющего напряжения U и, следовательно, кинетической энергии электронов ($mv^2/2 = eU$) коротковолновая граница этого тормозного излучения смещается (максимальная энергия излучаемых рентгеновских квантов $\hbar\omega$ равна кинетической энергии eU бомбардирующих электронов) и, кроме того, появляются узкие максимумы (характеристическое излучение). Длины волн этих дискретных линий зависят от того, какой химический элемент использован в качестве материала антикатода. Электрон, бомбардирующий антикатод, обладает большой кинетической энергией и может выбить электрон с внутренней оболочки атома. В результате атом оказывается возбужденным и может совершить квантовый переход в основное состояние с испусканием кванта электромагнитной энергии. Благодаря большой разности энергий основного состояния и состояния с возбуждением электрона внутренней оболочки для всех атомов с атомным номером Z порядка 10 или больше это излучение принадлежит рентгеновской области спектра.

Для регистрации рентгеновского излучения используют те же физические явления, что и при исследовании ультрафиолетовых лучей. Главным образом применяют фотохимические, фотоэлектрические и люминесцентные методы.

Верхний предел частот электромагнитных волн, которые могут испускать атомные системы, составляет около 10^{20} Гц. Излучение более высоких частот (гамма-излучение) испускается возбужденными атомными ядрами и при превращениях элементарных частиц.

Электромагнитные волны различных диапазонов получили разные названия и обнаруживают себя в совершенно несхожих физических явлениях. Но не следует забывать об общих свойствах электромагнитных волн: все упоминавшиеся виды излучения имеют единую природу и отличаются друг от друга только своими частотами и характерными длинами волн. Распространение всех видов излучения в вакууме подчиняется одним и тем же закономерностям.

Распространение электромагнитных волн в веществе в тех случаях, когда длина волны велика по сравнению с межатомными расстояниями, можно рассматривать

феноменологически без учета атомистического строения среды. Для этого уравнения Максвелла дополняются материальными уравнениями, в которых свойства среды учитываются введением соответствующих констант. Указанное условие выполняется для радиоволн и волн оптического диапазона. Поэтому электромагнитная теория света, основанная на макроскопической электродинамике Максвелла, объясняет в общих чертах все явления, связанные с распространением света. Однако она не способна описать процессы испускания и поглощения света, которые определяются более тонкими особенностями взаимодействия вещества с электромагнитным полем. Вопрос об излучении и поглощении света атомами относится не к одной только оптике, так как в него входит и механика самого атома. Наблюдаемые при этом спектральные закономерности раскрывают не только природу света, но и структуру атома.

Естественное углубление феноменологической электромагнитной теории дает классическая электронная теория, рассматривающая движение дискретных электрических зарядов в веществе и их взаимодействие с электромагнитным полем. Электронная теория вскрывает физическую сущность процессов, описываемых феноменологической теорией. Классическая электронная теория достигла больших успехов благодаря работам Г. А. Лоренца и других ученых. Она не утратила своего значения и в настоящее время, так как дает правильное качественное объяснение обширному кругу электромагнитных и оптических явлений с помощью простых и наглядных моделей. Ограниченность классической электронной теории обусловлена лежащим в ее основе предположением, что поведение электронов в атомах описывается классической механикой. Трудности этой теории привели к созданию квантовой теории, отражающей современные представления о строении вещества.

Квантовая теория оказала сильное влияние и на представления о природе света. В этой теории свободное электромагнитное поле можно рассматривать как совокупность частиц, называемых фотонами или световыми квантами. Каждый фотон характеризуется энергией $\varepsilon = \hbar\omega$ и импульсом $\mathbf{p} = \hbar\mathbf{k}$. Такое описание поля заменяет классическое описание с помощью напряженности электрического поля и индукции магнитного поля. При этом классическая волновая картина получается как предельный случай квантовой, соответствующей большому числу фотонов в одном состоянии.

На основе возникших в квантовой теории корпускулярных представлений удалось объяснить ряд явлений, связанных с превращением энергии света в энергию частиц, необъяснимых с помощью классической волновой теории. Это прежде всего фотоэффект и фотохимические процессы.

Волновые и корпускулярные свойства света не могут быть одновременно логически непротиворечиво объяснены классической физикой, ибо в ней понятия волны и частицы являются взаимоисключающими. В современной квантовой теории, построенной на основе относительности к средствам наблюдения, преодолеваются логические трудности, связанные с корпускулярно-волновым дуализмом. Свет обладает потенциальной возможностью проявлять и волновые, и корпускулярные свойства, но в чистом виде они могут проявиться лишь в разных опытах, проводимых при взаимоисключающих условиях. Эти свойства дополняют друг друга, ибо только их совокупность дает полное представление о свете.

Данная книга содержит описание как волновых, так и корпускулярных свойств света. Однако большее внимание уделено волновым свойствам. Обусловлено это тем, что большинство физических явлений, связанных с взаимодействием излучения и вещества, адекватно описывается так называемой полуклассической теорией. В этой теории поле оптического излучения рассматривается как классическое электромагнитное поле, подчиняющееся уравнениям Максвелла, тогда как поведение атомов вещества описывается квантовой механикой. Полуклассическая теория приводит к успеху при решении большинства задач оптики. Лишь в некоторых задачах, где необходим учет шумов (например, флуктуации лазерного излучения), нужно принимать во внимание не только дискретность процессов поглощения и испускания света атомными системами, но и сам факт квантования поля излучения (т. е. нужно использовать квантовую электродинамику). Интересно отметить, что даже фотоэффект, при объяснении которого в физику впервые было введено понятие фотона, может быть полностью описан в рамках полуклассической теории.

ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ В ВАКУУМЕ. ИСПУСКАНИЕ ВОЛН. КВАЗИМОНОХРОМАТИЧЕСКИЙ СВЕТ

● Основные свойства электромагнитных волн (в частности, света), распространяющихся в пустом пространстве, можно изучить исходя из фундаментальных законов электромагнитного поля, выражаемых уравнениями Максвелла. Наибольшей простотой отличаются плоские монохроматические волны. Более сложные волны (сферические, гауссовы и др.) в малых областях пространства в значительной мере неотличимы от плоских. Эти свойства заключаются в первую очередь в поперечности, поляризации и в определенном соотношении между напряженностями электрического и магнитного полей. Иными соотношениями полей характеризуются стоячие волны, образующиеся при наложении распространяющихся навстречу двух бегущих волн одинаковой частоты, амплитуды и поляризации.

● Испускание электромагнитных волн, согласно классической электродинамике, происходит при ускоренном движении электрических зарядов. Представление о характере поля излучения реальных источников света можно получить, рассматривая сначала простейшую модель источника света — осциллирующий диполь. Свободные колебания заряженного осциллятора затухают из-за излучения. Поэтому испускаемый им свет имеет вид цуга волн убывающей амплитуды и характеризуется не одной частотой, а спектральным контуром конечной ширины.

● Дальнейшее развитие модели реального источника света, когда он рассматривается как совокупность независимо излучающих осцилляторов, позволяет рассмотреть статистические свойства его излучения и получить представление о ширине и форме спектральных линий.

1.1. Плоские монохроматические электромагнитные волны в вакууме

Изучение свойств световых волн в вакууме целесообразно начать с простейшего в математическом отношении случая плоских монохроматических волн. Несмотря на ограниченную применимость такой идеализированной модели для описания реальных световых волн, она во многих случаях оказывается полезной. Для теоретического исследования свойств этих волн будем рассматривать их как одно из возможных решений уравнений Максвелла.

Напомним кратко уравнения Максвелла и их физическое содержание. Электромагнитное поле в вакууме в любой момент времени t определяется заданием в каждой точке $\mathbf{r}$ двух векторов: напряженности $\mathbf{E}(\mathbf{r}, t)$ электрического поля и индукции $\mathbf{B}(\mathbf{r}, t)$ магнитного поля. Через векторы $\mathbf{E}$ и $\mathbf{B}$ выражается сила $\mathbf{F}$, действующая в электромагнитном поле на пробный заряд Q, движущийся со скоростью $\mathbf{v}$:

$$\mathbf{F} = Q\mathbf{E} + Q\mathbf{v}\times\mathbf{B}. \tag{1.1}$$

Напряженность электрического поля в принципе может быть найдена измерением силы $\mathbf{F}$, действующей на неподвижный пробный заряд. С магнитным полем связана та часть силы, которая появляется только при движении заряда (сила Лоренца).

Источниками электромагнитного поля являются заряды и токи, для характеристики которых служат объемная *плотность заряда* ρ и вектор *плотности тока* $\mathbf{j}$. Связь электрического и магнитного полей с их источниками выражается следующими уравнениями:

$$\operatorname{div}\mathbf{E} = \rho/\varepsilon_0, \tag{1.2}$$

$$c^2\operatorname{rot}\mathbf{B} - \partial\mathbf{E}/\partial t = \mathbf{j}/\varepsilon_0. \tag{1.3}$$

Здесь $c \approx 3\cdot 10^8$ м/с — *электродинамическая постоянная*, связанная с электрической и магнитной постоянными соотношением $c = 1/\sqrt{\varepsilon_0\mu_0}$.

Чтобы увидеть, какие экспериментальные законы электромагнетизма выражаются этими уравнениями, перепишем их в интегральной форме. Проинтегрируем обе части уравнения (1.2) по объему V, ограниченному замкнутой поверхностью S, и преобразуем объемный интеграл в левой части в поверхностный с помощью математической теоремы Остроградского—Гаусса. В результате получим

$$\oint_S \mathbf{E}\,d\mathbf{S} = \frac{1}{\varepsilon_0}\int_V \rho\,dV \tag{1.4}$$

— поток напряженности электрического поля сквозь замкнутую поверхность S определяется полным зарядом, находящимся в ограниченном этой поверхностью объеме. Это *закон Гаусса*, эквивалентный в случае статического (т. е. создаваемого неподвижными зарядами) поля экспериментальному закону Кулона, описывающему взаимодействие неподвижных точечных зарядов.

Для получения интегральной формы уравнения (1.3) выберем замкнутый контур l и вычислим поток левой и правой частей (1.3) через произвольную (незамкнутую) поверхность S, опирающуюся на контур l. Поток ротора $\mathbf{B}$ преобразуем с помощью математической теоремы Стокса в циркуляцию вектора $\mathbf{B}$ по контуру l:

$$c^2\oint_l \mathbf{B}\,d\mathbf{l} = \frac{d}{dt}\int_S \mathbf{E}\,d\mathbf{S} + \frac{1}{\varepsilon_0}\int_S \mathbf{j}\,d\mathbf{S}. \tag{1.5}$$

Второе слагаемое в правой части (1.5) выражает справедливую в случае постоянных полей теорему о «магнитном напряжении», согласно которой циркуляция вектора $\mathbf{B}$ по замкнутому контуру определяется полным током I через поверхность, опирающуюся на этот контур. Эту теорему можно рассматривать как следствие закона Био—Савара—Лапласа, определяющего магнитное поле, создаваемое элементом постоянного тока.

Установленный Максвеллом первый член в правой части уравнения (1.5) говорит

о том, что изменяющееся электрическое поле $\partial \mathbf{E}/\partial t$ (ток смещения), так же как и ток проводимости $\mathbf{j}$, создает вихревое магнитное поле. Необходимость добавления члена с током смещения в случае переменных во времени полей видна из того, что только тогда уравнения (1.2) и (1.3) будут удовлетворять закону сохранения заряда. Чтобы показать это, возьмем дивергенцию от левой и правой частей уравнения (1.3). Учитывая, что дивергенция ротора любого векторного поля тождественно равна нулю, и изменяя во втором члене порядок дифференцирования по времени и по координатам, получаем

$$-\frac{\partial}{\partial t}\operatorname{div}\mathbf{E}=\frac{1}{\varepsilon_0}\operatorname{div}\mathbf{j}. \tag{1.6}$$

Подставляя сюда $\operatorname{div}\mathbf{E}$ из (1.2), приходим к уравнению, связывающему изменение плотности заряда с плотностью тока и выражающему закон сохранения заряда:

$$\frac{\partial \rho}{\partial t}+\operatorname{div}\mathbf{j}=0. \tag{1.7}$$

В интегральной форме это уравнение (уравнение непрерывности)

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V \rho \mathrm{d}V=-\oint_S \mathbf{j}\mathrm{d}\mathbf{S} \tag{1.8}$$

говорит о том, что изменение полного заряда в некотором объеме V может быть обусловлено только переносом заряда (потоком вектора плотности тока) через замкнутую поверхность S, ограничивающую этот объем.

Вторая пара уравнений Максвелла

$$\operatorname{div}\mathbf{B}=0, \tag{1.9}$$

$$\operatorname{rot}\mathbf{E}+\frac{\partial \mathbf{B}}{\partial t}=0 \tag{1.10}$$

не содержит источников электромагнитного поля ρ и $\mathbf{j}$. Так же как и первая пара (1.2) и (1.3), это линейные дифференциальные уравнения первого порядка, но в отличие от неоднородных уравнений (1.2) и (1.3) эти уравнения однородны. Интегральная форма уравнения (1.9)

$$\oint_S \mathbf{B}\mathrm{d}\mathbf{S}=0 \tag{1.11}$$

говорит о том, что поток индукции магнитного поля через любую замкнутую поверхность S равен нулю, т. е. линии индукции магнитного поля представляют собой замкнутые кривые. В интегральной форме уравнения (1.10)

$$\oint_l \mathbf{E}\mathrm{d}\mathbf{l}=-\frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{B}\mathrm{d}\mathbf{S} \tag{1.12}$$

мы узнаем экспериментально открытый Фарадеем закон электромагнитной индукции: э.д.с. в замкнутом контуре l определяется скоростью изменения потока магнитной индукции через поверхность S, ограниченную контуром l.

Покажем, что из уравнений Максвелла (1.2) — (1.3), (1.9) — (1.10) следует возможность существования связанных между собой изменяющихся во времени и в пространстве вихревых электрического и магнитного полей даже при отсутствии источников, т. е. при $\rho=0$ и $\mathbf{j}=0$. Такие поля и представляют собой электромагнитные волны в вакууме.

Запишем уравнения Максвелла, используя векторный дифференциальный оператор «набла»:

$$\nabla=\mathbf{i}\frac{\partial}{\partial x}+\mathbf{j}\frac{\partial}{\partial y}+\mathbf{k}\frac{\partial}{\partial z}. \tag{1.13}$$

Дивергенция любого вектора $\mathbf{a}$ с помощью оператора ∇ записывается как скалярное произведение $\nabla\mathbf{a}$, а ротор $\mathbf{a}$ — как векторное произведение $\nabla\times\mathbf{a}$. Полагая в уравнениях (1.2) и (1.3) $\rho=0$ и $j=0$, вместе с уравнениями (1.9) и (1.10) получаем

$$\nabla\mathbf{E}=0, \qquad (1.14) \qquad c^2\nabla\times\mathbf{B}=\dot{\mathbf{E}}, \qquad (1.15)$$

$$\nabla\mathbf{B}=0, \qquad (1.16) \qquad \nabla\times\mathbf{E}=-\dot{\mathbf{B}} \qquad (1.17)$$

(точкой над буквой обозначено, как обычно, дифференцирование по времени). Увидеть существование решений этой системы уравнений в виде волн можно, если из уравнений (1.15) и (1.17), включающих напряженность электрического поля и индукцию магнитного поля, получить уравнения, содержащие только одно из полей (например, $\mathbf{E}$).

Составим векторное произведение оператора ∇ на обе части уравнения (1.17):

$$\nabla\times(\nabla\times\mathbf{E})=-\nabla\times\dot{\mathbf{B}}. \qquad (1.18)$$

Преобразуем левую часть (1.18), раскрывая двойное векторное произведение по формуле* $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=\mathbf{B}(\mathbf{AC})-(\mathbf{AB})\,\mathbf{C}$. В получившемся выражении $\nabla\times(\nabla\times\mathbf{E})=\nabla(\nabla\cdot\mathbf{E})-\nabla^2\mathbf{E}$ первый член равен нулю, так как напряженность $\mathbf{E}$ удовлетворяет уравнению (1.14). В результате (1.18) принимает вид

$$\nabla^2\mathbf{E}=\nabla\times\dot{\mathbf{B}}. \qquad (1.19)$$

Чтобы выразить $\nabla\times\dot{\mathbf{B}}$ через напряженность электрического поля $\mathbf{E}$ с помощью (1.15), продифференцируем обе части (1.15) по времени и изменим в левой части порядок дифференцирования по координатам и времени:

$$c^2\nabla\times\dot{\mathbf{B}}=\ddot{\mathbf{E}}.$$

Подставляя отсюда $\nabla\times\dot{\mathbf{B}}$ в правую часть уравнения (1.19), получаем, что напряженность $\mathbf{E}$ электрического поля удовлетворяет *волновому уравнению*

$$\nabla^2\mathbf{E}-\frac{1}{c^2}\ddot{\mathbf{E}}=0. \qquad (1.20)$$

Аналогично можно показать, что точно такому же уравнению удовлетворяет и индукция $\mathbf{B}$ магнитного поля:

$$\nabla^2\mathbf{B}-\frac{1}{c^2}\ddot{\mathbf{B}}=0. \qquad (1.21)$$

Уравнение (1.20) имеет решения в виде бегущих волн, распространяющихся со скоростью c. Например, если все компоненты векторов

* При применении этой формулы к выражениям, содержащим операторный множитель ∇, нужно заботиться, чтобы он оставался впереди тех множителей, на которые он действует в исходном выражении.

$\mathbf{E}$ и $\mathbf{B}$ зависят только от одной пространственной координаты z (плоская волна), то решением уравнения

$$\frac{\partial^2 \mathbf{E}}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = 0$$

будет любая функция от единственного аргумента $t \pm z/c$:

$$\mathbf{E}(z, t) = \mathbf{E}(t \pm z/c),$$

в чем легко убедиться непосредственной подстановкой. Знак «—» соответствует волне, распространяющейся в положительном направлении оси z, знак «+» — в отрицательном.

Отметим, что любое решение системы уравнений (1.14) (1.17) обязательно удовлетворяет волновым уравнениям (1.20), (1.21), но обратное утверждение неверно, т. е. не всякое решение волновых уравнений дает электромагнитное поле, которое может существовать. Поясним это на простом примере. Волновое уравнение всегда имеет тривиальное нулевое решение, поэтому волновые уравнения допускают, например, решение в виде бегущей волны электрического поля без магнитного поля. Уравнения Максвелла такого решения не допускают.

Исследуем на основе уравнений Максвелла свойства *бегущих плоских монохроматических* электромагнитных волн. В таких волнах зависимость всех компонент векторов $\mathbf{E}$ и $\mathbf{B}$ от координат и времени имеет один и тот же вид и выражается гармонической функцией

$$E(\mathbf{r}, t) = E_{00}\cos(\mathbf{kr} - \omega t + \varphi). \qquad (1.22)$$

Под $E(\mathbf{r}, t)$ здесь можно понимать любую из проекций векторов $\mathbf{E}$ и $\mathbf{B}$. Амплитуда E_{00} и начальная фаза φ плоской монохроматической волны не зависят от $\mathbf{r}$ и t, т. е. одинаковы во всем пространстве во все моменты времени («однородная волна»)*. Никакие реальные волны этим свойством не обладают, поэтому образ плоской монохроматической волны представляет идеализацию, применимость которой к описанию реального волнового процесса зависит не только от рассматриваемого процесса, но и от характера решаемой задачи. Условия применимости этой идеализации в каждом конкретном случае требуют специального рассмотрения. Сейчас же необходимо заметить, что изучение свойств плоской монохроматической волны важно еще и потому, что любая электромагнитная волна может быть представлена в виде суперпозиции таких простых волн (благодаря линейности уравнений Максвелла сумма любых решений также является решением).

---

* Плоскую монохроматическую волну (1.22) можно рассматривать как частный случай гармонических волн общего вида

$$E(\mathbf{r}, t) = E_0(\mathbf{r})\cos[\omega t - g(\mathbf{r})],$$

где $E_0(\mathbf{r})$ и $g(\mathbf{r})$ зависят от положения $\mathbf{r}$ рассматриваемой точки. Поверхности постоянной фазы такой волны, вообще говоря, не совпадают с поверхностями постоянной амплитуды. В таком случае говорят, что волна *неоднородна*.

Аргумент косинуса в (1.22) называется *фазой* волны. Уравнение поверхности постоянной фазы (или *волновой поверхности*)

$$\mathbf{kr} - \omega t = \text{const}$$

определяет в пространстве плоскость, перпендикулярную вектору $\mathbf{k}$ (называемому *волновым вектором*). Эта плоскость перемещается в пространстве вдоль направления вектора $\mathbf{k}$ со скоростью

$$v = \omega/k, \tag{1.23}$$

где k — модуль волнового вектора, называемый *волновым числом*. Скорость перемещения поверхности постоянной фазы в пространстве называется *фазовой скоростью* волны. Период изменения напряженности поля в пространстве — это *длина волны* λ:

$$\lambda = 2\pi/k = v \cdot 2\pi/\omega = vT, \tag{1.24}$$

т. е. длина волны представляет собой то расстояние, на которое перемещается плоскость постоянной фазы за время, равное одному периоду колебаний $T = 2\pi/\omega$.

В дальнейшем для зависимости напряженности поля в волне от координат и времени вместо (1.22) удобно использовать комплексную запись

$$E(\mathbf{r}, t) = \mathrm{Re}\,[E_{00}\exp(i\varphi)\exp i(\mathbf{kr} - \omega t)].$$

Амплитуду E_{00} вместе с комплексным множителем $\exp(i\varphi)$ будем рассматривать как одно комплексное число $E_0 = E_{00}\exp(i\varphi)$. Его модуль равен амплитуде, а аргумент — начальной фазе колебаний в точке $\mathbf{r} = 0$. Знак «Re» при записи будем опускать, не забывая, однако, о том, что физический смысл имеет лишь вещественная часть используемых комплексных выражений:

$$E(\mathbf{r}, t) = E_0 e^{i(\mathbf{kr} - \omega t)}. \tag{1.25}$$

Комплексная запись особенно удобна потому, что при ее использовании дифференцирование напряженности поля по времени $\partial/\partial t$ сводится, как видно из (1.25), просто к умножению на $-i\omega$. Стоящее в показателе экспоненты в (1.25) скалярное произведение $\mathbf{kr}$ можно записать в виде $k_x x + k_y y + k_z z$, поэтому дифференцирование $E(\mathbf{r}, t)$ по координате x сводится к умножению $E(\mathbf{r}, t)$ на ik_x. Так как оператор ∇ означает дифференцирование по координатам, то его применение к напряженности поля сводится к умножению ее на вектор $i\mathbf{k}$. Поэтому для плоской монохроматической волны, у которой напряженность $\mathbf{E}(\mathbf{r}, t)$ электрического поля и индукция $\mathbf{B}(\mathbf{r}, t)$ магнитного поля записаны в комплексной форме (1.25), уравнения Максвелла (1.14)—(1.17) принимают следующий вид:

$$\mathbf{kE} = 0, \tag{1.26}$$

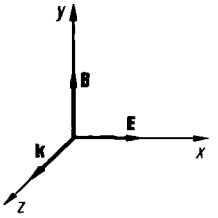
$$c^2\mathbf{k} \times \mathbf{B} = -\omega\mathbf{E}, \tag{1.27}$$

$$\mathbf{kB} = 0, \tag{1.28}$$

$$\mathbf{k} \times \mathbf{E} = \omega\mathbf{B}. \tag{1.29}$$

Из этих формул сразу следует *свойство поперечности* однородных плоских электромагнитных волн: из уравнений (1.26) и (1.28)

видно, что векторы **E** и **B** перпендикулярны направлению волны (вектору **k**), а из уравнений (1.27) и (1.29) — что векторы **E** и **B** ортогональны друг другу и образуют вместе с вектором **k** правую тройку векторов (рис. 1.1).



1.1 Взаимное расположение векторов **E**, **B** и **k**

Уравнения (1.26) и (1.28) — скалярные, а (1.27) и (1.29) — векторные, т. е. каждое из них эквивалентно трем скалярным. Но в случае монохроматических полей [когда все величины зависят от времени как $\exp(-i\omega t)$] из восьми уравнений Максвелла независимы только шесть: уравнение (1.26) является следствием (1.27), а (1.28) — следствием (1.29). В самом деле, подставляя **E** из (1.27) в (1.26), получаем тождество $\mathbf{k}(\mathbf{k}\times\mathbf{B})=0$. Точно так же при подстановке **B** из (1.29) в (1.28) приходим к тождеству $\mathbf{k}(\mathbf{k}\times\mathbf{E})=0$.

Для определения фазовой скорости монохроматических волн $v=\omega/k$ [см. (1.23)] нужно найти связь между частотой ω и модулем волнового вектора **k** (волновым числом). Подставим в уравнение (1.27) **B** из уравнения (1.29):

$$c^2\mathbf{k}\times(\mathbf{k}\times\mathbf{E})=-\omega^2\mathbf{E}.$$

Раскрывая двойное векторное произведение в левой части и учитывая, что, согласно (1.26), $\mathbf{kE}=0$, получаем условие

$$k^2=\omega^2/c^2, \tag{1.30}$$

при выполнении которого уравнения (1.26)—(1.29) имеют нетривиальное (ненулевое) решение. Это значит, что фазовая скорость $v=\omega/k$ для однородных монохроматических волн в вакууме равна входящей в уравнения Максвелла электродинамической постоянной c. Современное значение $c=2{,}9979246\cdot10^8$ м/с. В § 2.11 рассмотрены оптические методы определения этой важнейшей константы.

Подчеркнем, что скорость электромагнитных волн в вакууме не зависит от частоты.

В электромагнитной волне модули векторов **E** и **B** связаны между собой. Подставляя **k** из (1.30) в уравнение (1.27) или в (1.29), находим соотношение между E и B:

$$E=cB. \tag{1.31}$$

Отметим, что это соотношение, как и выражаемая рис. 1.1 связь между направлениями **E** и **B**, выполняется в любой точке в каждый момент времени.

Здесь мы рассмотрели простейшее решение уравнений Максвелла в пустоте — бегущую плоскую монохроматическую волну. В дальнейшем будут рассмотрены и другие решения. *Сферические* монохроматические волны, у которых поверхности постоянной фазы представляют собой концентрические сферы, изучаются в § 1.5. В отличие от плоской волны, амплитуда которой всюду одинакова, амплитуда сферической волны обратно пропорциональна расстоянию до центра.

57097

Другой важный частный случай — *гауссовы* волны (или гауссовы пучки, см. § 6.4), в которых распределение амплитуды по волновой поверхности описывается функцией Гаусса и имеет конечную ширину. Гауссовы волны могут служить математической моделью излучения оптических квантовых генераторов (лазеров).

Небольшой участок сферической волны вдали от ее центра можно приближенно рассматривать как плоскую волну (размеры этого участка должны быть малы по сравнению с расстоянием до центра). Поэтому рассмотренные здесь свойства плоских волн (фазовая скорость, поперечность, соотношение между **E** и **B**) локально (т. е. в каждой точке) справедливы и для сферических волн. То же относится и к небольшим (по сравнению с шириной поперечного распределения амплитуды) участкам гауссовых волн. Подчеркнем, что упомянутые свойства характерны только для бегущих волн. Стоячие волны (см. § 1.3) обладают существенно иными свойствами.

Контрольные вопросы

- Какие экспериментальные законы электромагнетизма содержатся в каждом из уравнений Максвелла?
- Какой физический смысл имеет член с $\partial \mathbf{E}/\partial t$ в уравнении Максвелла (1.3)?
- Эквивалентны ли волновые уравнения (1.20) и (1.21) системе уравнений Максвелла (1.14) — (1.17), из которой они выведены?
- Что такое однородная плоская волна?
- В чем заключается свойство поперечности однородных плоских волн?
- Как из уравнений Максвелла найти фазовую скорость монохроматических электромагнитных волн в вакууме?

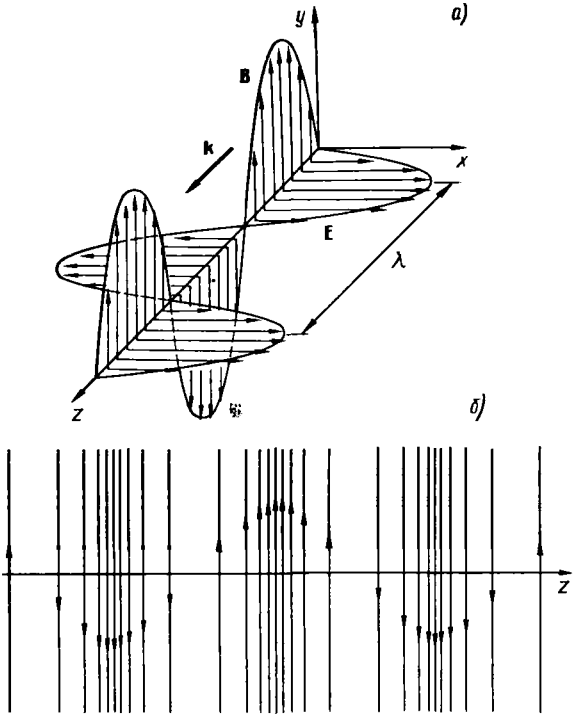
Задачи

1. Показать, исходя из уравнений Максвелла (1.14) — (1.17), что индукция **B** магнитного поля удовлетворяет волновому уравнению (1.21).
2. Показать, что выражение (1.22) описывает монохроматическую волну, поверхности постоянной фазы которой представляют собой плоскости, перпендикулярные вектору **k** и перемещающиеся вдоль **k** с $v=\omega/k$.

1.2. Поляризация плоских монохроматических волн

Выше на основе уравнений Максвелла было показано, что в бегущей плоской электромагнитной волне векторы **E** и **B** в каждой точке и в каждый момент времени образуют с волновым вектором **k** правую тройку векторов (рис. 1.1). В этом заключается свойство поперечности электромагнитных волн.

Выберем ось z системы координат вдоль волнового вектора **k**. Тогда у векторов **E** и **B** могут быть отличны от нуля только проекции на оси x и y. Уравнения Максвелла допускают, в частности, такое решение, когда у вектора **E** во всех точках и во все моменты времени отлична от нуля только одна проекция, например $E_x(z, t)$. Вследствие упомянутого выше свойства поперечности у вектора **B** отлична от нуля только проекция на ось y, т. е. $B_y(z,t)$. Эти проек-

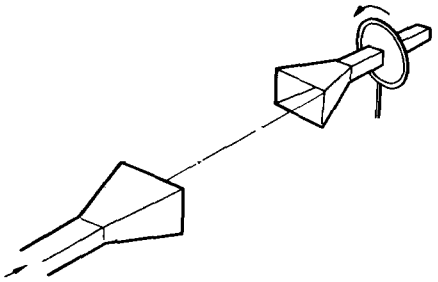


Линейно поляризованная электромагнитная волна

ции связаны между собой соотношением (1.31). Мгновенный «снимок» такой волны, показывающий векторы **E** и **B** в разных точках оси *z* в один момент времени, приведен на рис. 1.2, *а*. В этом случае говорят, что волна имеет *линейную*, или *плоскую, поляризацию.* Плоскость, в которой лежит вектор напряженности электрического поля волны и волновой вектор **k**, называют *плоскостью поляризации* или *плоскостью колебаний**. Чтобы представить себе изменение электрического и магнитного полей с течением времени, можно считать, что вся система векторов на рис. 1.2, *а* движется как целое вдоль оси *z* со скоростью *c*. На рис. 1.2, *б* электрическое поле такой линейно поляризованной плоской волны изображено с помощью силовых линий.

Линейную поляризацию электромагнитных волн легко продемонстрировать простыми опытами в микроволновом диапазоне. Источник (клистронный генератор) через волновод прямоугольного сечения с присоединенным к нему пирамидальным рупором (рис. 1.3) излучает электромагнитную волну линейной поляризации. Приемник состоит из такого же рупора и волновода, внутри которого перпенди-

* Такая терминология не вполне общепринята. В некоторых книгах (особенно старых) плоскостью поляризации называют плоскость, содержащую магнитный вектор **B**.

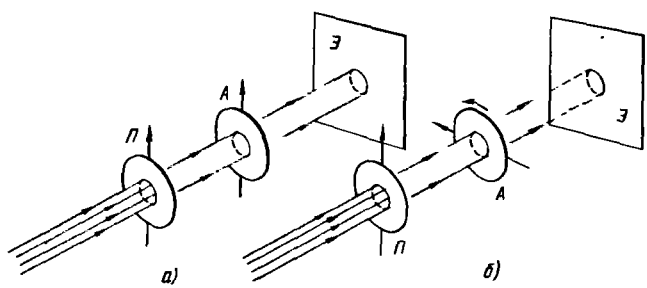


1.3

Установка для демонстрации свойств электромагнитных волн в СВЧ-диапазоне

кулярно продольной оси помещен линейный вибратор с кристаллическим детектором. Снимаемый с детектора сигнал усиливается и регистрируется осциллографом. При одинаковой ориентации рупоров излучателя и приемника регистрируемый сигнал максимален. При повороте приемника (или излучателя) на некоторый угол вокруг продольной оси сигнал уменьшается и исчезает совсем, когда этот угол достигает 90°. Детектор воспринимает только такие колебания, при которых вектор **E** имеет составляющую вдоль оси вибратора. Поэтому исчезновение сигнала при повороте приемника свидетельствует о линейной поляризации регистрируемой волны.

Излучение обычных источников света не поляризовано. Это так называемый *естественный свет*, в котором представлены все направления колебаний вектора **E** в плоскости, перпендикулярной направлению распространения. Физические процессы в источниках, приводящие к испусканию естественного света, рассмотрены в § 1.8. Линейно поляризованный свет получают, пропуская естественный через оптические *поляризаторы*. Существует много типов таких устройств. Их действие основано на различных физических принципах. Некоторые типы поляризаторов описаны ниже (см. § 3.2, § 4.4). С их помощью можно не только получить линейно поляризованный свет, но и выяснить, имеет ли исследуемое излучение линейную поляризацию. Выполняющее такую роль поляризационное устройство называют *анализатором*. Интенсивность пропускаемого через анализатор линейно поляризованного света при повороте анализатора изменяется от максимального значения, когда направление поляризации совпадает с направлением пропускания анализатора, до нуля, когда эти направления перпендикулярны. Схема таких опытов показана на рис. 1.4, *а, б*. Если свет не обладает линейной поляризацией, то при пропускании через анализатор *А* его



1.4

Схема опыта для наблюдения поляризации света

интенсивность не обращается в нуль ни при какой ориентации анализатора.

В отличие от обычных источников света излучение газового лазера, окна разрядной трубки которого наклонены на некоторый угол к ее оптической оси (угол Брюстера, см. § 3.2), обладает линейной поляризацией. Это можно продемонстрировать, пропуская излучение лазера через анализатор. При определенной ориентации анализатора наблюдается полное гашение проходящего излучения.

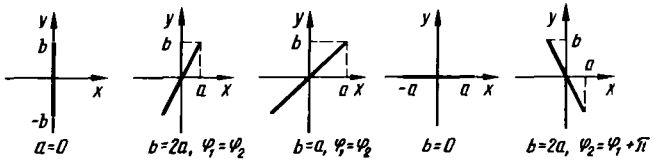
В рассмотренном примере линейно поляризованной волны предполагалось, что вектор **E** во всех точках направлен параллельно или антипараллельно оси x (см. рис. 1.2). Однако в пустом пространстве все направления в плоскости xy, перпендикулярной направлению распространения волны, эквивалентны. Поэтому вектор **E** может иметь любое направление в этой плоскости. Пусть наряду с волной, поляризованной вдоль оси x, в том же направлении распространяется другая волна той же частоты ω, но поляризованная вдоль оси y. Вследствие линейности уравнений Максвелла такое наложение (или суперпозиция) волн также является решением. В зависимости от разности фаз складываемых линейно поляризованных волн результирующая волна может иметь различную поляризацию.

Рассмотрим электрическое поле $\mathbf{E}(z, t)$ волны, возникающей при сложении двух волн одинаковой частоты с ортогональными направлениями линейной поляризации:

$$\begin{aligned} E_x(z, t) &= E_{0x}e^{i(kz-\omega t)} = ae^{-i\varphi_1}e^{i(kz-\omega t)}, \\ E_y(z, t) &= E_{0y}e^{i(kz-\omega t)} = be^{-i\varphi_2}e^{i(kz-\omega t)}. \end{aligned} \qquad (1.32)$$

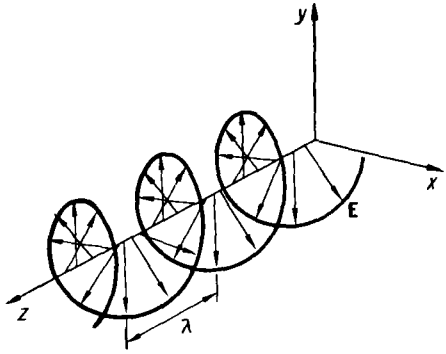
При одинаковых (или отличающихся на $n\pi$, где n — целое число) фазах φ_1 и φ_2 комплексных амплитуд E_{0x} и E_{0y} в каждой точке происходит сложение взаимно перпендикулярных колебаний в одной фазе, что дает колебание в новом направлении. Результирующая волна будет линейно поляризованной. Направление ее поляризации зависит от отношения амплитуд a и b. Различные случаи представлены на рис. 1.5.

Пусть теперь волна, поляризованная вдоль оси y, отстает по фазе на $\pi/2$ от волны, поляризованной вдоль оси x, т. е. $\varphi_1 - \varphi_2 = \pi/2$. Если амплитуды этих волн одинаковы ($a=b$), то вектор **E**



1.5

Сложение двух волн с ортогональными направлениями поляризации



1.6
Вектор **E** в разных точках оси z для волны с левой круговой поляризацией

в любой точке z будет вращаться в плоскости xy против часовой стрелки, оставаясь неизменным по модулю. Например, в точке $z = 0$

$$E_x(t) = a\cos(\omega t + \varphi_1), \qquad (1.33)$$

$$E_y(t) = a\cos(\omega t + \varphi_1 - \pi/2) = a\sin(\omega t + \varphi_1).$$

Такую волну называют *поляризованной по кругу* или *циркулярно поляризованной*. Когда при наблюдении навстречу волне вращение вектора **E** в фиксированной плоскости происходит, как в рассмотренном примере (1.33), против часовой стрелки, говорят о волне *левой* круговой поляризации*.

Попытаемся представить себе электрическое поле такой волны в разных точках оси z в один и тот же момент времени. В точке $z>0$ в некоторый момент t вектор **E** такой же, каким он был в точке $z=0$ в более ранний момент $t - z/c$. Поэтому концы векторов **E** для разных значений z лежат на винтовой линии (рис. 1.6), причем для левой круговой поляризации эта линия соответствует винту с «левой» нарезкой. Чтобы с помощью рис. 1.6 получить представление об изменении напряженности поля с течением времени, можно считать, что весь этот «винт», оставаясь на месте, вращается как целое вокруг оси z с угловой скоростью ω либо что он перемещается поступательно (без вращения) вдоль оси z со скоростью c.

Все сказанное о поведении вектора **E** в волне круговой поляризации можно отнести и к вектору **B**. В самом деле, в бегущей электромагнитной волне векторы **E** и **B** лежат в плоскости, перпендикулярной направлению распространения, в любой момент и в любой точке они перпендикулярны друг другу, а их модули связаны соотношением (1.31).

Правая круговая поляризация соответствует вращению вектора в фиксированной плоскости $z = \text{const}$ в направлении по часовой стрелке:

$$E_x(t) = a\cos(\omega t + \varphi_1),$$
$$E_y(t) = a\cos(\omega t + \varphi_1 + \pi/2) = -a\sin(\omega t + \varphi_1), \qquad (1.34)$$

т. е. $E_y(t)$ опережает по фазе волну $E_x(t)$ на $\pi/2$. «Мгновенный снимок» для такой волны будет отличаться от изображенного на рис. 1.6 тем, что концы векторов **E** лежат на винтовой линии с «правой» нарезкой.

На практике для превращения линейно поляризованного света

---

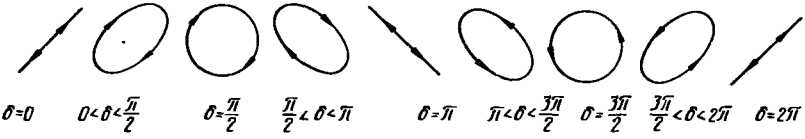
* В некоторых книгах встречаются противоположные определения, поэтому следует с осторожностью относиться к терминам «левая» и «правая» поляризация.

в свет круговой поляризации используют анизотропные кристаллические пластинки, в которых две волны с ортогональными направлениями линейной поляризации имеют различные фазовые скорости. Подбором толщины пластинки можно получить на выходе заданную разность фаз этих волн и тем самым требуемое состояние поляризации. Подробнее такие устройства рассмотрены в § 4.1. Для получения света круговой поляризации можно использовать также явление полного отражения света на границе прозрачных сред (см. § 3.3), при котором возникает разность фаз отраженных волн с ортогональными направлениями линейной поляризации.

Подведем некоторые итоги. При использовании комплексной записи (1.32) для складываемых волн с ортогональными направлениями линейной поляризации результирующая волна имеет линейную поляризацию, если отношение комплексных амплитуд E_{0y}/E_{0x} выражается вещественным числом. Направление поляризации составляет с осью x угол χ такой, что $\operatorname{tg}\chi = E_{0y}/E_{0x}$ (см. рис. 1.5). Результирующая волна имеет круговую поляризацию, если отношение комплексных амплитуд дается мнимым числом, по модулю равным единице. При $E_{0y}/E_{0x} = \exp(-i\pi/2) = -i$ будет волна правой круговой поляризации, при $E_{0y}/E_{0x} = i$ — левой.

В общем случае при произвольном значении комплексных амплитуд E_{0x} и E_{0y} в (1.32) в каждой точке пространства вектор напряженности электрического поля вращается в плоскости $z = \text{const}$, перпендикулярной направлению распространения волны, одновременно изменяясь периодически по модулю, так что конец его описывает эллипс. Ориентация осей и эксцентриситет этого эллипса определяются отношением амплитуд a/b и разностью фаз $\delta = \varphi_2 - \varphi_1$ складываемых волн (см. задачу 2). В таком случае говорят, что волна имеет *эллиптическую* поляризацию. Это наиболее общий вид поляризации монохроматической волны, переходящий при определенных условиях в линейную и круговую поляризации. Картины поляризации при одинаковых амплитудах и разных фазах складываемых взаимно перпендикулярных колебаний показаны на рис. 1.7.

Волну с произвольной (в общем случае эллиптической) поляризацией всегда можно разложить либо на сумму двух линейно поляризованных волн с ортогональными направлениями поляризации, либо на сумму двух поляризованных по кругу волн с правой и левой поляризациями. В зависимости от характера решаемой за-



1.7

Сложение взаимно перпендикулярных колебаний

дачи может оказаться предпочтительным или первое, или второе представление. В дальнейшем мы встретимся с примерами использования каждого из этих базисов. Так, при изучении распространения света в анизотропных средах (кристаллах) удобно разложить падающую волну на сумму двух линейно поляризованных волн, а при изучении естественного (см. § 2.9) и магнитного (см. § 2.8) вращений плоскости поляризации в веществе удобно использовать разложение на две волны с круговой поляризацией.

Контрольные вопросы

- □ В каком случае говорят, что волна имеет линейную поляризацию?
- □ Как на опыте можно определить, имеет ли исследуемый свет линейную поляризацию?
- □ Что такое правая круговая поляризация?
- □ Как зависит мгновенное значение напряженности $\mathbf{E}(z, t)$ электрического поля от z в волне с правой круговой поляризацией? Ответьте на тот же вопрос для магнитного поля $\mathbf{B}(z, t)$.
- □ Каким условиям должны удовлетворять две линейно поляризованные волны, распространяющиеся в одном направлении, чтобы при их сложении получалась волна левой круговой поляризации?

Задачи

1. Покажите, что линейно поляризованную волну с произвольным направлением поляризации можно представить как суперпозицию двух распространяющихся в том же направлении волн правой и левой круговой поляризаций. Как связаны амплитуды этих волн с амплитудой исходной волны?
2. При сложении двух линейно поляризованных волн (1.32) в общем случае возникает волна эллиптической поляризации. Выразить характеризующие эллипс поляризации большую и малую полуоси a_1 и a_2 и угол θ, определяющий его ориентацию (рис. 1.8), через амплитуды a, b и разность фаз $\delta = \varphi_2 - \varphi_1$.

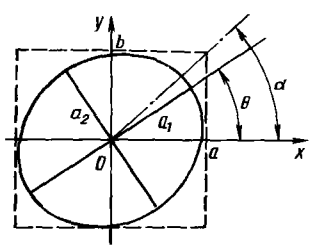
◇ Полуоси a_1 и a_2 эллипса колебаний связаны с амплитудами a и b соотношением $a_1^2 + a_2^2 = a^2 + b^2$. Форма эллипса колебаний характеризуется вспомогательным углом ψ: $\operatorname{tg}\psi = a_2/a_1$. Для нахождения ψ имеем соотношение

$$\sin 2\psi = (\sin 2\alpha)\sin\delta,$$

где $\operatorname{tg}\alpha = b/a$. Угол θ, характеризующий ориентацию эллипса колебаний, находится из соотношения

$$\operatorname{tg} 2\theta = (\operatorname{tg} 2\alpha)\cos\delta.$$

◇



1.8.
Эллипс, описываемый концом вектора **E**

3. Дайте характеристику эллипса колебаний магнитного вектора **B** для волны с эллиптической поляризацией. Сравните с эллипсом колебаний вектора **E**.

1.3. Стоячие электромагнитные волны

В изученных выше бегущих электромагнитных волнах электрическое и магнитное поля направлены перпендикулярно друг другу и в каждой пространственной точке изменяются с течением времени совершенно одинаково (например, в монохроматической волне они совершают гармоническое колебание в одинаковой фазе). Однако это свойство электромагнитных волн не универсально. Существенно иными свойствами обладают *стоячие волны*. Они образуются, в частности, в резонаторах оптических квантовых генераторов (лазеров). Для таких волн характерны пространственное разнесение и сдвиг во времени колебаний электрического и магнитного полей.

Стоячие волны возникают при наложении двух распространяющихся навстречу бегущих монохроматических волн одинаковой частоты, амплитуды и поляризации*. Такая картина получается, в частности, при полном отражении волны от границы. Условия, при которых отражение будет действительно полным, т. е. отраженная волна будет иметь такую же амплитуду, как и падающая, обсуждаются в § 3.3 и 5.7.

Рассмотрим структуру электромагнитного поля стоячей волны, созданной наложением встречных линейно поляризованных плоских волн. Выберем ось z в направлении распространения одной из волн, а ось x — в направлении поляризации. Начало отсчета расстояний на оси z выберем в точке, где колебания напряженности электрического поля обеих волн происходят в одинаковой фазе, а начало отсчета времени — в тот момент, когда эти напряженности в начале координат достигают максимума. При этом формулы для $\mathbf{E}_1(z, t)$ и $\mathbf{B}_1(z, t)$ принимают особенно простой вид:

$$\mathbf{E}_1=(E_0 e^{-i(\omega t-kz)},\ 0,\ 0),$$

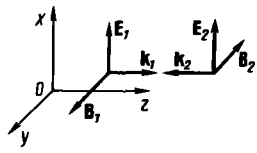
$$\mathbf{B}_1=(0,\ B_0 e^{-i(\omega t-kz)},\ 0). \qquad (1.35)$$

Эта волна распространяется в положительном направлении оси z. Для волны, распространяющейся навстречу, мы должны написать

$$\mathbf{E}_2=(E_0 e^{-i(\omega t+kz)},\ 0,\ 0),$$

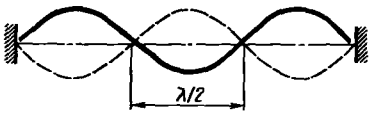
$$\mathbf{B}_2=(0,\ -B_0 e^{-i(\omega t+kz)},\ 0). \qquad (1.36)$$

Здесь учтено, что векторы $\mathbf{E}$, $\mathbf{B}$ и $\mathbf{k}$ в каждой из волн в соответствии с рис. 1.1 должны образовывать правую тройку векторов (рис. 1.9). Теперь из (1.35) и (1.36) легко найти результирующее электромагнитное поле $\mathbf{E}=\mathbf{E}_1+\mathbf{E}_2$,



1.9
Взаимная ориентация векторов $\mathbf{E}$, $\mathbf{B}$ и $\mathbf{k}$

* Это частный случай интерференции, другие многообразные проявления которой подробно рассматриваются в гл. 5



1.10

Стоячая волна в натянутом шнуре

$$\mathbf{B}=\mathbf{B}_1+\mathbf{B}_2;$$

$$\mathbf{E}=(2E_0e^{-i\omega t}\cos kz,\ 0,\ 0),$$

$$\mathbf{B}=(0,\ 2iB_0e^{-i\omega t}\sin kz,\ 0),$$

или в вещественном виде

$$\mathbf{E}=(2E_0\cos\omega t\cos kz,\ 0,\ 0),\quad \mathbf{B}=(0,\ 2B_0\sin\omega t\sin kz,\ 0). \tag{1.37}$$

Из этих формул видно, что вектор напряженности электрического поля результирующей волны в каждой точке совершает гармоническое колебание в направлении оси x с частотой ω, причем амплитуда колебаний изменяется от максимального значения $2E_0$ в плоскостях $z=(\pi/k)n=(\lambda/2)n$ ($n=0,\ \pm1,\ \pm2,\ ...$), называемых *пучностями* электрического поля, до нуля в плоскостях $z=(\pi/k)\times\times(n+{}^1/_2)=(\lambda/2)(n+{}^1/_2)$ ($n=0,\ \pm1,\ \pm2,\ ...$), называемых *узлами*. Фаза колебаний $\mathbf{E}$ во всех точках между соседними узлами одинакова, а колебания по разные стороны узла происходят в противофазе.

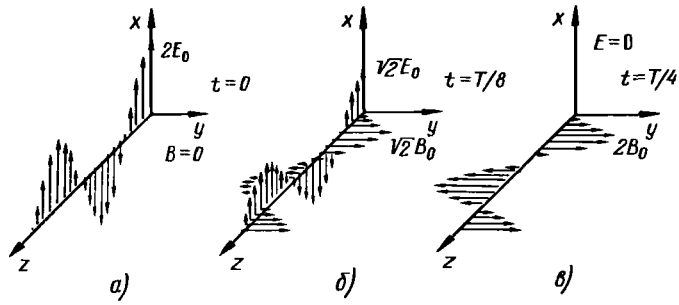
Здесь уместно вспомнить картину смещений в стоячей волне, возбуждаемой в закрепленном на концах натянутом шнуре (рис. 1.10). Такую стоячую волну можно рассматривать как одно из нормальных колебаний (мод) механической системы с распределенными параметрами, т. е. с бесконечным числом степеней свободы. Напомним, что при нормальном колебании в системе все ее элементы совершают чисто гармоническое движение с одной и той же характерной для данной моды частотой ω_n и с определенным соотношением амплитуд. Частоты нормальных мод закрепленного на концах шнура (или струны) образуют дискретный спектр и могут быть найдены из условия, что на длине шнура l укладывается целое число n полуволн: $l=n\lambda/2$, откуда

$$\omega_n=n\pi v/l,\ \nu_n=nv/(2l). \tag{1.38}$$

Здесь v — скорость распространения волн в шнуре.

Аналогично, стоячую электромагнитную волну между двумя параллельными идеально отражающими плоскостями, находящимися на расстоянии l друг от друга, можно рассматривать как нормальное колебание электромагнитного поля в таком плоском резонаторе. Частоты нормальных мод плоского резонатора образуют эквидистантный спектр и выражаются той же формулой (1.38), если скорость v в ней заменить на скорость света c.

Колебания магнитного поля [см. (1.37)] происходят в направлении, перпендикулярном напряженности электрического поля, и также представляют собой стоячую волну. Однако пучности и узлы стоячей волны магнитного поля сдвинуты вдоль оси z на четверть длины волны по отношению к пучностям и узлам электрического поля. Колебания магнитного поля отстают во времени на четверть периода от колебаний электрического поля. Так, например, в момент времени $t=0$ индукция магнитного поля [см. (1.37)] всюду равна нулю, а напряженность электрического поля максимальна и распределена в пространстве по закону $2E_0\cos kz$ (рис. 1.11, *а*). Спустя промежуток времени $T/8$ ($T=2\pi/\omega$ — период колебаний) напряженность электрического поля уменьшается до $\sqrt{2}\,E_0\cos kz$, а индукция магнитного поля, возрастая, достигает значения



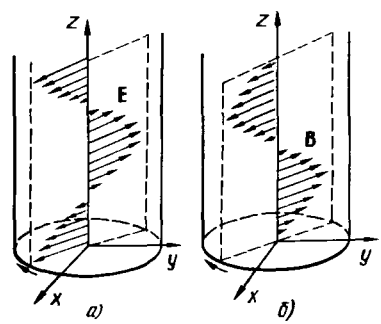
1.11

Электрическое и магнитное поля линейно поляризованной стоячей волны

$\sqrt{2}B_0 \sin kz$ (рис. 1.11, *б*). К моменту времени $T/4$ напряженность электрического поля всюду обращается в нуль, а индукция магнитного достигает максимального значения $2B_0 \sin kz$ (рис. 1.11, *в*).

Стоячая электромагнитная волна, как и бегущая, может иметь и иные, отличные от линейной состояния поляризации. Например, стоячая волна круговой поляризации получается при наложении распространяющихся навстречу волн круговой поляризации, имеющих одинаковые амплитуды и одинаковые направления вращения векторов поля (см. задачу 1). Это значит, что одна из волн имеет правую, а другая — левую круговую поляризацию. Такая картина возникает при отражении циркулярно поляризованной волны от идеального плоского зеркала.

Вектор напряженности электрического поля вращается в ту же сторону, что и в каждой из складываемых волн, однако во всех точках он направлен в каждый момент вдоль одной и той же прямой (рис. 1.12, *а*). Модуль его изменяется по косинусоидальному закону от максимального значения $2E_0$ в пучностях, т. е. в тех точках, где вращение векторов **E** в складываемых волнах происходит в одинаковой фазе, до нуля в узлах, где векторы **E** в складываемых волнах направлены противоположно в каждый момент времени. Вектор индукции магнитного поля результирующей стоячей волны вращается в каждой точке так же, как и вектор напряженности электрического поля. В отличие от бегущей волны векторы **E** и **B** в каждый момент времени лежат в одной плоскости. Простран-



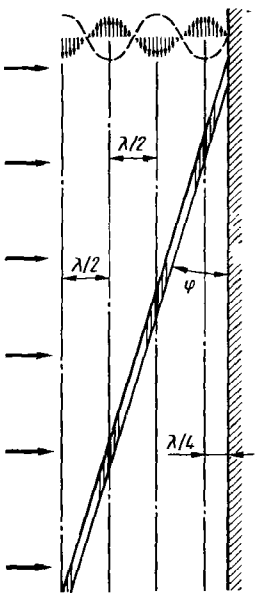
1.12

Пространственное распределение полей в циркулярно поляризованной стоячей волне

ственное распределение магнитного поля отличается сдвигом вдоль оси z на четверть длины волны, т. е. пучности магнитного поля совпадают с узлами электрического поля и наоборот (рис. 1.12, *б*).

Первые опыты по наблюдению стоячих световых волн были выполнены Винером (1890). В опытах Винера стоячая волна возникала при отражении от плоского металлического зеркала падающего нормально почти монохроматического света. Условия отражения от хорошего (идеального) проводника таковы, что первый узел электрического поля стоячей волны должен располагаться на поверхности зеркала. Для регистрации положения узлов и пучностей стоячей световой волны использовалось действие света на фотографическую эмульсию, в составе которой имеются светочувствительные кристаллические зерна бромистого серебра. Под действием света начинается разложение бромистого серебра, что приводит к почернению (после процесса химического проявления) тех участков фотоэмульсии, которые были подвержены освещению. В соответствии со слоистым распределением амплитуд колебаний электрического и магнитного полей в стоячей световой волне почернение фотоэмульсии должно происходить слоями. Из-за малого расстояния между пучностями и узлами ($\lambda/4$) такие слои располагаются очень близко друг к другу (на расстоянии $\lambda/2$).

Трудность наблюдения столь близких слоев Винер преодолел, расположив очень тонкий светочувствительный слой на поверхности стеклянной пластинки под очень малым углом φ (около 1′) к поверхности зеркала (рис. 1.13). Этот слой пересекается с плоскостями пучностей по параллельным прямым, расстояние между которыми, как видно из рис. 1.13, равно $\lambda/(2\sin\varphi)$. При $\lambda = 0{,}5$ мкм и $\varphi \approx 1'$ расстояние между полосами почернения на пластинке составляет около 1 мм.



1.13
Схема опытов Винера

В таких условиях нетрудно увидеть, что первая полоса почернения фотоэмульсии не совпадает с зеркалом, а отстоит от него на расстоянии $\lambda/4$ (по нормали). Как раз здесь располагается пучность электрического поля стоячей световой волны. Этот результат однозначно показывает, что фотохимическое действие света обусловлено электрическим полем световой волны.

Впоследствии Друде и Нернст в аналогичных опытах (1892) показали, что флуоресценция вещества под действием стоячей световой волны также максимальна в пучностях электрического поля. Затем Айвс (1933) обнаружил, что фотоэффект тоже вызывается электрическим полем световой волны.

Результаты всех этих опытов легко понять на основе представлений электронной теории. Если в процессе распространения световой волны ее электрическое и магнитное поля выступают как равноправные, то при взаимодействии с веществом основную роль играет электрическое поле. Любой физический процесс взаимодействия света с веществом сводится в первую очередь к действию поля световой волны на электроны, входящие в состав вещества. Действующая на электрон сила, выражаемая формулой (1.1), определяется главным образом электрическим полем световой волны. В самом деле, характеристики электрического и магнитного полей в световой волне связаны соотношением (1.31), поэтому обусловленная магнитным полем составляющая силы в (1.1) отличается множителем v/c от силы eE, действующей на электрон со стороны электрического поля. Обычно $v \ll c$, поэтому вынужденное светом движение зарядов в веществе определяется электрическим полем световой волны. Пространственное разнесение электрического и магнитного полей в стоячей волне дает возможность получить прямое экспериментальное подтверждение этих выводов.

Контрольные вопросы

Каким условиям должны удовлетворять частоты, амплитуды, фазы и состояния поляризации волн, распространяющихся навстречу для того, чтобы при их сложении возникла стоячая волна?

Что такое узлы и пучности стоячей волны? Опишите пространственное распределение электрического и магнитного полей в линейно поляризованной стоячей волне.

На каком расстоянии друг от друга находятся соседние узлы стоячей волны?

Какой частотный спектр имеют нормальные колебания электромагнитного поля в резонаторе, образованном плоскими параллельными идеальными зеркалами, находящимися на расстоянии l друг от друга?

Каким образом опыт Винера доказывает, что фотохимическое действие света обусловлено электрическим полем световой волны?

Почему во взаимодействии света с веществом основную роль играет электрическое поле световой волны? Как это объясняется электронной теорией?

Задача

Найти зависимость векторов электрического и магнитного полей от координат и времени в стоячей волне круговой поляризации, возникающей при наложении распространяющихся навстречу циркулярно поляризованных волн с одинаковым направлением вращения векторов электрического поля.

◇ Пусть волна, распространяющаяся вдоль оси z, имеет правую поляризацию:

$$\begin{aligned} \mathbf{E}_1(z, t) &= (E_0 e^{i(kz-\omega t)}, -iE_0 e^{i(kz-\omega t)}, 0), \\ \mathbf{B}_1(z, t) &= (iB_0 e^{i(kz-\omega t)}, B_0 e^{i(kz-\omega t)}, 0). \end{aligned} \tag{1.39}$$

Чтобы вращение вектора $\mathbf{E}_2$ встречной волны происходило в том же направлении, она должна иметь левую поляризацию:

$$\mathbf{E}_2(z, t)=(E_0e^{-i(kz+\omega t)}, -iE_0e^{-i(kz+\omega t)}, 0),$$
$$\mathbf{B}_2(z, t)=(-iB_0e^{-i(kz+\omega t)}, -B_0e^{-i(kz+\omega t)}, 0). \quad (1.40)$$

Для результирующей волны из (1.39) и (1.40) получаем

$$\mathbf{E}(z, t)=(2E_0e^{-i\omega t}\cos kz, -2E_0ie^{-i\omega t}\cos kz, 0),$$
$$\mathbf{B}(z, t)=(-2B_0e^{-i\omega t}\sin kz, 2B_0ie^{-i\omega t}\sin kz, 0),$$

или в вещественной форме

$$\mathbf{E}(z, t)=(2E_0\cos kz\cos\omega t, -2E_0\cos kz\sin\omega t, 0),$$
$$\mathbf{B}(z, t)=(-2B_0\sin kz\cos\omega t, 2B_0\sin kz\sin\omega t, 0). \quad (1.41)$$

Описываемое формулами (1.41) поведение векторов $\mathbf{E}$ и $\mathbf{B}$ иллюстрируется рисунком 1.12. ◇

1.4. Энергия электромагнитных волн

Электромагнитное поле обладает энергией. При распространении электромагнитных волн происходит перенос энергии поля в пространстве. Вопрос о переносимой световой волной энергии можно рассмотреть на основе уравнений Максвелла (1.14)—(1.17), описывающих электромагнитное поле в вакууме.

Напомним известные из курса электричества и магнетизма выражения для объемной плотности энергии электрического поля

$$w_э=\tfrac{1}{2}\varepsilon_0E^2 \quad (1.42)$$

и плотности энергии магнитного поля

$$w_м=1/(2\mu_0)B^2=\tfrac{1}{2}\varepsilon_0c^2B^2. \quad (1.43)$$

Отсюда для скорости изменения плотности энергии электромагнитного поля $w=w_э+w_м$ в данной точке можем получить [путем дифференцирования (1.42) и (1.43) по времени] следующее выражение:

$$\frac{d}{dt}(w_э+w_м)=\varepsilon_0(\mathbf{E}\dot{\mathbf{E}}+c^2\mathbf{B}\dot{\mathbf{B}}). \quad (1.44)$$

Правую часть (1.44) можно с помощью уравнений Максвелла преобразовать следующим образом. Обе части уравнения (1.15), содержащего $\dot{\mathbf{E}}$, умножим скалярно на $\mathbf{E}$, а обе части уравнения (1.17), содержащего $\dot{\mathbf{B}}$, умножим скалярно на $\mathbf{B}c^2$:

$$\mathbf{E}c^2(\nabla\times\mathbf{B})=\mathbf{E}\dot{\mathbf{E}}, \quad (1.45)$$
$$\mathbf{B}c^2(\nabla\times\mathbf{E})=-c^2\mathbf{B}\dot{\mathbf{B}}. \quad (1.46)$$

Теперь вычтем почленно уравнение (1.46) из (1.45). Тогда в правой части мы получим интересующее нас выражение (1.44) для скорости изменения плотности энергии электромагнитного поля:

$$\varepsilon_0c^2[\mathbf{E}(\nabla\times\mathbf{B})-\mathbf{B}(\nabla\times\mathbf{E})]=\frac{d}{dt}(w_э+w_м). \quad (1.47)$$

Легко показать, что выражение в квадратных скобках в левой части (1.47) представляет собой дивергенцию векторного произведения $\mathbf{B}\times\mathbf{E}$. В самом деле, применим к $\mathbf{B}\times\mathbf{E}$ дифференциальный оператор ∇. Чтобы воспользоваться правилом дифференцирования произведения, можно в смешанном векторно-скалярном произведении $\nabla(\mathbf{B}\times\mathbf{E})$ выполнить циклические перестановки сомножителей так, чтобы оператор ∇ действовал только на один из сомножителей. Это дает $\mathbf{E}(\nabla\times\mathbf{B})$ и $-\mathbf{B}(\nabla\times\mathbf{E})$ (во втором случае пришлось в векторном произведении переставить сомножители, чтобы оператор ∇ стоял перед $\mathbf{E}$, и одновременно изменить знак). Таким образом,

$$\nabla(\mathbf{B}\times\mathbf{E})=\mathbf{E}(\nabla\times\mathbf{B})-\mathbf{B}(\nabla\times\mathbf{E}). \tag{1.48}$$

С помощью этого тождества выражению (1.47) можно придать вид *уравнения непрерывности* для плотности энергии электромагнитного поля $w=w_э+w_м$ (подобно уравнению непрерывности (1.7) для плотности заряда, выражающему закон сохранения заряда):

$$dw/dt=-\mathrm{div}\mathbf{S}, \tag{1.49}$$

где вектор $\mathbf{S}$ имеет смысл плотности потока энергии электромагнитного поля и называется *вектором Пойнтинга:*

$$\mathbf{S}=\varepsilon_0c^2\mathbf{E}\times\mathbf{B}. \tag{1.50}$$

Уравнение (1.49), полученное как следствие уравнений Максвелла, выражает *закон сохранения энергии* для электромагнитного поля. Проинтегрируем обе части (1.49) по некоторому объему V, ограниченному замкнутой поверхностью σ. Интеграл по объему от $\mathrm{div}\mathbf{S}$ в правой части преобразуем с помощью математической теоремы Остроградского—Гаусса в интеграл по поверхности σ, ограничивающей этот объем:

$$\frac{d}{dt}\int_V w\,dV=-\oint_\sigma \mathbf{S}\,d\boldsymbol{\sigma}. \tag{1.51}$$

Уравнение (1.51), представляющее собой интегральную форму уравнения непрерывности (1.49), говорит о том, что изменение энергии электромагнитного поля в каком-то объеме V, не содержащем зарядов и токов (так как при его выводе предполагалось, что $\varrho=0$ и $\mathbf{j}=0$), равно потоку энергии в этот объем через охватывающую его замкнутую поверхность σ.

В *бегущей* электромагнитной волне происходит направленный перенос энергии электромагнитного поля в пространстве. Направление и интенсивность переноса энергии характеризуются вектором Пойнтинга (1.50). Для рассмотренных в § 1.1 плоских волн в вакууме векторы $\mathbf{E}$ и $\mathbf{B}$ в любой точке и в любой момент времени ортогональны друг другу и образуют вместе с вектором $\mathbf{k}$ правую тройку векторов (см. рис. 1.1). Поэтому направление вектора Пойнтинга $\mathbf{S}$ в таких волнах совпадает с направлением волнового вектора $\mathbf{k}$: энергия переносится в направлении, перпендикулярном поверхно-

стям постоянной фазы. Учитывая ортогональность векторов **E** и **B**, а также соотношение (1.31) между их модулями, для модуля плотности потока энергии из (1.50) получаем следующее выражение:

$$S=c\varepsilon_0 E^2=c\varepsilon_0 c^2 B^2. \tag{1.52}$$

Из соотношения (1.31) следует также, что объемная плотность энергии электрического поля $w_э$ (1.42) в бегущей электромагнитной волне в каждой точке и в любой момент времени равна плотности энергии магнитного поля $w_м$ (1.43). Поэтому выражаемую формулой (1.52) плотность потока энергии можно записать как произведение полной плотности энергии $w=w_э+w_м$ электромагнитного поля бегущей волны на скорость волны c:

$$S=c(w_э+w_м)=cw. \tag{1.53}$$

В монохроматической волне зависимость напряженности электрического поля от координат и времени выражается формулой (1.22). Подставляя $E(\mathbf{r}, t)$ из (1.22) в (1.52) и используя тождество $\cos^2\alpha=(1+\cos 2\alpha)/2$, для $S(\mathbf{r}, t)$ получаем

$$S(\mathbf{r}, t)={}^1/_2 c\varepsilon_0 E_0^2[1+\cos 2(\mathbf{kr}-\omega t)]. \tag{1.54}$$

При характерных для оптического диапазона высоких частотах ω ($\approx 10^{15}$ с$^{-1}$) колебания потока энергии волны в каждой точке, происходящие в соответствии с (1.54) на частоте 2ω, ненаблюдаемы и физический интерес представляет лишь *среднее по времени* значение S, называемое обычно *интенсивностью* света. Производя в (1.54) усреднение по времени, находим, что в монохроматической волне средняя плотность потока энергии $\langle S\rangle$ во всех точках одинакова:

$$\langle S\rangle={}^1/_2 c\varepsilon_0 E_0^2 \tag{1.55}$$

(угловые скобки обозначают усреднение по времени).

Совершенно аналогично в монохроматической волне в каждой точке осциллируют на частоте 2ω плотность энергии $w_э$ электрического поля [см. (1.42)], пропорциональная $E^2(\mathbf{r}, t)$, и равная ей плотность энергии $w_м$ магнитного поля [см. (1.43)]. Их средние по времени значения во всех точках одинаковы и пропорциональны квадрату амплитуды:

$$\langle w_э\rangle=\langle w_м\rangle={}^1/_4\varepsilon_0 E_0^2,$$
$$\langle w\rangle=\langle w_э\rangle+\langle w_м\rangle={}^1/_2\varepsilon_0 E_0^2. \tag{1.56}$$

Поэтому в формуле (1.55) правую часть можно выразить с помощью (1.56) через среднюю плотность энергии поля волны:

$$\langle S\rangle=c\langle w\rangle. \tag{1.57}$$

Это соотношение можно получить и непосредственно из (1.53), усредняя его за промежуток времени, равный периоду осцилляций S и w.

Таким образом, в световой волне энергия, переносимая в 1 с

через площадку 1 м^2, перпендикулярную направлению распространения волны, равна произведению скорости света c на объемную плотность w энергии электромагнитного поля волны.

В *стоячей* электромагнитной волне никакого направленного переноса энергии в среднем не происходит. Рассмотрим, например, стоячую волну линейной поляризации, электрическое и магнитное поля которой даются выражениями (1.37). У соответствующего такой волне вектора Пойнтинга **S** (1.50) отлична от нуля только z-проекция:

$$S_z = \varepsilon_0 c^2 (E_x B_y - E_y B_x) = \varepsilon_0 c^2 E_0^2 \sin 2\omega t \sin 2kz. \qquad (1.58)$$

При усреднении выражения (1.58) по времени получаем $\langle S_z \rangle = 0$. Сделаем замечание о *пространственной локализации* потока энергии, поверхностная плотность которого выражается формулой (1.50). Поток вектора Пойнтинга **S** через *замкнутую* поверхность, как видно из (1.51), дает энергию электромагнитного поля, выходящую в 1 с из объема, ограниченного этой поверхностью. Можно ли утверждать, что и в случае *незамкнутой* поверхности поток вектора Пойнтинга дает энергию, проходящую через поверхность в 1 с? Такая локализация потока энергии в пространстве отнюдь не вытекает из полученного для замкнутой поверхности соотношения (1.51). Поэтому применение вектора Пойнтинга к вычислению потока энергии через незамкнутую поверхность иногда приводит к парадоксам. Известный пример такого парадокса — непараллельные статические электрическое и магнитное поля, для которых $\mathbf{E} \times \mathbf{B} \neq 0$ и, следовательно, $\mathbf{S} \neq 0$, хотя поток вектора **S** через замкнутую поверхность конечно же равен нулю.

В отличие от статических полей в оптике иногда имеет смысл говорить о потоке энергии (т. е. потоке вектора **S**) через незамкнутую поверхность. Подробный анализ показывает, что такая локализация потока энергии допустима, если размеры площадки велики по сравнению с длиной волны. С помощью непрозрачного экрана с большим (по сравнению с длиной волны) отверстием можно выделить и измерить энергию, переносимую излучением через отверстие. Можно утверждать, что в падающей волне и при отсутствии непрозрачного экрана через площадку, совпадающую с отверстием, будет проходить та же энергия.

Способы регистрации электромагнитных волн оптического диапазона основаны на измерении потока переносимой волной энергии. Всем существующим приемникам света присуща бóльшая или меньшая *инерционность*, характеризуемая некоторой постоянной времени τ. Колебания интенсивности регистрируемого излучения, происходящие за времена, меньшие τ, приемник разрешить не в состоянии.

Чтобы измерить мгновенное значение напряженности электрического поля в световой волне, приемник должен был бы иметь время разрешения, которое малó по сравнению с периодом световых колебаний. Для видимого света этот период составляет примерно

10^{-15} с. Наименее инерционные приемники оптического излучения основаны на фотоэлектрическом эффекте (см. § 9.5). Они могут иметь время разрешения до 10^{-10} с. Это время много больше периода оптических колебаний. Поэтому все существующие приемники излучения оптического диапазона могут измерять только величины, *квадратичные* по напряженности поля и *усредненные* за времена, не меньшие времени разрешения приемника. Чтобы подчеркнуть это, их иногда называют квадратичными детекторами.

Раздел оптики, связанный с измерением световых потоков, называется *фотометрией* (см. § 1.10).

Для измерения интенсивности излучения с любой длиной волны можно использовать тепловое действие излучения. Поглощаемую чувствительным элементом такого приемника мощность падающего на него излучения можно измерить по превышению температуры чувствительного элемента над температурой окружающей среды (которая может быть довольно низкой, если в качестве среды взять, например, жидкий гелий). Стационарное состояние с определенной разностью температур устанавливается, когда в каждый отрезок времени за счет излучения чувствительный элемент получает столько же энергии, сколько теряет из-за теплопроводности и других причин.

Описываемый детектор должен иметь устройство для калибровки. Перекрывая доступ излучению, по вмонтированному в чувствительный элемент резистору R пропускают ток. Силу тока I подбирают такой, чтобы установилась та же разность температур чувствительного элемента и окружающей среды, что и при поглощении измеряемого излучения. В этих условиях мощность тока I^2R, рассеиваемая резистором, равна мощности излучения.

Детекторы подобного типа имеют много разновидностей. Их отклик определяется только падающим на чувствительный элемент потоком энергии и не зависит от спектрального состава излучения. Наибольшее распространение среди таких *неселективных приемников* получили *термоэлементы*, в которых для измерения разности температур поглощающего излучение чувствительного элемента и окружающей среды использована термопара. При небольших размерах термопар инерционность этих приемников может быть доведена до 10^{-2}—10^{-3} с. Несколько термоэлементов, соединенных последовательно для увеличения чувствительности, образуют *термостолбик*. Другой пример неселективного приемника — *термометр сопротивлений* (болометр). Его действие основано на изменении сопротивления тонкого слоя металла или полупроводника при нагревании энергией поглощаемого излучения.

К другому классу детекторов излучения относятся *фотоэлектрические* приемники. Это приемники, основанные на явлении внешнего фотоэффекта — *фотоэлементы* и *фотоумножители* (см. § 9.5) и внутреннего фотоэффекта — *фотосопротивления* и *фотодиоды*. В отличие от приемников, основанных на тепловом действии, где поглощаемая мощность излучения распределяется между всеми частицами чувствительного элемента, здесь она отдается лишь малой доле всех

частиц тела, а именно только некоторым входящим в его состав электронам. Поток высвобождаемых при фотоэффекте электронов (фототок) пропорционален потоку падающего излучения. Фотоэлектрическим приемникам свойственны высокая чувствительность и хорошие временны́е характеристики. Все эти приемники обладают *селективностью:* их чувствительность в значительной степени зависит от длины волны падающего излучения. Это нужно учитывать при измерениях потока энергии в разных участках спектра. Для абсолютных измерений интенсивности такие приемники требуют специальной калибровки.

Контрольные вопросы

- Как объемная плотность энергии электрического поля выражается через напряженность поля?
- Как следует понимать утверждение, что закон сохранения энергии электромагнитного поля выражается уравнением непрерывности (1.49)? Где при его выводе использованы уравнения Максвелла?
- Какие энергетические превращения описывает выражение (1.58)?
- В каких случаях с помощью вектора Пойнтинга можно вычислять поток энергии через незамкнутую поверхность?

Задачи

1. Докажите тождество (1.48), используя проекции векторов на оси декартовой системы координат.
2. Найдите плотность потока энергии $S_z(z, t)$ в стоячей волне круговой поляризации. Чему равно среднее значение S_z?
3. Средняя плотность потока солнечной радиации на поверхности Земли (солнечная постоянная) составляет 1,4 мВт/м^2 (при условии, что поглощения в атмосфере не происходит). Оцените амплитуду напряженности электрического поля световой волны, приходящей на Землю от Солнца. Ответ: $E_0 \approx 1$ кВ/м.
4. Две волны одинаковой частоты распространяются в одном направлении. Покажите, что плотность потока энергии результирующей волны равна сумме плотностей потоков $S_1 + S_2$ каждой из волн, если волны линейно поляризованы во взаимно перпендикулярных направлениях.

1.5. Излучение электромагнитных волн. Сферические волны

До сих пор мы изучали свойства электромагнитных волн в пустом пространстве, основываясь на уравнениях Максвелла (1.2) и (1.3) с равными нулю источниками ρ *и* $\mathbf{j}$, входящими в правые части этих уравнений. Такие уравнения без источников описывают распространение волн в пустоте, но не позволяют понять, как возникают элекромагнитные волны.

Испускание электромагнитных волн происходит при *ускоренном* движении электрических зарядов*. Простейшая модель источника света получается, если рассмотреть диполь, электрический момент которого $p(t)$ гармонически изменяется со временем. Такой диполь

* Это утверждение относится к движению заряда в вакууме. В веществе излучение возможно и при равномерном движении заряда, если скорость заряда больше фазовой скорости света в данной среде (эффект Вавилова—Черенкова, см. § 2.12).

можно представить себе как объединение движущегося заряда Q с находящимся вблизи него покоящимся зарядом — Q.

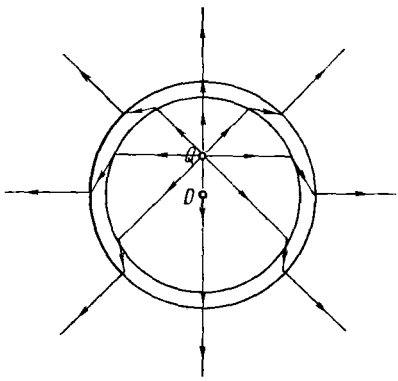
Точное решение на основе уравнений Максвелла задачи об излучении электромагнитных волн такой системой приводится в курсах электродинамики. Однако понять природу механизма испускания волн осциллирующим диполем и получить необходимые формулы можно с помощью предложенной Дж. Томсоном простой модели этого явления.

Рассмотрим точечный заряд Q, который до момента времени $t=0$ покоился в начале координат. Его электрическое поле изображалось радиальными силовыми линиями, выходящими из начала координат. Пусть в момент времени $t=0$ заряд под действием какой-то внешней силы начинает двигаться с постоянным ускорением a в направлении оси z. По истечении короткого промежутка времени τ действие этой силы прекращается, так что дальше заряд движется равномерно с той скоростью $v=a\tau$, которую он приобрел к концу периода ускорения. Будем считать, что ускорение a и время τ таковы, что $v \ll c$.

Представим себе картину силовых линий электрического поля, создаваемого совершающим такое движение зарядом, спустя промежуток времени t после начала движения, большой по сравнению с длительностью периода ускорения τ: $t \gg \tau$. В точках, лежащих за пределами сферы радиусом ct (с центром в начале координат), в момент времени t еще не известно о том, что в момент $t=0$ заряд начал движение. За пределами этой сферы поле такое же, каким оно было при неподвижном заряде (рис. 1.14). Напряженность этого поля равна

$$E(r) = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}. \qquad (1.59)$$

Внутри сферы радиусом $c(t-\tau)$ электрическое поле в момент времени t такое же, как и поле равномерно движущегося заряда, так как начиная с момента τ заряд уже движется с постоянной скоростью v. При $v \ll c$ это поле в момент времени t совпадает с полем неподвижного точечного заряда Q, находящегося в той точке, через которую в момент t проходит движущийся заряд. При $t \gg \tau$ можно считать, что движущийся заряд в момент времени t находится на расстоянии vt от начала координат (это отличается от истинного значения расстояния всего на $v\tau/2$).



1.14
Изломы на силовых линиях, вызванные ускоренным движением заряда Q

Таким образом, все изменения электрического поля, связанные с

ускоренным движением заряда в течение времени τ, в момент времени t заключены внутри тонкого шарового слоя толщиной $c\tau$, наружный радиус которого ct, а внутренний $c(t-\tau)$. Картину электрического поля в этой области внутри шарового слоя можно найти, учитывая непрерывность силовых линий. Для этого достаточно соединить соответствующие радиальные силовые линии (рис. 1.14). Вызванный ускоренным движением заряда излом силовых линий «убегает» от заряда со скоростью c. Изломы на силовых линиях между сферами $r=ct$ и $r=c(t-\tau)$ — это и есть поле излучения заряда, т. е. испускаемая зарядом волна, распространяющаяся со скоростью c от источника.

Рассмотрим одну из силовых линий, составляющую некоторый угол θ с направлением движения заряда (рис. 1.15). Вектор напряженности электрического поля в изломе разложим на две составляющие: радиальную $\mathbf{E}_\parallel$ и поперечную $\mathbf{E}_\perp$. Радиальная составляющая $\mathbf{E}_\parallel$ — это статическое поле, создаваемое неподвижным зарядом Q, находящимся в начале координат. Напряженность поля на расстоянии $r=ct$ задается выражением (1.59). Поперечная составляющая $\mathbf{E}_\perp$ — это напряженность электрического поля в волне, испущенной зарядом во время его движения с ускорением. Так как волна распространяется в радиальном направлении, то вектор $\mathbf{E}_\perp$ перпендикулярен направлению волны. Из рис. 1.15 видно, что

$$E_\perp/E_\parallel = v t \sin\theta/(c\tau).$$

Подставляя сюда $E_\parallel$ из (1.59) и учитывая, что $r=ct$, находим

$$E_\perp = Qv\sin\theta/(4\pi\varepsilon_0 c^3 t\tau). \qquad (1.60)$$

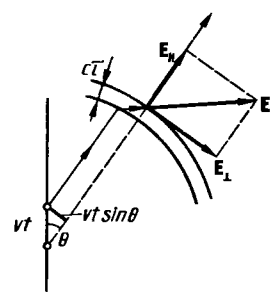
Отношение v/τ представляет собой ускорение a, с которым двигался заряд в течение промежутка времени от 0 до τ. Входящее в (1.60) время t можно выразить через расстояние r от заряда до точки наблюдения: $t=r/c$. Поэтому соотношение (1.60) можно переписать в виде

$$E_\perp = Qa\sin\theta/(4\pi\varepsilon_0 c^2 r). \qquad (1.61)$$

Приведенный здесь вывод этой формулы был основан на предположении, что любое электромагнитное возмущение распространяется в пустоте со скоростью c.

Прежде всего обратим внимание на то, что напряженность $E_\perp$ электрического поля волны убывает обратно пропорционально первой степени расстояния r от центра в отличие от напряженности $E_\parallel$ электростатического поля точечного заряда, которая убывает как $1/r^2$.

Далее отметим, что напряженность $E_\perp$ поля волны в момент времени t в точке



1.15
Вектор $\mathbf{E}$ в области излома силовой линии

наблюдения, находящейся на расстоянии r от источника, зависит от ускорения заряда a в более ранний момент времени $t=0$: волна, излученная в момент $t=0$, достигает точки наблюдения спустя время, равное r/c.

Предположим теперь, что заряд Q все время движется вдоль оси z с некоторым переменным ускорением $a(t)$, оставаясь при этом вблизи начала координат (например, совершает гармонические колебания). Тогда он будет излучать электромагнитные волны непрерывно. Напряженность электрического поля в точке наблюдения, находящейся на расстоянии r от начала координат, по-прежнему определяется формулой (1.61), причем поле $E_\perp$ в момент времени t зависит от ускорения a заряда в более ранний момент времени $t'=t-r/c$:

$$E_\perp(r, t)=Qa(t')\sin\theta/(4\pi\varepsilon_0 c^2 r). \qquad (1.62)$$

Для справедливости этой формулы существенно предположение о том, что совершающий ускоренное движение заряд все время остается вблизи начала координат, так как «время запаздывания» r/c в (1.62) принято неизменным и равным времени прохождения волны от начала координат в точку наблюдения.

Применим полученный результат к излучению заряда Q, совершающего гармоническое колебание с частотой ω:

$$z(t)=z_0\cos\omega t,\quad a(t)=\ddot{z}(t)=-\omega^2 z_0\cos\omega t. \qquad (1.63)$$

При таком движении заряд излучает монохроматическую волну. Поверхности постоянной фазы

$$\omega(t-r/c)=\omega t-kr=\text{const} \qquad (1.64)$$

($k=\omega/c$) представляют собой сферы с центрами в начале координат, расширяющиеся со скоростью $c=\omega/k$. Расстояние, на которое перемещается поверхность постоянной фазы за время, равное периоду колебаний $T=2\pi/\omega$, — это длина волны $\lambda=cT=2\pi/k$.

Формула (1.62) применима для поля изучения осциллирующего заряда, если амплитуда z_0 его осцилляций в (1.63) мала по сравнению с длиной волны: $z_0\ll\lambda$. Только тогда время запаздывания можно считать неизменным и полагать равным r/c. Это же условие можно сформулировать как требование, чтобы скорость заряда была много меньше скорости света ($v\ll c$). В этом случае его называют *электрическим дипольным осциллятором*, а испускаемые им волны — дипольным излучением.

Можно представить себе, что наряду с осциллирующим зарядом Q в начале координат находится покоящийся заряд $-Q$, образующий вместе с зарядом Q нейтральную систему — электрический диполь с осциллирующим дипольным моментом $p(t)=Qz(t)$. Добавление неподвижного заряда $-Q$ приведет к изменению радиального поля $E_\parallel$ (оно будет убывать с расстоянием быстрее, чем в рассмотренной выше картине поля, создаваемого одним зарядом Q).

На больших расстояниях от источника, где сравнительно медленно (как $1/r$) убывающее поперечное поле излучения $E_\perp$ значительно превосходит радиальное поле $E_\parallel$ (в волновой зоне, т. е. при $r \gg \lambda$), формула (1.62) в равной мере применима и к полю излучения диполя с дипольным моментом $p(t)=Qz(t)$:

$$E_\perp(r,\ t)=\frac{1}{4\pi\varepsilon_0}\frac{\ddot{p}(t')}{c^2 r}\sin\Theta \qquad (1.65)$$

При этом безразлично, чем обусловлены осцилляции дипольного момента: изменением расстояния между зарядами Q и $-Q$ по закону $z(t)=z_0\cos\omega t$ при неизменной их величине или изменением зарядов по закону $Q(t)=Q_0\cos\omega t$ при неизменном расстоянии z_0 между ними. Последний случай соответствует простым антеннам, применяемым в радиотехнике. Поле излучения в волновой зоне можно находить, заменяя антенну эквивалентным дипольным осциллятором.

Первый случай — колебания расстояния между зарядами — важен как классическая модель электромагнитного излучателя света в оптике. Например, в модели атома Томсона оптический электрон связан с атомом квазиупругой силой и может совершать гармонические колебания. Состоящий из таких атомов источник света можно заменить совокупностью элементарных дипольных осцилляторов. Оказывается, что электрический дипольный осциллятор как модель излучающей атомной системы в ряде случаев приводит к правильным результатам, подтверждающимся на опыте.

На больших расстояниях от источника, т. е. в волновой зоне, где $r \gg \lambda$, можно отдельные небольшие участки сферической волновой поверхности (1.64) рассматривать как плоскости. Если размеры этих участков велики по сравнению с длиной волны, то к ним применимы результаты, полученные выше при изучении плоских электромагнитных волн. В частности, мы сразу можем сказать, каким будет магнитное поле **B**. Если для каждой точки ввести волновой вектор **k**, направленный радиально из начала координат, то векторы **E**, **B** и **k** в этой точке образуют правую тройку векторов в соответствии с уравнением (1.29). Их взаимное расположение показано на рис. 1.16. Индукция $B(r, t)$ магнитного поля сферической электромагнитной волны в каждой точке связана с напряженностью $E(r, t)$ электрического поля (1.65) в этой же точке тем же соотношением (1.31), что и для плоской волны.

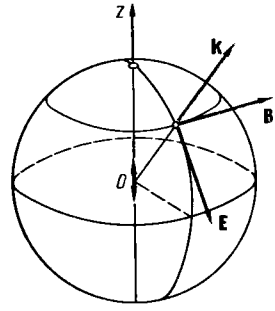


Рис. 1.16
Векторы **E** и **B** в волновой зоне

Как видно из рис. 1.16, вектор **E** направлен по касательной к меридиану, а вектор **B** — по касательной к параллели. Поэтому в сферической системе, орты которой направлены в сторону возрастания соот-

ветствующих координат (r, θ, φ), векторы $\mathbf{E}$ и $\mathbf{B}$ имеют следующие проекции:

$$E_r=0,\ E_\theta=\ddot{p}\sin\theta/(4\pi\varepsilon_0 c^2 r),\ E_\varphi=0,$$
$$B_r=0,\ B_\theta=0,\ B_\varphi=\ddot{p}\sin\theta/(4\pi\varepsilon_0 c^3 r). \qquad (1.66)$$

Эти формулы можно записать в векторном виде. Если ввести вектор дипольного момента $\mathbf{p}(t)$, единичный вектор $\mathbf{r}_1=\mathbf{r}/r$, то, как видно из рис. 1.16 и выражений (1.66),

$$\mathbf{B}=\frac{1}{4\pi\varepsilon_0}\frac{1}{c^3 r}\ddot{\mathbf{p}}(t')\times\mathbf{r}_1,$$
$$\mathbf{E}=c\mathbf{B}\times\mathbf{r}_1. \qquad (1.67)$$

Отметим, что поле излучения дипольного осциллятора, хотя и представляет собой сферическую волну, сферической симметрией не обладает. В волновой зоне поверхности постоянной фазы действительно сферические, но модули векторов $\mathbf{E}$ и $\mathbf{B}$ в разных точках такой сферы различны, ибо они, как видно из (1.67), зависят от полярного угла θ. Поле поперечной сферической волны не может быть сферически симметричным.

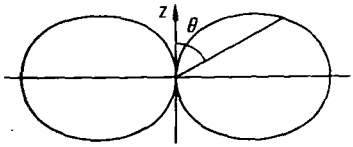
Поток энергии в волновой зоне осциллятора имеет в каждой точке радиальное направление. В самом деле, из рис. 1.16 ясно, что направление вектора Пойнтинга $\mathbf{S}$ (1.50) в каждой точке волновой зоны совпадает с направлением радиуса-вектора $\mathbf{r}$. Найдем значение плотности потока энергии в точке с координатами $\mathbf{r}$, θ для диполя с гармонически изменяющимся дипольным моментом $p(t)=p_0\cos\omega t$:

$$S(r,\theta,t)=\varepsilon_0 cE^2(r,\theta,t)=\frac{1}{(4\pi)^2\varepsilon_0}\frac{\omega^4 p_0^2}{c^3 r^2}\sin^2\theta\cos^2(\omega t-kr). \qquad (1.68)$$

Усредняя выражение (1.68) по времени, получаем интенсивность излучения диполя на расстоянии r:

$$\langle S(r,\theta)\rangle=\omega^4 p_0^2\sin^2\theta/(32\pi^2\varepsilon_0 c^3 r^2). \qquad (1.69)$$

Зависимость интенсивности от направления выражается в (1.69) множителем $\sin^2\theta$. Максимальная интенсивность наблюдается при $\theta=\pi/2$, т. е. в экваториальной плоскости: максимум интенсивности соответствует направлению, перпендикулярному оси диполя. Вдоль оси диполя ($\theta=0$) энергия не излучается. Угловое распределение излучаемой осциллирующим диполем энергии показано на рис. 1.17 с помощью «диаграммы направленности». Длина отрезка, проведенного из начала координат до пересечения с линией $r=\sin^2\theta$, пропорциональна интенсивности распространяющейся в данном направлении волны. Распределение интенсивности по направлениям в пространстве характе-



1.17
Угловое распределение энергии

ризуется поверхностью, которая получается вращением кривой на рис. 1.17 вокруг оси z.

Полную энергию, излучаемую диполем за 1 с по всем направлениям (поток излучения), можно найти, вычисляя поток $\langle S\rangle$ через поверхность сферы радиусом R с центром в осцилляторе. Разобьем сферу на кольца координатными поверхностями $\theta=\text{const}$ и $\theta+d\theta=\text{const}$. Площадь такого кольца равна $2\pi R^2\sin\theta d\theta$, а значение $\langle S\rangle$ во всех его точках одинаково. Поэтому полная излучаемая мощность

$$P=\frac{\omega^4 p_0^2}{16\pi\varepsilon_0 c^3}\int_0^{\pi}\sin^3\theta d\theta=\frac{\omega^4 p_0^2}{12\pi\varepsilon_0 c^3}. \qquad (1.70)$$

Полученный результат заслуживает подробного обсуждения. Излучаемая осциллятором мощность пропорциональна квадрату амплитуды его дипольного момента и четвертой степени частоты, т. е. обратно пропорциональна четвертой степени длины волны. Этот закон играет большую роль в теории рассеяния света. Столь сильной зависимостью интенсивности излучения от длины волны объясняется, например, голубой цвет неба (короткие волны рассеиваются сильнее, чем длинные) и красный цвет Солнца на закате, когда при прохождении через большую толщу атмосферы голубые лучи рассеиваются из прямого пучка гораздо сильнее, чем красные.

Выражаемый формулой (1.70) поток излучения осциллятора через поверхность сферы не зависит от ее радиуса R — через любую охватывающую осциллятор замкнутую поверхность протекает за 1 с одинаковая энергия. В окружающем осциллятор пространстве нет ни проводников, ни электрических зарядов, ввиду чего излучаемая им электромагнитная энергия не может переходить в другие формы энергии и должна без потерь переноситься с волной в отдаленные области пространства, нигде не накапливаясь и не исчезая. Это объясняет характер зависимости напряженности $E(r)$ электрического поля в формуле (1.65) от расстояния до источника. Чтобы общее излучение через сферическую поверхность не зависело от ее радиуса, плотность потока $S(r)$ должна убывать обратно пропорционально r^2, так как поверхность сферы пропорциональна r^2. С другой стороны, плотность потока S пропорциональна E^2. Следовательно, напряженность $E(r)$ должна убывать обратно пропорционально расстоянию r.

Осциллятор совершает незатухающие колебания лишь в том случае, когда эти колебания поддерживаются каким-либо внешним источником. Без такого источника колебания будут затухать даже при движении в абсолютно пустом пространстве, так как осциллятор теряет энергию на излучение (*радиационное затухание*). Хотя никаких сил сопротивления, никакой «вязкости» в обычном смысле этого слова здесь нет, затухание колебаний можно описать, вводя в уравнение движения излучающего заряда эффективную силу трения таким образом, чтобы потеря энергии на излучение могла быть представлена как средняя работа этой силы.

Используя полученное выше выражение (1.70) для излучаемой осциллятором мощности, легко сделать оценку *времени жизни* атома в возбужденном состоянии.

Воспользуемся простой классической моделью атома, согласно которой оптический электрон — частица с зарядом $Q=-e$ и массой m — связан в атоме квазиупругой силой, так что при возбуждении атома он совершает собственные колебания с определенной частотой ω : $z(t)=z_0\cos\omega t$. Незначительным изменением частоты вследствие затухания здесь можно пренебречь. Энергия такого осциллятора состоит из кинетической энергии $W_к=m\dot{z}^2(t)/2$ и потенциальной $W_п=m\omega^2 z^2(t)/2$, средние значения которых равны между собой. Полная энергия осциллятора

$$W=W_к+W_п={}^1/_2 m\omega^2 z_0^2 \qquad (1.71)$$

пропорциональна квадрату амплитуды. Излучаемая осциллятором мощность $P_{изл}$, представляющая собой скорость уменьшения его энергии — $\mathrm{d}W/\mathrm{d}t$, в соответствии с (1.70) также пропорциональна квадрату амплитуды. Выражая z_0^2 через энергию W из (1.71) и подставляя в правую часть (1.70), получаем, что скорость уменьшения энергии осциллятора пропорциональна его энергии:

$$-\mathrm{d}W/\mathrm{d}t=2\gamma W, \qquad (1.72)$$

где $\gamma=e^2\omega^2/(4\pi\varepsilon_0\cdot 3mc^3)$. (1.73)

Из уравнения (1.72) следует, что энергия возбуждения осциллятора уменьшается вследствие потерь на излучение по экспоненциальному закону:

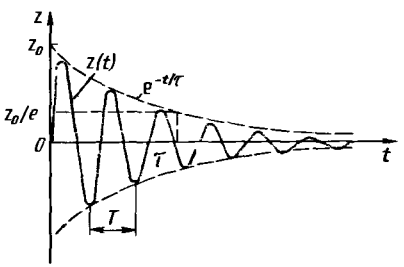
$$W(t)=W(0)\mathrm{e}^{-2\gamma t}=W(0)\mathrm{e}^{-t/\tau_э}.$$

Здесь $\tau_э=1/(2\gamma)$ — *время радиационного затухания*, в течение которого энергия осциллятора уменьшается в e раз. Амплитуда z_0 колебаний осциллятора также убывает экспоненциально (рис. 1.18):

$$z_0(t)=z_0(0)\mathrm{e}^{-\gamma t}=z_0(0)\mathrm{e}^{-t/\tau}. \qquad (1.74)$$

Длительность этого процесса характеризуется временем затухания амплитуды $\tau=1/\gamma$ (временем жизни колебаний), которое в два раза превышает время затухания энергии: $\tau=2\tau_э$. Время τ определяет продолжительность цуга волн, испускаемых возбужденным осциллятором. Используя (1.73), находим число полных колебаний, совершаемых осциллятором за время τ:

$$\tau/T=\omega/(2\pi\gamma)=6\varepsilon_0 mc^3/(\mathrm{e}^2\omega). \qquad (1.75)$$



1.18
Затухание колебаний излучающего электрона

Принимая для $\omega=2\pi c/\lambda$ значение, соответствующее $\lambda=0{,}5$ мкм (видимый свет), из (1.75) нахо-

дим, что за время затухания осциллятор совершает около 10 млн. колебаний. Период этих колебаний составляет примерно 10^{-15} с, поэтому время жизни возбужденного состояния атома, обусловленное радиационным распадом, по порядку величины равно 10^{-8} с. Хотя эта оценка получена для простой классической модели излучающего атома, результат по порядку величины согласуется с наблюдаемыми на опыте значениями.

Таким образом, излучаемая отдельным атомом волна не является строго монохроматической. Она представляет собой постепенно затухающий цуг, в котором содержится около 10^7 волн. Пространственная протяженность такого цуга составляет несколько метров. Разумеется, приведенные оценки справедливы для невзаимодействующих атомов в пустом пространстве (разреженный газ), так как среднее время жизни атома в возбужденном состоянии может определяться и другими причинами, например столкновениями атомов друг с другом или со стенками сосуда (разрядной трубки).

Непосредственное экспериментальное определение времени радиационного затухания было впервые осуществлено В. Вином для атомов водорода с помощью каналовых лучей. Через узкие отверстия (каналы) в катоде разрядной трубки возбужденные атомы водорода вылетают в закатодное пространство, где поддерживается очень высокий вакуум. Высвечивающиеся атомы в пучке движутся практически без соударений. По убыванию интенсивности вдоль светящегося пучка при известной скорости* атомов можно было определить время затухания $\tau_э$, которое по порядку величины совпало с приведенной выше оценкой (10^{-8} с).

В современной экспериментальной физике используют другие, более совершенные методы измерения времени радиационного затухания. Некоторые из них рассмотрены ниже (см. § 5.4).

Контрольные вопросы

В выводе выражения для электрического поля ускоренно движущегося заряда были неявно использованы уравнения Максвелла и некоторые следствия из ннх. Еще раз внимательно проследите за всеми этапамн вывода и укажите, какие уравнения и в каких местах были использованы.

- Какую поляризацию имеют волны, излучаемые осциллирующим диполем?
- Что называют волновой зоной поля нзлучення осциллирующего диполя? Почему полученные выражения для поля излучения справедливы только в волновой зоне?
- Объясните завнсимость напряженности поля излучения в сферической волне от расстояния на основе закоиа сохранення энергин.
- Как, используя классическую модель атома, оценнть время его жизни в возбужденном состоянии?
- По какому закону измеияется в результате излучения энергия возбужденного атома? амплитуда колебаний оптнческого электрона?

* Скорость атомов в пучке определялась по доплеровскому смещению спектральных лиинй прн наблюдеиии вдоль направления каналовых лучей.

Задача

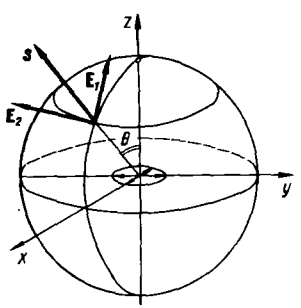
Точечный заряд Q движется по окружности радиусом a с угловой скоростью ω, причем $\omega a \ll c$. Найти пространственное распределение излучаемой таким «ротатором» энергии. Какой характер имеет поляризация излучаемых волн?

◇ Движение заряда Q можно разложить на два гармонических колебания одинаковой частоты ω и амплитуды a, происходящих во взаимно перпендикулярных направлениях со сдвигом по фазе на $\pi/2$. Поэтому поле излучения ротатора представляет собой суперпозицию полей двух осциллирующих диполей.

Начало координат выберем в центре ротатора, ось z направим перпендикулярно его плоскости. Ось z является осью симметрии рассматриваемой системы. Направление оси x можно выбрать так, чтобы точка наблюдения, направление на которую образует угол θ с осью z, лежала в плоскости xz (рис. 1.19). Пусть заряд Q движется против часовой стрелки. Тогда

$$x(t) = a\cos\omega t, \quad y(t) = a\sin\omega t.$$

Направления колебаний векторов напряженности $\mathbf{E}_1$ и $\mathbf{E}_2$ электрического поля в волновой зоне для волн, испускаемых каждым диполем, взаимно перпендикулярны (рис. 1.19), а по фазе колебания сдвинуты на $\pi/2$. Амплитуда E_2 колебаний поля (от диполя, осциллирующего вдоль оси y) не зависит от угла θ, а амплитуда E_1 в соответствии с (1.66) пропорциональна $\cos\theta$. Поэтому в произвольной точке «северного полушария» на рис. 1.19 результирующая волна имеет левую эллиптическую поляризацию.



К излучению заряда, движущегося по окружности

Оси эллипса колебаний в каждой точке ориентированы вдоль «меридиана» и «параллели». Эллиптическая поляризация для «северного полюса», т. е. положительного направления оси z, превращается в круговую, ибо в этом направлении обе волны имеют одинаковые амплитуды. Для точек наблюдения в «южном полушарии» поляризация будет правой.

В экваториальной плоскости (в направлении оси x) амплитуда одной из волн обращается в нуль, поэтому излучение имеет линейную поляризацию; при наблюдении вдоль оси x движение заряда по окружности неотличимо от осцилляций вдоль оси y.

Рассмотренный пример показывает, что понятие «поляризация» относится к поведению волны в данной точке, т. е. состояние поляризации, вообще говоря, различно в разных точках поля. Волна может иметь линейную поляризацию в одних точках и круговую или эллиптическую — в других. Только в некоторых случаях, например для однородной плоской волны, состояние поляризации всюду одинаково.

Поток энергии результирующей волны аддитивно складывается из потоков энергии каждой из волн, так как эти волны для любого направления распространения имеют ортогональные направления поляризации (см. задачу 4 к § 1.4). Поэтому интенсивность излучения ротатора в направлении, составляющем угол θ с осью z, пропорциональна $1+\cos^2\theta$:

$$\langle S(\theta)\rangle \sim 1 + \cos^2\theta.$$

1.6. Спектральное разложение излучения

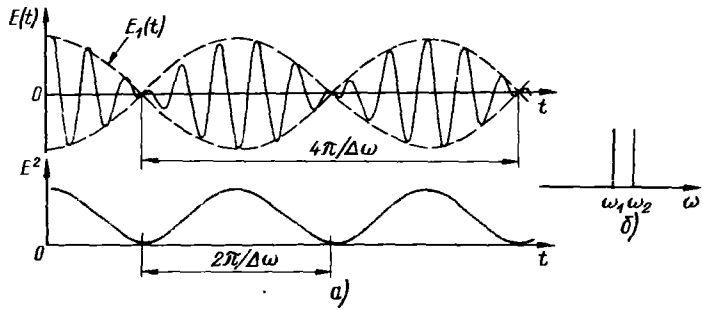
Рассмотренные выше монохроматические электромагнитные волны представляют собой идеализацию. Свет от реального физического источника никогда не бывает строго монохроматическим. Любой реальный физический процесс ограничен во времени в отличие от гармонического колебания, продолжающегося бесконечно долго. Таким образом, ограниченность во времени и монохроматичность исключают друг друга.

Реальную электромагнитную волну можно представить в виде наложения монохроматических волн с различными частотами. Такая возможность обусловлена *принципом суперпозиции:* при наличии нескольких волн поля этих волн просто накладываются друг на друга; общая напряженность поля двух источников равна сумме напряженностей полей, создаваемых каждым источником в отдельности. Этот известный из опыта факт находит свое математическое выражение в том, что описывающие электромагнитное поле уравнения Максвелла *линейны.* Линейные уравнения обладают тем свойством, что любая суперпозиция решений (сумма, линейная комбинация) также является решением.

Таким образом, принцип суперпозиции позволяет исследовать сложную электромагнитную волну, заменяя ее суммой (в общем случае бесконечной) монохроматических составляющих. С математической точки зрения такая замена означает разложение функции в ряд или интеграл Фурье.

Целесообразность разложения сложной волны на сумму именно монохроматических составляющих обусловлена не только тем, что монохроматические волны — это наиболее простые волны и их свойства хорошо известны. Сложную функцию можно представить как сумму других функций самыми разнообразными способами. Возможно разложение не только по синусам и косинусам, но и по другим функциям. С математической точки зрения все такие разложения одинаково допустимы. Целесообразность разложения именно на монохроматические составляющие связана с *физикой,* с возможностью выделения в эксперименте отдельных монохроматических составляющих. В экспериментальной оптике *спектральный прибор,* например спектрограф с призмой или дифракционной решеткой, производит *фактическое разложение* сложного излучения на монохроматические составляющие и позволяет вести экспериментальный контроль монохроматичности, ибо синусоидальная волна дает в таком приборе резкую отдельную линию. Поэтому синусоидальные функции оказываются для таких устройств физически выделенными по сравнению с различными полными системами других функций.

Рассмотрим сначала самый простой в математическом отношении пример немонохроматической волны, получающейся при сложении двух монохроматических волн одинаковой линейной поляризации и одинаковой амплитуды E_0, распространяющихся в одном направлении. Пусть эти волны имеют различные частоты ω_1 и ω_2.



1.20

Напряженность поля и интенсивность при сложении двух монохроматических волн (*а*) и частотный спектр этого колебания (*б*)

Напряженность результирующего поля этих волн в некоторой точке представляет собой сумму гармонических колебаний с частотами ω_1 и ω_2. С помощью тригонометрического тождества ее можно преобразовать к следующему виду:

$$E(t)=E_0\cos\omega_1 t+E_0\cos\omega_2 t=2E_0\cos\left(\frac{\omega_2-\omega_1}{2}t\right)\cos\left(\frac{\omega_2+\omega_1}{2}t\right) \qquad (1.76)$$

Будем предполагать, что разность частот $\Delta\omega=\omega_2-\omega_1$ мала по сравнению со средней частотой $\omega=(\omega_2+\omega_1)/2$: $\Delta\omega\ll\omega$. Тогда правую часть в (1.76) можно рассматривать как «быстрые» колебания со средней частотой ω, амплитуда $E_1(t)$ которых «медленно» изменяется со временем:

$$E(t)=E_1(t)\cos\omega t, \qquad E_1(t)=2E_0\cos(\Delta\omega/2)t. \qquad (1.77)$$

График $E(t)$ показан на рис. 1.20, *а*.

В течение нескольких периодов «быстрых» колебаний $\cos\omega t$ амплитуда $E_1(t)$ изменяется незначительно. В таких случаях говорят, что $E(t)$ представляет собой «почти гармоническое» колебание с медленно изменяющейся амплитудой — *амплитудно-модулированное колебание.*

Интенсивность монохроматической волны пропорциональна квадрату амплитуды напряженности. Если применить это понятие к «почти монохроматической» модулированной волне (1.77), то мы получим, что ее интенсивность периодически изменяется с частотой $\Delta\omega$:

$$E_1^2(t)=2E_0^2(1+\cos\Delta\omega t). \qquad (1.78)$$

Малоинерционный приемник излучения, реагирующий на интенсивность («квадратичный детектор»), должен зарегистрировать эти периодические (с периодом $2\pi/\Delta\omega$, рис. 1.20) изменения интенсивности. Но инерционный приемник с постоянной времени, большой по сравнению с периодом модуляции, зарегистрирует среднее зна-

чение интенсивности, которое, как видно из (1.78), равно сумме интенсивностей отдельных монохроматических составляющих*.

Иногда ставят вопрос о «реальности» разложения колебания на синусоидальные составляющие. Что реально в соотношении (1.76) — левая часть, т. е. совокупность чисто гармонических колебаний, или правая, часть, т. е. «почти гармоническое» колебание с медленно меняющейся амплитудой? В такой форме вопрос лишен смысла, так как (1.76) — математическое тождество, т. е. слева и справа стоит одно и то же. Никаким способом нельзя установить, создается поле $E(t)$ двумя различными монохроматическими источниками с частотами ω_1 и ω_2 или одним источником, испускающим волну с частотой ω, амплитуда которой периодически изменяется (рис. 1.20, *а*).

Вопрос приобретает смысл, когда его ставят в связи с приемником, воспринимающим колебания. Для малоинерционного приемника целесообразно использовать представление $E(t)$ в виде колебания одной частоты $\omega=(\omega_1+\omega_2)/2$ с медленно меняющейся амплитудой. Для спектрального аппарата или резонатора, настроенного на определенную частоту, целесообразно представление $E(t)$ в виде совокупности монохроматических колебаний с частотами ω_1 и ω_2.

В качестве другого примера рассмотрим волну, порождаемую колебанием $E_2(t)\cos\omega t$, амплитуда которого промодулирована по синусоидальному закону $E_2(t)=E_0(1+m\cos\Omega t)$ с «глубиной» $m(m<1)$ и частотой $\Omega\ll\omega$. Легко видеть, что такая волна может быть представлена как совокупность трех строго монохроматических волн с частотами ω, $\omega+\Omega$ и $\omega-\Omega$:

$$E(t)=E_0(1+m\cos\Omega t)\cos\omega t=E_0\cos\omega t+{}^1/_2mE_0\cos(\omega+\Omega)t+{}+{}^1/_2mE_0\cos(\omega-\Omega)t. \qquad (1.79)$$

О совокупности монохроматических составляющих говорят как о *спектре* немонохроматической волны. Интенсивность модулированной волны, усредненная за промежуток времени, равный периоду модуляции, пропорциональна $E_0^2(1+m^2/2)$ и равна сумме интенсивностей монохроматических составляющих с частотами ω, $\omega+\Omega$ и $\omega-\Omega$.

Этот пример ясно показывает, что изменение амплитуды со временем приводит к нарушению монохроматичности волны и появлению новых частот в ее спектре.

В случае иного, более сложного, чем в разобранных примерах, закона изменения амплитуды напряженности во времени физическая сущность разложения волны в спектр, т. е. на монохроматические

* Заметим, что при точном равенстве частот $\omega_1=\omega_2$ интенсивность результирующей волны не равна сумме интенсивностей составляющих. Эта интенсивность не изменяется со временем и в зависимости от соотношения начальных фаз может иметь любое значение от нуля до учетверенной интенсивности отдельной волны.

составляющие, остается прежней, хотя задача отыскания этих составляющих становится более сложной и требует применения теоремы Фурье.

Простейший случай соответствует периодическому изменению напряженности поля с некоторым периодом T. При этом возможно разложение в обычный *ряд Фурье*. Оно содержит дискретные частоты, являющиеся целыми кратными основной частоты $\omega=2\pi/T$:

$$E(t)=\sum_{n=-\infty}^{\infty} E_n e^{-i\omega nt}. \tag{1.80}$$

Комплексные коэффициенты Фурье E_n этого разложения определяются по известному виду самой функции $E(t)$ интегралами

$$E_n=\frac{1}{T}\int_{-T/2}^{T/2} E(t)e^{in\omega t}dt. \tag{1.81}$$

Ввиду вещественности $E(t)$ очевидно, что $E_{-n}=E_n^*$. При возведении суммы (1.80) в квадрат и усреднении по времени квадраты всех слагаемых и произведения всех слагаемых с различными частотами обращаются в нуль из-за наличия в них осциллирующих синусоидальных множителей, за исключением произведений вида $E_nE_{-n}=|E_n|^2$. Отсюда следует, что средняя интенсивность волны равна сумме интенсивностей монохроматических компонент:

$$\langle E^2(t)\rangle=\sum_{n=-\infty}^{\infty}|E_n|^2=2\sum_{n=1}^{\infty}|E_n|^2. \tag{1.82}$$

При переходе к суммированию только по положительным n мы учли, что для света $E(t)$ не имеет постоянной составляющей и поэтому в (1.80) $E_n=0$ при $n=0$.

Для непериодических полей разложение в спектр дает непрерывный набор различных частот. Здесь нужно использовать представление в виде *интеграла Фурье:*

$$E(t)=\int_{-\infty}^{\infty} E_\omega e^{-i\omega t}\frac{d\omega}{2\pi}. \tag{1.83}$$

Чтобы такое разложение было возможно, функция $E(t)$ должна удовлетворять определенным требованиям, которые в физических задачах обычно выполняются. Входящая в подынтегральное выражение непрерывная функция частоты E_ω определяется по известной функции $E(t)$ соотношением

$$E_\omega=\int_{-\infty}^{\infty} E(t)e^{i\omega t}dt. \tag{1.84}$$

При этом $E_{-\omega}=E_\omega^*$, так как функция $E(t)$ вещественна. Поэтому

фактически достаточно знать комплексную функцию E_ω только при положительных значениях ω. Как видно из (1.83), $E_\omega d\omega/(2\pi)$ определяет вклад в $E(t)$ монохроматических составляющих из интервала частот ω, $\omega + d\omega$.

Чтобы выразить полную энергию волны через интенсивность ее компонент Фурье, вычислим интеграл от $E^2(t)$ по времени:

$$\int_{-\infty}^{\infty} E^2(t)dt = \int_{-\infty}^{\infty} E(t)\left[\int_{-\infty}^{\infty} E_\omega e^{-i\omega t}\frac{d\omega}{2\pi}\right] dt.$$

Здесь мы воспользовались формулой (1.83). Изменяя теперь порядок интегрирования по ω и t и используя формулу (1.84), получаем

$$\int_{-\infty}^{\infty} E^2(t)dt = \int_{-\infty}^{\infty} E_\omega\left[\int_{-\infty}^{\infty} E(t)e^{-i\omega t}dt\right]\frac{d\omega}{2\pi} =$$

$$= \int_{-\infty}^{\infty} E_\omega E_{-\omega}\frac{d\omega}{2\pi} = \int_{-\infty}^{\infty} |E_\omega|^2\frac{d\omega}{2\pi} = 2\int_0^{\infty} |E_\omega|^2\frac{d\omega}{2\pi}. \qquad (1.85)$$

Таким образом, полная энергия немонохроматической волны выражается через интеграл по положительным частотам от ее *спектральной плотности,* характеризующей распределение энергии волны по спектру частот. Отметим, что термином «спектр» в физике пользуются несколько вольно, вкладывая в него порой разный смысл. Иногда его относят просто к набору частот (дискретному или непрерывному), входящих в состав немонохроматического излучения, иногда — к распределению энергии (интенсивности) излучения по этим частотам, характеризуемому спектральной плотностью $2|E_\omega|^2$, а иногда — к фурье-образу E_ω математической функции $E(t)$, описывающей немонохроматическое излучение. В то время как E_ω в соответствии с формулой (1.83) полностью определяет функцию $E(t)$, знание спектральной плотности энергии $2|E_\omega|^2$ еще не позволяет восстановить функцию $E(t)$. Дело в том, что в энергетическом спектре $2|E_\omega|^2$ уже не содержится информация о *фазах* монохроматических составляющих. Поэтому данное поле $E(t)$ характеризуется вполне определенным спектром, но одному и тому же спектру могут соответствовать *разные* функции $E(t)$.

В заключение заметим еще раз, что *математическому* разложению немонохроматической волны в ряд или интеграл Фурье для нахождения спектральной плотности ее энергии можно сопоставить реальный *физический процесс* — экспериментальное измерение спектра такой волны с помощью соответствующего анализатора (спектрального прибора). Принцип действия некоторых спектральных приборов рассматривается в § 6.6.

Контрольные вопросы

В чем заключается физическое содержание принципа суперпозиции? Какое свойство уравнений Максвелла обеспечивает выполнение принципа суперпозиции?

В чем преимущество разложения произвольной электромагнитной волны на синусоидальные волны по сравнению с разложениями по полным системам других функций?

В каких случаях выражения (1.76) или (1.79) удобно рассматривать как почти гармоническое колебание с медленно изменяющейся амплитудой («модулированное» колебание) и в каких — как сумму нескольких монохроматических колебаний?

Каким спектром [коэффициентами E_n при разложении в ряд Фурье (1.80)], характеризуется колебание вида $E(t)=E_0\cos\omega t$? $E(t)=E_0\sin\omega t$? $E(t)=E_0\cos(\omega t+\varphi)$?

Каким спектром характеризуется колебание вида $E(t)=A\cos\Omega t\cos\omega t$? $E(t)=E_0(1+0{,}5\cos\Omega t)\cos(10\Omega t)$?

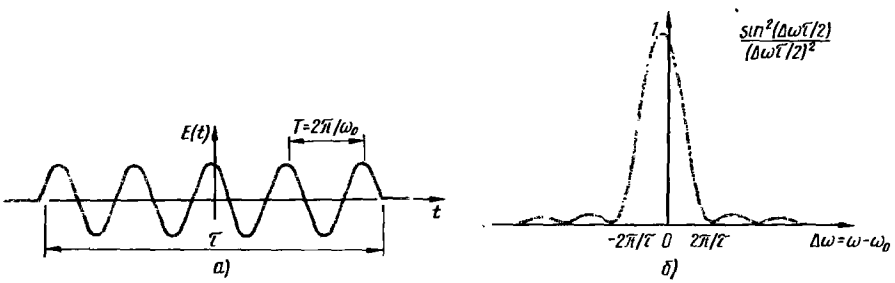
Можно ли по измеренному идеальным прибором энергетическому спектру $|E_\omega|^2$ восстановить зависимость от времени напряженности $E(t)$ электрического поля волны, воздействовавшей на спектральный прибор?

1.7 Квазимонохроматический свет

Исследуем спектральный состав немонохроматической электромагнитной волны, представляющей собой отрезок синусоиды:

$$E(t)=\begin{cases} E_0\cos\omega_0 t & (-\tau/2<t<\tau/2), \\ 0 & (|t|>\tau/2). \end{cases} \qquad (1.86)$$

Описываемое этой формулой колебание длится не бесконечно долго, а конечное время τ (рис. 1.21, *а*). Поэтому его спектр образован не единственной частотой ω_0, а непрерывным набором частот. Мы увидим, что при большой по сравнению с отдельным «периодом» $T=2\pi/\omega_0$ длительности колебаний $\tau\gg T$ спектр в основном сосредоточен в небольшом интервале частот $\Delta\omega$ вблизи

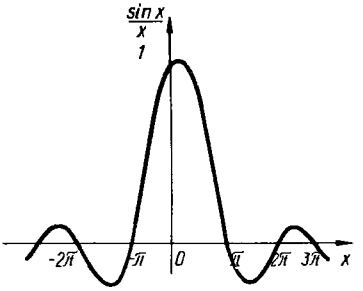


Синусоидальный цуг (*а*) и соответствующее ему распределение энергии по частотам (*б*)

значения ω_0, причем монохроматическая составляющая с частотой ω_0 имеет наибольшую амплитуду.

Пространственная картина волны, порождаемой колебанием (1.86), представляет собой цуг синусоидальных волн длины $\lambda = cT$, имеющий конечную протяженность $l = c\tau$.

Для нахождения спектра подставим $E(t)$ из (1.86) в формулу (1.84). Вычисление интеграла удобно произвести, выразив $\cos\omega_0 t$ через показательные функции по формуле Эйлера:



1.22
График фуикции $\sin x/x$

$$E_\omega = \frac{1}{2} E_0 \int_{-\tau/2}^{\tau/2} (e^{-i\omega_0 t} + e^{i\omega_0 t}) e^{i\omega t} dt =$$

$$= \frac{1}{2} E_0 \tau \left[\frac{\sin(\omega - \omega_0)\tau/2}{(\omega - \omega_0)\tau/2} + \frac{\sin(\omega + \omega_0)\tau/2}{(\omega + \omega_0)\tau/2} \right]. \quad (1.87)$$

На рис. 1.22 показан график функции $\sin x/x$. Ее главный максимум расположен при $x=0$, где она имеет значение, равное 1. Функция обращается в нуль при $x = \pm\pi, \pm 2\pi, \ldots$. В промежутках она имеет второстепенные максимумы и минимумы: значение $(\sin x/x)^2$ на интервале $(\pi, 2\pi)$ не превосходит 0,05; на интервале $(2\pi, 3\pi)$ не превосходит 0,02 и т. д. Сопоставляя график $\sin x/x$ с формулой (1.87), заключаем, что конечный отрезок синусоиды имеет фурье-компоненты, в основном сосредоточенные вблизи значений ω_0 и $-\omega_0$ в частотных интервалах порядка ширины главного максимума $2\pi/\tau$.

Для характеристики спектрального распределения энергии (или интенсивности), как это видно из формулы (1.85), достаточно рассмотреть ход функции $|E_\omega|^2$ при положительных частотах $\omega > 0$. Пусть рассматриваемый цуг содержит много «периодов», т. е. $\tau \gg T = 2\pi/\omega_0$. Это значит, что расстояние ω_0 от начала координат до главных максимумов функции E_ω (1.87) велико по сравнению с шириной этих максимумов $2\pi/\tau$. Поэтому в области положительных частот $\omega > 0$ функция E_ω (1.87) практически определяется только своим первым слагаемым. Таким образом, длинный цуг синусоидальных волн характеризуется следующим распределением энергии по спектру:

$$2|E_\omega|^2 = \frac{1}{2} E_0^2 \tau^2 \frac{\sin^2[(\omega - \omega_0)\tau/2]}{[(\omega - \omega_0)\tau/2]^2}. \quad (1.88)$$

График этой функции приведен на рис. 1.21, *б*. Он дает представление о *контуре спектральной линии* рассматриваемого излучения. Максимум спектральной плотности соответствует значению $\omega = \omega_0$. Большая часть энергии цуга приходится на монохроматические составляющие, лежащие в пределах этого главного максимума, т. е. между частотами, отстоящими от ω_0 на $2\pi/\tau$.

Ширину спектральной линии можно определить как интервал частот $\Delta\omega$ между значениями, при которых спектральная плотность равна половине максимальной плотности («ширина на половине высоты»). Так как $(\sin x/x)^2 = {}^1/_2$ при $|x| = 1{,}39$, то приближенно можно положить

$$\Delta\omega\tau \approx 2\pi, \quad \Delta\nu\tau \approx 1, \tag{1.89}$$

где $\nu = \omega/(2\pi)$. Чем больше длительность цуга синусоидальных волн τ, тем уже соответствующий ему спектральный интервал $\Delta\omega$ (или $\Delta\nu$). Ширина спектральной линии $\Delta\nu$ приблизительно равна обратной длительности колебаний $1/\tau$. Это важное соотношение между шириной спектра и длительностью колебаний имеет общий характер: мы увидим, что оно сохраняется и в том случае, когда огибающая цуга имеет более сложную форму. Его можно записать также в виде соотношения между протяженностью волнового цуга в пространстве $l = c\tau$ и интервалом волновых чисел $\Delta k = \Delta\omega/c$ монохроматических компонент, входящих в состав цуга: $\Delta k l \approx 2\pi$.

Когда ширина спектра $\Delta\omega$ мала по сравнению со средней частотой $\omega_0 (\Delta\omega \ll \omega_0)$, излучение называют *квазимонохроматическим.* Излучение в виде достаточно длинного цуга синусоидальных волн или в виде хаотической последовательности таких цугов дает пример квазимонохроматического излучения. Отдельные спектральные линии в излучении разреженных газов представляют собой квазимонохроматический свет. Такой свет можно также выделить из излучения источников, дающих непрерывный спектр (Солнце, раскаленные тела), с помощью монохроматоров — приборов, осуществляющих спектральное разложение. Наибольшей степенью монохроматичности (характеризуемой отношением $\omega_0/\Delta\omega$ или $\lambda_0/\Delta\lambda$) обладает излучение стабилизированных по частоте газовых лазеров.

Классическая модель оптического излучения возбужденного атома (см. § 1.5) также дает пример квазимонохроматического света. Напряженность поля в волне, испускаемой затухающим осциллятором, изменяется по закону

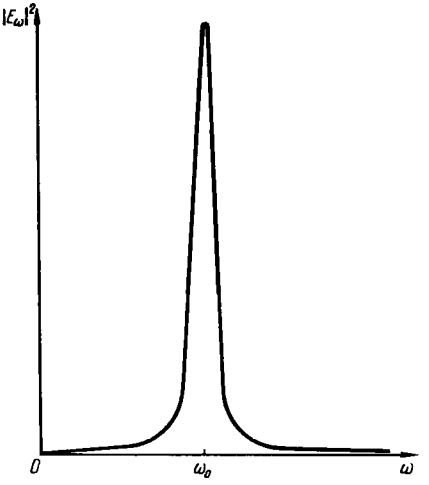
$$E(t) = \begin{cases} 0 & (t<0), \\ E_0 e^{-\gamma t} \cos\omega_0 t & (t>0), \end{cases} \tag{1.90}$$

где постоянная затухания γ определяется соотношением (1.73). Для нахождения спектра излучения подставим $E(t)$ в (1.84). При вычислении интеграла удобно $\cos\omega_0 t$ выразить через показательные функции::

$$E_\omega = \frac{1}{2} E_0 \left\{ \int_0^\infty e^{[-\gamma + i(\omega-\omega_0)]t} dt + \int_0^\infty e^{[-\gamma + i(\omega+\omega_0)]t} dt \right\} =$$

$$= \frac{1}{2} E_0 \left\{ \frac{i(\omega-\omega_0)+\gamma}{(\omega-\omega_0)^2+\gamma^2} + \frac{i(\omega+\omega_0)+\gamma}{(\omega+\omega_0)^2+\gamma^2} \right\}. \tag{1.91}$$

Рассматривая вещественную часть этой функции, замечаем, что она имеет два максимума в точках $\omega=\omega_0$ и $\omega=-\omega_0$. При $\gamma\ll\omega_0$ максимумы очень острые и только в окрестности этих пиков функция E_ω заметно отлична от нуля. Это относится как к вещественной, так и к мнимой части E_ω. Поэтому при $\gamma\ll\omega_0$ в интересующей нас области положительных частот вкладом второго слагаемого из (1.91) в функцию E_ω можно пренебречь. В результате для спектральной плотности энергии излучения затухающего осциллятора из (1.91) в случае слабого затухания находим



1.23
Лоренцевский контур спектральной линии

$$2|E_\omega|^2=\frac{1}{2}E_0^2\frac{1}{(\omega-\omega_0)^2+\gamma^2}=$$

$$=\frac{1}{2}E_0^2\tau^2\frac{1}{1+(\omega-\omega_0)^2\tau^2}. \qquad (1.92)$$

Описываемая выражением (1.92) форма спектральной линии излучения называется *лоренцевским контуром* (рис. 1.23). Кривая имеет резкий максимум при $\omega=\omega_0$, т. е. на частоте собственных колебаний в отсутствие затухания. Уширение спектра излучаемых частот обусловлено радиационным затуханием свободных колебаний осциллятора. Интенсивность излучения уменьшается вдвое для частот, отличающихся от ω_0 на $\gamma=1/\tau$. Отсюда для ширины линии на половине высоты находим $\Delta\omega=2\gamma=2/\tau$. Это значит, что в случае затухающего осциллятора ширина полосы излучаемых частот $\Delta\nu$ связана с характерной длительностью цуга τ тем же соотношением (1.89) $\Delta\nu\tau\sim1$: чем меньше длительность процесса испускания, тем шире спектр частот. Так как $\Delta\omega=2\gamma\ll\omega_0$, то излучаемый свет является квазимонохроматическим. На рис. 1.23 масштаб не выдержан — ширина лоренцевского контура сильно преувеличена.

Рассмотренный пример позволяет оценить обусловленную радиационным затуханием *естественную ширину* спектральных линий излучения свободных атомов. Так как время жизни возбужденного состояния τ составляет около 10^{-8} с (см. § 1.5), то для естественной ширины получаем $\Delta\nu\sim10^8$ Гц. В шкале длин волн оценка естественной ширины спектральной линии дает $\Delta\lambda\sim10^{-5}$ нм.

Приведенные выше примеры соответствовали квазимонохроматическому свету, получаемому из монохроматического при медленном изменении (модуляции) его амплитуды: $E(t)=E_0(t)\cos(\omega t+\varphi)$, где $E_0(t)$ — медленно (по сравнению с $\cos\omega t$) изменяющаяся

функция времени. Но немонохроматичность может быть вызвана также изменением (модуляцией) фазы: $\varphi=\varphi(t)$. Поэтому для более общего случая квазимонохроматического света колебание напряженности электрического поля может быть записано в виде

$$E(t)=E_0(t)\cos[\omega t+\varphi(t)],$$

где $E_0(t)$ и $\varphi(t)$ — медленно изменяющиеся функции времени.

Контрольные вопросы

- ☐ Постройте графики функций E_ω и $|E_\omega|^2$ для отрезка синусоидального колебания $E(t)$ (1.86).
- ☐ При каких условиях для характеристики спектрального распределения энергии вторым слагаемым в формуле (1.87) можно пренебречь?
- ☐ В каком приближении формула (1.92) дает спектральную плотность энергии излучения затухающего осциллятора?
- ☐ Рассмотрите связь лоренцевского контура (1.92) с резонансной кривой, характеризующей установившиеся колебания затухающего осциллятора под действием синусоидальной внешней силы.

Задача

Найти спектральную плотность энергии для квазимонохроматического цуга синусоидальных волн, амплитуда напряженности которых медленно изменяется по колоколообразному (гауссову) закону:

$$E(t)=E_0e^{-t^2/\tau^2}\cos\omega_0 t.$$

Здесь $\tau\gg T_0=2\pi/\omega_0$ характеризует длительность волнового цуга. Показать, что и в этом примере длительность цуга и ширина спектрального интервала $\Delta\nu$ связаны соотношением $\Delta\nu\tau\sim 1$.

$$2|E_\omega|^2=\frac{1}{2}E_0^2\pi\tau^2\exp\left\{-\frac{1}{2}\tau^2(\omega-\omega_0)^2\right\}$$

Форма спектральной линии гауссова с шириной на половине высоты $\Delta\nu=\sqrt{2\ln 2}/(\pi\tau)$.

1.8. Спектральные линии. Поляризация квазимонохроматического света

До сих пор мы рассматривали спектральный состав излучения, представляющего собой *одиночный* волновой цуг конечной длительности. При регистрации такого цуга идеальным спектральным прибором (прибором бесконечно большой разрешающей силы) должны получиться спектральный контур (1.88) в случае отрезка синусоиды вида (1.86) и лоренцевский контур (1.92) в случае одиночного затухающего цуга вида (1.90).

Рассмотрим теперь спектральный состав излучения реального источника света, состоящего из большого числа атомов — элементарных излучателей. Точнее, будем рассматривать *модель* такого

излучения. В точке наблюдения (например, там, где находится приемник) происходит сложение колебаний электромагнитного излучения, приходящего в эту точку от разных элементарных излучателей макроскопического источника. Несмотря на то что вопрос о спектральном составе результирующего излучения очень сложен, ответ на него оказывается достаточно простым. Поэтому сначала сформулируем результат и лишь затем приведем его обоснование.

В основе математической модели излучения обычного (нелазерного) источника света лежит *статистическая гипотеза* о том, что в случае *спонтанного* излучения различные атомы источника испускают отдельные цуги волн независимо друг от друга в *случайные* моменты времени. Фазы колебаний электромагнитного поля в излучении различных атомов *не скоррелированы* друг с другом. Поэтому оказывается, что распределение интенсивности излучения всех атомов источника в такой некогерентной суперпозиции определяется *суммированием распределений интенсивности* для индивидуальных атомов. В частности, если цуги волн, испускаемые различными элементарными излучателями в случайные моменты времени, одинаковы (или отличаются амплитудами), то спектр излучения источника как целого будет таким же, как и распределение интенсивности для изолированного излучателя (атома).

Перейдем к обоснованию сформулированного выше результата. Предположим, что через точку наблюдения в течение интервала времени T, необходимого для наблюдения, проходит большое число волновых цугов конечной длины. Допустим, что все эти волновые цуги, испущенные как одним и тем же атомом в разные моменты времени, так и разными атомами, идентичны по форме, т. е. колебания напряженности электрического поля, вызываемые этими цугами в точке наблюдения, описываются одной и той же функцией $E(t-t_k)$. В одних условиях это может быть функция вида (1.90), в других — (1.86). Физические условия, приводящие к той или иной модели, обсуждаются ниже. Отдельные цуги различаются только моментами прохождения через точку наблюдения t_k^*. Пусть пока и направление линейной поляризации излучения во всех цугах одинаково (при отказе от этого предположения результирующее излучение источника представляет собой неполяризованный, т. е. «естественный», свет).

Предположим, что за интервал времени $(-T/2,\ T/2)$ через точку наблюдения проходит n цугов. Тогда по принципу суперпозиции полное поле $F(t)$ в точке наблюдения в момент времени t можно записать в виде суммы n слагаемых:

$$F(t)=\sum_{k=1}^{n} E(t-t_k). \tag{1.93}$$

---

* В частном случае (1.90) момент t_k соответствует прохождению начала цуга, в случае (1.86) — прохождению середины цуга.

Слагаемые, соответствующие отдельным цугам, разложим в интеграл Фурье:

$$F(t)=\sum_{k=1}^{n} e^{i\omega t_k}\int_{-\infty}^{\infty} E_\omega e^{-i\omega t}\frac{d\omega}{2\pi}. \tag{1.94}$$

Если рассматривать эту формулу как разложение полной напряженности поля в интеграл Фурье

$$F(t)=\int_{-\infty}^{\infty} F_\omega e^{-i\omega t}\frac{d\omega}{2\pi}, \tag{1.95}$$

то, как видно из сравнения (1.94) и (1.95),

$$F_\omega=E_\omega\sum_{k=1}^{n} e^{i\omega t_k}. \tag{1.96}$$

Интересующая нас средняя интенсивность полного поля излучения (1.93) за время наблюдения T

$$I\sim\frac{1}{T}\int_{-T/2}^{T/2} F^2(t)dt\approx\frac{1}{T}\int_{-\infty}^{\infty} F^2(t)dt. \tag{1.97}$$

Пределы интегрирования здесь распространены до бесконечности, так как при большом (по сравнению с длительностью отдельных цугов) интервале времени наблюдения T можно пренебречь частичным срезанием некоторых цугов краями интервала. Далее можно воспользоваться формулой (1.85), чтобы выразить полную интенсивность через интеграл по всем частотам от ее спектральной плотности:

$$I\sim\frac{1}{T}\int_{\infty}^{\infty} F^2(t)dt=\frac{2}{T}\int_{0}^{\infty}|F_\omega|^2\frac{d\omega}{2\pi}. \tag{1.98}$$

Таким образом, для распределения энергии результирующего излучения (1.93) по спектру из (1.96) находим

$$\frac{2}{T}|F_\omega|^2=\frac{2}{T}|E_\omega|^2\left|\sum_{k=1}^{n} e^{i\omega t_k}\right|^2=$$
$$=\frac{2}{T}|E_\omega|^2\sum_{i,k=1}^{n} e^{i\omega(t_i-t_k)}. \tag{1.99}$$

Двойную сумму по i, k в (1.99) можно преобразовать, выделив в ней слагаемые с $i=k$:

$$n+2\sum_{i<k}\cos\omega(t_i-t_k). \tag{1.100}$$

Вспомним теперь, что моменты прохождения отдельных цугов t_k распределены согласно статистической гипотезе *случайным* об-

разом, поэтому одинакова вероятность появления равных по модулю положительных и отрицательных значений косинуса у разных слагаемых в (1.100). Это значит, что вторая сумма в (1.100) при большом числе слагаемых в среднем равна нулю и вся двойная сумма в (1.99) сводится к n, т. е. к числу цугов, проходящих через точку наблюдения за интервал $(-T/2, T/2)$. Так как $2|E_\omega|^2$ характеризует спектральный состав излучения отдельного цуга, то из (1.99) с учетом сказанного следует, что спектральная плотность интенсивности излучения всего источника, усредненная за время наблюдения T, равна спектральной плотности энергии излучения отдельного цуга, умноженной на среднее число цугов n/T, проходящих за единицу времени. При этом безразлично, налагаются ли отдельные цуги друг на друга или они следуют настолько редко, что между ними есть свободные промежутки. Важно лишь, чтобы за все время наблюдения T проходило большое число цугов $n \gg 1$.

Рассмотренную выше математическую модель источника света можно усложнить, отказавшись от предположения об идентичности волновых цугов, испущенных различными атомами. Оказывается, что и в этом случае спектральное распределение интенсивности излучения всех атомов источника находится суммированием распределений интенсивности излучения отдельных атомов. Спектр излучения источника как целого будет таким же, как и распределение интенсивности (1.92) для изолированного атома, если обеспечить, во-первых, отсутствие соударений атомов, и во-вторых, их движение с одинаковой скоростью в направлении наблюдения. Эти условия выполняются для свечения пучка атомов в вакууме, возбуждаемого скрещенным с ним электронным пучком, при наблюдении в направлении, перпендикулярном движению атомов. Таким образом удается получить излучение, ширина спектра которого определяется радиационным временем жизни возбуждаемого атома (естественная ширина).

В других условиях наблюдаемая на опыте ширина спектральных линий обусловлена, как правило, вторичными явлениями. Прежде всего укажем на уширение линий, вызванное *столкновениями* излучающих атомов с окружающими их атомами и молекулами. При определенной плотности газовой среды эффективное время жизни $\tau_{уд}$ излучающего атома в возбужденном состоянии может оказаться меньше радиационного времени $\tau_{изл}$ ($\approx 10^{-8}$ с). По классическим представлениям, столкновения нарушают процесс колебаний возбужденных осцилляторов, поэтому протяженность излучаемого волнового цуга, как и длительность колебаний, уменьшается. Если характерное время между столкновениями много меньше времени радиационного затухания, то изменением амплитуды на протяжении отдельного цуга можно пренебречь. Тогда спектр излучаемого некоторым атомом «оборванного» в результате столкновения волнового цуга можно аппроксимиро-

вать выражением (1.88). Если считать, что промежутки времени между соударениями распределены экспоненциально (пуассоновский процесс), то результирующая спектральная линия имеет такую же лоренцевскую форму (1.92), как при радиационном уширении, но с характерным средним временем $\tau_{уд}$ между соударениями (см. задачу):

$$I(\omega) \sim 1/[1 + (\omega - \omega_0)^2 \tau_{уд}^2]. \qquad (1.101)$$

В типичном газоразрядном источнике света $\tau_{уд}$ порядка 10^{-9} с, поэтому ширина линии (1.101) $\Delta\nu \sim 10^9$ Гц. Если время $\tau_{уд}$ одного порядка с временем радиационного затухания $1/\gamma$, то можно показать, что и в этом случае спектр характеризуется лоренцевским контуром с максимумом при $\omega = \omega_0$ и шириной $2(\gamma + n_1)$, где $n_1 = 1/\tau_{уд}$ — среднее число соударений в единицу времени.

Другая причина уширения спектральных линий — эффект Доплера. Спектр излучения, испущенного движущимся атомом, в лабораторной системе отсчета сдвинут по частоте. Излучающие атомы в источнике совершают хаотическое тепловое движение, и полный спектр излучения источника определяется наложением сдвинутых друг относительно друга одинаковых спектральных распределений отдельных атомов. В случае свечения газоразрядной плазмы низкого давления столкновения излучающих атомов происходят редко, и эти спектральные распределения обусловлены радиационным затуханием, т. е. даются сдвинутыми лоренцевскими контурами (1.92). Наложение этих контуров дает спектральную линию излучения источника с шириной, зависящей от температуры. Эта *доплеровская ширина* для водорода при комнатной температуре почти в 500 раз больше естественной.

Прежде чем исследовать форму спектральной линии, обусловленную эффектом Доплера, отметим принципиальное отличие между доплеровским уширением и рассмотренными выше радиационным и столкновительным уширениями. Это различие заключается в следующем. Радиационное и столкновительное уширение обусловлены тем, что *каждый* атом излучает цуг волн ограниченной длительности, характеризуемый определенным спектром частот. Поэтому излучению отдельного атома соответствует *весь* профиль спектральной линии, так что невозможно приписать определенную частотную компоненту внутри излучаемой линии какому-либо отдельному атому источника (или группе атомов). Такой тип уширения обычно называют *однородным*. В случае доплеровского уширения излучению разных атомов соответствуют различные части профиля спектральной линии источника, т. е. возможна идентификация определенной группы атомов по интервалу излучаемых частот в пределах контура линии. Этот тип уширения называют *неоднородным*.

Выражение для формы спектральной линии, обусловленной эффектом Доплера, получим в приближении, когда однородная

(например, естественная) ширина много меньше доплеровской. Другими словами, будем считать, что неподвижные атомы излучают практически на одной частоте ω_0. Для движущихся атомов наблюдаемые частоты сдвинуты. При этом имеет значение лишь проекция скорости атома на направление наблюдения. Если эта проекция равна v, то наблюдаемая частота излучения (при $v \ll c$) определяется выражением

$$\omega = \omega_0(1 + v/c). \tag{1.102}$$

Вероятность того, что в состоянии теплового равновесия при температуре T атом массой m имеет проекцию скорости в интервале между v и $v + \mathrm{d}v$, определяется максвелловской функцией распределения:

$$f(v)\mathrm{d}v = \left(\frac{m}{2\pi kT}\right)^{1/2} \exp\left(-\frac{mv^2}{2kT}\right)\mathrm{d}v. \tag{1.103}$$

Строго говоря, распределение Максвелла (1.103) справедливо только при тепловом равновесии. Тем не менее отклонение от него для излучающих атомов в газовом разряде обычно незначительно.

Подставляя в (1.103) v из (1.102), получаем вероятность того, что наблюдаемая частота лежит в интервале от ω до $\omega + \mathrm{d}\omega$:

$$p(\omega)\mathrm{d}\omega = \frac{c}{\omega_0}\left(\frac{m}{2\pi kT}\right)^{1/2} \exp\left[-\frac{mc^2}{2kT}\left(\frac{\omega - \omega_0}{\omega_0}\right)^2\right]\mathrm{d}\omega.$$

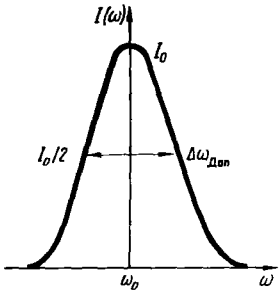
Это есть та доля полного числа возбужденных атомов источника, которая излучает в направлении наблюдения на частоте от ω до $\omega + \mathrm{d}\omega$. Акты спонтанного испускания различных атомов происходят независимо, т. е. их излучение некогерентно. Поэтому полная интенсивность излучения источника $I(\omega)\mathrm{d}\omega$ в интервале от ω до $\omega + \mathrm{d}\omega$ пропорциональна числу атомов, излучающих в этом интервале. Таким образом, *доплеровское уширение* приводит к распределению энергии в спектре, выражаемому следующей формулой:

$$I(\omega) = I_0 \exp\left[-\frac{mc^2}{2kT}\left(\frac{\omega - \omega_0}{\omega_0}\right)^2\right], \tag{1.104}$$

где I_0 — спектральная плотность интенсивности излучения в центре линии на частоте ω_0. Описываемый выражением (I.104) контур спектральной линии имеет колоколообразную форму с быстро (экспоненциально) спадающими крыльями (рис. 1.24). Он называется *гауссовым*, так как совпадает с кривой нормального закона распределения Гаусса. Ширину доплеровской линии $\Delta\omega_{\text{доп}}$ определим из (I.104) как разность частот, при которых интенсивность равна половине ее максимального значения. Полагая $I(\omega) = I_0/2$, находим

$$\Delta\omega_{\text{доп}} = 2\omega_0\left(\frac{2kT}{mc^2}\right)^{1/2}\sqrt{\ln 2} = 2\sqrt{\ln 2}\,\frac{v_m}{c}\omega_0 = 1{,}67\,\frac{v_m}{c}\omega_0, \tag{1.105}$$

где $v_m = \sqrt{2kT/m}$ — наивероятнейшая скорость атома при температуре T источника. Подставляя в (1.105) числовые значения констант,



1.24
Гауссов контур спектральной линии

получаем для оцеики доплеровской ширины спектральной линии следующую простую формулу:

$$\Delta\omega_{\text{доп}}/\omega_0 = \Delta\nu_{\text{доп}}/\nu_0 = \Delta\lambda_{\text{доп}}/\lambda_0 = $$
$$= 7{,}2\cdot 10^{-7}\sqrt{T/M}. \qquad (1.106)$$

Здесь M — атомная масса. Например, для линии неона при $\lambda_0 = 632{,}8$ нм, $M = 20$, $T = 300$ К из (1.106) находим $\Delta\lambda_{\text{доп}} = 1{,}6\cdot 10^{-3}$ нм. Это на два порядка превосходит естественную ширину спектральной линии ($\Delta\lambda_{\text{изл}} = 1{,}17\cdot 10^{-5}$ нм).

Эффект Доплера дает основной вклад в уширение спектральных линий в разреженных газах. Этот эффект важен также для понимания процессов в газовых лазерах: при низких давлениях рабочего газа в разрядной трубке эффектом Доплера определяется спектральный контур линии усиления.

В рассмотренной модели источника, содержащего большое число спонтанно излучающих атомов, предполагалось, что в точку наблюдения испущенные отдельными атомами волновые цуги приходят в состоянии линейной поляризации с одним и тем же направлением колебаний. Такое условие выполняется, если излучение на пути от источника проходит через поляризатор, пропускающий волны с определенным направлением колебаний. Результирующее излучение, возникающее при наложении многих цугов, будет немонохроматическим (имеющим спектр конечной ширины), но по-прежнему обладает линейной поляризацией с тем же направлением колебаний. Если отказаться от предположения о том, что излучение пропускается через поляризатор, то мы получим модель *неполяризованного*, или *естественного*, излучения.

По классической электронной теории, оптический электрон в атоме связан квазиупругой силой, пропорциональной его смещению из положения равновесия, так что при возбуждении он совершает колебания с определенной частотой ω_0. Такая система обладает сферической симметрией, т. е. колебания электрона в зависимости от начальных условий могут происходить в любом направлении. Поэтому ясно, что поляризация излучения в отдельных волновых цугах, испущенных различными атомами, зависит в точке наблюдения от соотношения амплитуд и фаз колебаний излучающего электрона по двум взаимно перпендикулярным направлениям. В общем случае поляризация излучения в отдельных цугах будет эллиптической с произвольными ориентацией и эксцентриситетом эллипса колебаний. Эти характеристики эллипса колебаний сохраняются на протяжении одного цуга, но случайным образом изменяются от одного цуга к другому.

Результирующее электромагнитное поле в точке наблюдения, возникающее в результате наложения волновых цугов, испущенных

различными атомами, будет обладать определенным состоянием поляризации (т. е. определенным эллипсом колебаний) лишь на протяжении промежутка времени, малого по сравнению с длительностью отдельного цуга. В силу случайного характера поляризации спонтанного излучения разных атомов состояние поляризации результирующего излучения спустя промежуток времени порядка длительности одного цуга уже никак не связано с состоянием поляризации в начале этого промежутка.

Такое излучение и называют неполяризованным. Важно отметить, что отсутствие поляризации не является его «внутренним» свойством, а проявляется лишь как результат инерционности любой существующей аппаратуры для измерения состояния поляризации. Можно считать, что в неполяризованном излучении все направления колебаний в плоскости, перпендикулярной направлению распространения, представлены с одинаковой вероятностью, т. е. для такого излучения имеется *осевая симметрия*. Иначе можно сказать так: при разложении колебаний в неполяризованном излучении на два любых взаимно перпендикулярных направления амплитуды этих квазимонохроматических колебаний одинаковы, а фазы не скоррелированы друг с другом.

Рассмотренными здесь примерами линейно поляризованного и естественного квазимонохроматического света не исчерпываются все возможные случаи. Линейно поляризованный квазимонохроматический свет, как и монохроматический, можно с помощью определенных устройств превратить в эллиптически поляризованный (в частном случае в циркулярно поляризованный). Кроме того, возможен *частично поляризованный свет*, который можно рассматривать как смесь естественного света с поляризованным.

Контрольные вопросы

- Как связана форма спектральной линии макроскопического источника, содержащего большое число независимо излучающих атомов, со спектральными контурами излучения, соответствующими отдельным цугам? Какая статистическая гипотеза лежит в основе рассмотренной математической модели излучения макроскопического источника?
- Какие условия должны быть выполнены, чтобы можно было наблюдать излучение со спектральным контуром, определяемым радиационным затуханием?
- Какой порядок величины имеет естественная ширина спектральных линий?
- В чем отличие однородного и неоднородного уширения спектральных линий?
- Объясните механизм возникновения доплеровского уширения спектральных линий.
- Какой порядок величины имеет доплеровская ширина спектральных линий? Как она зависит от температуры источника?
- Объясните, почему излучение источника, содержащего большое число атомов, может быть неполяризованным, несмотря на то что волновые цуги, испускаемые отдельными атомами, характеризуются определенным состоянием поляризации.

Задача

Рассмотрим математическую модель уширеиия спектральной линии, обусловлеиного столкиовениями излучающих атомов. Будем считать, что излучаемые отдельными атомами волновые цуги характеризуются одиой и той же средней частотой ω_0, но вследствие испытываемых атомами столкновений имеют различную длительность. Можно считать, что длительности отдельных цугов распределены по экспоненциальному закону: вероятность того, что некоторый цуг имеет длительность, заключенную в промежутке от t до $t+\mathrm{d}t$, равна

$$p(t)\mathrm{d}t=\frac{1}{\tau}\,\mathrm{e}^{-t/\tau}\mathrm{d}t.$$

Здесь τ — средняя длительность цуга (среднее время между соударениями):

$$\langle t\rangle=\int_0^\infty tp(t)\mathrm{d}t=\frac{1}{\tau}\int_0^\infty \mathrm{e}^{-t/\tau}\,t\mathrm{d}t=\tau.$$

Найти форму спектральной линии излучения, представляющего собой хаотическую последовательность таких цугов.

◇ Будем считать, что среднее время между соударениями много меньше радиационного времени жизни. Тогда затуханием колебаний на протяжении отдельного цуга можно пренебречь и принять для него форму отрезка синусоиды (1.86). Распределение энергии по спектру для такого цуга характеризуется выражеиием (1.88). Спектральную линию всего источника получим как наложение спектров отдельных цугов с учетом распределения цугов по длительности:

$$I(\omega)\sim\frac{1}{\tau}\int_0^\infty \mathrm{e}^{-\frac{t}{\tau}}\,t^2\,\frac{\sin^2[(\omega-\omega_0)t/2]}{[(\omega-\omega_0)t/2]^2}\,\mathrm{d}t\sim\frac{1}{1+(\omega-\omega_0)^2\tau^2}.$$

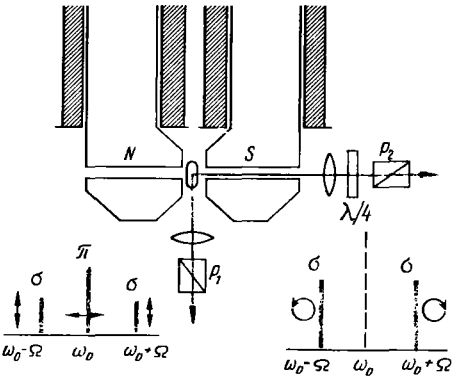
Спектр описывается лоренцевским контуром с шириной на половине высоты, определяемой средним временем между соударениями.

1.9. Эффект Зеемана

В 1896 г. голландский физик П. Зееман открыл явление расщепления спектральных линий под действием внешнего магнитного поля, приложенного к источнику. Это открытие установило несомненную связь между испусканием света и движением электрических зарядов в атоме. Объяснение наблюдаемых явлений было дано Лоренцем и сыграло важную роль в обосновании и развитии электронной теории. Эффект Зеемана позволил определить знак и значение удельного заряда частиц, ответственных за излучение света. Таким образом впервые было получено доказательство того, что эти частицы — электроны.

Схема экспериментальной установки для наблюдения явления Зеемана показана на рис. 1.25. Источник света с линейчатым спектром (газоразрядная трубка) находится в однородном магнитном поле между полюсами мощного электромагнита. По оси сердечника просверлен канал для того, чтобы иметь возможность вести наблюдение не только поперек магнитного поля, но и вдоль него. Излучение проходит через анализаторы P_1 и P_2 (николи) и кристаллическую пластинку $\lambda/4$ для исследования поляризации и затем направляется в спектрограф большой разрешающей силы. Исходная одиночиая

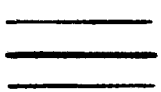
спектральная линия при наблюдении поперек поля расщепляется на три линейно поляризованные компоненты. Центральная линия соответствует той же частоте ω_0, что и при отсутствии магнитного поля, а направление поляризации параллельно магнитному полю (π-компонента). Боковые линии вдвое меньшей интенсивности смещены по шкале частот в противоположные стороны на одинаковые расстояния и поляризованы ортогонально центральной линии, т. е. перпендикулярно индукции магнитного поля (σ-компоненты). Фотография спектральной линии неона $\lambda = 613{,}3$ нм, расщепленной магнитным полем 3,3 Тл на три компоненты, приведена на рис. 1.26. При наблюдении вдоль магнитного поля боковые компоненты имеют такое же смещение, а их интенсивность вдвое меньше интенсивности исходной линии. Центральная несмещенная компонента вообще отсутствует. Смещенные компоненты имеют круговую поляризацию. Если свет распространяется в направлении индукции магнитного поля, то боковая компонента с меньшей частотой имеет правую, а с большей — левую поляризацию. При изменении направления индукции на противоположное круговая поляризация смещенных компонент также меняется на противоположную.



Расщепление спектральной линии в эффекте Зеемана

Характер расщепления одиночной спектральной линии в магнитном поле и состояние поляризации отдельных компонент можно объяснить на основе рассмотренной в § 1.8 модели источника света как совокупности одинаковых классических изотропных осцилляторов. В отсутствие магнитного поля излучение такой системы не поляризовано и имеет одну и ту же частоту ω_0 при наблюдении в любом направлении (точнее, характеризуется одинаковым спектральным контуром с центром на частоте ω_0). Влияние внешнего постоянного магнитного поля B на движение излучающего электрона можно рассмотреть, пренебрегая затуханием его колебаний. В уравнение свободных колебаний электрона нужно теперь включить еще силу Лоренца:

$$\ddot{\mathbf{r}} + \omega_0^2 \mathbf{r} = -(e/m)\dot{\mathbf{r}} \times \mathbf{B}. \qquad (1.107)$$



Расщепление линии неона $\lambda = 613{,}3$ нм в магнитном поле 3,3 Тл при наблюдении поперек поля (фотографический негатив)

Введем, по определению, вектор угловой

скорости Ω (ларморовскую частоту):

$$\Omega = -e/(2m)\mathbf{B}. \tag{I.108}$$

Теперь (I.107) можно записать в виде

$$\ddot{\mathbf{r}} - 2\dot{\mathbf{r}} \times \Omega + \omega_0^2 \mathbf{r} = 0. \tag{I.109}$$

Классическая теория эффекта Зеемана сводится к нахождению решений этого уравнения. Если записать его в проекциях на оси декартовой системы координат (ось z удобно направить вдоль индукции $\mathbf{B}$ магнитного поля, тогда $\Omega_z = -\Omega$), то получится система однородных дифференциальных уравнений:

$$\ddot{x} + 2\Omega\dot{y} + \omega_0^2 x = 0, \quad \ddot{y} - 2\Omega\dot{x} + \omega_0^2 y = 0, \quad \ddot{z} + \omega_0^2 z = 0. \tag{I.110}$$

Из последнего уравнения видно, что одно из возможных решений этой системы — гармоническое колебание электрона вдоль оси z. Такое колебание происходит с той же частотой ω_0, что и при отсутствии магнитного поля: Это и понятно, так как при движении вдоль поля сила Лоренца равна нулю. Существование такого решения объясняет несмещенную π-компоненту в эффекте Зеемана. При наблюдении поперек поля π-компонента имеет наибольшую интенсивность (равную половине интенсивности исходной линии) и линейно поляризована вдоль оси z, т. е. вдоль магнитного поля. При наблюдении вдоль поля ее интенсивность обращается в нуль — осциллирующий заряд в этом направлении не излучает.

Первые два уравнения в (I.110) «зацепляются» между собой. Но если вместо искомых функций $x(t)$ и $y(t)$ ввести другие функции $\xi_\pm = x \pm iy$, то для них из (I.110) получаются несвязанные уравнения: умножая второе уравнение на i и складывая с первым, получаем

$$\ddot{\xi}_+ - 2i\Omega\dot{\xi}_+ + \omega_0^2 \xi_+ = 0. \tag{I.111}$$

Будем искать вещественные функции $x(t)$ и $y(t)$. Тогда $x = \mathrm{Re}\,\xi_+$, $y = \mathrm{Im}\,\xi_+$. Подставим в (I.111) ξ_+ в виде $\xi_+(t) = \xi_0 e^{-i\omega t}$. Для ω получается следующее характеристическое уравнение:

$$\omega^2 + 2\Omega\omega - \omega_0^2 = 0.$$

Его корни $\omega_{1,2} = -\Omega \pm \sqrt{\omega_0^2 + \Omega^2}$. Ларморовская частота Ω, определяемая формулой (I.108), много меньше собственной частоты электрона ω_0 даже в очень сильных магнитных полях. Например, при $B = 10$ Тл (I.108) дает $\Omega \approx 10^{12}\ \mathrm{с}^{-1}$, а ω_0 для видимого света ($\lambda = 500$ нм) $\approx 4 \cdot 10^{15}\ \mathrm{с}^{-1}$. Поэтому $(\Omega/\omega_0)^2 \approx 10^{-7}$ и вторым членом в подкоренном выражении можно пренебречь, т. е. $\omega_{1,2} \approx \pm\omega_0 - \Omega$.

Первому корню $\omega_1 = \omega_0 - \Omega$ соответствует решение $\xi_+(t) = \xi_0 e^{i\omega_1 t}$, т. е.

$x(t) = \xi_0 \cos(\omega_0 - \Omega)t$, $y(t) = -\xi_0 \sin(\omega_0 - \Omega)t$.

Это решение описывает движение электрона по окружности в плоскости xy в направлении по часовой стрелке с угловой скоростью $\omega_0 - \Omega$. Такую же частоту имеет излучаемая им волна. В результате возникает компонента, смещенная в сторону низких частот на $\Omega = [e/(2m)]B$. При наблюдении поперек поля (в плоскости xy) она имеет линейную поляризацию (см. задачу 1 к § 1.5). Направление поляризации лежит в плоскости xy, т. е. перпендикулярно индукции магнитного поля. Свет с частотой $\omega_0 - \Omega$, распространяющийся вдоль магнитного поля, имеет правую круговую поляризацию.

Второму корню характеристического уравнения $\omega_2 = -\omega_0 - \Omega$ соответствует решение $\xi_+(t) = \xi^{i(\omega_0 + \Omega)t}$, т. е.

$x(t) = \xi_0 \cos(\omega_0 + \Omega)t$, $y(t) = \xi_0 \sin(\omega_0 + \Omega)t$.

Здесь электрон движется в плоскости xy по окружности против часовой стрелки. Поэтому компонента, смещенная в сторону высоких частот, при распространении вдоль магнитного поля имеет левую круговую поляризацию.

Исследуя характер круговой поляризации смещенных компонент при наблюдении вдоль магнитного поля, можно установить *знак заряда* излучающей частицы: волна левой круговой оляризации должна смещаться в сторону высоких частот для отрицательной частицы и в сторону низких — для положительной. Соответствие с опытом получается, если считать, что заряд излучающей частицы *отрицателен*. По величине расщепления можно установить *удельный заряд* e/m. Опыт дает такое же значение, как и измерения e/m по отклонению электронных пучков в электрическом и магнитном полях. Это можно рассматривать как доказательство того, что оптическое излучение атомов обусловлено движением электронов.

Дальнейшие исследования показали, что описанный выше тип расщепления в магнитном поле, называемый *нормальным* эффектом Зеемана, характерен только для *одиночных* (синглетных) спектральных линий. Большинство наблюдаемых спектральных линий представляет собой *мультиплеты*, состоящие из двух или нескольких тесно расположенных компонент. Примером может служить желтая линия натрия, состоящая из двух линий D_1 и D_2 ($\lambda_1 = 589{,}6$ нм, $\lambda_2 = 589{,}0$ нм), причем интенсивность D_2-линии вдвое больше. В магнитном поле D_1-линия расщепляется на четыре компоненты, D_2-линия — на шесть компонент. Интенсивность отдельных π- и σ-компонент такова, что смесь всех линий при наблюдении в любом направлении дает неполяризованный свет.

Расщепление мультиплетов может быть и более сложным. Для объяснения сложного эффекта Зеемана необходима квантовая теория. Причина сложного характера расщепления обусловлена тем, что электрон помимо заряда и связанного с его орбитальным движением магнитного момента обладает еще и собственным моментом

импульса (спином) и связанным с ним магнитным моментом. Спиновый момент электрона и его проявления в оптических спектрах не имеют адекватной классической интерпретации. Квантовая теория атома дает исчерпывающее объяснение сложного (или аномального) эффекта Зеемана, как и сложной структуры спектральных линий при отсутствии внешнего магнитного поля. Для простого (нормального) эффекта Зеемана в случае синглетных спектральных линий квантовая теория приводит к тем же результатам, что и классическая.

Если учитывать причины, вызывающие уширение линий, то каждая компонента, как и вся спектральная линия при отсутствии магнитного поля, характеризуется некоторой конечной шириной. Частоты ω_0, $\omega_0 \pm \Omega$ соответствуют положению центров этих уширенных компонент.

Влияние приложенного к источнику постоянного электрического поля на спектр испускаемого света было обнаружено Штарком в 1913 г. В случае водорода происходит расщепление спектральных линий на несколько компонент, пропорциональное напряженности электрического поля (линейный эффект Штарка). Расщепление спектральных линий атомов, содержащих более одного электрона, пропорционально квадрату электрического поля (квадратичный эффект Штарка).

В рамках классической модели источника как совокупности заряженных осцилляторов объяснить эффект Штарка невозможно, так как электрическое поле не влияет на собственную частоту гармонического осциллятора. Как и в случае аномального эффекта Зеемана, для объяснения эффекта Штарка требуется привлечение квантовых законов, определяющих строение атома.

Эффект Штарка дает вклад в неоднородное уширение спектральных линий газоразрядных источников, так как электрические поля ионов, образующихся в разряде, влияют на частоту света, испускаемого атомами, оказавшимися в этих случайных локальных полях.

Контрольные вопросы

- ☐ Покажите, что постоянное магнитное поле с индукцией **B** оказывает на излучающий электрон такое же влияние, как и вращение источника с угловой скоростью $\boldsymbol{\Omega} = -[e/(2m)]\mathbf{B}$.
- ☐ Почему несмещенная компонента отсутствует при наблюдении вдоль поля?
- ☐ Как электронная теория объясняет поляризацию и относительную интенсивность компонент при наблюдении поперек поля?
- ☐ Покажите, что сложение всех компонент дает неполяризованный свет при наблюдении в любом направлении. Воспользуйтесь тем, что при отсутствии поля все направления колебаний электрона равновероятны.
- ☐ Каким образом эффект Зеемана позволяет определить знак заряда частицы, ответственной за испускание света атомами?

1.10. Основы фотометрии

Способы измерения энергетических величин, характеризующих оптическое излучение, были рассмотрены в конце § 1.4. Здесь мы кратко сформулируем основные понятия фотометрии, приведем соотношения между фотометрическими величинами и их единицы.

Поток излучения Φ_σ — это мощность излучения, переносимого электромагнитными волнами через некоторую поверхность σ, усредненная за промежуток времени, значительно превышающий период колебаний. Единица потока излучения — ватт (Вт). Поток излучения через поверхность σ связан с *интенсивностью* $\langle \mathbf{S} \rangle$ (средней по времени поверхностной плотностью потока энергии) соотношением

$$\Phi_\sigma = \int_\sigma \langle \mathbf{S} \rangle \, d\sigma$$

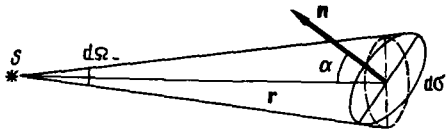
(здесь $\mathbf{S}$ — вектор Пойнтинга, см. § 1.4). Наряду с величинами Φ_σ и $\langle \mathbf{S} \rangle$, характеризующими излучение, вводят еще ряд энергетических величин для характеристики освещаемой поверхности и самого источника излучения.

Энергетической освещенностью (облученностью) E некоторой поверхности называют отношение потока излучения, приходящегося на элементарный участок поверхности, к площади этого участка: $E = d\Phi/d\sigma$. Когда излучение падает перпендикулярно поверхности, ее освещенность равна интенсивности $\langle S \rangle$. При наклонном падении (под углом θ) освещенность уменьшается: $E = \langle S \rangle \cos\theta$. Энергетическая освещенность выражается в тех же единицах, что и интенсивность (Вт/м^2).

Полная энергия излучения, падающая на 1 м^2 поверхности за некоторое время t, называется *энергетической экспозицией* H. Этой величиной определяется почернение фотоэмульсии при фотографической регистрации излучения, широко используемой во многих оптических и спектральных приборах. При неизменной освещенности $H = Et$. Экспозиция выражается в джоулях на квадратный метр (Дж/м^2).

В тех случаях, когда излучение распространяется в виде сферической волны от *точечного источника**, для характеристики пространственно-угловой плотности потока излучения вводят *энергетическую силу света*, или силу излучения. Силой излучения $J(\theta, \varphi)$ в некотором направлении, задаваемом углами θ и φ сферической системы координат, называется отношение потока излучения $d\Phi$ в элементарный телесный угол $d\Omega$, содержащий данное направление, к этому углу: $J(\theta, \varphi) = d\Phi/d\Omega$. Размерность силы излучения совпадает с размерностью потока, но в наименовании ее единицы (Вт/ср) указывается единица телесного угла — стерадиан. Полный поток излучения (т. е. поток через замкнутую поверхность, окружающую

* *Точечным* называют источник, размеры которого малы по сравнению с расстоянием от него до точки наблюдения.



1.27
К вычислению освещенности поверхности

источник) получается интегрированием энергетической силы света по всем направлениям:

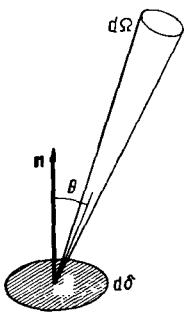
$$\Phi = \int J\,d\Omega = \int_0^{2\pi} d\varphi \int_0^{\pi} J(\theta,\ \varphi) \sin\theta d\theta.$$

Если источник излучает равномерно по всем направлениям, то сила света J не зависит от θ и φ. Тогда $\Phi = 4\pi J$. Полный поток излучения характеризует данный источник; этот поток нельзя увеличить никакими оптическими системами. Действие таких систем сводится лишь к перераспределению потока в пространстве, к большей концентрации его по некоторым направлениям (прожектор). При этом сила света $J(\theta, \varphi)$ возрастает по одним направлениям и уменьшается по другим.

Создаваемая излучением точечного источника освещенность поверхности обратно пропорциональна квадрату расстояния r от источника до поверхности. В самом деле, $E = d\Phi/d\sigma$, $d\Phi = J\,d\Omega$, а телесный угол $d\Omega$, под которым видна площадка $d\sigma$ из источника, равен $d\Omega = d\sigma \cos\alpha/r^2$. Здесь α — угол, образованный нормалью к площадке с направлением на источник (рис. 1.27). Таким образом,

$$E = J\,d\Omega/d\sigma = J\cos\alpha/r^2.$$

Для характеристики *протяженного источника* вводят еще две энергетические величины: яркость и светимость. Пусть $d\Phi$ — поток излучения, исходящего от площадки $d\sigma$ источника в телесный угол $d\Omega$ по направлению, образующему угол θ с нормалью к поверхности (рис. 1.28). Отношение этого потока к телесному углу и к видимой по данному направлению площади выделенного участка источника $d\sigma\cos\theta$ называется *энергетической яркостью* источника: $B(\theta) = d\Phi/(d\sigma\cos\theta d\Omega)$. Единица энергетической яркости 1 Вт/(м^2·ср).



1.28
К определению яркости излучающей поверхности

Из всех фотометрических характеристик источника света яркость наиболее непосредственно связана со зрительными ощущениями, так как освещенность изображения источника света на сетчатке глаза пропорциональна яркости излучающей поверхности по направлению наблюдения.

Если сила излучения $J(\theta) = d\Phi/d\Omega$, характеризующая элемент $d\sigma$ поверхности протяженного источника, пропорциональна видимой по данному направлению площади этого элемента $d\sigma\cos\theta$, то говорят, что источник удовлетворяет *закону Ламберта*. Яркость ламбертова источника $B(\theta) = J(\theta)/(d\sigma\cos\theta)$ не зависит от θ, т. е. одинакова по всем направлениям. Примером такого идеального «косинусного» излучателя может служить черное тело (см.

§ 9.1). В случае несамосветящихся поверхностей ламбертовым источником с хорошей точностью можно считать мутную среду или матовую поверхность, каждый участок которой рассеивает падающий на нее свет равномерно во все стороны, например освещенный изнутри колпак светильника из молочного стекла: его поверхность выглядит одинаково яркой при наблюдении под любым углом. Солнце практически не отличается от ламбертова источника; яркость его поверхности почти не зависит от направления. Поэтому Солнце выглядит как плоский диск почти равномерной яркости.

Энергетической светимостью R называется отношение потока излучения, исходящего от элемента поверхности источника по всем направлениям, к площади этого элемента. Выражение для R через яркость излучающей поверхности можно получить, интегрируя поток энергии по всем направлениям в пределах телесного угла 2π:

$$R=\int B(\theta)\cos\theta\,d\Omega=\int_0^{2\pi}d\varphi\int_0^{\pi/2}B(\theta)\cos\theta\sin\theta\,d\theta.$$

Для источника, удовлетворяющего закону Ламберта, $B=\text{const}$, и связь между светимостью и яркостью становится особенно простой: $R=\pi B$. Единица энергетической светимости — 1 Вт/м^2.

Наиболее старый метод измерения энергии излучения в видимой области спектра — визуальный. Здесь приемником излучения служит глаз, а основным способом количественных измерений — визуальное уравнивание яркости двух фотометрических полей: стандартного и измеряемого. При таких измерениях играет роль только та часть энергии излучения, которая непосредственно вызывает световое ощущение. Чувствительность «среднего» глаза к монохроматическому излучению разных длин волн характеризуется *спектральной световой эффективностью,* или *видностью* (см. кривую на переднем форзаце). Очевидно, что при измерениях энергии светового излучения, основанных на зрительных ощущениях, обычные энергетические характеристики излучения оказываются недостаточными. В таких случаях применяют специальные *световые величины,* базирующиеся на использовании установленного международным соглашением стандартного источника (*светового эталона*) с определенным распределением энергии по спектру. В качестве эталонного выбрано излучение абсолютно черного тела (см. § 9.1) при температуре затвердевания чистой платины (2042 К). Основной светотехнической единицей (входящей в число основных единиц СИ) установлена единица силы света *I кандела* (от лат. candela — свеча). Кандела (кд) — это сила света, испускаемого с 1/60 см^2 поверхности эталонного источника в направлении нормали.

За единицу светового потока принимается *люмен* (лм) — световой поток от точечного источника силой света 1 кд, распространяющийся в пределах телесного угла 1 ср. Единица освещенности *люкс* (лк) — освещенность поверхности, на 1 м^2 которой падает световой поток 1 лм. Точечный источник силой света 1 кд создает

освещенность 1 лк на поверхности, расположенной в 1 м от него перпендикулярно лучам. Яркость $B(\theta)=J/(d\sigma\cos\theta)$ выражается в канделах на квадратный метр (кд/м^2), светимость $R=d\Phi/d\sigma$ — в люменах на квадратный метр (лм/м^2).

Вследствие зависимости чувствительности глаза от длины волны световому потоку 1 лм соответствует разная мощность в зависимости от спектрального состава излучения. Для монохроматического излучения с длиной волны $\lambda=555$ нм, соответствующей максимальной чувствительности глаза, *механический эквивалент света* составляет 0,0016 Вт/лм. Это минимальная мощность излучения в ваттах, способная создать световой поток 1 лм в наиболее воспринимаемой глазом спектральной области.

Во всех фотометрах, предназначенных для визуального сравнения различных источников, роль глаза сводится к установлению равенства освещенностей двух граничащих друг с другом полей. Для достижения этого равенства используют разнообразные приемы ослабления освещенности, создаваемоей более сильным источником (изменение расстояния, нейтральные фильтры переменной толщины, системы поляризационных призм). При равенстве освещенностей граница между полями исчезает и они сливаются в одно поле. Когда оба поля имеют одинаковый цвет, глаз фиксирует равенство их освещенностей с высокой точностью.

Для сравнения источников с разным спектральным составом излучения (гетерохромная фотометрия) равенство освещенностей имеющих разный цвет фотометрических полей устанавливают на основе некоторых психофизиологических особенностей зрения (например, совпадение предельной частоты исчезновения мельканий при освещении прерывистым светом рассматривается как признак равенства освещенностей полей разного цвета).

Контрольные вопросы

Сформулируйте определения основных энергетических характеристик излучения: потока энергии излучения, энергетической освещенности, силы света, яркости, светимости. В каких единицах их выражают?

Что такое точечный источник? Как зависит от расстояния создаваемая им освещенность?

При каком условии протяженный источник называется ламбертовым? Приведите примеры ламбертовых источников.

Почему наряду с энергетическими характеристиками излучения необходимо вводить светотехнические? Как определяется основная светотехническая единица — кандела?

РАСПРОСТРАНЕНИЕ СВЕТА В ИЗОТРОПНЫХ СРЕДАХ

● При изучении распространения электромагнитных волн и, в частности, света в среде нужно учитывать, что любое вещество содержит электроны и ядра, входящие в состав атомов и молекул. Под действием электрического поля падающей волны заряды среды начинают совершать колебательное движение и сами становятся источниками электромагнитных волн. Эти вторичные волны от всех элементов объема вещества налагаются друг на друга и вместе с падающей волной образуют в итоге полное электромагнитное поле в веще́стве, которое и вызывает вынужденное движение входящих в состав вещества зарядов.

● Последовательная реализация этого подхода приводит к интегральным уравнениям. Опираясь на них можно показать, как внешнее электромагнитное поле, распространяющееся в пустом пространстве между зарядами среды со скоростью света в вакууме, точно компенсируется и заменяется в веществе результирующим вторичным возмущением, распространяющимся с меньшей скоростью. Эффект сложения падающей волны со вторичными волнами эквивалентен изменению фазовой скорости света в веществе.

● Значительно более простым оказывается решение этой задачи не на основе имеющих весьма прозрачное физическое содержание интегральных уравнений, а на основе дифференциальных уравнений для средних полей в веществе. Именно такой подход и используется в данной главе. Но все же отметим, что достигаемые при использовании дифференциальных уравнений математические упрощения несколько затеняют физическую картину возникновения преломленной волны как результата наложения первоначальной волны и вновь излученных в веществе волн.

2.1. Уравнения Максвелла для волн в веществе

Основные представления о распространении света в веществе могут быть систематизированы в рамках классической электронной теории, получившей наибольшее развитие к началу XX в. в трудах Г. А. Лоренца. В этой теории электрические свойства вещества объясняются на основе простейшей модели атома, предложенной Дж. Томсоном, в которой электроны удерживаются в положении равновесия квазиупругой силой. Хотя дальнейшее изучение строения атома привело к отказу от этой модели, связанные с ней представления о макроскопических электромагнитных свойствах вещества дошли до нашего времени без особых изменений. Это означает, что основные результаты электронной теории, относящиеся к распространению света в веществе, не связаны непосредственно с ее примитивными модельными представлениями.

Философский аспект подобной ситуации заключается в том, что успешное объяснение наблюдаемых явлений с помощью некоторой модели еще не может служить решающим ее подтверждением. В то же время полученные на основе определенной модели результаты могут иметь общее значение, выходящее за ее рамки. История физики знает немало таких примеров.

Согласно представлениям классической электронной теории, вещество рассматривается как совокупность заряженных частиц, размещенных в вакууме на расстояниях, больших по сравнению с размерами самих частиц. Электромагнитное поле сильно меняется на расстоянии между входящими в состав вещества заряженными частицами. Однако эти быстрые изменения поля в пространстве в обычных условиях на опыте не наблюдаются.

В самом деле, опыт показывает, что свет распространяется в воздухе, стекле и многих других веществах в значительной мере так же, как и в вакууме, т. е. макроскопически эти вещества представляются непрерывными и однородными. Непрерывными представляются и поля распространяющихся в них электромагнитных волн. Поэтому интерес представляют *средние* значения характеризующих вещество физических величин, таких, как плотность заряда или плотность тока, и средние значения напряженностей электромагнитных полей. Усреднение должно производиться по элементам объема, содержащим макроскопически большое число атомов или молекул, т. е. большим по сравнению со средним расстоянием между частицами. В то же время линейный размер этих элементов объема должен быть много меньше характерного размера макроскопических неоднородностей, мерой которых может служить длина электромагнитной волны. Удовлетворяющие таким условиям элементы объема принято называть *физически бесконечно малыми*.

Электрическое и магнитное поля волны остаются почти постоянными на расстояниях порядка $\lambda/(2\pi)$, что для видимого света составляет примерно 10^{-7} м. В объеме $(10^{-7})^3$ м^3 содержится около 10^8 атомов конденсированного вещества и около $3 \cdot 10^4$ молекул

любого газа при нормальных условиях. Поэтому во всем оптическом диапазоне, включая и короткие ультрафиолетовые волны, вещество обычно проявляет себя как сплошная однородная среда. Лишь в особых условиях, например при распространении света в верхней атмосфере, плотность которой очень мала, или в веществе, находящемся в состоянии, близком к критическому, среда не может рассматриваться как вполне однородная: флуктуации плотности в объемах порядка $[\lambda/(2\pi)]^3$ становятся существенными, что приводит к рассеянию преимущественно коротковолнового излучения. В диапазоне рентгеновских лучей, когда длина волны становится сравнимой с межатомными расстояниями, макроскопический подход к веществу как сплошной среде становится, вообще говоря, неприменим.

Изучение распространения света в веществе облегчается следующими обстоятельствами. Электроны в атомах находятся в движении, но их скорость v обычно много меньше скорости света c. Поэтому пропорциональная v/c сила Лоренца (1.1), действующая на электрон со стороны магнитного поля световой волны, оказывается пренебрежимо малой по сравнению с силой, действующей со стороны электрического поля. Сила Лоренца существенна при объяснении давления излучения (см. § 3.5).

Наконец, внутренние электрические поля в атомах, создаваемые зарядами ядер и электронов, обычно значительно превосходят напряженность поля излучения. Исключение в этом отношении составляет лишь сфокусированное лазерное излучение высокой интенсивности, о распространении которого речь пойдет в гл. 10. Поэтому в обычных условиях излучение лишь незначительно возмущает состояние электронов в веществе. Такие возмущения можно рассмотреть в линейном приближении. Линейность отклика вещества на действие электрического поля волны позволяет воспользоваться методом фурье-анализа. Это значит, что достаточно рассмотреть отклик вещества на плоские монохроматические волны различных частот.

Действие электрического поля электромагнитной волны на электрон в атоме вызывает его смещение из положения равновесия. Относительное смещение отрицательного и положительного зарядов проявляется в том, что атом приобретает *дипольный момент*. Вещество оказывается поляризованным*. Макроскопической характеристикой *поляризованности* вещества служит вектор **P**, который равен отношению векторной суммы дипольных моментов всех атомов в физически бесконечно малом элементе среды к объему этого элемента.

Наличие поляризованности вещества означает возникновение в нем поляризационных зарядов и токов, обусловленных связанными в атомах электронами. Эти заряды и токи необходимо учитывать в полных уравнениях Максвелла при нахождении электромагнит-

* Понятие поляризации вещества в электрическом поле не имеет ничего общего с поляризацией электромагнитных волн.

ного поля. Поэтому для описания электромагнитных волн в веществе нужно искать решение уравнений Максвелла (1.2)—(1.3), когда заряды и токи ρ и $\mathbf{j}$ не равны нулю, а выражаются через поляризованность $\mathbf{P}$.

Когда вещество поляризовано неоднородно, т. е. вектор $\mathbf{P}$ меняется от точки к точке, то физически бесконечно малый элемент объема приобретает не только дипольный момент, но и отличный от нуля полный заряд. Макроскопически этот *поляризационный заряд* характеризуется объемной плотностью ρ, которая выражается через скорость изменения вектора $\mathbf{P}$ в пространстве следующим соотношением:

$$\rho = -\nabla \mathbf{P}. \qquad (2.1)$$

Изменение поляризованности с течением времени означает, что создающие ее заряды вещества движутся, т. е. возникает *поляризационный ток*. Он характеризуется вектором плотности тока $\mathbf{j}$, который равен скорости изменения вектора $\mathbf{P}$:

$$\mathbf{j} = d\mathbf{P}/dt. \qquad (2.2)$$

Соотношение (2.2) можно пояснить следующим образом. Каждый движущийся элементарный заряд вещества q вносит в ток свой вклад, равный произведению заряда на его скорость $\mathbf{v}$. Если в единице объема содержится N таких зарядов (движущихся с одинаковой скоростью), то они создают плотность тока $\mathbf{j} = Nq\mathbf{v}$. Но $\mathbf{v} = d\mathbf{r}/dt$, поэтому $\mathbf{j} = d(Nq\mathbf{r})/dt = d\mathbf{P}/dt$.

Входящие в правые части неоднородных уравнений Максвелла (1.2) и (1.3) плотность заряда ρ и плотность тока $\mathbf{j}$ содержат вклад как внешних источников, создающих интересующее нас поле излучения, так и рассмотренный выше вклад поляризованности вещества, выражаемый соотношениями (2.1) и (2.2). Если, как и в § 1.1, интересоваться распространяющимися от внешнего источника волнами в тех местах, где внешних источников уже нет, то в правые части уравнений Максвелла (1.2) и (1.3) войдут только обусловленные поляризованностью ρ и $\mathbf{j}$ из (2.1) и (2.2):

$$\nabla \mathbf{E} = -\frac{1}{\varepsilon_0} \nabla \mathbf{P}, \qquad (2.3)$$

$$c^2 \nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon_0} \frac{\partial \mathbf{P}}{\partial t}. \qquad (2.4)$$

Этим уравнениям можно формально придать такой же вид, как и уравнениям Максвелла для вакуума, если ввести вместо напряженности $\mathbf{E}$ электрического поля вектор электрической индукции $\mathbf{D}$ по следующему определению:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}. \qquad (2.5)$$

При использовании вектора $\mathbf{D}$ уравнения (2.3) и (2.4) записываются в виде

$$\nabla \mathbf{D} = 0, \qquad (2.6)$$

$$\varepsilon_0 c^2 \nabla \times \mathbf{B} - \partial \mathbf{D}/\partial t = 0. \qquad (2.7)$$

Вторая пара уравнений Максвелла (1.9) и (1.10) не содержит

источников ρ и $\mathbf{j}$, и поэтому для полей в веществе имеет тот же вид, что и для полей в вакууме:

$$\nabla \mathbf{B}=0, \quad (2.8) \qquad \nabla \times \mathbf{E}+\partial \mathbf{B}/\partial t=0. \quad (2.9)$$

Рассматривая распространение электромагнитных волн в среде, мы говорили о влиянии на вещество только электрического поля волны. Если атомы или молекулы среды обладают магнитными моментами, ориентирующее действие магнитного поля приводит к появлению отличного от нуля суммарного магнитного момента у физически бесконечно малых элементов объема. Макроскопически такая *намагниченность* вещества характеризуется вектором $\mathbf{M}$ — средней объемной плотностью суммарного магнитного момента. Процессы ориентации магнитных моментов атомов или молекул происходят *медленно* по сравнению с периодом колебаний магнитного поля в электромагнитной волне оптического диапазона (см. § 2.7). Поэтому в оптике всегда можно считать $\mathbf{M}=0$. Для описания магнитного поля световой волны в веществе, как и в вакууме, достаточно одного лишь вектора $\mathbf{B}$*.

В дальнейшем мы будем рассматривать решения уравнений Максвелла (2.6)—(2.9), описывающие плоские монохроматические волны в веществе.

Контрольные вопросы

Что такое физически бесконечно малые элементы объема?
Какие условия должны выполняться для того, чтобы можно было использовать это понятие?

2.2. Материальные уравнения. Плоские монохроматические волны в изотропной среде

Фундаментальные уравнения Максвелла (2.6)—(2.9) для электромагнитного поля в веществе еще не составляют полной системы уравнений. Их необходимо дополнить соотношениями, характеризующими свойства рассматриваемой среды. Такие соотношения устанавливают связь между векторами $\mathbf{E}$ и $\mathbf{D}$ (или между $\mathbf{E}$ и $\mathbf{P}$) и называются *материальными уравнениями.*

В быстропеременных электромагнитных полях, заметно изменяющихся за времена, сравнимые с временем установления (релаксации) электрической поляризованности вещества, связь между $\mathbf{E}$ и $\mathbf{P}$ становится, вообще говоря, довольно сложной. В отличие от медленно меняющихся полей значение поляризованности $\mathbf{P}$ в некоторый момент времени отнюдь не определяется значением напряженности $\mathbf{E}$ в тот же момент, а зависит от значений функции $\mathbf{E}(t)$

* Влияние магнитного поля волны на поляризованность среды в области оптических частот неотличимо от эффектов пространственной дисперсии (см. § 2.9).

во все предыдущие моменты времени. Только в монохроматических полях связь между $\mathbf{E}$ и $\mathbf{P}$ оказывается сравнительно простой. Поэтому целесообразно сначала рассмотреть распространение в среде монохроматических волн определенной частоты ω. Чтобы для таких волн можно было использовать макроскопическое описание поля, длина волны λ, имеющая порядок c/ω, должна быть много больше среднего расстояния между атомами или молекулами среды.

Зависимость напряженности электрического поля плоской монохроматической волны от координат и времени может быть записана в виде

$$\mathbf{E}(\mathbf{r}, t)=\mathbf{E}_0 e^{i(\mathbf{kr}-\omega t)}. \tag{2.10}$$

Волновой вектор $\mathbf{k}$ перпендикулярен поверхностям постоянной фазы и характеризует направление волны, а его модуль k (волновое число) обратно пропорционален длине волны: $\lambda=2\pi/k$. В пределах элемента объема, малого по сравнению с длиной волны, электрическое поле волны (2.10) можно считать однородным* и изменяющимся со временем по гармоническому закону:

$$\mathbf{E}(t)=\mathbf{E}_0 e^{-i\omega t}. \tag{2.11}$$

Если амплитуда $\mathbf{E}_0$ напряженности поля волны много меньше напряженности внутриатомных электрических полей, отклик вещества на поле волны можно считать линейным. Это значит, что поляризованность среды в пределах рассматриваемого элемента объема однородна и совершает вынужденные колебания под действием поля $\mathbf{E}(t)$ по гармоническому закону с частотой внешнего воздействия ω и амплитудой $\mathbf{P}_0$, пропорциональной $\mathbf{E}_0$:

$$\mathbf{P}(t)=\mathbf{P}_0 e^{-i\omega t}=\varepsilon_0\chi(\omega)\mathbf{E}_0 e^{-i\omega t} \tag{2.12}$$

Направление вектора $\mathbf{P}$ в изотропной среде, где нет физически выделенных направлений, совпадает с направлением вектора $\mathbf{E}$. Поэтому коэффициент пропорциональности $\chi(\omega)$ между $\mathbf{P}$ и $\mathbf{E}$, называемый *диэлектрической восприимчивостью*, в изотропной среде является скаляром**. В однородной среде, свойства которой всюду одинаковы, восприимчивость $\chi(\omega)$ не зависит от пространственных переменных.

При вынужденных колебаниях электронов вещества под действием электрического поля волны их движение, создающее поляризованность, вообще говоря, происходит с отставанием по фазе от колебаний напряженности электрического поля. Это запаздывание по фазе в формуле (2.12) проявляется в том, что восприимчивость $\chi(\omega)$ — *комплексная* величина.

В тех случаях, когда можно пренебречь затуханием элемен-

---

* Влияние неоднородности электрического поля монохроматической волны обсуждается в § 2.9.

** В анизотропных средах (кристаллах) направление $\mathbf{P}$ в общем случае не совпадает с направлением $\mathbf{E}$, т. е. диэлектрическая восприимчивость $\chi(\omega)$ является тензором (см. § 4.2).

тарных осцилляторов, отставания по фазе не будет и восприимчивость выражается вещественной величиной. Так будет для частот, далеких от собственной частоты осцилляторов, т. е. в спектральной области прозрачности среды.

Характеризующая распространение монохроматической волны восприимчивость $\chi(\omega)$ зависит от частоты волны ω. Об этой зависимости говорят как о *законе дисперсии* восприимчивости. Вид функции $\chi(\omega)$ определяется структурой вещества. Ее можно рассчитать в рамках той или иной идеализированной модели. Сравнительно простой оказывается модель для вещества с малой плотностью (газы, плазма), рассматриваемая в классической электронной теории дисперсии (см. § 2.3). В плотном материале (конденсированные среды) атомы расположены тесно и сильно взаимодействуют друг с другом. Собственные частоты ω_0 и коэффициенты затухания γ атомных электронов в плотном веществе из-за этих взаимодействий будут иными, чем у свободных атомов. Кроме того, локальное поле, действующее на отдельный атом в плотном веществе, отличается от среднего макроскопического поля E. Все это приводит к тому, что точное вычисление функции $\chi(\omega)$ (теория дисперсии) для плотного вещества представляет собой трудную задачу, которая решена только для некоторых особенно простых веществ.

Однако составить ясное физическое представление о многих явлениях, связанных с распространением света в веществе, можно и не решая задачи вычисления функции восприимчивости $\chi(\omega)$. Для этого можно воспользоваться уравнениями Максвелла и *феноменологической* теорией, в которой *материальное уравнение* (2.12), связывающее поляризованность $\mathbf{P}$ с напряженностью $\mathbf{E}$ поля монохроматической волны, считается заданным (например, полученным путем экспериментальных измерений на разных частотах ω).

В рамках такой феноменологической теории удобно перейти от уравнений Максвелла (2.3)—(2.4), содержащих поляризованность $\mathbf{P}$, к иной их форме (2.6)—(2.7), аналогичной соответствующим уравнениям для вакуума. При этом вместо $\mathbf{E}$ и $\mathbf{P}$ в них фигурирует единственный вектор $\mathbf{D}$, определяемый соотношением (2.5). Поскольку вектор $\mathbf{P}$ такой заменой из дальнейшего рассмотрения исключается, целесообразно вместо функции восприимчивости $\chi(\omega)$, связывающей $\mathbf{P}$ и $\mathbf{E}$, ввести *диэлектрическую проницаемость* $\varepsilon(\omega)$, связывающую векторы $\mathbf{D}$ и $\mathbf{E}$ в монохроматической волне частоты ω:

$$\mathbf{D}(t)=\mathbf{D}_0 e^{-i\omega t}=\varepsilon_0\varepsilon(\omega)\mathbf{E}_0 e^{-i\omega t}=\varepsilon_0\varepsilon(\omega)\mathbf{E}(t). \qquad (2.13)$$

В излагаемой теории соотношение (2.13) играет роль материального уравнения. Чтобы выразить функцию $\varepsilon(\omega)$ через $\chi(\omega)$, подставим в (2.5) $\mathbf{E}$ из (2.11), $\mathbf{P}$ из (2.12) и $\mathbf{D}$ из (2.13). В результате получим

$$\varepsilon(\omega)=1+\chi(\omega). \qquad (2.14)$$

В изотропной однородной среде $\varepsilon(\omega)$, как и $\chi(\omega)$, является скаляром и не зависит от пространственных координат. В случае комплексной восприимчивости $\chi(\omega)$ диэлектрическая проницаемость также будет комплексной:

$$\varepsilon(\omega)=\varepsilon'(\omega)+i\varepsilon''(\omega). \tag{2.15}$$

Ее вещественная часть $\varepsilon'(\omega)$ определяет ту составляющую индукции $\mathbf{D}$, которая изменяется в фазе с напряженностью $\mathbf{E}$ электрического поля, а мнимая часть $\varepsilon''(\omega)$ — составляющую $\mathbf{D}$, которая отстает от $\mathbf{E}$ на $\pi/2$. Ниже мы увидим, что вещественная часть ε ответственна за преломление света на границе среды, а мнимая — за поглощение света в веществе.

Отметим, что между вещественной и мнимой частями $\varepsilon(\omega)$ существуют интегральные соотношения (формулы Крамерса — Кроинга), которые являются прямым следствием чрезвычайно общего физического принципа причинности. Они позволяют вычислить, например, функцию $\varepsilon'(\omega)$, если известна для данного вещества функция $\varepsilon''(\omega)$ во всем диапазоне частот ω. Поэтому для полной характеристики оптических свойств среды в феноменологической теории достаточно знать (например, из эксперимента) только одну из этих функций, скажем $\varepsilon''(\omega)$.

Выразим в уравнениях Максвелла (2.6)—(2.7) индукцию $\mathbf{D}$ через напряженность $\mathbf{E}$ поля с помощью материального соотношения (2.13), согласно которому в монохроматической волне $\mathbf{D}=\varepsilon(\omega)\mathbf{E}$:

$$\nabla\mathbf{E}=0, \tag{2.16}$$

$$c^2\nabla\times\mathbf{B}=\varepsilon(\omega)\,\partial\mathbf{E}/\partial t. \tag{2.17}$$

Нас интересует решение в виде плоской монохроматической волны, в которой зависимость характеристик электрического и магнитного полей от координат и времени выражается формулой (2.10). Для таких функций дифференцирование по времени сводится к умножению на $-i\omega$, а применение оператора дифференцирования по координатам ∇ — к умножению на $i\mathbf{k}$. Поэтому уравнения (2.16)—(2.17), а также вторая пара уравнений Максвелла (2.8)—(2.9) превращаются из дифференциальных уравнений в алгебраические:

$$\mathbf{kE}=0, \tag{2.18}$$

$$c^2\mathbf{k}\times\mathbf{B}=-\varepsilon(\omega)\omega\mathbf{E}, \tag{2.19}$$

$$\mathbf{kB}=0, \tag{2.20}$$

$$\mathbf{k}\times\mathbf{E}=\omega\mathbf{B}. \tag{2.21}$$

Эти уравнения* однородны, и из условия существования их ненулевого решения можно получить допустимые при данной частоте ω значения волнового числа k, определяющего фазовую скорость волны. Подставив $\mathbf{B}$ из (2.21) в (2.19), найдем

$$\mathbf{k}\times(\mathbf{k}\times\mathbf{E})=-\varepsilon(\omega)(\omega^2/c^2)\mathbf{E}. \tag{2.22}$$

* Для монохроматических полей уравнения (2.18) и (2.20) следуют из уравнений (2.19) и (2.21) соответственно: при подстановке $\mathbf{E}$ из (2.19) в (2.18), как и при подстановке $\mathbf{B}$ из (2.21) в (2.20), получаем тождества. Поэтому здесь только шесть независимых уравнений.

Раскрывая двойное векторное произведение в левой части и учитывая, что, согласно (2.18), $\mathbf{kE}=0$, находим условие

$$k^2=\varepsilon(\omega)\omega^2/c^2, \tag{2.23}$$

при выполнении которого уравнения (2.18)—(2.21) имеют ненулевое решение. В общем случае, когда $\varepsilon(\omega)$ — комплексная величина, k^2 тоже становится комплексным. Здесь могут представиться разные случаи. Запишем вектор $\mathbf{k}$ в виде

$$\mathbf{k}=\mathbf{k}'+i\mathbf{k}'',$$

где $\mathbf{k}'$ и $\mathbf{k}''$ — вещественные векторы. Подставляя в (2.23) $k^2=k'^2-k''^2+2i\mathbf{k}'\mathbf{k}''$, находим, что вектор $\mathbf{k}$ может быть вещественным только тогда, когда величина $\varepsilon(\omega)$ вещественна и положительна. Но даже и в этом случае $\mathbf{k}$ все же может быть комплексным, если только $\mathbf{k}'\mathbf{k}''=0$ (например, при полном отражении, см. § 3.3).

В случае комплексного вектора $\mathbf{k}$ волна (2.10), вообще говоря, не будет плоской в обычном смысле этого слова. В самом деле, записав (2.10) в виде

$$\mathbf{E}(\mathbf{r}, t)=\mathbf{E}_0 e^{-\mathbf{k}''\mathbf{r}} e^{i(\mathbf{k}'\mathbf{r}-\omega t)}, \tag{2.24}$$

убеждаемся, что плоскости постоянной фазы перпендикулярны вектору $\mathbf{k}'$, в то время как поверхности постоянной амплитуды представляют собой плоскости, перпендикулярные вектору $\mathbf{k}''$, в направлении которого происходит затухание волны. Если векторы $\mathbf{k}'$ и $\mathbf{k}''$ имеют различные направления, то волну (2.24) называют *неоднородной* плоской волной.

Рассмотрим наиболее важные частные случаи.

Во многих веществах электромагнитные волны в широком диапазоне частот распространяются практически без ослабления. Примером может служить прохождение видимого света сквозь стекло, воздух, воду и другие прозрачные газы и жидкости. В случае таких *непоглощающих сред* диэлектрическая проницаемость $\varepsilon(\omega)$ для рассматриваемых частот вещественна и положительна. Волновой вектор при этом также веществен и, как следует из (2.2[3]), по модулю равен

$$k=\frac{\omega}{c}\sqrt{\varepsilon(\omega)}=\frac{\omega}{c}n(\omega), \tag{2.25}$$

где $n=\sqrt{\varepsilon}$ — *показатель преломления* среды. Это очень важная характеристика оптических свойств вещества. Через показатель преломления выражается, в частности, *фазовая скорость* монохроматической волны; как видно из (2.10),

$$v=\omega/k=c/n(\omega). \tag{2.26}$$

Зависимость фазовой скорости (или показателя преломления) от частоты (или от длины волны) называется *дисперсией**. Наличие дисперсии позволяет в эксперименте «рассортировать» монохроматические компоненты сложной волны, так как при наклонном падении на границу вещества эти компоненты в преломленном излучении распространяются в различных направлениях. На этом принципе действуют призменные спектрографы и монохроматоры.

Дисперсия проявляется не только в эффекте преломления на границе вещества. Она сильно влияет и на прямолинейное распространение немонохроматического излучения в объеме вещества, так как различные монохроматические компоненты имеют отличающиеся скорости, т. е. как бы «идут не в ногу». При этом огибающая модулированной волны перемещается с так называемой групповой скоростью, не совпадающей с фазовой скоростью. Подробно эти эффекты рассматриваются в § 2.11.

В случае вещественного волнового вектора $\mathbf{k}$ из уравнений (2.18) и (2.20) следует, что векторы электрического и магнитного полей лежат в плоскости, перпендикулярной вектору $\mathbf{k}$, т. е. световая волна поперечна. Из (2.19) или (2.21) видно, что векторы $\mathbf{E}$ и $\mathbf{B}$ перпендикулярны друг другу и вместе с вектором $\mathbf{k}$ образуют правую тройку векторов (как орты $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ правой системы координат). Величины векторов $\mathbf{E}$ и $\mathbf{B}$ в каждой точке и в любой момент времени связаны соотношением

$$\sqrt{\varepsilon(\omega)}E=cB. \qquad (2.27)$$

Для каждого конденсированного вещества имеются полосы частот, в которых оно поглощает излучение. В поглощающей среде волновой вектор становится комплексным, но может при этом иметь определенное направление, т. е. векторы $\mathbf{k}'$ и $\mathbf{k}''$ параллельны друг другу. Такая волна будет плоской в буквальном смысле, так как для нее поверхности постоянных значений напряженности поля представляют собой плоскости, перпендикулярные направлению распространения. Чтобы подчеркнуть это, ее называют *однородной* плоской волной. Иначе говоря, в такой волне плоскости постоянной фазы и плоскости постоянной амплитуды параллельны друг другу.

При параллельных $\mathbf{k}'$ и $\mathbf{k}''$ можно ввести комплексную «длину» волнового вектора, записав $\mathbf{k}$ в виде $k\mathbf{e}_k$, где $\mathbf{e}_k$ — единичный вектор в направлении $\mathbf{k}$. Тогда из (2.23) имеем $k=\sqrt{\varepsilon(\omega)}\,\omega/c$. Комплексную величину $\sqrt{\varepsilon(\omega)}$ обычно пишут в виде $n+i\varkappa$ с вещественными n и $\varkappa$ и называют *комплексным показателем преломления*. Поэтому $k=(n+i\varkappa)\omega/c$. Выбирая ось z в направлении вектора $\mathbf{k}$ и подставляя

* Слово «дисперсия» означает разброс, отклонение, рассеяние. Это слово используется для обозначения различных понятий. Например, в статистике дисперсией называют среднюю квадратичную ошибку или меру разброса значений случайной величины от среднего. В оптике под дисперсией понимают круг явлений, берущий начало в известных опытах Ньютона по различному преломлению цветных лучей в призме. Под дисперсией в узком смысле этого слова понимают зависимость показателя преломления от частоты $n(\omega)$ (или от длины волны λ).

комплексную величину k в (2.10), получаем для этого случая зависимость напряженности электрического поля волны от координат и времени:

$$E(z, t)=E_0 e^{-\varkappa\omega z/c} e^{-i\omega(t-nz/c)}. \qquad (2.28)$$

Вещественная часть показателя преломления n, как и в случае прозрачной среды, определяет фазовую скорость волны $v=c/n$. Мнимая часть $\varkappa$ называется *показателем затухания*. Она характеризует ослабление волны по мере ее распространения: как видно из (2.28), на расстоянии $\Delta z=c/\omega=\lambda_0/(2\pi)$, где λ_0 — длина волны данной частоты ω в вакууме, амплитуда приобретает множитель $e^{-\varkappa}$.

Заметим, что затухание волны не обязательно связано с истинным поглощением электромагнитной энергии: диссипация энергии происходит лишь тогда, когда мнимая часть $\varepsilon(\omega)$ отлична от нуля, а коэффициент $\varkappa$ может быть отличен от нуля и при вещественном (отрицательном) $\varepsilon(\omega)$. Именно так дело обстоит для плазмы при $\omega<\omega_p$, где ω_p — плазменная частота (см. § 2.3). Фактически это означает, что излучение при $\varepsilon(\omega)<0$ не может проникнуть в вещество и происходит полное отражение волны на границе.

Векторы электрического и магнитного полей в затухающей плоской волне перпендикулярны друг другу и направлению распространения, а их величины по-прежнему связаны соотношением (2.27), но только с комплексным значением $\varepsilon(\omega)$. Если записать $\sqrt{\varepsilon}$ в виде

$$\sqrt{\varepsilon} = \sqrt{n^2+\varkappa^2}\,\exp\left[i\,\mathrm{arctg}\,(\varkappa/n)\right],$$

то становится ясно, что колебания магнитного поля отстают по фазе от колебаний электрического поля на угол $\mathrm{arctg}\,(\varkappa/n)$ в отличие от волн в вакууме или прозрачной среде, где колебания полей в каждой точке происходят синфазно.

Основываясь на уравнениях Максвелла (2.6)—(2.9) для средних полей в веществе, можно показать, что плотность потока энергии и в этом случае характеризуется вектором Пойнтинга (1.50), хотя выражение для закона сохранения энергии электромагнитного поля в среде имеет иной вид, чем выражение (1.49) или (1.51) для вакуума. Для волны с определенным направлением вектора $\mathbf{k}$ (т. е. при параллельных $\mathbf{k}'$ и $\mathbf{k}''$) вектор Пойнтинга направлен вдоль $\mathbf{k}$. Интенсивность (среднее по времени значение плотности потока энергии) пропорциональна квадрату амплитуды напряженности поля, и в поглощающей среде, характеризуемой комплексным показателем преломления $n+i\varkappa$, убывает вдоль направления волны по закону

$$\langle S(z)\rangle = S_0 e^{-2\varkappa\frac{\omega}{c}z} = S_0 e^{-\alpha z}. \qquad (2.29)$$

Экспоненциальное уменьшение интенсивности по мере проникновения излучения в среду, выражаемое соотношением (2.29), называют *законом Бугера*. Величина $\alpha=2\varkappa\omega/c$, называемая *коэффициентом поглощения*, характеризует быстроту этого уменьшения: на расстоянии $l=1/\alpha$ интенсивность уменьшается в e раз. В материалах,

которые мы считаем прозрачными, расстояние l («глубина проникновения») обычно много больше толщины материала.

Если мнимая часть $\varepsilon(\omega)$ отлична от нуля, затухание интенсивности излучения обусловлено поглощением энергии и определяется работой, совершаемой электрическим полем волны $\mathbf{E}$ над поляризационными токами в веществе, плотность $\mathbf{j}$ которых равна $d\mathbf{P}/dt$. Эта работа в единице объема в единицу времени равна $\mathbf{Ej}$, и ее среднее значение отлично от нуля, если у $\mathbf{j}=d\mathbf{P}/dt$ есть составляющая, изменяющаяся в фазе с $\mathbf{E}(t)$. Такая составляющая существует, когда мнимая часть восприимчивости χ (и, следовательно, мнимая часть ε) отлична от нуля.

Усиливающим (активным) средам соответствуют отрицательные значения показателя затухания $\varkappa$ и коэффициента поглощения α (см. § 2.5).

Контрольные вопросы

1. Что такое диэлектрическая восприимчивость $\chi(\omega)$?
2. Почему материальное уравнение, связывающее поляризованность $\mathbf{P}(t)$ с вызывающим ее электрическим полем $\mathbf{E}(t)$, справедливо только для монохроматических полей?
3. Какой физический смысл имеет вещественная часть диэлектрической восприимчивости $\chi'(\omega)$? мнимая часть $\chi''(\omega)$?
4. В чем заключается физическое содержание условия (2.23)?
5. При каком условии волновой вектор $\mathbf{k}$ плоской волны в среде будет вещественным? Может ли вектор $\mathbf{k}$ быть комплексным при вещественном $\varepsilon(\omega)$?
6. Как расположены плоскости постоянной фазы и плоскости постоянной амплитуды в волне, характеризуемой комплексным волновым вектором $\mathbf{k}=\mathbf{k}'+i\mathbf{k}''$? Как связаны между собой векторы электрического и магнитного полей монохроматической волны в веществе?
7. Какой физический смысл имеют вещественная и мнимая части комплексного показателя преломления?
8. По какому закону изменяется интенсивность излучения по мере его распространения в среде?
9. Что такое коэффициент поглощения и как он связан с показателем затухания?
10. При каком условии затухание волны обусловлено поглощением энергии?

2.3. Классическая электронная теория дисперсии

В § 2.2 были введены материальные уравнения, связывающие поляризованность среды $\mathbf{P}$ (или электрическую индукцию $\mathbf{D}$) с напряженностью $\mathbf{E}$ электрического поля монохроматической волны, что позволило рассмотреть свойства монохроматических плоских волн в веществе. Для обоснования феноменологических материальных уравнений и нахождения явного вида входящей в них диэлектрической восприимчивости $\chi(\omega)$ [или диэлектрической проницаемости $\varepsilon(\omega)$] в той или иной среде, т. е. для «объяс-

нения» дисперсии, необходима микроскопическая теория взаимодействия электромагнитного поля волны с веществом.

Микроскопическая теория исходит из некоторой идеализированной *модели* строения вещества. Наибольшей простотой отличается модель газообразной среды, так как для нее в первом приближении можно не учитывать взаимодействие атомов или молекул и, кроме того, считать, что действующее на отдельный атом поле совпадает со средним полем электромагнитной волны. В таких условиях для получения макроскопического материального уравнения достаточно рассмотреть действие поля волны на изолированный атом.

Входящие в состав атома электроны можно разделить на внешние, или *оптические*, и электроны внутренних оболочек. Принятая терминология обусловлена тем, что с излучением оптического диапазона взаимодействуют практически только внешние электроны. Собственные частоты электронов внутренних оболочек столь велики, что поле световой волны на них практически не влияет и лишь для рентгеновского излучения внутренние электроны становятся существенными.

В классической теории дисперсии оптический электрон в атоме рассматривается как затухающий дипольный осциллятор, характеризуемый определенной собственной частотой ω_0 и постоянной затухания γ, так что уравнение его движения в поле $\mathbf{E}(t)$ световой волны имеет вид

$$\ddot{\mathbf{r}} + 2\gamma\dot{\mathbf{r}} + \omega_0^2\mathbf{r} = -(e/m)\mathbf{E}(t), \tag{2.30}$$

где $\mathbf{r}$ — смещение электрона из положения равновесия.

Уравнение (2.30) основано на предположении о том, что к движению электрона в атоме применимы законы классической механики. С позиций современной физики это предположение совершенно неоправданно. В атомных масштабах движение электрона подчиняется законам квантовой, а не классической механики. Поэтому может сложиться впечатление, что классическая теория дисперсии лишена физического содержания и представляет лишь исторический интерес. Однако это не так. Дело в том, что для описанной выше модели гармонического осциллятора квантовая теория дисперсии приводит к таким же результатам, что и классическая. Для такой модели различие в *принципах* (классическая или квантовая теория) применительно к вопросам *дисперсии* оказывается несущественным, хотя в других вопросах (например, средняя энергия гармонического осциллятора в состоянии теплового равновесия, см. § 9.2) есть различие.

Другое дело, что сама *модель* дипольного осциллятора (восходящая к предложенной Дж. Томсоном простейшей модели атома), используемая в классической теории дисперсии, выглядит чрезмерно упрощенной в свете современных представлений о строении атома. Например, для атома водорода правильные, согласующиеся с опытом результаты получаются, когда мы считаем, что на электрон действует кулоновская сила, обратно пропорциональная квадрату расстоя-

ния до центра ($\sim 1/r^2$), а не квазиупругая, пропорциональная расстоянию ($\sim r$). Никаких квазиупругих сил и сил трения, пропорциональных скорости, в атомах нет. Строение атомов и молекул определяется кулоновскими силами взаимодействия электронов и ядер. Однако классическая физика оказалась не в состоянии объяснить на основе этих сил структуру и даже само существование атомов и молекул как устойчивых образований. Это и не удивительно, так как уравнения классической физики получены на основе макроскопического опыта и в масштабах атома оказываются за пределами своей применимости.

Правильную теорию атома дает квантовая механика. Поэтому и последовательная теория дисперсии, использующая реалистическую модель среды, должна быть квантовой. Однако ее изложение выходит за рамки данной книги. Мы вынуждены здесь ограничиться упрощенной моделью атома как гармонического осциллятора, для которой квантовая теория, как уже говорилось, приводит к тем же результатам, что и классическая. Такой подход оправдывается тем, что последовательная квантовая теория дисперсии, учитывающая реальную структуру атома, дает аналогичный результат, хотя и с некоторыми особенностями, о которых сказано ниже.

Входящая в уравнение (2.30) *собственная частота* ω_0 атомного электрона может быть рассчитана только на основе квантовой теории атома. В рамках классической теории дисперсии ее следует рассматривать как формально введенную постоянную. В эксперименте значение ω_0 определяет частоту линии поглощения в спектре исследуемого вещества (см. ниже).

Постоянная затухания γ, характеризующая в (2.30) силу «сопротивления», пропорциональную скорости электрона, содержит вклад, обусловленный *радиационным затуханием:* в классической теории осциллирующий электрон обязательно излучает. Другие причины затухания (например, взаимодействие с другими атомами и соударения) связаны с *диссипацией* энергии электромагнитного поля, т. е. с ее превращением в другие формы (в теплоту). Такое диссипативное затухание можно считать истинным поглощением и включить его вклад в константу γ. Относительная роль разных членов в уравнении (2.30) зависит от рассматриваемой области частот. Например, при частотах ω, далеких от собственной частоты ω_0 осциллятора, затуханием, как правило, можно пренебречь.

Заметим, что в уравнении (2.30) осциллятор предполагается изотропным, т. е. квазиупругая сила — $m\omega_0^2\mathbf{r}$ одинакова при смещении электрона в любом направлении.

В монохроматической волне действующее на осциллятор поле $\mathbf{E}(t)$ в правой части (2.30) изменяется со временем синусоидально:

$$\mathbf{E}(t)=\mathbf{E}_0 e^{-i\omega t}. \qquad (2.31)$$

Нас интересует *частное* решение уравнения (2.30), описывающее *установившиеся* вынужденные колебания осциллятора. Эти колеба-

ния под действием синусоидальной внешней силы также будут синусоидальными, и их частота совпадает с частотой вынуждающей силы. Поэтому решение уравнения (2.30) для смещения электрона $\mathbf{r}(t)$ можно искать в виде

$$\mathbf{r}(t)=\mathbf{r}_0 e^{-i\omega t}.$$

Амплитуду $\mathbf{r}_0$ найдем, подставляя $\mathbf{r}(t)$ и его производные в левую часть (2.30) и действующее поле $\mathbf{E}(t)$ из (2.31) в правую часть. В результате получим

$$\mathbf{r}(t)=\frac{-e/m}{\omega_0^2-\omega^2-2i\omega\gamma}\mathbf{E}(t).$$

Индуцированный действующим полем $\mathbf{E}$ дипольный момент атома $\mathbf{p}(t)$ равен $-e\mathbf{r}(t)$ и, следовательно, пропорционален напряженности:

$$\mathbf{p}(t)=\frac{e^2/m}{\omega_0^2-\omega^2-2i\omega\gamma}\mathbf{E}(t). \tag{2.32}$$

Зависящий от ω коэффициент пропорциональности между $\mathbf{p}$ и $\mathbf{E}$ в (2.32) называется *атомной поляризуемостью* $\alpha(\omega)$:

$$\mathbf{p}=\varepsilon_0\alpha(\omega)\mathbf{E}. \tag{2.33}$$

Если N — концентрация атомов вещества, то поляризованность $\mathbf{P}$ равна $N\mathbf{p}$. Ограничимся пока случаем достаточно *разреженной среды* (газы или пары), чтобы действующее на осциллятор поле $\mathbf{E}(t)$ в (2.32) и (2.33) можно было считать совпадающим со средним макроскопическим полем, которое входит в уравнения Максвелла и в соотношение (2.12), определяющее диэлектрическую восприимчивость $\chi(\omega)$. Подставляя в (2.12) $\mathbf{P}=N\mathbf{p}$ с $\mathbf{p}$ из (2.33) и отождествляя $\mathbf{E}$ из (2.33) с $\mathbf{E}_0 e^{-i\omega t}$ в (2.12), находим восприимчивость:

$$\chi(\omega)=N\alpha(\omega)=\frac{Ne^2/(m\varepsilon_0)}{\omega_0^2-\omega^2-2i\omega\gamma}. \tag{2.34}$$

Диэлектрическая проницаемость $\varepsilon(\omega)$ связана с $\chi(\omega)$ соотношением (2.14). Подставляя в него $\chi(\omega)$ из (2.34), получаем $\varepsilon(\omega)$ для рассматриваемой модели разреженной среды, содержащей N осцилляторов в 1 м^3,

$$\varepsilon(\omega)=1+\frac{Ne^2/(m\varepsilon_0)}{\omega_0^2-\omega^2-2i\omega\gamma}. \tag{2.35}$$

Для упрощения записи дальнейших формул удобно ввести характеризующую модель среды константу ω_p согласно следующему определению:

$$\omega_p^2=Ne^2/(m\varepsilon_0). \tag{2.36}$$

Она имеет размерность частоты. Теперь $\varepsilon(\omega)$ из (2.35) можно записать в виде

$$\varepsilon(\omega)=1+\omega_p^2/(\omega_0^2-\omega^2-2i\omega\gamma). \tag{2.37}$$

Вследствие затухания атомных осцилляторов диэлектрическая

проницаемость оказывается комплексной. Выделяя в (2.37) вещественную и мнимую части $\varepsilon'+i\varepsilon''$, можно получить выражения для зависящих от частоты показателя преломления $n(\omega)$ и показателя затухания $\varkappa(\omega)$. Эти выражения весьма громоздки, поэтому анализ физических результатов электронной теории дисперсии проведен ниже для сравнительно простых частных случаев.

Контрольные вопросы

1. Какая модель среды рассматривается в классической теории дисперсии?
2. Поясните физический смысл отдельных членов в уравнении (2.30).

2.4. Дисперсия вдали от линии поглощения

При частотах, далеких от собственной частоты ω_0 атомных осцилляторов, где выполняется условие $2\omega\gamma \ll |\omega_0^2 - \omega^2|$, мнимой частью в (2.37) можно пренебречь. Тогда для зависимости показателя преломления от частоты получаем следующую приближенную формулу:

$$n^2(\omega) = \varepsilon(\omega) = 1 + \omega_p^2/(\omega_0^2 - \omega^2). \tag{2.38}$$

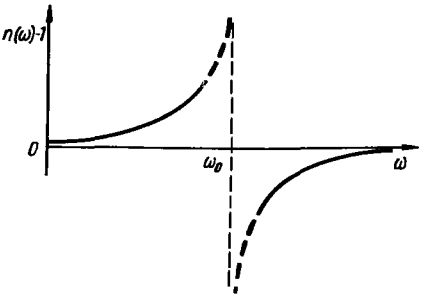
При достаточно малой концентрации N (газы или пары) вдали от собственной частоты ω_0 (когда $\omega_p^2 \ll |\omega_0^2 - \omega^2|$) показатель преломления n близок к единице, т. е. второй член в (2.38) мал по сравнению с первым. Применяя приближенную формулу $\sqrt{1+x} \approx 1 + x/2$ (при $x \ll 1$), из (2.38) получаем

$$n(\omega) \approx 1 + \frac{Ne^2}{2m\varepsilon_0} \frac{1}{\omega_0^2 - \omega^2}. \tag{2.39}$$

Описываемая этой формулой зависимость показателя преломления от частоты показана на рис. 2.1. При тех значениях частоты, где формула (2.39) применима, показатель преломления, как видно из рис. 2.1, возрастает с увеличением частоты. Такой характер зависимости $n(\omega)$ называют *нормальной дисперсией*.

Для низких частот ($\omega < \omega_0$) показатель преломления [см. (2.39)] больше единицы, т. е. фазовая скорость c/n волны в среде меньше скорости света в пустоте. Это значит, что измененная средой волна отстает по фазе от падающей. Если же частота света больше собственной частоты осцилляторов ($\omega > \omega_0$), то $n < 1$ и фазовая скорость волны в среде $v = c/n$ оказывается больше скорости света в вакууме, т. е. измененная волна по фазе опережает падающую. Никакого противоречия с теорией относительности здесь нет. Теория относительности утверждает, что скорость материальных тел и *скорость сигнала* не могут превышать c. Понятие показателя преломления применимо к *монохроматической* волне, имеющей бесконечную протяженность в пространстве и во времени, т. е. к *уста-*

новившимся выиужденным колебаниям осцилляторов среды. Монохроматическая волна не может служить для передачи сигнала. Сигнал равносилен неустановившемуся процессу, и из одного факта $n<1$ ничего нельзя заключить о скорости сигнала (см. § 2.11). Тот факт, что при $\omega>\omega_0$ установившаяся волна в среде по фазе опережает падающую волну, удивителен не более, чем то, что маятник при установившихся вынужденных колебаниях движется противоположно действующей на него силе, если частота этой силы больше его собственной.



2.1

Зависимость показателя преломления разрежеиной среды от частоты ω вдали от собственной частоты атомных осцилляторов ω_0

Многие вещества имеют собственные частоты ω_0 (и связанные с ними полосы поглощения) в далекой ультрафиолетовой части спектра. Поэтому для частот ω во всей видимой области справедливо неравенство $\omega\ll\omega_0$ (начальный участок левой ветви кривой на рис. 2.1). Рассматривая $(\omega/\omega_0)^2$ как малый параметр, преобразуем правую часть (2.38) следующим образом:

$$n^2(\omega)=1+\frac{\omega_p^2}{\omega_0^2}\left(1+\frac{\omega^2}{\omega_0^2}\right). \qquad (2.40)$$

Переходя в (2.40) от частоты ω к длине волны (в вакууме) $\lambda=2\pi c/\omega$, получаем простую формулу для $n^2(\lambda)$, удобную для сравнения с опытными данными:

$$n^2(\lambda)=1+A(1+B/\lambda^2), \qquad (2.41)$$

где

$$A=\omega_p^2/\omega_0^2, \qquad (2.42) \qquad B=4\pi^2c^2/\omega_0^2. \qquad (2.43)$$

Выражаемая формулой (2.41) зависимость показателя преломления от длины волны (с некоторыми эмпирическими константами A и B) была предсказана Френелем и Коши задолго до появления электронной теории дисперсии. Во многих случаях она дает удовлетворительное описание экспериментальных данных. Сравнение теоретической зависимости (2.41) с экспериментально наблюдаемой позволяет определить значения констант A и B для конкретной среды. При этом появляется возможность проверки электронной теории дисперсии, так как константы A и B можно оценить по (2.42) и (2.43). Для такой оценки нужно знать концентрацию N атомов и собственную частоту ω_0. В тех случаях, когда сведения о частоте ω_0 отсутствуют, можно оценить отношение B/A, которое [см. (2.42), (2.43)] не зависит от ω_0, и полученную оценку сравнить с опытными дан-

ными. Подобные оценки во многих случаях дают удовлетворительное совпадение (в пределах 10—20%) с экспериментальными значениями.

Формула (2.38), как уже отмечалось, справедлива для разреженной среды, так как при ее выводе предполагалось, что действующее на отдельный атом поле совпадает со средним макроскопическим полем **E**. В *плотном веществе* (жидкости, твердые тела) это предположение несправедливо, так как напряженности полей, создаваемых дипольными моментами соседних атомов, индуцированными внешним полем, могут быть одного порядка с **E**. Напряженность этого дополнительного поля можно приближенно найти, представив себе, что рассматриваемый атом находится в центре сферической полости, вырезанной в окружающем его диэлектрике. Аналогичная задача встречается в электростатике диэлектриков. Можно показать, что *локальное поле* $\mathbf{E}_{\text{лок}}$ в сферической полости, вырезанной в поляризованном диэлектрике, связано со средним полем **E** и поляризованностью **P** соотношением

$$\mathbf{E}_{\text{лок}} = \mathbf{E} + \mathbf{P}/(3\varepsilon_0). \tag{2.44}$$

Строго говоря, напряженность поля, действующего на атом, совпадает с напряженностью поля в полости (2.44) только в случае изотропного материала или кубического кристалла. Хотя соотношение (2.44) получено в электростатике, его можно использовать и для электрического поля электромагнитной волны, пока длина ее много больше размеров полости, т. е. среднего расстояния между соседними атомами. Считая, что входящее в (2.33) поле, действующее на рассматриваемый атом, совпадает с локальным полем (2.44), для зависимости показателя преломления от частоты мы придем вместо (2.38) к формуле, полученной Г. Лоренцом и Л. Лоренцем:

$$(n^2-1)/(n^2+2) = \omega_p^2/[3(\omega_0^2-\omega^2)]. \tag{2.45}$$

Для разреженной среды, показатель преломления которой близок к единице, в знаменателе левой части можно положить $n^2+2 \approx 3$, и (2.45) сводится к выражению (2.38). В конденсированной среде помимо учтенного в (2.45) отличия действующего на отдельный атом поля от среднего существенны и другие усложнения, обусловленные тесным расположением атомов. Из-за взаимодействия соседних атомов собственные частоты ω_0 атомных осцилляций оказываются сильно сдвинутыми и «размазанными», а постоянная затухания γ возрастает, т. е. значения ω_0 и γ оказываются иными, чем у свободных атомов.

Правая часть формулы (2.45) пропорциональна концентрации N атомов, а следовательно, и плотности вещества ρ. Поэтому величина

$$\frac{1}{\rho}\frac{n^2-1}{n^2+2}, \tag{2.46}$$

называемая *удельной рефракцией,* для данного вещества не должна изменяться при изменении плотности. Во многих случаях это условие хорошо выполняется. Например, для воздуха при увеличении его давления от нормального атмосферного в 200 раз удельная рефракция (на фиксированной частоте) остается постоянной с точностью около 10^{-3}. Иногда удельная рефракция почти не изменяется даже при переходе вещества из газообразного состояния в жидкое. Так, при ожижении кислорода удельная рефракция изменяется менее чем на 1%; при конденсации водяного пара — всего на 3%, хотя плотность ρ изменяется при этом на три порядка. Имеется, однако, немало случаев, когда постоянство удельной рефракции не соблюдается.

Контрольные вопросы

1. Фазовая скорость электромагиитных волн в веществе при $n<1$ превышает скорость света в вакууме. Почему это не противоречит теории относительности?
2. Как изменяется показатель преломления данного вещества при изменении его плотности?

2.5. Аномальная дисперсия

Вернемся к выражению (2.37) для диэлектрической проницаемости $\varepsilon(\omega)$, справедливому для достаточно разреженной среды, и исследуем теперь на его основе дисперсию *вблизи собственной частоты* ω_0 атомного осциллятора. В этом случае (при $\omega \approx \omega_0$) пренебрегать членом с постоянной затухания γ в (2.37) нельзя и $\varepsilon(\omega)$ оказывается комплексной.

Показатель преломления $\sqrt{\varepsilon(\omega)}=n+i\varkappa$ также будет комплексным. Для нахождения его вещественной (n) и мнимой ($\varkappa$) частей приравняем $n^2-\varkappa^2$ вещественной части $\varepsilon(\omega)$ из (2.37), а $2n\varkappa$ — мнимой части $\varepsilon(\omega)$:

$$n^2-\varkappa^2=1+\frac{\omega_p^2(\omega_0^2-\omega^2)}{(\omega_0^2-\omega^2)^2+4\omega^2\gamma^2}, \quad n\varkappa=\frac{\omega_p^2\omega\gamma}{(\omega_0^2-\omega^2)^2+4\omega^2\gamma^2}. \tag{2.47}$$

При частотах ω, близких к собственной частоте ω_0, в формулах (2.47) возможны упрощения: при $\omega \approx \omega_0$ всюду, кроме $(\omega_0^2-\omega^2)$, можно заменить ω на ω_0, а $(\omega_0^2-\omega^2)$ преобразуется следующим образом:

$$\omega_0^2-\omega^2=(\omega_0+\omega)(\omega_0-\omega)\approx 2\omega_0\Delta\omega, \tag{2.48}$$

где $\Delta\omega=\omega_0-\omega$. В результате формулы (2.47) принимают вид

$$n^2-\varkappa^2\approx 1+\frac{\omega_p^2}{2\omega_0}\frac{\Delta\omega}{(\Delta\omega)^2+\gamma^2}, \quad n\varkappa=\frac{\omega_p^2}{4\omega_0}\frac{\gamma}{(\Delta\omega)^2+\gamma^2}. \tag{2.49}$$

Сделаем, наконец, еще одно упрощающее предположение, которое позволит получить сравнительно простые выражения для $n(\omega)$

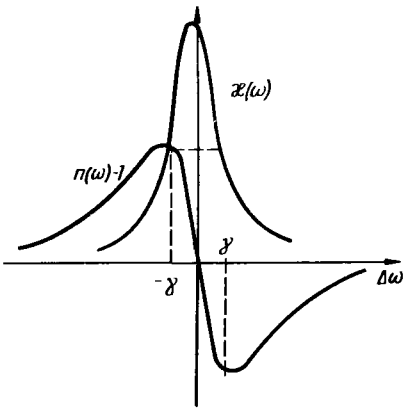
и $\varkappa(\omega)$, а именно: будем считать среду настолько разреженной, что максимальное значение $n\varkappa$ в (2.49), которое достигается при $\Delta\omega=0$, мало́ по сравнению с единицей: $\omega_p^2/(4\omega_0\gamma)\ll 1$. Тогда и максимальное значение второго слагаемого в выражении для $n^2-\varkappa^2$ много меньше единицы. Пренебрегая членами, квадратичными по малому параметру $\omega_p^2/(4\omega_0\gamma)$, из (2.49) находим

$$n(\omega)\approx 1+\frac{\omega_p^2}{4\omega_0}\,\frac{\Delta\omega}{(\Delta\omega)^2+\gamma^2}, \qquad \varkappa(\omega)\approx \frac{\omega_p^2}{4\omega_0}\,\frac{\gamma}{(\Delta\omega)^2+\gamma^2}. \tag{2.50}$$

Благодаря простому виду этих формул мы легко можем до конца проанализировать их физическое содержание. И хотя они имеют ограниченную применимость из-за сделанных при их выводе предположений, такой анализ позволит и для общего случая составить качественное представление о характере зависимости от частоты показателей преломления и затухания вблизи линии поглощения.

Графики функций $n(\omega)-1$ и $\varkappa(\omega)$ приведены на рис. 2.2. Зависимость показателя затухания от частоты характеризует *спектральный контур линии поглощения*. В рассматриваемой модели среды, состоящей из неподвижных затухающих атомных осцилляторов, этот контур (как и спектральный контур линии излучения, см. § 1.8) имеет *лоренцевскую* форму с максимумом при $\Delta\omega=0$, т. е. при $\omega=\omega_0$. Ширина максимума на половине высоты равна 2γ, она растет с увеличением постоянной затухания.

График зависимости показателя преломления от частоты называют *дисперсионной кривой*. Показатель преломления сначала растет с увеличением частоты, достигая максимального значения при $\Delta\omega=-\gamma$, т. е. при $\omega=\omega_0-\gamma$. Затем $n(\omega)$ *убывает* при возрастании частоты и после перехода через центр линии поглощения становится меньше единицы. Минимум $n(\omega)$ расположен при $\Delta\omega=\gamma$ (т. е. при $\omega=\omega_0+\gamma$) При дальнейшем увеличении частоты показатель преломления возрастает, асимптотически приближаясь к 1.



2.2
Дисперсионная кривая и контур поглощения

Дисперсионную кривую на рис. 2.2 можно сравнить с графиком зависимости $n(\omega)$ (см. рис. 2.1) для частот, далеких от собственной частоты ω_0. Мы видим, что полученная при $\gamma=0$ кривая с разрывом в точке $\omega=\omega_0$ трансформировалась при учете поглощения в непрерывную кривую с падающим участком вблизи ω_0.

Когда частота света приближается к ω_0, основной эффект обусловлен мнимой частью показателя

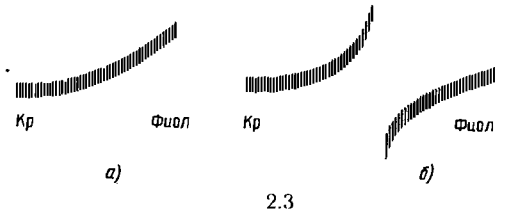
преломления $\varkappa(\omega)$, что связано с *поглощением* света (абсорбцией). Такое избирательное поглощение приводит, например, к появлению темных линий в спектре солнечного излучения. Свет, излучаемый поверхностью Солнца, имеет сплошной (непрерывный) спектр. При прохождении сквозь солнечную атмосферу излучение сильно поглощается на частотах, равных собственным частотам входящих в состав атмосферы атомов. Измеряя эти частоты, можно определить химический состав солнечной атмосферы. Такими методами установлено, что химические элементы на Солнце и звездах не отличаются от земных.

Уменьшение показателя преломления с частотой, которое происходит в пределах ширины спектрального контура линии поглощения, называют *аномальной дисперсией*. Экспериментально она была открыта Леру в 1860 г. в опытах по преломлению белого света призмой, наполненной парами иода. Оказалось, что синие лучи преломляются меньше красных, т. е. показатель преломления убывает с частотой.

Систематические экспериментальные исследования аномальной дисперсии были выполнены Кундтом, установившим, в частности, что аномальный ход дисперсии всегда сопровождается *поглощением*.

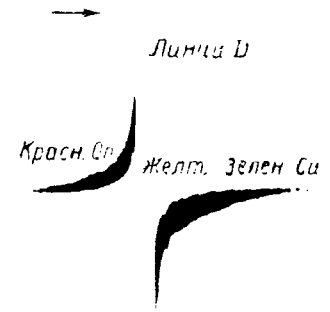
Происхождение названия «аномальная дисперсия» связано с тем, что у всех прозрачных веществ, которые были к тому времени изучены, показатель преломления возрастает с частотой, т. е. дисперсия «нормальная», получившая теоретическое объяснение в работах Френеля и Коши. После создания электронной теории дисперсии стало ясно, что аномальная дисперсия должна наблюдаться у всех веществ в тех областях спектра, где имеется сильное поглощение.

Для изучения аномальной дисперсии Кундт использовал метод *скрещенных призм*, впервые примененный еще Ньютоном. Одна из призм (с вертикальным ребром) стеклянная и обладает нормальной дисперсией. Она разлагает проходящий через нее узкий пучок белого света в цветную горизонтальную полоску («спектр»). Вторая призма (с горизонтальным ребром) изготовлена из исследуемого вещества. Она смещает каждую точку цветной полоски по вертикали. Показатель преломления и, следовательно, смещение по вертикали зависят от длины волны. Поэтому цветная полоска становится наклонной и искривляется, воспроизводя характер зависимости показателя преломления второй призмы от длины волны. Если дисперсия материала второй призмы нормальная, то полоска монотонно поднимается от



2.3

Наблюдаемая методом скрещенных призм картина в случае нормальной (*а*) и аномальной (*б*) дисперсии



2.4
Дисперсия вблизи желтой линии поглощения натрия, наблюдаемая в опыте Кундта—Вуда

красного конца спектра к фиолетовому (рис. 2.3, *а*). В случае аномальной дисперсии цветная полоска в области поглощения разрывается, а ее края искривляются в противоположные стороны (рис. 2.3, *б*) в соответствии с характером дисперсионной кривой $n(\omega)$ на рис. 2.2. Падающий участок дисперсионной кривой наблюдать трудно из-за сильного поглощения.

В эффектном демонстрационном опыте Кундта, иллюстрирующем зависимость показателя преломления от частоты в парах натрия вблизи линии поглощения, роль второй призмы играло конусообразное пламя газовой горелки, в которое вводился натрий. В результате создаваемый первой (стеклянной) призмой спектр разрывался в месте расположения желтой *D*-линии поглощения паров натрия и изгибался, как на рис. 2.3, *б*.

Этот опыт был усовершенствован Вудом, использовавшим в качестве второй призмы горизонтальную кювету с парами натрия, плотность которых уменьшается по высоте. Для создания неравномерной по высоте плотности паров верхняя часть кюветы охлаждается, а нижняя, где лежит кусочек металлического натрия, подогревается горелкой или электрической спиралью. Такой столб паров натрия действует на проходящий пучок света как призма с горизонтальным ребром, вызывая отклонение по вертикали, а стеклянная призма с вертикальным ребром разлагает пучок в горизонтальный спектр. Наблюдаемая на экране зависимость показателя преломления от частоты вблизи *D*-линии поглощения натрия терпит разрыв (рис. 2.4).

Количественные измерения дисперсии для проверки теории лучше производить на разреженных газах и парах металлов, так как тогда имеется возможность работать как можно ближе к центру линии поглощения. При малой плотности паров изменения показателя преломления малы, и для их измерения наилучшим является интерференционный метод «крюков», предложенный Д. С. Рождественским (см. § 5.6).

Классическая дисперсионная формула (2.35) и ее предельный случай (2.50) для разреженных газов с большой точностью описывают наблюдаемую на опыте зависимость показателя преломления $n(\omega)$ и показателя затухания $\varkappa(\omega)$ от частоты вблизи отдельных линий поглощения. Опыт показывает, что даже для одноатомных газов, атомы которых имеют только *один* оптический электрон, существует *несколько* линий поглощения. Формулу (2.35) можно обобщить так, чтобы она описывала ход показателя преломления в широкой

области спектра, содержащей несколько спектральных линий на частотах ω_{0k}:

$$n^2(\omega)=1+\frac{Ne^2}{m\varepsilon_0}\sum_k \frac{f_k}{\omega_{0k}^2-\omega^2-2i\omega\gamma_k}. \qquad (2.51)$$

Величины f_k, называемые *силами осцилляторов*, удовлетворяют правилу сумм $\sum_k f_k = 1$.

К такой же дисперсионной формуле приводит и квантовая теория. Однако в квантовой теории собственные частоты ω_{0k} уже не рассматриваются как эмпирические постоянные, определяемые из самой кривой дисперсии (т. е. из фактического положения спектральных линий), а приобретают вполне определенный физический смысл. При отсутствии внешних полей атом имеет некоторый набор стационарных состояний, в которых его энергия принимает дискретные значения $E_0, E_1, E_2, \ldots, E_k, \ldots$. Эти уровни энергии могут быть рассчитаны методами квантовой механики. При переходе атома из одного состояния в другое происходит испускание (или поглощение) света с частотой, определяемой правилами Бора:

$$\hbar\omega_{ik}=E_k-E_i.$$

Если все атомы среды находятся в основном состоянии, которому соответствует наинизшая энергия E_0, то в дисперсионную формулу входят только слагаемые с частотами ω_{0k}, соответствующими переходам атома из основного состояния в возбуждеиные. Вклад каждого возбужденного состояния в атомную поляризуемость определяется силой осциллятора f_k. Сила осциллятора f_k пропорциональна вероятности спонтанного перехода из соответствующего возбужденного состояния в основное.

При определенных условиях (например, при электрическом разряде в газе) часть атомов среды находится в возбужденных состояниях. Дисперсионная формула тогда содержит резонансные члены с частотами $\omega_{ik}=(E_k-E_i)/\hbar$:

$$n^2(\omega)=1+\frac{e^2}{m\varepsilon_0}\sum_{i,k}\frac{N_i f_{ik}}{\omega_{ik}^2-\omega^2-2i\omega\gamma_{ik}}, \qquad (2.52)$$

где $f_{ik}=-f_{ki}$ — сила осциллятора для перехода между уровнями энергии E_i и E_k; N_i — концентрация атомов, находящихся в состоянии с энергией E_i; $\Sigma_i N_i = N$, где N — полная концентрация атомов.

Особый интерес представляет неравиовесный случай, когда для какой-либо пары уровней с энергиями E_i и $E_k (E_i > E_k)$ выполняется условие $N_i > N_k$. Такая *инверсная населенность* создается с помощью специальных средств в активной среде оптических квантовых усилителей и генераторов — лазеров (см. § 9.4). Сила осциллятора f_{ik} положительна при $i<k$ и отрицательна при $i>k$. Поэтому в случае инверсной населенности ($N_i > N_k$) дисперсия в окрестности частоты $\omega \approx \omega_{ik}$ (соответствующей переходу между этой парой уровней), определяемая двумя слагаемыми с N_i и N_k, *отрицатель-*

ная. Отрицательным будет и показатель затухания $\varkappa(\omega)$, что соответствует *усилению* волн с частотами $\omega \approx \omega_{ik}$ вследствие преобладания *вынужденного испускания* над поглощением. Отрицательные дисперсия и поглощение не находят объяснения в рамках классической теории дисперсии.

Контрольные вопросы

1. В каких случаях дисперсию называют нормальной и в каких — аномальной?
2. Спектральный контур линии поглощения $\varkappa(\omega)$ в формуле (2.51) имеет лоренцевскую форму. С какими свойствами принятой при ее выводе модели среды это связано?
3. Каким будет контур $\varkappa(\omega)$, если допустить, что атомы рассматриваемой разреженной среды совершают хаотическое тепловое движение?

2.6. Дисперсия в металлах и плазме. Показатель преломления рентгеновских лучей

В металлах некоторые из электронов не связаны с каким-либо определенным атомом; это «свободные» электроны, ответственные за электрическую проводимость металла. В отличие от рассмотренных выше оптических электронов в атомах диэлектрика на свободные электроны не действует «квазиупругая» сила, привязывающая их к какому-то отдельному атому, но сила «трения» характеризующая сопротивление движению электрона, остается. Поэтому уравнение (2.30) классической теории дисперсии и все следствия из него можно применить к свободным электронам, положив обусловленную квазиупругой силой собственную частоту ω_0 равной нулю.

Электроны проводимости участвуют в тепловом движении и все время изменяют свое положение. В результате оказывается, что действующее на них электрическое поле в среднем как раз равно макроскопическому полю **E**. Следовательно, мы должны взять формулу (2.37), полученную для разреженной среды (без поправки на отличие локального поля от среднего), и положить в ней $\omega_0 = 0$:

$$\varepsilon(\omega) = n^2 = 1 - \omega_p^2/[\omega(\omega + 2i\gamma)]. \qquad (2.53)$$

Константа ω_p (2.36) зависит от концентрации N свободных электронов и называется в данном случае плазменной частотой*. Постоянную затухания γ в (2.53) можно оценить, выразив ее через удельную проводимость металла для постоянного тока (см. задачу).

Формула (2.53) для показателя преломления в металлах предсказывает совершенно разный характер распространения волн в областях низких и высоких частот. При *низких* частотах, когда $\omega \ll \gamma$

* Название связано с тем, что ω_p определяет частоту собственных коллективных колебаний в плазме, концентрация электронов в которой N.

(для меди это соответствует электромагнитным волнам длиной порядка 1 мм и более), формула (2.53) приводит к комплексному показателю преломления с одинаковыми вещественной и мнимой частями $n=\varkappa\gg 1$. Такие волны проникают в глубь металла на расстояние, которое много меньше длины волны в вакууме (скин-эффект). Коэффициент отражения R (см. § 3.4) для них близок к единице, т. е. они практически полностью отражаются от поверхности.

В противоположном случае *высоких* частот, удовлетворяющих неравенству $\omega\gg\gamma$, в формуле (2.53) можно пренебречь мнимым слагаемым $2i\gamma$ по сравнению с ω и для диэлектрической проницаемости получается вещественное выражение

$$\varepsilon(\omega)=n^2(\omega)=1-\omega_p^2/\omega^2. \qquad (2.54)$$

Такой же результат получается и непосредственно из уравнения (2.30) при $\omega_0=0$, если в нем пренебречь диссипативным членом $2\gamma\dot{\mathbf{r}}$ по сравнению с инерционным $\ddot{\mathbf{r}}$. При высоких частотах характер дисперсионных явлений в металлах обусловлен инерцией свободных электронов: за промежуток времени между двумя актами рассеяния, который в среднем равен $\tau=1/(2\gamma)$, электрон успевает совершить много вынужденных колебаний, так как при $\omega\gg\gamma$ их период $T\ll\tau$.

Из формулы (2.54) видно, что плазменная частота ω_p имеет смысл своего рода критической частоты. При $\omega<\omega_p$ диэлектрическая проницаемость отрицательна, а показатель преломления чисто мнимый. Это значит, что волны с $\omega<\omega_p$ (но $\omega\gg\gamma$) не могут распространяться в металле из-за сильного затухания, причем это затухание не связано с поглощением (т. е. диссипацией) энергии. В самом деле, диэлектрическая проницаемость вещественна (а истинное поглощение происходит только при $\operatorname{Im}\varepsilon\neq 0$), да и выражение (2.54) для $\varepsilon(\omega)$ получается при пренебрежении диссипативным членом в уравнении движения электрона. Фактически при $\omega<\omega_p$ происходит *полное отражение* падающей волны от среды. При чисто мнимом показателе преломления коэффициент отражения равен единице (см. § 3.4).

При $\omega>\omega_p$ показатель преломления становится вещественным, а металл — прозрачным для излучения. Обычно плазменная частота у металлов попадает в область рентгеновских лучей, но для некоторых металлов область прозрачности начинается с ультрафиолетовых лучей. Например, у натрия длина волны, соответствующая граничной частоте ω_p, составляет 210 нм, что хорошо согласуется с теоретической оценкой ω_p по формуле (2.36) на основе известной концентрации N свободных электронов. Прозрачность щелочных металлов в ультрафиолетовой области спектра была обнаружена на опыте Вудом в 1933 г.

Для промежуточных частот ($\omega\approx\gamma$) нужно пользоваться полным выражением (2.53), а не его предельными формами. В этом случае у показателя преломления отличны от нуля зависящие от частоты вещественная и мнимая части. Это значит, что волны разных частот

при распространении в металле по-разному затухают. Очень тонкие слои металла прозрачны даже для видимого света. Например, тонкий слой золота, полученный напылением в вакууме на стеклянную подложку, пропускает видимый свет, но сильно поглощает инфракрасное излучение. Экспериментальные методы определения оптических констант металлов основаны на исследовании поляризации отраженного света (см. § 3.4).

Уравнения (2.53) или (2.54), описывающие дисперсию электромагнитных волн в среде со свободными электронами, в равной мере применимы к электронам проводимости в металлах и к свободным электронам в плазме, например в ионосферной плазме. Полученные выше выражения (при надлежащих значениях N и γ) можно использовать для объяснения характера распространения радиоволн в ионосфере Земли. Граничная частота здесь попадает в радиодиапазон, поэтому волны длиной порядка 10 м и более отражаются ионосферой, что широко используется для радиосвязи, тогда как ультракороткие (УКВ) свободно проходят сквозь нее. Это обстоятельство открывает возможность радиолокации Луны и планет и жизненно важно для радиоастрономии, использующей технику ультракоротких волн. Исследование частотной зависимости отражения радиоволн дает хороший метод изучения ионосферы, в частности определения N по критической частоте.

В случае очень высоких частот ($\omega \to \infty$) диэлектрическая проницаемость $\varepsilon(\omega)$ любого вещества стремится к единице: при очень быстрых изменениях напряженности поля процессы поляризации не успевают происходить. Предельный вид функции $\varepsilon(\omega)$ при больших частотах, справедливый для любых тел (безразлично — металлов или диэлектриков), можно установить, рассматривая электроны вещества как свободные, пренебрегая их взаимодействием друг с другом и с ядрами атомов. Для этого частота ω поля должна быть велика по сравнению с собственными частотами ω_0 электронов в атомах данного вещества. Пренебрегая ω_0 по сравнению с ω, для $\varepsilon(\omega)$ получаем из (2.38) такое же выражение (2.54), как и в металлах:

$$\varepsilon(\omega) = n^2(\omega) = 1 - \omega_p^2/\omega^2, \qquad (2.55)$$

с той разницей, что в $\omega_p^2 = Ne^2/(m\varepsilon_0)$ под N нужно понимать не концентрацию электронов проводимости, а *полное* число электронов в единице объема вещества. Область применимости этой формулы начинается от далекого ультрафиолета у самых легких элементов и от рентгеновских частот у более тяжелых элементов.

При частотах, соответствующих рентгеновскому излучению, перестает выполняться условие $\lambda \gg a$ (a — среднее расстояние между атомами среды). Поэтому, строго говоря, макроскопическое описание поля здесь неприменимо и среду нельзя рассматривать как непрерывную. Нужно исходить из рассеяния рентгеновского излучения на отдельных электронах, распределенных в пространстве с некоторой плотностью $N(x, y, z)$. В кристаллах эта функция координат будет трехмерно периодической, отражая упорядочение расположение атомов в узлах кристаллической решетки.

Когда длина волны меньше пространственного периода решетки, при определенных условиях возможно появление волн, распространяющихся в направлениях, сильно отличающихся от направления падающей волны. Это явление подобно образованию дифракционных максимумов при падении света на оптическую дифракционную решетку (см. § 6.5). Однако если интересоваться распространением рентгеновского излучения в веществе в направлении, близком к направлению падающей волны, то зависимость плотности числа электронов $N(x, y, z)$ от координат становится несущественной и вместо нее можно рассматривать усредненную по объему величину N — полную концентрацию электронов. Поэтому для преломления на малые углы, несмотря на нарушение условия $\lambda \gg a$, диэлектрическая проницаемость $\varepsilon(\omega)$ и показатель преломления $n(\omega)$ сохраняют свой обычный смысл и для рентгеновского излучения.

Из приведенной формулы видно, что показатель преломления рентгеновских лучей меньше единицы, хотя и очень мало отличается от нее. Его можно измерить, наблюдая предельный угол полного отражения рентгеновских лучей при переходе из воздуха в среду. Для $\lambda \approx 0{,}1$ нм в стекле $n = 1 - 5 \cdot 10^{-6}$.

Широкое применение рентгеновских лучей в медицине и в технике основано именно на том, что показатель преломления для них практически не отличается от единицы. Глубина проникновения рентгеновских лучей в металлах больше, чем для видимого света, но во многих других веществах она даже отдаленно не приближается к тем громадным глубинам проникновения, которых можно достичь в видимой или инфракрасной области. Прозрачная для видимого света атмосфера Земли полностью поглощает приходящее из космоса рентгеновское излучение (рентгеновская астрономия стала возможной только при выведении телескопов на спутниках за пределы атмосферы). Аналогично обстоит дело и в таких средах, как вода и стекло. Но видимый свет, для которого показатели преломления этих сред имеют значения около 1,5, чрезвычайно чувствителен к внутренним граничным поверхностям. В таких неоднородных средах, как, например, мышцы и другие ткани организма, происходит диффузное отражение света на многочисленных граничных поверхностях, разделяющих отдельные области, что делает эти среды непрозрачными для видимого света. Рентгеновские лучи, для которых во всех средах $n \approx 1$, как бы не замечают этих граничных поверхностей. Поэтому шапка мыльной пены совершенно не прозрачна для видимого света (дает на экране черную тень) и полностью прозрачна для рентгеновских лучей.

Контрольные вопросы

1. При каких частотах ω электромагнитные волны могут распространяться в металлах?
   Какой предельный вид имеет функция $\varepsilon(\omega)$ при высоких частотах?
   Каким свойством рентгеновских лучей обусловлено их широкое применение в медицине?

Задача

Выразить постоянную затухания γ в уравнении (2.30) через удельную проводимость σ.

Основной механизм, который дает вклад в испытываемое электроном проводнмости сопротивление движению, — это рассеянне на отклоненнях действующего на него поля в металле от идеальной пространственной периодичности. Эти отклонения обусловлены дефектами кристаллической структуры (примеси, вакансии, дислокации, граннцы зерен) и тепловымн колебаниями ионов. Все эти факторы существенны не только при осцнллирующем движении электрона в поле световой волны, но и при движении в постоянном электрическом поле, где они обеспечивают постоянную скорость упорядоченного движения (дрейфа) электронов в направлении поля, накладывающегося на хаотическое тепловое движение.

Скорость дрейфа $\mathbf{v}=\dot{\mathbf{r}}$ можно найти из уравнения (2.30), считая в нем $\omega_0=0$ и $\mathbf{E}$ в правой части постоянной. При установившемся движении в постоянном поле $\dot{\mathbf{v}}=\ddot{\mathbf{r}}=0$, поэтому

$$\mathbf{v}=-[e/(2\gamma m)]\,\mathbf{E}. \tag{2.56}$$

Каждый электрон проводимости дает в ток вклад $-e\mathbf{v}$. Полная плотность тока $\mathbf{j}=-N e\mathbf{v}$. Учнтывая (2.56), получаем

$$\mathbf{j}=\frac{Ne^2}{2\gamma m}\mathbf{E}. \tag{2.57}$$

В (2. 57) легко узнать дифференциальную форму закона Ома: $\mathbf{j}=\sigma\mathbf{E}$. Сравнивая эти два выражения, получаем

$$\gamma=\frac{Ne^2}{2m\sigma}. \tag{2.58}$$

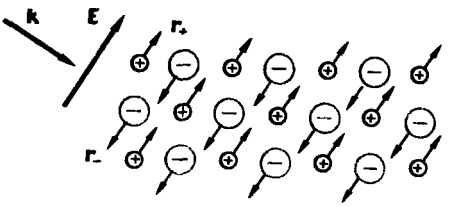
Поэтому постоянная затухания γ может быть выражена через удельную проводимость σ, значение которой для каждого металла дают измерения на постоянном токе.

2.7. Дисперсия в ионных кристаллах. Ориентационная дисперсия

Рассмотрим вклад ионов в дисперсию диэлектрической проницаемости (или показателя преломления) на примере типичного ионного кристалла — каменной соли NaCl. Этот кристалл имеет кубическую решетку, образуемую однозарядными положительными ионами натрия и отрицательными ионами хлора. Под действием внешнего электрического поля ионы Na^+ и Cl^- смещаются из своих равновесных положений в узлах кристаллической решетки в противоположные стороны. В результате каждый элемент объема среды приобретает дипольный момент. Поэтому в ионных кристаллах (а также любом веществе, содержащем молекулы с ионной связью) вклад в поляризованность $\mathbf{P}$ дают не только оптические электроны, но и сами ионы, смещаясь как целое из своих равновесных положений.

Длина волны излучения оптического диапазона ($\lambda\approx 10^{-5}$ см) много больше постоянной решетки ($a\approx 10^{-8}$ см), поэтому в пределах элемента объема, содержащего макроскопически большое число ионов, поле волны можно считать однородным. Тогда смещения $\mathbf{r}_+$ всех положительных ионов из своих равновесных положений одинаковы (рис. 2.5). То же самое можно сказать и о смещениях отрицательных ионов $\mathbf{r}_-$. Действующая на каждый ион

возвращающая сила пропорциональна относительному смещению положительной и отрицательной подрешеток: $\mathbf{F}_+ = -\beta(\mathbf{r}_+ - \mathbf{r}_-)$, $\mathbf{F}_- = -\mathbf{F}_+$. Уравнения движения ионов (без учета затухания) можно записать в виде



2.5
Смещение ионов кристалла в поле электромагнитной волны

$$M_+\ddot{\mathbf{r}}_+ = -\beta(\mathbf{r}_+ - \mathbf{r}_-) + e\mathbf{E}, \tag{2.59}$$

$$M_-\ddot{\mathbf{r}}_- = \beta(\mathbf{r}_+ - \mathbf{r}_-) - e\mathbf{E}. \tag{2.60}$$

Здесь M_+ и M_- — массы положительного и отрицательного ионов.

Сложив почленно эти уравнения, получим $M_+\mathbf{r}_+ + M_-\mathbf{r}_- = 0$ — центр масс системы остается на месте. Для относительного смещения положительных и отрицательных ионов $\mathbf{r} \equiv \mathbf{r}_+ - \mathbf{r}_-$ из уравнений (2.59) и (2.60) находим (нужно первое уравнение разделить на M_+, второе — на M_- и произвести почленное вычитание)

$$\ddot{\mathbf{r}} + \omega_i^2\mathbf{r} = e\mathbf{E}/\mu, \tag{2.61}$$

где $1/\mu = 1/M_+ + 1/M_-$ — приведенная масса двух ионов противоположного знака; $\omega_i = (\beta/\mu)^{1/2}$ — частота собственных колебаний, возникающих при относительном смещении положительной и отрицательной подрешеток (длинноволновых оптических колебаний). Индуцированный дипольный момент $\mathbf{p}$ пары соседних ионов противоположного знака равен $e\mathbf{r}$. Решая уравнение (2.61) для монохроматического поля частоты ω, получаем, что $\mathbf{r}$ и, следовательно, момент $\mathbf{p} = e\mathbf{r}$ пропорционален $\mathbf{E}$. Коэффициент пропорциональности $\alpha_i(\omega)$ — ионная поляризуемость (в расчете на пару ионов) — имеет такой же вид, как и для оптических электронов, только вместо массы электрона в $\alpha_i(\omega)$ входит приведенная масса μ двух ионов и вместо собственной частоты ω_0 электрона (приходящейся обычно на ультрафиолетовую часть спектра) входит ω_i — частота оптических колебаний ионов:

$$\alpha_i(\omega) = [e^2/(\mu\varepsilon_0)]/(\omega_i^2 - \omega^2). \tag{2.62}$$

Вследствие большой массы ионов по сравнению с массой электрона ($\mu \gg m$) их собственная частота ω_i много меньше ω_0 и обычно соответствует инфракрасной области спектра (порядка 10^{13} Гц). Как видно из (2.62), инфракрасное излучение с частотами ω, близкими к ω_i, вызывает особенно сильную ионную поляризацию, т. е. сильно взаимодействует с оптическими колебаниями кристаллической решетки. Конечно, непосредственно вблизи ω_i выражение (2.62) неприменимо: для правильного описания резонанса в уравнение (2.61) нужно включить член, учитывающий затухание в движении ионов.

Действующее на ионы электрическое поле в (2.61) в кубическом кристалле можно отождествить с локальным полем $\mathbf{E}_{\text{лок}}$, связанным со средним макроскопическим полем соотношением (2.44). Учитывая вклады электронов и ионов в поляризованность среды, теперь вместо (2.45) получим следующее выражение:

$$\frac{n^2-1}{n^2+2}=\frac{N_e e^2/(3m\varepsilon_0)}{\omega_0^2-\omega^2}+\frac{N_i e^2/(3\mu\varepsilon_0)}{\omega_i^2-\omega^2}.$$

Этому выражению (после некоторых преобразований) можно придать вид обычной дисперсионной формулы:

$$n^2-1=C_0/(\tilde{\omega}_0^2-\omega^2)+C_i/(\tilde{\omega}_i^2-\omega^2). \qquad (2.63)$$

Ионные кристаллы, как правило, прозрачны в видимой области спектра, так как полюсы резонансных членов в правой части (2.63), обусловленных электронами и ионами, и связанные с ними полосы поглощения находятся соответственно в ультрафиолетовой и инфракрасной областях спектра. Но зависимость показателя преломления от частоты в видимой области существенно определяется этими членами, хотя сами резонансные часто́ты $\tilde{\omega}_0$ и $\tilde{\omega}_i$ находятся за ее пределами. Эти часто́ты, а также постоянные C_0 и C_i в (2.63) могут быть найдены по измерениям показателя преломления в видимой области (при нескольких значениях частоты ω). Полагая затем в (2.63) $\omega=0$, можно получить статическое значение диэлектрической проницаемости: $\varepsilon(0)=n^2(0)=1+C_0/\tilde{\omega}_0^2+C_i/\tilde{\omega}_i^2$. Так как $\omega_i \ll \omega_0$, основную роль здесь играет член с ω_i, т. е. главный вклад в $\varepsilon(0)$ обусловлен ионной поляризуемостью. Замечательно, что найденное таким образом из оптических измерений в видимой области статическое значение ε вполне удовлетворительно согласуется с измерениями диэлектрической проницаемости ионных кристаллов электрическими методами.

Некоторые ионные кристаллы, в целом прозрачные в видимой области, имеют характерную окраску. Такая окраска говорит о наличии сравнительно узких полос поглощения в видимой области. Эти полосы обусловлены наличием примесных ионов и других точечных дефектов. Например, окрашенный в розовый цвет рубин представляет собой кристалл корунда Al_2O_3, в котором некоторые ионы алюминия Al^{3+} замещены ионами хрома Cr^{3+}. Окраска рубина связана с полосами поглощения примесного иона (чистый корунд бесцветен). Ионные кристаллы с подходящими примесями используются как материал для активной среды твердотельных лазеров.

В щелочно-галоидных кристаллах типа NaCl полосы поглощения возникают при наличии вакансий. Нагревание кристалла NaCl в атмосфере паров натрия приводит к некоторому избытку ионов натрия и обусловливает образование вакантных узлов в подрешетке хлора. При быстром охлаждении кристалла («закаливании») вакансии оказываются «замороженными». На место отсутствующего отрицательного иона хлора для компенсации заряда может быть захвачен электрон. Образуется F-центр. Собственная частота электрона в F-центре приходится на видимую область спектра. В результате кристалл NaCl окрашивается в желтовато-коричневый, а KCl — в голубой цвет. Центры окраски в щелочно-галоидных кристаллах могут быть созданы также облучением образца рентгеновскими или гамма-лучами.

Релаксационная (ориентационная) *дисперсия* проявляется в средах с полярными молекулами — жидкостях и растворах. Полярные молекулы обладают собственным постоянным дипольным моментом, обусловленным несимметричным распределением электрических зарядов внутри молекулы. При отсутствии внешнего поля поляризованность **P** среды равна нулю, так как дипольные моменты молекул ориентируются хаотически. При наложении электрического поля на каждую молекулу действует пара сил, поворачивая дипольный момент молекулы вдоль по полю. Тепловое движение препятствует этой ориентации. В среднем получается преимущественное направление моментов по полю, и поляризованность будет отлична от нуля, а при слабых полях — пропорциональна напряженности поля. Ясно, что рассматриваемый механизм поляризации дает сильную зависимость диэлектрической восприимчивости (и проницаемости) от температуры, выражающуюся в уменьшении χ и ε с ростом температуры.

В переменном электрическом поле частотой ω ориентационная поляризованность будет изменяться с той же частотой. Пока период изменения напряженности поля $T=2\pi/\omega$ много больше *времени релаксации* τ (т. е. времени установления равновесия), значение $\varepsilon(\omega)$ такое же, как и в постоянном поле. Но при увеличении частоты, когда T приближается к τ, полное установление среднего дипольного момента не успевает произойти. Тогда $\varepsilon(\omega)$ будет меньше своего статического значения. Поэтому в области частот $\omega \sim 1/\tau$ возникает зависимость диэлектрической проницаемости от частоты, т. е. *дисперсия*. При этом обязательно будет и *абсорбция*, так как такие процессы необратимы: поворот дипольных моментов молекул происходит с трением и сопровождается диссипацией энергии поля, т. е. выделением теплоты.

Времена релаксации для процессов ориентации диполей изменяются в широких пределах и сильно зависят от температуры. Например, для воды при комнатной температуре $\tau \sim 10^{-11}$ с, для льда при $-20°C$ $\tau \sim 10^{-3}$ с. Если дипольные моменты молекул велики, то и значения ε при низких частотах $\omega \ll 1/\tau$ оказываются большими. Так, у воды при комнатной температуре $\varepsilon=81$, $n=\sqrt{\varepsilon}=9$. При частотах $\omega \gg 1/\tau$ ориентационный механизм практически не дает вклада в восприимчивость, т. е. молекулы ведут себя так, как если бы они не были полярными. Для воды при оптических частотах диэлектрическая проницаемость составляет всего лишь 1,77 (показатель преломления $n=\sqrt{\varepsilon}=1{,}33$). Переход от статических значений ε к значениям, соответствующим оптическим частотам, происходит в области частот $\omega \sim 1/\tau$. Для воды это примерно 10^{10} Гц, что соответствует диапазону сантиметровых радиоволн (микроволны). На оптических частотах рассматриваемый механизм поляризации никакой роли не играет.

Здесь уместно объяснить, почему в оптике магнитную проницаемость μ можно считать равной единице для всех веществ. Механизм намагничивания парамагнетиков состоит в ориентации магнитных моментов атомов и молекул во внешнем

магнитном поле. Характерное время установления равновесной намагниченности того же порядка, что и время установления ориентационной поляризованности в электрическом поле. Эти времена заведомо велики по сравнению с оптическими периодами. Тем более это относится к временам релаксации ферромагнитных процессов. Поэтому нужно полагать $\mu = 1$ уже при частотах гораздо более низких, чем оптические.

Контрольные вопросы

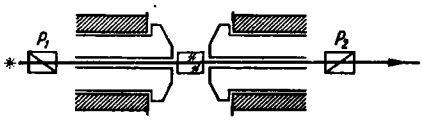
1. В чем заключается физическая причина поглощения инфракрасного излучения в ионных кристаллах?
2. Почему значения диэлектрической проницаемости воды для радиоволн и для видимого света столь сильно отличаются?
3. Чем объясняется, что постоянный дипольный момент молекул и связанная с ним ориентационная дисперсия не играют никакой роли в оптике?

2.8. Поворот направления линейной поляризации в магнитном поле (эффект Фарадея)

Влияние внешнего постоянного магнитного поля на излучение дипольного осциллятора (эффект Зеемана) было рассмотрено в § 1.9. Классическая электронная теория приводит к выводу, что свет, излучаемый состоящим из таких изотропных осцилляторов источником, при наблюдении вдоль магнитного поля вместо одной узкой спектральной линии на частоте ω_0 образует две компоненты на частотах $\omega_0 \pm \Omega$, где $\Omega = [e/(2m)]B$ — ларморовская частота. Смещенные компоненты имеют левую и правую циркулярную поляризацию в отличие от естественного света исходной спектральной линии. Для одиночных спектральных линий (синглетов) эксперимент полностью подтверждает предсказания классической теории.

Расщепление собственной частоты осциллятора в магнитном поле должно проявляться не только в испускании, но и в *поглощении* света: максимумы поглощения света левой и правой круговых поляризаций при распространении вдоль магнитного поля приходятся на частоты $\omega_0 \pm \Omega$. Линейно поляризованный и естественный свет можно разложить на сумму двух циркулярно поляризованных волн, поэтому в поглощении должны наблюдаться сразу две линии на частотах $\omega_0 \pm \Omega$ (обратный эффект Зеемана).

Поглощение и дисперсия связаны друг с другом. Поэтому следует ожидать, что при распространении вдоль магнитного поля в среде, содержащей дипольные осцилляторы, волны левой и правой круговых поляризаций будут характеризоваться разными значениями показателя преломления, т. е. будут иметь неодинаковые фазовые скорости. Когда эти волны представляют собой составляющие одной линейно поляризо-



2.6
Схема наблюдения эффекта Фарадея

ванной волны, различие их скоростей проявляется в *изменении* (повороте) *направления* ее *поляризации* по мере распространения в среде вдоль линий магнитного поля.

Поворот плоскости поляризации при прохождении света через вещество, находящееся в магнитном поле, впервые наблюдал Фарадей в 1846 г. Это был первый эксперимент, продемонстрировавший связь оптических и магнитных явлений. Для наблюдения эффекта Фарадея исследуемое вещество помещается между полюсами электромагнита (рис. 2.6). Линейно поляризованный свет пропускается вдоль магнитного поля через канал, просверленный в сердечнике. Анализатор P_2 устанавливается так, чтобы при отсутствии магнитного поля свет не проходил (скрещенные поляризаторы P_1 и P_2). При включении электромагнита свет проходит через анализатор. Поворотом анализатора его можно погасить. Это значит, что свет остается линейно поляризованным и здесь наблюдается именно поворот направления линейной поляризации.

Количественные исследования показали, что угол поворота φ пропорционален длине пути l света в веществе и индукции B внешнего магнитного поля:

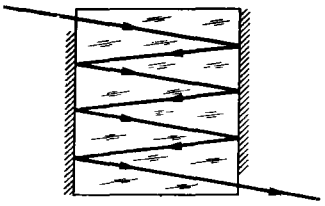
$$\varphi = \rho l B. \qquad (2.64)$$

Коэффициент пропорциональности ρ, характеризующий исследуемое вещество, зависит от частоты света. Он называется *постоянной Верде*. Поворот условно считают положительным, если для наблюдателя, смотрящего вдоль магнитного поля, он происходит вправо (по часовой стрелке). Большинство веществ характеризуются положительным вращением.

Направление поворота плоскости поляризации для каждого тела не зависит от направления распространения света, а связано только с направлением магнитного поля. Это свойство позволяет увеличить суммарный угол поворота удлинением пути света в образце за счет многократных отражений на посеребренных поверхностях образца (рис. 2.7).

В последние годы эффект Фарадея широко применяется в научных исследованиях. В кольцевых лазерах с его помощью можно добиться того, что бегущим навстречу волнам будут соответствовать отличающиеся частоты генерации (см. § 8.4). Малая инерционность эффекта (время установления меньше 10^{-9} с) позволяет использовать его для модуляции света, для создания оптического затвора и т. п.

Рассмотрим теоретическое истолкование эффекта Фарадея на основе классической электронной теории. Как уже отмечалось, физическая причина поворота направления поляризации состоит в различии фазовых скоростей волн левой и правой круговых поляризаций при их распространении вдоль магнитного поля в веществе. Пусть в падающей на образец линейно поляризованной



2.7
Усиление эффекта при многократных отражениях

волне колебания вектора $\mathbf{E}$ происходят вдоль направления AA (рис. 2.8). Тогда в точке $z=0$, где распространяющаяся вдоль оси z волна входит в образец, векторы $\mathbf{E}_+$ и $\mathbf{E}_-$ циркулярно поляризованных волн, на которые можно разложить падающую волну, вращаются с угловой скоростью ω в противоположных направлениях, сохраняя в каждый момент симметричное расположение относительно направления AA (рис. 2.8, *а*), т. е. углы φ_+ и φ_- при $z=0$ равны друг другу: $\varphi_+(0, t)=\varphi_-(0, t)=\omega t+\text{const}$. Обозначим фазовую скорость волны левой круговой поляризации, в которой вращение вектора $\mathbf{E}_+$ происходит против часовой стрелки, через v_+.

В точке z направление вектора $\mathbf{E}_+$ в момент времени t, характеризуемое углом $\varphi_+(z, t)$, будет таким же, как и в точке $z=0$ в более ранний момент времени $t-z/v_+$:

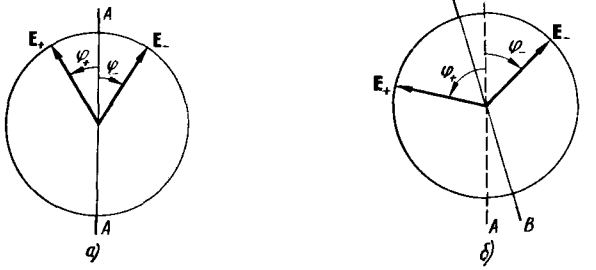
$$\varphi_+(z, t)=\varphi_+(0, t-z/v_+)=\omega(t-z/v_+)+\text{const}.$$

Совершенно аналогично для волны правой круговой поляризации, распространяющейся со скоростью v_-, имеем

$$\varphi_-(z, t)=\varphi_-(0, t-z/v_-)=\omega(t-z/v_-)+\text{const}.$$

Векторы $\mathbf{E}_+$ и $\mathbf{E}_-$ в точке z вращаются с угловой скоростью ω в противоположных направлениях и в сумме дают колебание вдоль направления BB (рис. 2.8, *б*), относительно которого они в каждый момент расположены симметрично. Так как $\varphi_+(z, t)\neq\varphi_-(z, t)$, то направление BB поляризации в точке z повернуто относительно исходного направления AA на угол $\varphi=(\varphi_+-\varphi_-)/2$:

$$\varphi={}^1/_2[\omega(t-z/v_+)-\omega(t-z/v_-)]-{}^1/_2\omega z(1/v_- - 1/v_+)=$$
$$=[\omega/(2c)]\ (n_- - n_+)z, \qquad (2.65)$$



2.8
Направление поляризации в точке $z=0$ (AA) и в точке $z=l$ (BB)

где n_+ — показатель преломления для циркулярно поляризованной волны, в которой вращение вектора $\mathbf{E}$ происходит в том же направлении, в котором идет ток в обмотке создающего поле электромагнита, и n_- — для волны с противоположным направлением вращения.

В модели вещества, которую использует классическая электронная теория дисперсии, оптические электроны в атомах рассматриваются как изотропные осцилляторы, характеризуемые единственной собственной частотой ω_0. При этом показатель преломления среды $n(\omega)$ не зависит от характера поляризации света. Все формулы для $\varepsilon(\omega)$ и $n(\omega)$, полученные в § 2.3—2.5, одинаково применимы для света любой поляризации. Но при помещении среды во внешнее магнитное поле $\mathbf{B}$ происходит расщепление собственных частот осциллятора. Нормальными модами осциллятора в магнитном поле являются колебания вдоль направления поля с прежней частотой ω_0 и круговые движения в плоскости, перпендикулярной индукции магнитного поля, с частотами $\omega_\pm = \sqrt{\omega_0^2 + \Omega^2} \pm \Omega \approx \omega_0 \pm \Omega$ (см. § 1.9), причем большая частота $\omega_+ = \omega_0 + \Omega$ соответствует вращению электрона в том же направлении, в котором идет ток, создающий магнитное поле.

Круговые движения электронов в атомных осцилляторах возбуждаются при распространении в среде циркулярно поляризованной волны. Поэтому в выражениях, определяющих показатель преломления для распространяющихся вдоль магнитного поля циркулярно поляризованных волн, резонансные члены будут максимальны при совпадении частоты ω с собственными частотами круговых движений $\omega_\pm$.

Найдем показатель преломления вдали от этих частот, где затуханием осцилляторов можно пренебречь. Уравнение вынужденного движения электрона в поле световой волны $\mathbf{E}(t)$ при наличии постоянного внешнего поля $\mathbf{B}$ имеет вид

$$\ddot{\mathbf{r}} + \omega_0^2 \mathbf{r} = -(e/m)[\mathbf{E}(t) + \dot{\mathbf{r}} \times \mathbf{B}].$$

Дальше действуем так же, как и в § 1.9. Вводим ларморовскую угловую скорость Ω согласно (1.99) и соответствующий член, содержащий $\dot{\mathbf{r}}$, переносим в левую часть уравнения:

$$\ddot{\mathbf{r}} - 2\dot{\mathbf{r}} \times \Omega + \omega_0^2 \mathbf{r} = -(e/m)\mathbf{E}(t). \qquad (2.66)$$

Направим ось z вдоль магнитного поля $\mathbf{B}$. Тогда вектор Ω будет иметь проекции $(0, 0, -\Omega)$. Вектор напряженности $\mathbf{E}(t)$ электрического поля для распространяющейся параллельно $\mathbf{B}$ циркулярно поляризованной монохроматической волны частотой ω имеет проекции

$$E_x = E_0 \cos\omega t, \quad E_y = \pm E_0 \sin\omega t, \quad E_z = 0,$$

где верхний знак в E_y соответствует волне с вектором $\mathbf{E}_+$, вращающимся против часовой стрелки, нижний — волне с векто-

ром $\mathbf{E}_-$. Заметим, что направление распространения волны (вдоль $\mathbf{B}$ или навстречу) при этом никакой роли не играет. Записывая (2.66) в проекциях на оси x и y, получаем систему уравнений:

$$\ddot{x}+2\Omega\dot{y}+\omega_0^2 x=-(e/m)E_0\cos\omega t,$$

$$\ddot{y}-2\Omega\dot{x}+\omega_0^2 y=\pm(-e/m)E_0\sin\omega t. \qquad (2.67)$$

Дальше удобно вместо вещественных функций $x(t)$ и $y(t)$ ввести комплексную функцию $\xi(t)=x+iy$, так что $x=\mathrm{Re}\,\xi$, $y=\mathrm{Im}\,\xi$. Умножая второе уравнение (2.67) на i и складывая с первым, получаем

$$\ddot{\xi}-2i\Omega\dot{\xi}+\omega_0^2\xi=(-e/m)E_0 e^{\pm i\omega t}. \qquad (2.68)$$

Установившееся движение электрона описывается частным решением этого уравнения, имеющим вид $\xi(t)=\xi_0\exp(\pm i\omega t)$. Для амплитуды ξ_0 из (2.68) находим $\xi_0=E_0/(\omega_0^2-\omega^2\pm 2\omega\Omega)$. Отделяя вещественную и мнимую части в $\xi(t)$, получаем

$$x(t)=\frac{(-e/m)E_x(t)}{\omega_0^2-\omega^2\pm 2\Omega\omega},\quad y(t)=\frac{(-e/m)E_y(t)}{\omega_0^2-\omega^2\pm 2\Omega\omega}. \qquad (2.69)$$

Отсюда находим индуцированный дипольный момент $\mathbf{p}=-e\mathbf{r}(t)$ и атомную поляризуемость $\alpha_\pm(\omega)$ в поле циркулярно поляризованных волн:

$$\alpha_\pm=[e^2/(m\varepsilon_0)]/(\omega_0^2-\omega^2\pm 2\Omega\omega). \qquad (2.70)$$

Учитывая то, что ларморовская частота Ω даже в очень сильных магнитных полях много меньше частоты ω для оптического диапазона ($\Omega/\omega\sim 2\cdot 10^{-4}$ при $B=10$ Тл и $\lambda=500$ нм) знаменатель в (2.70) можно приближенно записать в виде $\omega_0^2-(\omega\mp\Omega)^2$. Отсюда следует, что атомная поляризуемость $\alpha_\pm$ в поле циркулярно поляризованных волн (2.70) может быть получена из выражения для атомной поляризуемости $\alpha(\omega)$ при отсутствии магнитного поля (т. е. при $\Omega=0$)

$$\alpha(\omega)=[e^2/(m\varepsilon_0)]/(\omega_0^2-\omega^2)$$

простой заменой ω на $\omega\mp\Omega$: $\alpha_\pm=\alpha(\omega\mp\Omega)$. То же самое справедливо и для восприимчивости $\chi_\pm$, диэлектрической проницаемости $\varepsilon_\pm$, показателя преломления $n_\pm$. Таким образом,

$$n_\pm(\omega)=n(\omega\mp\Omega)\approx n(\omega)\mp\Omega\,dn(\omega)/d\omega.$$

Подставляя n_--n_+ в (2.65), получаем выражение для угла поворота направления поляризации:

$$\varphi=\frac{\omega}{2c}(n_--n_+)z=\frac{\omega}{c}\Omega\frac{dn(\omega)}{d\omega}z. \qquad (2.71)$$

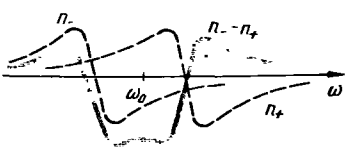
В области нормальной дисперсии показатель преломления возрастает с частотой, т. е. $dn/d\omega>0$, и поворот направления поляризации происходит в ту сторону, в которую идет ток в обмотке электромагнита, создающего поле. Это справедливо как для света,

распространяющегося вдоль магнитного поля, так и навстречу. Угол φ, как видно из (2.71), пропорционален индукции B магнитного поля и пройденному расстоянию $z=l$. Таким образом, электронная теория дисперсии объясняет эмпирический закон (2.64) и позволяет по известной зависимости $n(\omega)$ вычислить постоянную Верде ρ:

$$\rho=\frac{e}{2mc}\omega\frac{dn}{d\omega}=-\frac{e}{2mc}\lambda\frac{dn}{d\lambda}. \qquad (2.72)$$

Для закона дисперсии $n(\lambda)$, описываемого формулой Коши (2.41), зависимость постоянной Верде от длины волны, согласно (2.72), имеет вид $\rho(\lambda)\sim 1/\lambda^2$, что находится в полном соответствии с экспериментальными данными.

В непосредственной близости к собственной частоте осцилляторов ω_0 эффект Фарадея описывается более сложными закономерностями. В уравнении движения осциллирующего электрона (2.66) необходимо учитывать затухание. Можно показать (см. задачу), что для циркулярно поляризованных волн, распространяющихся вдоль магнитного поля, дисперсионная кривая и спектральный контур линии поглощения имеют для данной среды тот же вид, что и при отсутствии магнитного поля, отличаясь только сдвигом по шкале частот на Ω вправо для волны с положительным направлением вращения вектора $\mathbf{E}$ и на Ω влево — для волны с противоположным направлением вращения $\mathbf{E}$. На рис. 2.9 штриховыми линиями показаны графики функций $n_+(\omega)$ и $n_-(\omega)$, а их разность $n_- - n_+$ — сплошной линией. Видно, что в окрестности ω_0 дважды изменяется знак эффекта Фарадея: в интервале частот $\sim 2\Omega$ вблизи ω_0 поворот направления поляризации происходит в отрицательную сторону, а вне этого интервала — в положительную. Однако следует иметь в виду, что в данном случае эффект не сводится только к повороту направления поляризации падающей волны. В окрестности ω_0 существенно поглощение света, причем при данном значении ω коэффициенты затухания $\varkappa_\pm(\omega)$ для циркулярно поляризованных составляющих падающей волны имеют разные значения (круговой дихроизм). Поэтому после прохождения через образец амплитуды этих составляющих не равны и при их сложении получается эллиптически поляризованный свет.



2.9

Зависимость показателей преломления от частоты для циркулярно поляризованных волн в магнитном поле

Важно сознавать, что в эффекте Фарадея магнитное поле влияет на состояние поляризации света лишь *косвенно*, изменяя характеристики среды, в которой распространяется свет. В вакууме магнитное поле никакого влияния на свет не оказывает.

Обычно угол поворота направления поляризации очень мал, но благодаря высокой чувствительности эксперимен-

тальных методов измерения состояния поляризации эффект Фарадея лежит в основе совершенных оптических методов определения атомных констант.

Контрольные вопросы

- В чем заключается физическая причина поворота плоскости поляризации при распространении света вдоль магнитного поля в веществе? Почему это не происходит в вакууме?
- Как угол поворота направления поляризации выражается через показатели преломления для циркулярно поляризованных волн?
- Что представляют собой нормальные моды изотропного осциллятора в магнитном поле?
- При отражении волна с правой круговой поляризацией превращается в волну с левой поляризацией и наоборот. Почему же при отражении линейно поляризованной волны, состоящей из двух циркулярно поляризованных, направление поворота плоскости поляризации остается прежним?
- Как, зная показатель преломления среды $n(\omega)$ при отсутствии магнитного поля, найти показатели преломления для циркулярно поляризованных волн, распространяющихся вдоль магнитного поля?
- В каком направлении происходит поворот плоскости поляризации в области нормальной дисперсии?
- Почему распространяющийся вдоль магнитного поля линейно поляризованный свет становится эллиптически поляризованным, если его частота близка к линии поглощения?

Задача

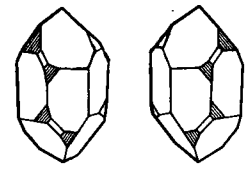
Найти показатели преломления для монохроматических циркулярно поляризованных волн, распространяющихся вдоль магнитного поля в разреженной среде, содержащей N атомных осцилляторов в единице объема. Каждый осциллятор характеризуется зарядом e, массой m, собственной частотой ω_0 и постоянной затухания γ. Рассмотреть область частот вблизи собственной частоты осцилляторов $\omega \approx \omega_0$.

Ответ: выражения для показателя преломления и показателя затухания циркулярно поляризованных монохроматических волн получаются из соответствующих выражений (2.49), справедливых в отсутствие магнитного поля, заменой $\Delta\omega \to \Delta\omega \pm \Omega$. Другими словами, дисперсионная кривая и спектральный контур линии поглощения в магнитном поле для волны с положительным направлением вращения сдвинуты на Ω по шкале частот вправо, а с отрицательным — на Ω влево: $n_{\pm}(\omega) = n(\omega \mp \Omega)$, $\varkappa_{\pm}(\omega) = \varkappa(\omega \mp \Omega)$.

2.9. Естественное вращение направления поляризации

В некоторых случаях распространение линейно поляризованного света в веществе сопровождается поворотом направления поляризации вокруг оси пучка даже при отсутствии внешнего магнитного поля. Это явление в отличие от магнитного вращения направления поляризации называют естественным вращением или *естественной оптической активностью* (гиротропией).

Естественная активность была открыта Араго в 1811 г. для кристаллов кварца. Позднее Био обнаружил поворот направления поляризации света в растворах сахара. Теперь известно много естественно-активных веществ, хотя у большинства из них это явление выражено очень слабо.



2.10
Внешняя форма монокристаллов право- и левовращающего кварца

Для наблюдения оптической активности слой исследуемого вещества помещают между установленными на темноту (скрещенными) поляризаторами. Проходящий при этом свет можно снова погасить поворотом одного из поляризаторов. Это значит, что после прохождения через слой активного вещества свет остается линейно поляризованным, но направление его поляризации повернуто на некоторый угол относительно первоначального. Угол поворота направления поляризации зависит от длины волны света, а для данной длины волны пропорционален толщине l слоя активного вещества: $\varphi = \alpha(\omega)l$. Например, в кварце при распространении вдоль оптической оси желтый свет натрия ($\lambda = 589{,}3$ нм) испытывает поворот направления поляризации 21,73 угл. град. на миллиметр.

Опыт показывает, что при изменении направления распространения света на противоположное поворот плоскости поляризации происходит в обратную сторону. Поэтому при прохождении сквозь активную среду, отражении от зеркала и вторичном прохождении через ту же среду назад направление линейной поляризации восстанавливается. Этим естественное вращение отличается от магнитного, где имело значение направление вектора индукции внешнего магнитного поля, а не направление распространения света. Для естественного вращения направление (знак) поворота связывают с направлением света: активную среду называют правовращающей, если для наблюдателя, смотрящего навстречу световому пучку, поворот происходит по часовой стрелке, и левовращающей в противоположном случае. Определенный таким образом знак вращения не меняется при изменении направления света, т. е. свойство среды быть право- или левовращающей не зависит от направления луча, подобно тому как свойство винтовой линии быть правой или левой не зависит от того, с какой стороны на нее смотреть.

Кристаллический кварц встречается в двух модификациях: для одной поворот направления поляризации при распространении вдоль оптической оси* происходит вправо, для другой — влево (на такой же угол). Для кристаллической решетки кварца характерно отсутствие центра и плоскостей симметрии. Две модификации по своей симметрии и внешней форме кристаллов (рис. 2.10) отличаются друг от друга, как правая рука отличается от левой или как

* Кварц представляет собой одноосный кристалл (см. § 4.2). Распространение света в нем вдоль оптической оси происходит так же, как в изотропной среде (т. е. возможна любая ориентация плоскости поляризации).

винт с правой нарезкой от такого же винта с левой нарезкой. Никаким перемещением их нельзя совместить друг с другом. Одна модификация является зеркальным изображением второй. Вращательная способность кварца связана с его кристаллической структурой, так как плавленый кварц не обладает оптической активностью.

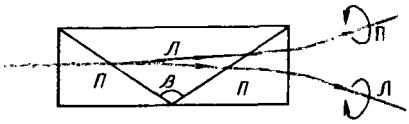
Для оптически активных жидкостей и аморфных тел эффект вращения обусловлен асимметричным строением молекул. Согласно представлениям стереохимии, асимметрия органических молекул связана со свойством атома углерода образовывать направленные валентные связи с другими атомами или атомными группами (радикалами). В простейшем случае (например, метан CH_4) атом углерода находится в центре правильного тетраэдра, образованного другими атомами. Но если все четыре вершины тетраэдра заняты разными радикалами, то возможны две изомерные модификации, представляющие собой зеркальные изображения друг друга. Оптическая активность возникает тогда, когда в веществе преобладают молекулы только одной модификации.

Поворот направления линейной поляризации можно объяснить, предположив, что в оптически активной среде монохроматические волны правой и левой круговой поляризаций имеют различные фазовые скорости $v_п$ и $v_л$, т. е. различные показатели преломления $n_п$ и $n_л$. Для угла φ поворота плоскости поляризации можно получить (точно так же, как в § 2.8) следующее выражение:

$$\varphi = (\omega l/2)(1/v_л - 1/v_п) = [\omega/(2c)](n_л - n_п)\, l. \qquad (2.73)$$

Знак φ в (2.73) определен в соответствии со сформулированным выше правилом ($\varphi > 0$ в случае правого вращения).

Возможность существования в активной среде циркулярно поляризованных волн, распространяющихся с различными скоростями, была непосредственно показана экспериментально Френелем с помощью специально изготовленной сложной призмы (рис. 2.11), состоящей из трех призм: двух торцовых из правовращающего кварца (П) и средней с тупым преломляющим углом из левовращающего кварца (Л). Оптические оси всех призм параллельны основанию. Падающий на торцовую грань линейно поляризованный пучок света расщепляется на два пучка с правой и левой круговыми поляризациями. В правовращающем кварце $n_п < n_л$, в левовращающем, наоборот, $n_п > n_л$. Поэтому луч правой круговой поляризации при преломлении на внутренних гранях отклоняется в сторону основания средней призмы, а левой — в сторону вершины. Состояние круговой поляризации



2.11
Призма Френеля

выходящих из призмы Френеля пучков непосредственно проверяется с помощью пластинки $\lambda/4$ и анализатора.

Теоретическое объяснение различия фазовых скоростей волн правой и левой круговых поляризаций в оптически активной среде может быть дано только при учете структуры и конечного размера ее молекул. При этом существенно, что действующее на электроны поле волны $\mathbf{E}(\mathbf{r}, t)$ для каждого момента времени t в разных точках протяженной молекулы различно. Индуцированный неоднородным полем волны дипольный момент молекулы зависит от значений $\mathbf{E}(\mathbf{r})$ на всем протяжении молекулы [а не в одной точке, как это предполагалось в теории дисперсии, где действующее на электрон поле считалось однородным и вместо $\mathbf{E}\exp i(\mathbf{kr} - \omega t)$ мы писали $\mathbf{E}_0 \exp(-i\omega t)$].

Характерный масштаб неоднородности поля — это длина волны λ. Отношение размеров a молекул к длине световой волны λ имеет порядок 10^{-3}. Для многих оптических проблем пренебрежение величинами, содержащими это малое отношение a/λ, вполне допустимо так как учет таких величин привел бы лишь к малым поправкам, не внося ничего существенно нового. Но существуют эффекты, которые целиком определяются этими малыми величинами порядка a/λ. С этой точки зрения проблема естественного вращения плоскости поляризации представляет принципиальный интерес, требуя выхода за рамки нулевого приближения по a/λ.

Влияние неоднородности поля волны на индуцированный дипольный момент молекулы с макроскопической точки зрения означает, что поляризованность среды $\mathbf{P}(\mathbf{r})$ в каждой точке зависит от значения напряженности $\mathbf{E}(\mathbf{r})$ не только в той же точке $\mathbf{r}$, но и в соседних точках области порядка молекулярных размеров. Другими словами, связь между $\mathbf{P}$ и $\mathbf{E}$ имеет нелокальный характер. То же самое относится, очевидно, и к связи между вектором индукции $\mathbf{D}$ и напряженностью $\mathbf{E}$ поля. Для учета этой нелокальности достаточно представить $\mathbf{E}$ в виде разложения в ряд Тэйлора по смещениям $\Delta\mathbf{r}(\Delta x, \Delta y, \Delta z)$ из рассматриваемой точки $\mathbf{r}$ и ограничиться первыми членами разложения:

$$E_i(\mathbf{r}+\Delta\mathbf{r}) = E_i(\mathbf{r}) + \sum_k \Delta x_k \left(\frac{\partial E_i}{\partial x_k}\right)_{\mathbf{r}} +$$

$$+ \frac{1}{2}\sum_{k,\,l} \Delta x_k \Delta x_l \left(\frac{\partial^2 E_i}{\partial x_k \partial x_l}\right)_{\mathbf{r}} + \ldots$$

Производная $\partial E_i/\partial x_k$ имеет порядок E/λ, а Δx_k порядка размеров молекулы a, так что член $\Delta x_k(\partial E_i/\partial x_k) \sim (a/\lambda)E$ и его отношение к первому члену разложения порядка a/λ. Аналогично, следующий член разложения будет иметь порядок $(a/\lambda)^2 E$, и т. д.

Разложению E по степеням a/λ в рамках феноменологической теории соответствует материальное уравнение, в котором вектор

индукции $\mathbf{D}$ зависит не только от вектора $\mathbf{E}$, но и от его пространственных производных:

$$D_i = \varepsilon(\omega) E_i + \sum_{k,l} \gamma_{ikl}(\omega) \frac{\partial E_k}{\partial x_l} + \\ + \sum_{k,l,m} \alpha_{iklm}(\omega) \frac{\partial^2 E_k}{\partial x_l \partial x_m} + \ldots \tag{2.74}$$

Таким образом, оптические свойства среды характеризуются диэлектрической проницаемостью $\varepsilon(\omega)$ и тензорами третьего и четвертого рангов $\gamma_{ikl}(\omega)$ и $\alpha_{iklm}(\omega)$. В однородной среде они не зависят от пространственных координат, а об их зависимости от частоты монохроматического поля говорят как о частотной (или временно́й) дисперсии.

В монохроматической плоской волне зависимость $\mathbf{D}$ и $\mathbf{E}$ от координат и времени имеет вид $\exp i(\mathbf{kr} - \omega t)$. При этом дифференцирование $\mathbf{E}$ по координате x_l сводится к умножению $\mathbf{E}$ на ik_l и (2.74) приводится к виду

$$D_i = \sum_k \varepsilon_{ik}(\omega, \mathbf{k}) E_k, \tag{2.75}$$

где

$$\varepsilon_{ik}(\omega, \mathbf{k}) = \varepsilon(\omega)\, \delta_{ik} + i \sum_l \gamma_{ikl}(\omega) k_l - \sum_{l,m} \alpha_{iklm}(\omega) k_l k_m. \tag{2.76}$$

Здесь δ_{ik} — символ Кронекера ($\delta_{ik} = 1$ при $i = k$, $\delta_{ik} = 0$, если $i \neq k$). Связь векторов $\mathbf{D}$ и $\mathbf{E}$ (2.75) принимает формально локальный характер. Нелокальность этой связи проявляется в том, что диэлектрическая проницаемость $\varepsilon_{ik}(\omega, \mathbf{k})$ зависит не только от частоты ω света, но и от волнового вектора $\mathbf{k}$ (т. е. от длины волны $\lambda = 2\pi/k$). Об этой зависимости говорят как о *пространственной дисперсии* в отличие от временно́й дисперсии, отражающей нелокальность связи между $\mathbf{D}$ и $\mathbf{E}$ во времени.

Вследствие малости параметра a/λ эффекты пространственной дисперсии в оптике малы. Они становятся существенными лишь тогда, когда приводят к качественно новым явлениям. В средах, не обладающих центром симметрии, второй член в (2.76), имеющий порядок a/λ, приводит, как показано ниже, к небольшому различию фазовых скоростей волн правой и левой круговых поляризаций, т. е. к естественной активности. При наличии центра симметрии этот член обращается в нуль и эффекты пространственной дисперсии могут быть обусловлены лишь третьим членом в (2.76), имеющим порядок $(a/\lambda)^2$. Пример такого эффекта — слабая оптическая анизотропия кубических кристаллов, на возможность существования которой Лоренц обратил внимание еще в 1878 г. Из-за малости эффекта $[(a/\lambda)^2 \sim 10^{-6}]$ наблюдать его трудно. Экспериментально он был обнаружен лишь в 1960 г. Е. Ф. Грос-

сом и А. А. Каплянским по зависимости поглощения света от поляризации в кубическом кристалле закиси меди Cu_2O.

Естественная оптическая активность может наблюдаться лишь в средах без центра инверсии, для которых тензор $\gamma_{ikl}(\omega)$ отличен от нуля. В изотропной среде (жидкость или раствор) асимметричные молекулы расположены хаотически и выбор направления осей x, y и z совершенно произволен. Основываясь на эквивалентности направлений осей x, y и z в соотношении (2.74), связывающем макроскопические величины $\mathbf{E}$ и $\mathbf{D}$, можно показать (см. задачу), что в изотропной среде тензор $\gamma_{ikl}(\omega)$ имеет вид $\gamma(\omega)e_{ikl}$, где $\gamma(\omega)$ — скаляр, а e_{ikl} — полностью антисимметричный тензор третьего ранга ($e_{ikl}=0$, если среди индексов i, k, l имеются одинаковые, и $e_{ikl}=\pm 1$ в зависимости от того, получены индексы i, k, l из x, y, z четным или нечетным числом перестановок). В таком случае материальное уравнение (2.75) можно записать в виде

$$\mathbf{D}=\varepsilon(\omega)\mathbf{E}-i\gamma(\omega)\mathbf{k}\times\mathbf{E}. \qquad (2.77)$$

Если молекулы жидкости обладают центром инверсии (т. е. не имеют стереоизомеров), то она симметрична не только по отношению к любому повороту, но и по отношению к отражению (инверсии) в любой точке, и для нее $\gamma(\omega)=0$. Только тогда, когда жидкость содержит молекулы двух модификаций (представляющих зеркальные изображения друг друга) в разных количествах, она не обладает центром симметрии и $\gamma(\omega)\neq 0$. В такой гиротропной среде у вектора $\mathbf{D}$, как видно из (2.77), есть небольшая составляющая, перпендикулярная $\mathbf{E}$ и направлению волнового вектора $\mathbf{k}$. По модулю она отличается от составляющей вдоль $\mathbf{E}$ множителем порядка a/λ. При вещественных $\varepsilon(\omega)$ и $\gamma(\omega)$ (прозрачная среда) эта составляющая сдвинута по фазе на четверть периода.

Покажем, что в среде, удовлетворяющей материальному уравнению (2.77), волны правой и левой круговых поляризаций характеризуются различными показателями преломления $n_{\text{П}}\neq n_{\text{Л}}$. Обратимся к уравнениям Максвелла (2.7) и (2.9). В случае плоской монохроматической волны, когда зависимость $\mathbf{E}$ и $\mathbf{D}$ от $\mathbf{r}$ и t имеет вид $\exp i(\mathbf{kr}-\omega t)$, они из дифференциальных превращаются в алгебраические:

$$\varepsilon_0 c^2\mathbf{k}\times\mathbf{B}=-\omega\mathbf{D}, \qquad (2.78) \qquad \mathbf{k}\times\mathbf{E}=\omega\mathbf{B}. \qquad (2.79)$$

Исключим из них индукцию магнитного поля, подставляя $\mathbf{B}$ из (2.79) в (2.78):

$$\varepsilon_0 c^2\mathbf{k}\times(\mathbf{k}\times\mathbf{E})=-\omega^2\mathbf{D}.$$

Преобразуем двойное векторное произведение, учитывая, что $\mathbf{kE}=0$, и выразим в правой части $\mathbf{D}$ через $\mathbf{E}$ с помощью материального уравнения (2.77). В результате получим однородное уравнение для электрического поля $\mathbf{E}$:

$$(c^2k^2-\varepsilon\omega^2)\mathbf{E}+i\gamma\omega^2\mathbf{k}\times\mathbf{E}=0. \qquad (2.80)$$

Выберем ось z вдоль направления распространения волны, т. е. вдоль вектора $\mathbf{k}$. В проекциях на оси x и y (2.80) принимает вид

$$\begin{aligned}(c^2k^2 - \varepsilon\omega^2)E_x - i\gamma\omega^2kE_y = 0,\\ i\gamma\omega^2kE_x + (c^2k^2 - \varepsilon\omega^2)E_y = 0.\end{aligned} \tag{2.81}$$

Ненулевое решение этой однородной системы уравнений для E_x и E_y существует, когда ее определитель равен нулю. Отсюда получаем уравнение для нахождения k:

$$c^2k^2 - \varepsilon(\omega)\omega^2 = \pm\,\gamma\omega^2k. \tag{2.82}$$

В нулевом приближении малую правую часть в (2.82) можно положить равной нулю, что дает $k_0 = \sqrt{\varepsilon}\omega/c$, $n_0 = \sqrt{\varepsilon}$. В следующем приближении

$$k_\pm = k_0 \pm \gamma\omega^2/(2c^2),$$

$$n_\pm = n_0 \pm \gamma\omega/(2c). \tag{2.83}$$

Знаку «+» в (2.83) соответствует решение системы (2.81), у которого $E_y/E_x = -i$, т. е. волна правой круговой поляризации, знаку «—» — левой. Поэтому $n_л - n_п = n_- - n_+ = -\gamma\omega/c$. У двух изомерных модификаций вещества знак постоянной γ противоположен. При $\gamma > 0$ будет $n_л - n_п < 0$ и среда в соответствии с (2.73) характеризуется левым вращением, при $\gamma < 0$ — правым.

Таким образом, феноменологическая теория на основе материального уравнения (2.77) дает объяснение естественному вращению направления поляризации. Задача микроскопической теории оптической активности состоит в расчете константы $\gamma(\omega)$, определяющей угол поворота, и нахождении ее частотной зависимости (дисперсии) для той или иной модели гиротропной среды.

Естественное вращение плоскости поляризации можно проиллюстрировать наглядным демонстрационным опытом в СВЧ-диапазоне. Модель активной среды устроена следующим образом: пружинки диаметром 6—7 мм и длиной около 10 мм из медной изолированной проволоки (активные «молекулы»), хаотически ориентированные, заполняют небольшую картонную коробку. Рупор излучателя, питаемого клистронным генератором, создает линейно поляризованную монохроматическую волну ($\lambda = 3$ см). Такой же рупор приемника повернут вокруг продольной оси на угол $\pi/2$ относительно излучателя («скрещен» с излучателем). Поэтому приемник никакого сигнала не регистрирует. Введение коробки с пружинками на пути волны между излучателем и приемником приводит к появлению сигнала. Поворотом рупора приемника на некоторый угол вокруг продольной оси можно этот сигнал погасить. Это доказывает, что поляризация волны после прохождения через модель среды остается линейной, но с другим направлением поляризации. Если взять коробку, заполненную хаотически ориентированными пружинками таких же размеров, но навитыми в противоположную

сторону, то и направление поляризации поворачивается на такой же угол в противоположную сторону. Этот опыт моделирует правое и левое вращения направления поляризации двумя модификациями молекул (стереоизомерами) одного и того же макроскопически изотропного вещества.

Контрольные вопросы

- □ В чем отличие естественной оптической активности от магнитного вращения плоскости поляризации?
- □ В каком случае активная среда называется правовращающей?
- □ Чему равна разность показателей преломления для желтого света ($\lambda = 589{,}3$ нм) правой и левой круговых поляризаций в кварце?
- □ При каком условии макроскопически изотропная среда не имеет центра симметрии?
- □ Почему в призме Френеля пучок линейно поляризованного света расщепляется на два циркулярно поляризованных?
- □ Почему в неоднородном внешнем поле связь между **P** и **E** имеет нелокальный характер?
- □ Что такое пространственная дисперсия?
- □ Рассмотрите с помощью уравнений Максвелла и материального уравнения (2.77) взаимное расположение векторов **E**, **B**, **D** и **k** плоской волны, распространяющейся в гиротропной среде.

Задачи

1. Показать, что в изотропной среде связь векторов **E** и **D** с точностью до величин порядка $(a/\lambda)^2$ выражается соотношением

$$D_i = \varepsilon E_i + \sum_{k,l} \gamma_{ikl} \frac{\partial E_k}{\partial x_l},$$

причем у тензора γ_{ikl} все компоненты, имеющие два одинаковых индекса, обращаются в нуль (при любом значении третьего индекса), например $\gamma_{xxy} = 0$, $\gamma_{xxx} = 0$. Показать, что $\gamma_{xyz} = -\gamma_{yxz} = \gamma_{yzx}$ и т. д., т. е. что при перестановке любой пары индексов компонента γ_{ikl}, у которой все индексы различны, меняет знак. Показать, что при наличии центра симметрии все компоненты тензора γ_{ikl} равны нулю.

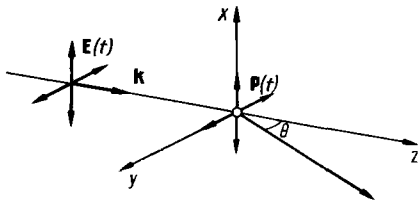
Указание. Рассмотреть поворот системы координат вокруг одной из осей, например оси x, на угол 90°. При этом $y \to z$, $E_y \to E_z$, $z \to -y$, $E_z \to -E_y$. Отсюда $\gamma_{xxy} = \gamma_{xxz}$; $\gamma_{xxz} = -\gamma_{xxy}$, т. е. $\gamma_{xxy} = \gamma_{xxz} = 0$ и т. д. Для среды с центром симметрии рассмотреть преобразование инверсии.

2. Используя приведенные в тексте данные, найдите угол, на который разведены поляризованные по правому и левому кругу лучи, выходящие из призмы Френеля (рис. 2.11). Угол $\beta = 152°$, показатель преломления в кварце при распространении вдоль оптической оси $n = 1{,}54$.

◇ Ответ: $\approx 4'$.

2.10. Рассеяние света

Электрическое поле распространяющейся в веществе световой волны раскачивает входящие в состав атомов и молекул электроны, и они становятся центрами вторичных сферических волн, излучаемых во все стороны. Поэтому распространение света в веществе должно,



2.11
К зависимости поляризации и интенсивности рассеянного света от угла рассеяния θ

поляризуемость $\alpha_1(\omega)$ частицы зависит от диэлектрической проницаемости $\varepsilon_1(\omega)$ материала частицы и пропорциональна ее объему:

$$\alpha_1 = [(\varepsilon_1 - 1)/(\varepsilon_1 + 2)]\, 4\pi a^3 \quad (2.84)$$

Напряженность $\mathbf{E}'$ электрического поля вторичной волны, излучаемой осциллирующим дипольным моментом частицы, на большом по сравнению с ее размерами расстоянии $r \gg a$ определяется формулами (1.66). Направление $\mathbf{E}'$ связано с направлением $\mathbf{p}$ так, как показано на рис. 1.17. Если падающий свет линейно поляризован, то колебания вектора $\mathbf{p}$ происходят в той же плоскости и при наблюдении перпендикулярно первичному пучку в этой плоскости рассеянного излучения не будет, так как диполь вдоль своей оси не излучает. Поместив на пути первичного пучка поляризатор и поворачивая его вокруг оси пучка, при наблюдении перпендикулярно пучку увидим изменение интенсивности рассеянного света вплоть до полного исчезновения его при совпадении направления колебаний в первичном пучке с направлением наблюдения.

Естественный падающий свет можно представить как некогерентную смесь двух волн одинаковой интенсивности, линейно поляризованных во взаимно перпендикулярных направлениях, например вдоль осей x и y на рис. 2.14. Поэтому и дипольный момент $\mathbf{p}$ рассеивающей частицы будет совершать колебания вдоль осей x и y. При наблюдении перпендикулярно первичному пучку, т. е. вдоль оси $y\,(\theta = \pi/2)$, рассеянный свет будет полностью поляризован, так как распространяющееся в этом направлении излучение обусловлено только колебаниями $\mathbf{p}$ вдоль оси x. По мере изменения угла θ от значения $\pi/2$ (в обе стороны) к поляризованному вдоль оси x рассеянному свету неизменной интенсивности примешивается не когерентный с ним свет, поляризованный в плоскости yz (рис. 2.14), интенсивность которого изменяется как $\cos^2\theta$. В результате степень поляризации рассеянного света постепенно уменьшается, обращаясь в нуль для $\theta = 0$ и $\theta = \pi$, а его интенсивность изменяется как $1 + \cos^2\theta$. Этим объясняется индикатриса рассеяния естественного света, приведенная на рис. 2.13.

Для нахождения интенсивности рассеянного одной частицей света $I(r, \theta)$ подставим в (1.66) $\ddot{\mathbf{p}} = \varepsilon_0 \alpha_1 \omega^2 \mathbf{E}$. Кроме того, в случае неполяризованного первичного пучка в выражении для интенсивности рассеянного света, как только что было выяснено, нужно заменить $\sin^2\theta$ на $(1 + \cos^2\theta)/2$. В результате получим

$$I(r, \theta) = \frac{1}{(4\pi)^2} \frac{\omega^4 \alpha_1^2(\omega)}{\varepsilon^4 r^2} \frac{1}{2} (1 + \cos^2\theta) I_0, \quad (2.85)$$

где I_0 — интенсивность первичного пучка. В том случае, когда рас-

сеивающие сферические частицы находятся на расстояниях, больших по сравнению с длиной волны, и распределены по объему беспорядочно, относительные фазы рассеянных ими волн в точке наблюдения имеют случайные значения. Поэтому полная интенсивность рассеянного света равна сумме интенсивностей волн, рассеянных отдельными частицами. Когда расстояние до точки наблюдения велико по сравнению с размерами рассеивающего объема, расстояние r в (2.85) для всех частиц можно считать одинаковым и полная интенсивность рассеянного света получается из (2.85) умножением на число частиц N в объеме V.

Основная зависимость интенсивности рассеянного света от частоты определяется множителем ω^4 в (2.85). В самом деле, если материал взвешенных сферических частиц прозрачен в видимой области, зависимость поляризуемости $\alpha_1(\omega)$ от частоты, обусловленную дисперсией $\varepsilon_1(\omega)$ в (2.84), можно не принимать во внимание по сравнению с сильной зависимостью $\sim\omega^4$. Интенсивность рассеянного мутной средой света оказывается обратно пропорциональной четвертой степени длины волны. Этот результат известен как *закон Рэлея*. Коротковолновое излучение рассеивается сильнее, поэтому при рассеянии белого света получается голубой оттенок, а прошедший свет, обедненный голубыми лучами, имеет красноватый оттенок.

Рассмотренные закономерности рассеяния света не выполняются, если размеры рассеивающих частиц сравнимы с длиной волны. Зависимость интенсивности от длины волны становится менее заметной. Рассеянный в поперечном направлении свет будет поляризован лишь частично, причем степень его поляризации зависит от размеров и формы частиц. Рассеяние вперед становится преобладающим, и индикатриса сохраняет симметрию лишь относительно направления первичного пучка.

Теория рассеяния света на сферических частицах, размеры которых могут быть порядка или больше длины волны, была впервые разработана Дж. Ми в 1908 г. Рассеяние Ми можно рассматривать как дифракцию (см. § 6.3) плоской волны на одинаковых однородных сферах, хаотически распределенных в однородной среде и находящихся друг от друга на расстояниях, больших по сравнению с длиной волны.

Рассеяние Ми имеет много практических приложений: изучение атмосферной пыли и аэрозолей, влияние облаков и тумана на распространение света, теория радуги. Измерение характеристик рассеянного света лежит также в основе оптических методов исследования коллоидных суспензий (нефелометрия).

Опыт показывает, что рассеяние света происходит не только в мутных средах, но и в тщательно очищенных от посторонних примесей или включений жидкостях и газах. Рассеяние в *чистом* веществе, как правило, слабое. Но тем не менее оно представляется принципиально важным явлением. Физическая причина оптической неоднородности в идеально чистых средах была указана М. Смолуховским в 1908 г. Из-за хаотического характера теплового движения молекул в среде возникают *флуктуации плотности* и, следовательно, флуктуации показателя преломления. Такой тип рассеяния света называют *молекулярным*. Флуктуации плотности особенно велики вблизи критического состояния, где сжимаемость среды становится очень большой. Поэтому в критической точке наблюдается интен-

сивное рассеяние света (критическая опалесценция). Стеклянная ампула с эфиром при достижении критического состояния дает на экране совершенно черную тень.

Количественная теория молекулярного рассеяния света была построена Эйнштейном в 1910 г. Здесь мы ограничимся простейшим случаем слабого рассеяния в идеальных газах. Наиболее известный пример — молекулярное рассеяние солнечного света в земной атмосфере, которым объясняется голубой цвет и поляризация свечения неба. Эти эффекты легко наблюдаемы благодаря большой толщине слоя рассеивающего газа.

Полное гашение вторичных волн, распространяющихся во всех направлениях, кроме направления падающей волны, происходит лишь тогда, когда равны амплитуды вторичных волн от одинаковых элементов объема ΔV среды. Из-за флуктуаций числа молекул газа в таких элементах объема точного равенства амплитуд вторичных волн не будет. Электрическое поле вторичной волны от i-го элемента объема в точке наблюдения можно представить в виде суммы $\mathbf{E}_i+\delta\mathbf{E}_i$, где $\mathbf{E}_i$ — напряженность поля вторичной волны при условии, что число молекул в этом элементе равно своему среднему значению (одинаковому для всех элементов), а $\delta\mathbf{E}_i$ — изменение напряженности, вызванное флуктуацией числа молекул (дополнительным дипольным моментом i-го элемента). Сумма напряженностей $\mathbf{E}_i$ полей когерентных волн от всех элементов объема равна нулю для любого направления (кроме $\theta=0$), поэтому для нахождения результирующего поля нужно сложить $\delta\mathbf{E}_i$. Интенсивность рассеянной волны

$$I\sim\Big(\sum_i \delta\mathbf{E}_i\Big)^2=\sum_{i,j}\delta\mathbf{E}_i\delta\mathbf{E}_j=\sum_{i\neq j}\delta\mathbf{E}_i\delta\mathbf{E}_j+\sum_i(\delta\mathbf{E}_i)^2.$$

Из-за независимого характера флуктуаций числа молекул идеального газа в разных элементах объема двойная сумма с $i\neq j$ обращается в нуль. Поэтому интенсивность рассеянного света равна сумме интенсивностей вторичных волн, обусловленных дипольными моментами, возникшими за счет флуктаций числа молекул в каждом элементе объема. Дополнительный дипольный момент i-го элемента объема равен $\varepsilon_0\alpha\mathbf{E}\delta N_i$, где α — поляризуемость молекулы, а δN_i — отклонение числа молекул в этом элементе от среднего значения. Интенсивность $I_i(r,\theta)$ рассеянного им света можно найти, повторяя рассуждения, которые привели к формуле (2.85). В результате получим выражение, отличающееся от (2.85) заменой поляризуемости $\alpha_1(\omega)$ взвешенной частицы на $\alpha\delta N_i$. Поэтому $I_i(r,\theta)\sim(\delta N_i)^2$.

Интенсивность света, рассеянного всеми элементами объема, можно получить, заменив $(\delta N_i)^2$ его средним значением $\langle(\delta N_i)^2\rangle$ и умножив на число q таких элементов во всем рассеивающем объеме V: $q=V/\Delta V$. Для идеального газа средний квадрат флуктуации числа частиц в выделенном малом элементе объема ΔV равен,

как известно, среднему числу частиц в этом элементе: $\langle (\delta N_i)^2 \rangle = \overline{N}\Delta V$, где N — концентрация молекул. Таким образом, $\langle (\delta N_i)^2 \rangle q = \overline{N}V$ и для полной интенсивности рассеянного идеальным газом света получаем

$$I(r, \theta) = \frac{1}{(4\pi)^2} \frac{\omega^4 \alpha^2}{c^4 r^2} \frac{1}{2}(1 + \cos^2\theta) N V I_0. \qquad (2.86)$$

Этот результат заслуживает обсуждения. Оказывается, идеальный газ рассеивает свет так, что интенсивности вторичных волн, испускаемых отдельными молекулами, просто складываются, несмотря на когерентность этих вторичных волн. Простой результат (2.86) получается только для идеального газа, когда средний квадрат флуктуации числа молекул в некотором объеме равен среднему числу молекул в этом объеме. Поэтому измерение интенсивности рассеянного газом света позволяет «сосчитать» рассеивающие молекулы и тем самым определить постоянную Авогадро N_A. Совпадение N_A со значениями, полученными другими методами, может служить доказательством молекулярного характера рассеяния света. Лишь в предельном случае сильно разреженного газа, в условиях, когда среднее расстояние между молекулами не мало по сравнению с длиной волны света, среду в пределах малых элементов объема уже нельзя рассматривать как сплошную. Тогда рассеяние действительно происходит независимо на каждой молекуле, подобно тому, как на взвешенных посторонних частицах в мутной среде.

Удобной характеристикой отдельных рассеивающих центров служит *эффективное сечение рассеяния* σ, определяемое как отношение потока энергии, рассеиваемой по всем направлениям, к интенсивности падающего излучения. Эта величина имеет размерность площади. Если в падающем излучении мысленно выделить поток, приходящийся на «мишень» такой площади σ, то как раз столько энергии и рассеивает центр по всем направлениям. Для нахождения этой энергии нужно интенсивность $I(r, \theta)$ излучения, рассеиваемого одной частицей, проинтегрировать по поверхности сферы радиусом r. Учитывая, что

$$\int \frac{1}{2}(1 + \cos^2\theta)\, d\Omega = \frac{1}{2}\int_0^{2\pi} d\varphi \int_0^{\pi} (1 + \cos^2\theta) \sin\theta\, d\theta = \frac{8\pi}{3},$$

из (2.85) находим

$$\sigma = \frac{1}{6\pi} \frac{\omega^4}{c^4} \alpha^2(\omega). \qquad (2.87)$$

Рассмотрим рассеяние света заряженным осциллятором, который моделирует поведение оптического электрона атома (или моле-

кулы) в классической теории дисперсии. Его поляризуемость $\alpha(\omega)$ была найдена в § 2.3 [см. (2.32)]. Вдали от собственной частоты ω_0 затуханием можно пренебречь, положив $\gamma=0$. Подставляя α в (2.87), получаем

$$\sigma=8\pi r_0^2\omega^4/[3(\omega_0^2-\omega^2)^2], \tag{2.88}$$

где

$$r_0=e^2/(4\pi\varepsilon_0 mc^2). \tag{2.89}$$

В случае рассеяния света свободным электроном в формуле (2.88) следует положить $\omega_0=0$. При этом частота ω выпадает, т. е. сечение рассеяния σ становится константой: $\sigma=(8\pi/3)r_0^2$. Оно пропорционально площади круга радиусом r_0. По этой причине $r_0=2{,}8\cdot10^{-13}$ см называют *классическим радиусом электрона.*

При рассеянии света в воздухе собственные частоты осцилляторов значительно больше частоты видимого света. Поэтому ω^2 в знаменателе (2.88) можно пренебречь по сравнению с ω_0^2 и сечение рассеяния оказывается пропорциональным четвертой степени частоты. Синий свет, частота которого примерно *вдвое* больше красного, рассеивается в *шестнадцать* раз интенсивнее. Этим, например, объясняется голубой цвет неба.

При приближении частоты ω света к собственной частоте ω_0 осциллятора сечение рассеяния, как видно из (2.88), резонансно возрастает. При $\omega\approx\omega_0$ пренебрегать затуханием уже нельзя. Можно показать (см. задачу 1), что при $\omega=\omega_0$ сечение $\sigma=\frac{2}{3}\pi r_0^2\omega^2/\gamma^2$. Если константа γ обусловлена радиационным затуханием осциллятора (сильно разреженный газ), то сечение рассеяния приближается к λ^2: в условиях резонанса эффективный размер «мишени», которую представляет атом для падающего излучения, в масштабах самого атома ($\approx10^{-8}$ см) становится огромным ($\approx10^{-5}$ см). Этот пример показывает, что локализация потока энергии световой волны для незамкнутой поверхности, размеры которой меньше длины волны, не имеет смысла (см. замечание в § 1.4).

Описываемое явление называется *резонансной флуоресценцией.* В стеклянный сосуд, откачанный до высокого вакуума и заполненный парами металла (например, натрия) с очень малой плотностью, направляют пучок света от газоразрядной лампы с парами того же металла. Сосуд начинает излучать по всем направлениям яркий свет той же частоты, что и у возбуждающей лампы (желтый в случае натрия). Энергия падающего пучка почти целиком может перейти в энергию рассеянного излучения. Как впервые показал Вуд, для этого необходима лишь достаточно низкая плотность паров металла, чтобы константа γ определялась радиационным затуханием, а не столкновениями, при которых часть приобретаемой ато-

мом световой энергии безызлучательно превращается в энергию хаотического теплового движения атомов. По той же причине в сосуде должно быть возможно меньше инородных молекул. В противном случае резонансная флуоресценция может ослабиться или даже совсем исчезнуть.

Выше было показано, что благодаря поперечности световой волны при наблюдении под прямым углом к направленню первичного пучка естественного света ($\theta=\pi/2$ на рис. 2.14) рассеянный свет должен быть полностью линейно поляризован в перпендикулярной первичному пучку плоскости. Однако при рассеянии в газе или жидкости с анизотропными молекулами поляризация рассеянного света обычно не бывает полиой. Объясняется это тем, что направленне вектора индуцированного падающей волной дипольного момента анизотропной молекулы не совпадает, вообще говоря, с направлением электрического поля волны. Деполяризация рассеянного света будет выражена тем сильнее, чем больше анизотропия поляризуемости молекул среды.

Изменение спектрального состава излучения при рассеянии обусловлено зависимостью эффективного сечения от частоты. Быстрое возрастание $\sigma\sim\omega^4$ в соответствии с формулой (2.87) (это относится и к рассеянию в мутных средах, и к молекулярному рассеянию, когда поляризуемость $\alpha(\omega)$ почти не зависит от частоты; исключение составляет только резонансная флуоресценция) приводит к тому, что распределение энергии в рассеянном свете отличается от первичного относительно большей величиной энергии в высокочастотной части спектра. Для монохроматического падающего излучения во всех рассмотренных выше случаях рассеянный свет характеризуется той же частотой. Такой тип рассеяния называют *рэлеевским.*

Однако при молекулярном рассеянии света в среде, содержащей многоатомные молекулы, в спектре рассеянного излучения наблюдаются добавочные линии (сателлиты), сопровождающие каждую из спектральных линий первичного света. Это явление было открыто в 1928 г. Л. И. Мандельштамом и Г. С. Ландсбергом в Москве и Раманом в Индии. Оно называется *комбинационным рассеянием света.* Происхождение сателлитов связано с *модуляцией* рассеянного света низкочастотными колебаниями атомов, образующих молекулу рассеивающей среды.

Поляризуемость молекулы зависит, вообще говоря, от расположения составляющих ее атомов. При колебаниях атомов поляризуемость изменяется около среднего значения α_0 в такт с этими колебаниями:

$$\alpha(t)=\alpha_0+F(t).$$

Часто́ты этих колебаний соответствуют инфракрасной области спектра (10^{12}—10^{13} Гц), т. е. изменения $\alpha(t)$ происходят медленно по сравнению с колебаниями электрического поля в падающем свете ($\approx10^{15}$ Гц). Поэтому можно считать, что в монохроматическом поле падающей волны изменение дипольного момента молекулы происходит по закону

Сравнивая это выражение с выражением для эффективного сечения σ, найденным в задаче 1, убеждаемся, что $\alpha = N\sigma = \eta$. Это значит, что мнимая часть показателя преломления $\varkappa$ в (2.50) учитывает в этом случае эффект рассеяния. ◇

3. Найти закон ослабления интенсивности параллельного пучка монохроматического света за счет молекулярного рассеяния в идеальном газе, показатель преломления n которого мало отличается от единицы.

◇ При рассеянии на флуктуациях плотности в идеальном газе интенсивность рассеянного света в соответствии с (2.86) равна сумме интенсивностей вторичных волн, рассеянных отдельными молекулами. Поэтому, как и в предыдущей задаче, $I(z) = I_0 \exp(-\eta z)$, где $\eta = N\sigma$, а для сечения σ можно воспользоваться формулой (2.87), выразив в ней поляризуемость молекулы α через показатель преломления:

$$n = \sqrt{\varepsilon} = \sqrt{1 + N\alpha} \approx 1 + N\alpha/2,$$

откуда $\alpha = 2(n-1)/N$. Подставляя α в (2.87), находим

$$\eta = \frac{2\omega^4}{3\pi c^4}\frac{(n-1)^2}{N}.$$

Эта формула была получена Рэлеем в 1881 г. ◇

4. Почему мы отчетливо видим облака (или туман), в то время как водяной пар, из которого они образовались при понижении температуры, совершенно не был виден, хотя, безусловно, этот пар содержался в воздухе до образования облаков (или тумана)?

Рассмотрим N молекул водяного пара. Пока они находятся в газообразной фазе, суммарное сечение рассеяния света этими молекулами пропорционально N. Предположим, что они сконденсировались в одну капельку тумана (размером много меньше длины волны света). Соответствующее сечение рассеяния определяется квадратом ее поляризуемости [см. (2.87)] и, следовательно, пропорционально N^2. Те же молекулы после конденсации рассеивают свет приблизительно в N раз сильнее. ◇

2.11. Скорость света. Фазовая и групповая скорости

Скорость света c в вакууме — одна из важнейших мировых констант. Согласно современным представлениям, выдающееся значение этой универсальной постоянной обусловлено тем, что она определяет *предельную скорость* распространения любых взаимодействий и сигналов, любых силовых полей независимо от их физической природы. Эта скорость одинакова во всех системах отсчета и обусловлена структурой пространства и времени, а не конкретным видом носителя сигнала — электромагнитными волнами. В частности, она определяет и скорость гравитационных волн в вакууме.

Универсальный характер скорости света в вакууме находится в глубоком противоречии с классическими представлениями о пространстве и времени, согласно которым при переходе от одной системы отсчета к другой скорость любого движения изменяется на величину относительной скорости этих систем. Установление абсолютной скорости потребовало формирования новых — релятивистских — представлений о пространстве и времени, составляющих

суть теории относительности. Эти представления утверждают относительный (т. е. зависящий от системы отсчета) характер одновременности событий, промежутков времени и пространственных расстояний. Ограничившись этими общими замечаниями о роли скорости света в физической теории, перейдем к обсуждению экспериментальных методов ее определения.

Скорость света впервые была измерена в 1676 г. датским астрономом Ремером, который правильно связал наблюдаемые нерегулярности затмений спутников Юпитера с временем прохождения света по диаметру орбиты Земли и получил таким образом правильное по порядку величины значение скорости света. Ныне наиболее кратко суть этого метода можно уяснить путем сравнения с эффектом Доплера: наблюдаемый период обращения спутника становится больше, когда Земля при движении по орбите удаляется от Юпитера, и меньше, когда приближается. Для относительного изменения периода можно использовать элементарное выражение $\Delta T/T = \pm v/c$, справедливое, когда относительная скорость источника и наблюдателя v много меньше скорости света. Знак плюс соответствует увеличению расстояния, минус — уменьшению.

Другой астрономический метод определения скорости света основан на явлении аберрации света (см. § 8.2), которое было открыто английским астрономом Брэдли в 1725—1728 гг. Это явление заключается в кажущемся смещении положений звезд, вызываемом движением Земли по орбите. Звезды, расположенные в направлении нормали к плоскости орбиты Земли, описывают в течение года на небесной сфере окружности с угловым диаметром около 41″. В соответствии с теорией этот угловой диаметр равен $2v/c$, что позволяет определить c. Наиболее точные измерения аберрации дают $c = 2{,}999 \cdot 10^8$ м/с.

Измерение скорости света от земного источника в лабораторных условиях впервые было выполнено Физо в 1849 г. Пучок света прерывался зубчатым колесом, вращавшимся перед источником света, и отражался от зеркала, находившегося на расстоянии около 9 км. Если за время движения светового импульса до зеркала и обратно колесо повернется на такой угол, что на месте прорезей окажутся зубья, вернувшийся свет не попадет в окуляр и поле зрения окажется темным. При вдвое большей угловой скорости вернувшийся световой импульс проходит через следующую прорезь и наблюдатель видит источник. Очевидно, что в этом случае для определения скорости света нужно разделить путь от колеса до зеркала и обратно на время поворота колеса на один зубец. Современная модификация метода Физо основана на прерывании света с помощью практически безынерционного оптического затвора (конденсатора Керра, см. § 4.5). Это позволяет значительно повысить точность, несмотря на сокращение длины базиса до нескольких метров.

Другой лабораторный метод измерения скорости света, основанный на использовании отражения света от вращающегося зеркала,

был впервые осуществлен Фуко в 1862 г. и значительно усовершенствован Майкельсоном в 1926 г. В установке Майкельсона за время прохождения отраженным световым импульсом расстояния до неподвижного зеркала и обратно восьмигранное зеркало успевало повернуться на $^1/_8$ оборота так, чтобы отражающая поверхность заменялась соседней.

Измерения скорости производились не только для видимого света, но и в других диапазонах электромагнитных волн. В частности, для радиосигналов измерения времени прохождения известного расстояния дали значение скорости, совпадающее в пределах ошибок измерения со скоростью видимого света, что вновь подтверждает справедливость электромагнитной теории света.

Все упомянутые выше способы измерения скорости света основаны на прерывании света, т. е. на использовании *сигналов*, передаваемых с помощью *модулированных* световых волн. Поэтому фактически здесь измеряется не фазовая скорость электромагнитных волн, а скорость светового импульса. В вакууме эти скорости одинаковы, но в среде, обладающей дисперсией, монохроматические составляющие разных частот, входящие в состав модулированной волны, распространяются с разными скоростями, что приводит к совершенно особому характеру распространения импульса. В простейшем случае слабой дисперсии импульс распространяется без искажений с так называемой групповой скоростью (см. ниже).

Для непосредственного измерения *фазовой* скорости электромагнитных волн требуются принципиально иные методы. В области радиочастот можно использовать стоячие волны в объемном резонаторе. Теория позволяет связать размеры резонатора и его резонансную частоту с фазовой скоростью электромагнитных волн. В случае плоского резонатора длины l эта связь имеет вид $\nu_k = kc/(2l)$, где k — целое число (см. § 1.3). Измеряя в вакууме частоту резонанса непосредственным сравнением с эталоном частоты и длину резонатора — сравнением с эталоном длины интерференционными методами, Л. Эссен в 1950 г. получил для скорости электромагнитных волн длиной около 10 см значение $c = (299792{,}5 \pm 1)$ км/с.

В 1972 г., когда были разработаны прецизионные методы измерения частоты в оптическом диапазоне, значение скорости света было определено в Национальном бюро стандартов США К. Ивенсоном с сотрудниками на основе независимых измерений длины волны и частоты света. В качестве источника использовался стабилизированный гелий-неоновый лазер, генерирующий близкое к монохроматическому инфракрасное излучение ($\lambda = 3{,}39$ мкм). Частота ν этого излучения измерялась сравнением с цезиевым эталоном частоты (времени) с использованием методов нелинейной оптики. Длина волны λ этого же излучения измерялась с большой точностью интерферометрическим сравнением с эталоном длины, т. е. с длиной волны оранжевой линии криптона-86. Для этого пришлось исполь-

зовать ряд вспомогательных монохроматических источников с промежуточными значениями частоты и длины волны. Несмотря на длинную цепочку измерений, таким методом удалось получить значение скорости света в вакууме с громадной точностью: $c=\lambda\nu=$ $=(299792456{,}2\pm1)$ м/с. Фактически этот метод можно рассматривать как сравнение *частоты* излучения оранжевой линии криптона (длина волны которого лежит в основе *эталона длины*) с частотой *эталона времени* на основе цезия-133. Опыт показывает, что точность такого сравнения лимитируется точностью первичного эталона длины. В связи с этим в метрологии на повестку дня встает вопрос об отказе от независимо выбранных эталонов длины и времени и о переходе на *единый эталон*. Длина волны излучения эталонного источника должна лежать в основе определения единицы длины, а частота (период) — в основе единицы времени. При этом числовое значение c будет получено не как результат *измерений* (неизбежно содержащий некоторую погрешность), а введено *по определению* (т. е. *точно*) на основе международного соглашения. Разумеется, это значение должно быть выбрано так, чтобы обеспечивалась преемственность с прежними эталонами длины и времени.

Обратимся теперь к вопросу о *скорости света в веществе*. Лабораторные методы с использованием короткого базиса позволяют определять скорость света в различных средах. Так, для воды Майкельсон нашел $c/v=1{,}33$, что находится в хорошем соответствии со значением показателя преломления воды, полученным из измерений на основе преломления света на границе. Во многих случаях значение n, полученное как отношение синусов угла падения и угла преломления, хорошо согласуется со значением, найденным из измерения скорости света. Однако в некоторых случаях возникают серьезные расхождения. Например, для сероуглерода Майкельсон нашел $c/v=1{,}76$, тогда как обычное определение показателя преломления дает $n=1{,}64$. Объяснение расхождений было дано Рэлеем, выяснившим сложный характер самого понятия скорости волнового движения. С этим понятием связан широкий круг проблем, касающихся кинематики волнового движения и поэтому общих для волн разной физической природы.

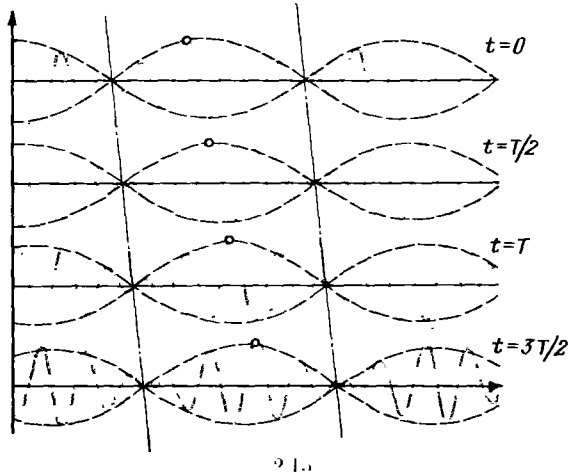
Понятие скорости возникло в механике для описания движения *частицы* (материальной точки). В волновом движении происходит перенос *состояния* (т. е. значений поля) из одного места в другое. В общем случае понятие скорости здесь неоднозначно. Остановимся на этом вопросе подробнее.

В среде без дисперсии (для электромагнитных волн, строго говоря, только вакуум дает пример такой «среды», хотя вода, воздух и многие газы характеризуются очень слабой дисперсией в видимой области) всякое возмущение распространяется без изменения своей формы, и введение скорости волнового движения не вызывает затруднений. Но в среде с *дисперсией* возмущение по мере распространения *деформируется*, и понятие скорости становится

неопределенным. В таком случае требуется сначала сказать, что мы будем *по определению* называть скоростью распространения. Например, для движения изменяющего свою форму облака не существует однозначного понятия скорости: можно говорить и о скорости его переднего края («фронта»), и о скорости его центра тяжести, и т. д. Аналогично обстоит дело со скоростью волнового движения в диспергирующей среде. Здесь приходится вводить ряд понятий: скорость группы, скорость энергии, скорость фронта, скорость сигнала и т. д.

Понятие фазовой скорости, использовавшееся во всем предыдущем изложении, при наличии дисперсии применимо только к монохроматической волне, бесконечно протяженной в пространстве и во времени. Но такая волна непригодна для передачи сигнала, и сама постановка подобных вопросов требует отказа от монохроматической идеализации. В любом опыте (в том числе и в описанных выше опытах по измерению скорости света в веществе) мы всегда имеем более или менее сложный импульс или, как говорят, «волновой пакет», ограниченный в пространстве и во времени. При определенных условиях деформация («расплывание») волнового пакета происходит медленно и можно говорить о его скорости как о скорости какой-либо точки пакета, например точки максимальной амплитуды. Стокс впервые обратил внимание на то, что скорость пакета будет отличаться от фазовой скорости любой из составляющих его монороматических волн.

Для лучшего уяснения сущности этого вопроса рассмотрим модулированную волну, возникающую в результате наложения двух монохроматических волн одинаковой амплитуды с близкими значениями длины волны λ и $\lambda+\Delta\lambda$. В результирующей волне максимумы ам-



Группы волн, возникающие при сложении двух волн с длинами λ и $\lambda+\Delta\lambda$

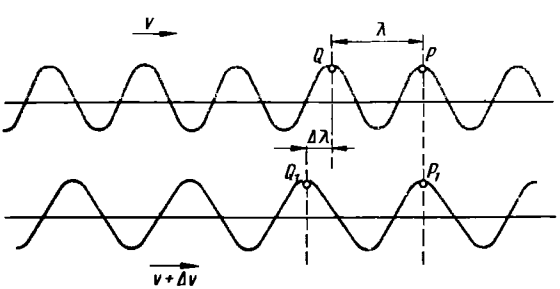
плитуды будут в тех местах, где «горб» одной волны совпадает с горбом другой. Там, где горб одной волны совпадает с «впадиной» другой, результирующая волна имеет нулевую амплитуду. Все вместе это выглядит как последовательность отдельных *волновых групп* (рис. 2.15). При $\Delta\lambda \ll \lambda$ каждая группа содержит большое число периодов.

Если скорости складываемых волн одинаковы, то результирующая модулированная волна распространяется с той же скоростью, не изменяя своей формы. При наличии дисперсии монохроматические составляющие имеют несколько различающиеся скорости v и $v+\Delta v$. Поэтому взаимное расположение их горбов и впадин меняется с течением времени (рис. 2.15). И хотя общий вид результирующей волны остается прежним, центры отдельных групп с течением времени смещаются относительно положения отдельных горбов и впадин, т. е. движутся с иной скоростью, нежели складываемые монохроматические волны. Скорость движения центров этих групп называют *групповой скоростью*. Найдем ее.

Будем для определенности считать, что скорость монохроматических волн растет с увеличением длины волны (нормальная дисперсия). Тогда нижняя волна на рис. 2.16, имеющая длину $\lambda+\Delta\lambda$, обгоняет верхнюю волну с длиной λ. Пусть в какой-то момент времени совпадают горбы P и P_1, т. е. центр группы волн приходится на точку P. Через некоторое время τ горб P_1 обгонит P, но зато совпадут горбы Q и Q_1. Это значит, что центр группы волн за это время сместился назад на одну длину волны λ и совпадает с точкой Q. Поэтому скорость u центра группы меньше фазовой скорости верхней волны на $\lambda/\tau : u=v-\lambda/\tau$. Время τ, в течение которого горб Q_1 догоняет Q, как легко видеть из рис. 2.16, равно $\Delta\lambda/\Delta v$. Поэтому выражение для групповой скорости (*формула Рэлея*) в пределе $\Delta\lambda \to 0$ принимает вид

$$u = v - \lambda \, dv/d\lambda. \qquad (2.90)$$

Когда фазовая скорость монохроматических волн v не зависит от длины волны (т. е. отсутствует дисперсия), $dv/d\lambda=0$ и групповая скорость совпадает с фазовой. При нормальной дисперсии



2.16

К выводу формулы для групповой скорости волн

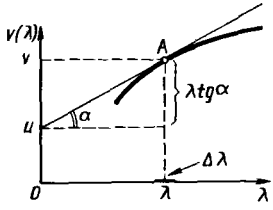
$dv/d\lambda > 0$ и групповая скорость меньше фазовой: $u < v$. В области аномальной дисперсии $dv/d\lambda < 0$ и формула (2.90) дает $u > v$.

Например, в случае гравитационных (т. е. обусловленных силой тяжести) волн на поверхности воды в глубоком водоеме фазовая скорость пропорциональна $\sqrt{\lambda}$: $v = C_1\sqrt{\lambda}$. Применяя формулу Рэлея, получаем $u = v/2$ — центр группы таких волн движется вдвое медленнее, чем отдельные горбы и впадины. В случае коротких ($\lambda < 1{,}7$ см) капиллярных волн, обусловленных силами поверхностного натяжения, $v = C_2/\sqrt{\lambda}$ и $u = {}^3/_2 v$ — центр группы бежит в полтора раза быстрее.

Во всех опытах по измерению скорости света, основанных на прерывании (модуляции) света, измеряется именно групповая скорость. Это относится и к астрономическим методам Ремера и Брэдли, хотя здесь свет распространяется в вакууме, где нет дисперсии и групповая скорость совпадает с фазовой. В опытах Майкельсона с водой и сероуглеродом измерялась групповая скорость, но для воды в видимой области значение $dv/d\lambda$ настолько мало, что практически $u = v$, поэтому и получается $c/u \approx c/v = n$. В сероуглероде $\lambda\, dv/d\lambda$ дает заметный вклад и $u < v = c/n$, что и обнаружил Майкельсон. Тщательное измерение зависимости $n(\lambda)$ для сероуглерода показывает, что найденная Майкельсоном величина действительно сооответствует групповой скорости, выражаемой формулой Рэлея.

Существует простой графический способ нахождения групповой скорости по кривой $v(\lambda)$, предложенный Эренфестом. Как ясно из рис. 2.17, отрезок, отсекаемый на оси ординат касательной к кривой $v(\lambda)$, проведенной в какой-либо точке A этой кривой, равен как раз $v - \lambda\, dv/d\lambda$, т. е. групповой скорости при данной длине волны λ.

Формула Рэлея (2.90) справедлива не только для скорости перемещения огибающей бесконечной череды волновых групп, которая получается при сложении двух монохроматических волн. При определенных условиях она характеризует также скорость движения центра *одиночного* волнового пакета, образованного непрерывным набором монохроматических составляющих. Эти условия касаются как самого волнового возмущения, так и свойств среды, в которой оно распространяется: для тех длин волн из интервала $\Delta\lambda$, которые входят в состав рассматриваемого возмущения, фазовая скорость $v(\lambda)$ монохроматических волн в среде должна с хорошим приближением представлять линейную функцию длины волны. Другими словами, групповая скорость (2.90) в пределах интервала длин волн $\Delta\lambda$ должна быть постоянной. Смысл этого условия легко понять из рис. 2.17: если огибающая суммы двух монохроматических составляющих из интервала $\Delta\lambda$ перемещается с одной и той



2.17
Графическое определение групповой скорости

же скоростью u, то и весь импульс будет перемещаться с этой скоростью, не изменяя своей формы. В таких условиях групповая скорость (2.90) может служить адекватной характеристикой скорости волнового процесса.

Рассмотрим вопрос о *скорости распространения энергии*, переносимой электромагнитной волной. Пусть в среде имеется монохроматическая волна. Энергия пропорциональна квадрату напряженности поля, а так как одинаковые значения напряженности (например, максимумы) перемещаются с фазовой скоростью, то, казалось бы, с этой же скоростью будет распространяться и энергия. Однако в действительности такое утверждение бессодержательно, так как оно не может быть проверено экспериментально. При распространении неограниченной в пространстве монохроматической волны средняя за период энергия произвольного элемента объема будет оставаться постоянной и ни о каком наблюдаемом потоке энергии сквозь какую-либо поверхность говорить не приходится. Изменение энергии в каком-то объеме среды можно констатировать лишь тогда, когда мы имеем ограниченный волновой цуг или группу волн. Поэтому движение энергии следует отождествлять с перемещением *изменений амплитуды*, и в тех случаях, когда групповая скорость имеет смысл (т. е. волновой цуг или импульс распространяется не расплываясь), она совпадает со скоростью переноса энергии:

$$\langle S \rangle = \langle w \rangle u.$$

Формула Рэлея (2.90) удобна для вычисления групповой скорости в тех случаях, когда фазовая скорость задана как функция длины волны: $v=v(\lambda)$. Однако часто бывает удобно вместо λ использовать волновое число (модуль волнового вектора) $k=2\pi/\lambda$ и рассматривать $v=v(k)$. Тогда вместо (2.90) можно написать

$$u=v+k\,\mathrm{d}v/\mathrm{d}k. \tag{2.91}$$

Если выразить здесь фазовую скорость через частоту $v=\omega/k$ и рассматривать частоту как функцию волнового числа $\omega=\omega(k)$, то от (2.91) придем к другой (эквивалентной) формуле для групповой скорости:

$$u=\mathrm{d}\omega(k)/\mathrm{d}k. \tag{2.92}$$

Еще одна эквивалентная форма соотношения (2.90) полезна тогда, когда свойства среды характеризуются показателем преломления, заданным как функция частоты $n=n(\omega)$:

$$u=c/[n+\omega(\mathrm{d}n/\mathrm{d}\omega)]. \tag{2.93}$$

Пусть, например, модулированная волна или отдельный волновой импульс (с достаточно узким спектром) распространяется в прозрачной разреженной среде, показатель преломления которой

мало отличается от единицы. Вдали от собственной частоты ω_0 атомных осцилляторов их затуханием можно пренебречь, и зависимость показателя преломления от частоты выражается формулой (2.39):

$$n(\omega) \approx 1 + \omega_p^2/[2(\omega_0^2 - \omega^2)].$$

При $\omega > \omega_0$ фазовая скорость $v = c/n$ оказывается больше скорости света в вакууме, что, как уже отмечалось, не противоречит теории относительности, так как фазовая скорость характеризует бесконечно протяженную монохроматическую волну, которая не может служить для передачи сигнала. Если волну с частотой $\omega > \omega_0$ промодулировать, то распространение модуляции (т. е. изменений амплитуды, которые можно рассматривать как сигналы) будет происходить с групповой скоростью. Вычисление u по формуле (2.93) дает

$$u = c\left[1 + \frac{\omega_p^2(\omega_0^2 + \omega^2)}{2(\omega_0^2 - \omega^2)^2}\right]^{-1} < c.$$

Эта величина меньше скорости света в вакууме как при $\omega < \omega_0$, так и при $\omega > \omega_0$, где фазовая скорость больше c.

Другой пример — распространение волн в среде со свободными электронами (плазма, металлы). Здесь в соответствии с (2.53) $n(\omega) = \sqrt{1 - \omega_p^2/\omega^2}$. Поэтому в области прозрачности, т. е. при $\omega > \omega_p$, фазовая скорость всегда больше c, а для групповой скорости в результате вычисления по формуле (2.93) получаем

$$u = c\sqrt{1 - \omega_p^2/\omega^2} < c,$$

причем $uv = c^2$. И здесь распространение изменений амплитуды происходит со скоростью, меньшей c.

Однако в некоторых случаях и групповая скорость, вычисляемая по формуле Рэлея (2.90) или любой из эквивалентных ей формул (2.91) — (2.93), оказывается больше c. Так будет, например, в области аномальной дисперсии, в чем можно убедиться, воспользовавшись формулой (2.51) для $n(\omega)$ вблизи собственной частоты атомных осцилляторов. Здесь тоже нет противоречия с теорией относительности, ибо групповая скорость отнюдь не во всех случаях выражает скорость сигнала. Как уже отмечалось, волновой импульс в процессе распространения не изменяет своей формы и перемещается с групповой скоростью, если фазовые скорости его монохроматических составляющих зависят от частоты по линейному закону, т. е. групповая скорость постоянна в пределах занимаемой импульсом полосы частот. Когда изменение показателя преломления на этом интервале частот происходит нелинейно, импульс по мере распространения неизбежно деформируется. Кроме таких фазовых искажений, связанных с непостоянством групповой скорости в пре-

делах занимаемой импульсом полосы частот, причиной деформации импульса может быть зависящее от частоты поглощение, особенно существенное в области аномальной дисперсии (или усиление в случае активной среды). При этом по мере распространения импульса происходит искажение его энергетического спектра; изменяется и частота, соответствующая максимуму спектра (несущая частота), так что на больших расстояниях импульс может стать совершенно не похожим на то, что было послано. В поглощающей среде в определенных условиях потери «выедают» преимущественно заднюю часть импульса, что увеличивает скорость перемещения его максимума. В усиливающей среде передняя часть импульса проходит при наибольшей инверсии населенностей (см. § 9.3) и усиливается в большей мере, чем задняя, проходящая по «обедненной» среде. В таких условиях на конечном интервале расстояний максимум огибающей может перемещаться со скоростью, превосходящей скорость света в вакууме, или даже с отрицательной скоростью, но это не противоречит ни релятивистской причинности, ни здравому смыслу, так как эта скорость связана с «внутренней» перестройкой импульса и не может быть использована для передачи сигнала. Что касается групповой скорости, определяемой формулой Рэлея, то она в таких условиях утрачивает определенное физическое содержание.

В случае импульса с резко ограниченным передним краем, впереди которого никакого возмущения нет, можно говорить о скорости распространения его *фронта* (т. е. переднего края). Оказывается, что скорость такого фронта в любой среде совпадает со скоростью света в вакууме. В этом легко убедиться на основе электронной теории. К моменту прихода фронта электрон атомного осциллятора находится в покое и благодаря своей инерции не сразу начнет излучать вторичные волны (из-за которых изменяются структура и скорость результирующей волны). Поэтому фронт любого сигнала проходит через среду со скоростью c, как через пустое пространство, и действие среды сказывается лишь позади фронта. Отсюда понятно, что скорость сигнала, как бы она ни была определена, всегда меньше или равна скорости света в вакууме в согласии с требованием теории относительности.

Скорость сигнала (или *сигнальная скорость*) в случае среды с дисперсией требует специального определения, на что впервые было указано Зоммерфельдом. Физически содержательное определение сигнальной скорости должно учитывать не только свойства среды, но и чувствительность приемников излучения в рассматриваемой области спектра. Дело в том, что в процессе распространения импульс может стать очень пологим, так что его переднему краю будет соответствовать слишком малая энергия, а все приемники излучения имеют конечную чувствительность. Поэтому понятие скорости фронта соответствует идеализированному случаю предельно чувствительного регистрирующего прибора.

Контрольные вопросы

- ☐ Опишите осиовные астрономическне и лабораторные методы нзмереиия скорости света.
- ☐ Какие методы дают значение фазовой, а какие — групповой скорости?
- ☐ С каким свойством скорости света связана возможиость введения единого эталона длины и времени?
- ☐ Какой физический смысл имеет групповая скорость?
- ☐ Как групповая скорость выражается через фазовую скорость монохроматических волн?
- ☐ С какой скоростью происходит перенос энергии в электромагнитной волне?
- ☐ В каких случаях групповая скорость волн утрачивает физический смысл?
- ☐ С какой скоростью распространяется передний фронт светового возмущения в среде?
- ☐ Доказать эквивалентность формул (2.90) — (2.93).
- ☐ Показать, что групповая скорость может быть вычислена по формуле

$$u=v\left(1+\frac{\lambda}{n}\frac{\mathrm{d}n}{\mathrm{d}\lambda}\right).$$

2.12. Излучение Вавилова — Черенкова

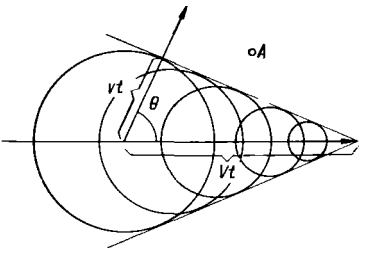
Электрическое поле падающей волны вызывает вынужденные колебания электронов, входящих в состав атомов и молекул среды, и они становятся источниками вторичных волн. В однородной среде эти вторичные волны когерентны, и в результате их сложения между собой и с падающей волной образуется преломленная волна.

Вынужденное движение электронов среды может быть вызвано не только падающей волной, но и иными способами, например электрическим полем движущейся в среде заряженной частицы (электрона, протона и т. п.). При этом отдельные элементы объема среды вдоль траектории заряда также становятся когерентными источниками электромагнитных волн. Если заряд движется *равномерно* со скоростью V, меньшей фазовой скорости света в среде $v(V<v)$, то волны, исходящие от разных частей траектории заряда, при сложении полностью гасят друг друга. Так будет для волн, распространяющихся в любом направлении.

Условие полного погашения волн перестает выполняться тогда, когда скорость заряда *превосходит* фазовую скорость света в среде. При $n>1$ скорость света $v=c/n<c$ и движение заряда со скоростью V в интервале $v<V<c$ не противоречит теории относительности. Движущийся с такой скоростью заряд встречает лежащие на его пути электроны среды раньше, чем к ним может прийти излучение, испущенное возбужденными перед этим электронами. Здесь мы вправе ожидать явлений, подобных тем, что хорошо известны в акустике: снаряд, движущийся со скоростью, превышающей скорость звука в среде, обгоняет созданную им волну давления и оставляет за собой скачок давления в виде «конуса Маха»

(рис. 2.18). Нормаль к фронту акустической волны составляет угол θ с направлением движения снаряда, определяемый отношением скорости звука v к скорости снаряда V:

$$\cos\theta = v/V. \qquad (2.94)$$

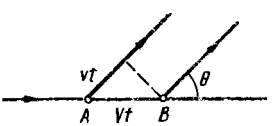


2.18
Фронт акустической волны при движении снаряда со сверхзвуковой скоростью

Аналогичное оптическое явление — излучение света при равномерном движении заряда в среде со скоростью $V>v=c/n$ — было открыто в 1934 г. С. И. Вавиловым и П. А. Черенковым и теоретически объяснено в 1937 г. И. Е. Таммом и И. М. Франком. За его открытие и объяснение Черенкову, Тамму и Франку была присуждена Нобелевская премия 1958 г. по физике.

Для определения направления волны излучения Вавилова — Черенкова рассмотрим два любых одинаковых элемента объема A и B на траектории заряда (рис. 2.19). Элемент B испускает точно такую же волну, что и элемент A, но с запаздыванием на время t, в течение которого заряд проходит расстояние AB. Эти волны придут в удаленную точку наблюдения одновременно, если направление на нее определяется таким углом θ, что $\cos\theta=v/V$, ибо волне из A потребуется как раз на t больше времени для прихода в точку наблюдения. Все когерентные волны от разных элементов объема среды на траектории заряда, распространяющиеся в этом направлении, имеют одинаковую фазу и при сложении (интерференции) усиливают друг друга. Для любого другого направления всегда можно разделить все элементы объема вдоль траектории на такие пары, чтобы испускаемые ими волны были в противофазе и гасили друг друга.

Условие, определяющее направление излучения Вавилова — Черенкова, совпадает с условием (2.94) для направления акустической волны. Это и не удивительно, так как оно имеет чисто кинематический характер и поэтому справедливо для волн любой природы. Отметим еще раз, что оно выполняется только при $V>v$ — скорость источника должна быть больше скорости волн в среде. Фронт излучаемой волны представляет собой конус, вершина которого перемещается вместе с зарядом со скоростью V (см. рис. 2.18). Приемник излучения, находящийся в какой-либо точке A, зафиксирует отдельную вспышку света в тот момент, когда фронт пройдет через эту точку.



2.19
Направление излучения Вавилова — Черенкова

Отметим, что для возникновения излучения существенна нестационарность возму-

щения среды полем пролетающей частицы. Если бы вместо одной частицы текла непрерывная заряженная жидкость, то и при скорости $V>c/n$ излучение Вавилова — Черенкова не возникло бы. В случае пучка частиц статистические нерегулярности приводят к взаимной некогерентности излучений, создаваемых отдельными частицами, так что их интенсивности просто складываются.

Выше молчаливо предполагалось, что в среде отсутствует дисперсия. Существующая в реальных средах зависимость показателя преломления и, следовательно, фазовой скорости света от частоты $v=c/n(\omega)$ приводит к тому, что при заданной скорости V заряда черенковское излучение возможно только на частотах ω, для которых $n(\omega)>c/V$. Поэтому спектр излучения Вавилова — Черенкова лежит преимущественно в видимой области, обрываясь на высоких частотах, когда из-за приближения показателя преломления к единице условие $n(\omega)>c/V$ перестает выполняться. В частности, ни при какой скорости заряда в спектре черенковского излучения не может быть рентгеновских лучей, так как для них $n<1$. Излучаемый свет поляризован так, что напряженность электрического поля лежит в плоскости, образуемой лучом и направлением движения заряда.

Энергия излучаемого света черпается из кинетической энергии движущегося заряда, скорость которого должна при этом уменьшаться. Но само по себе черенковское излучение отнюдь не связано с торможением электрона, ибо оно должно иметь место и при *равномерном* движении заряда со скоростью $V>v=c/n$. Конечно, такое равномерное движение вследствие потерь энергии на излучение не может быть движением по инерции, но в принципе эти потери всегда можно компенсировать с помощью постоянной внешней силы. Если энергия частицы велика по сравнению с потерями на излучение, то роль этих потерь становится совершенно несущественной*. Поэтому приближение, при котором движение частицы, несмотря на излучение, считается равномерным, имеет широкую область применимости.

Излучение будет происходить и тогда, когда заряд движется не в самом веществе, а вдоль оси проделанного в нем канала, параллельно краям узкой щели или просто параллельно границе. Для возникновения волн в среде в таких условиях кроме прежнего условия $V>v$ требуется еще, чтобы диаметр канала или расстояние от траектории заряда до границы были меньше длины волны излучения.

Практическое применение излучение Вавилова — Черенкова получило в физике высоких энергий для регистрации быстрых заряженных частиц. На нем основано действие так называемых черен-

---

* Потери энергии вследствие излучения обычно гораздо меньше так называемых ионизационных потерь, связанных с возбуждением и ионизацией атомов вещества пролетающей частицей.

ковских счетчиков, с помощью которых можно разделить релятивистские частицы, обладающие одинаковым импульсом, но разной массой и, следовательно, разной скоростью.

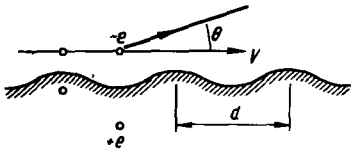
Может ли возникать излучение при равномерном движении заряженной частицы со скоростью, *меньшей* скорости света в среде? Из предыдущего рассмотрения напрашивается отрицательный вывод: при $V<c/n$ вторичные волны от разных элементов объема вдоль траектории гасят друг друга для любого направления наблюдения и только движение заряда с *ускорением* может привести к излучению*. Однако в среде для возникновения излучения важно не изменение самой скорости V заряда, а изменение отношения этой скорости к фазовой скорости света V/v. Отсюда ясно, что излучение возникает и при постоянной скорости V заряда, если вдоль его траектории изменяется показатель преломления. Полного гашения когерентных волн от элементов объема в подобных условиях не происходит.

Именно такая ситуация складывается, например, при пересечении равномерно движущейся заряженной частицей границы вакуума со средой или границы раздела двух сред. На возникающее при этом *переходное излучение* было указано В. Л. Гинзбургом и И. М. Франком в 1944 г., а первые экспериментальные результаты (с использованием протонов с энергией от 1 до 5 МэВ) были опубликованы только в 1959 г., хотя, как теперь ясно, такое свечение уже давно наблюдалось на анодах рентгеновских трубок. В случае пересечения зарядом границы металла происхождение переходного излучения становится особенно наглядным. Когда заряженная частица находится вблизи плоской поверхности металла, напряженность электрического поля вне металла совпадает с напряженностью поля диполя, образованного зарядом и его «электрическим изображением». В момент пересечения поверхности металла заряд и его изображение одновременно исчезают (поле заряда, находящегося в металле, практически полностью экранируется электронами проводимости). Возникающее при этом переходное излучение будет (вне металла) точно таким же, как при мгновенной остановке в одной точке двух движущихся навстречу зарядов противоположного знака, т. е. эквивалентно соответствующему тормозному излучению.

Аналогичное явление можно наблюдать и при равномерном движении заряда (с досветовой скоростью) параллельно границе. Излучение возникает, если граница обладает некоторым рельефом, т. е. не представляет собой идеальную плоскость. В опытах Смита и Парселла (1953) электроны с энергией $eU=300$ кэВ двигались в вакууме над поверхностью металлической дифракционной решетки с периодом $d=1{,}67$ мкм перпендикулярно ее штрихам (рис. 2.20).

---

* Излучение, возникающее при столкновении заряженных частиц, называют *тормозным*, а при ускорении под действием магнитного поля — *магнитотормозным* или (для релятивистских частиц) *синхротронным*.



2.20
Излучение при движении заряда вдоль рельефной поверхности

Отрицательный заряд электрона вместе с индуцированным положительным зарядом на поверхности металла (который можно заменить «изображением» электрона) образуют диполь. Расстояние между обоими зарядами и, следовательно, дипольный момент при равномерном движении электрона со скоростью V изменяются периодически с периодом $T = d/V$, где d — расстояние между штрихами решетки. В результате возникает дипольное излучение. Так как излучающий диполь движется со скоростью V, в лабораторной системе отсчета частота ω' и длина волны λ' излучаемого им света зависят (вследствие эффекта Доплера, см. § 8.3) от направления наблюдения:

$$\omega' = \omega/(1 - V\cos\theta/c), \qquad \lambda' = d(c/V - \cos\theta).$$

Для $\theta \approx 15°$ расчет по этой формуле дает $\lambda' = 0{,}5$ мкм, что соответствует видимой области. Траектория электронов имеет вид светящейся полосы, цвет которой изменяется в зависимости от угла наблюдения θ.

Контрольные вопросы

- ☐ При каких условиях равномерное движение заряда сопровождается излучением?
- ☐ Какое влияние оказывает дисперсия среды на черенковское излучение? Почему в его спектре не может быть рентгеновских лучей?
- ☐ Может ли возникать излучение при равномерном движении заряда в среде со скоростью, меньшей скорости света?

ОТРАЖЕНИЕ И ПРЕЛОМЛЕНИЕ СВЕТА НА ГРАНИЦЕ

● Когда свет достигает границы раздела двух сред с разными оптическими свойствами (или границы среды с вакуумом), он частично проходит во вторую среду, изменяя направление в случае наклонного падения, и частично возвращается в первую среду. Направление отраженного и преломленного света описывается хорошо известными законами геометрической оптики. Однако эти законы ничего не говорят о поляризации и интенсивности отраженного и преломленного света. В данной главе ответ на эти вопросы (как и вывод известных законов отражения и преломления) дается на основе электромагнитной теории света.

● Появление преломленной и отраженной световых волн на границе раздела сред обусловлено теми же физическими причинами, что и изменение фазовой скорости волны при ее распространении в неограниченной среде по сравнению со скоростью света в вакууме: электрическое поле падающей волны раскачивает входящие в состав вещества среды заряженные частицы, которые становятся источниками вторичных волн. Задача нахождения отраженной и преломленной волн, возникающих в результате сложения этих когерентных вторичных волн, может быть решена в рамках макроскопической электродинамики, т. е. с помощью уравнений Максвелла и феноменологических материальных уравнений, в которых среды рассматриваются как сплошные, а их оптические свойства задаются показателями преломления. При этом законы отражения и преломления, а также выражаемые формулами Френеля соотношения между амплитудами и фазами падающей, отраженной и преломленной волн получаются как следствие граничных условий для электромагнитного поля, вытекающих из уравнений Максвелла.

3.1. Законы отражения и преломления света

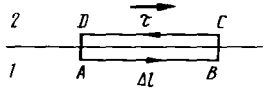
Закономерности отражения и преломления света относятся к наиболее ранним экспериментальным открытиям в оптике. Закон отражения был известен еще Архимеду; открытие закона преломления связывают с именами Снеллиуса и Декарта.

Для получения этих законов на основе электромагнитной теории рассмотрим идеализированный случай бесконечной плоской границы раздела двух неподвижных однородных изотропных сред, каждая из которых занимает целое полупространство. Пусть в одной из этих сред задана приходящая из бесконечности плоская монохроматическая волна. Эта падающая на границу волна, поверхности постоянной фазы которой представляют собой неограниченные плоскости, порождает волновой процесс в обеих средах, который мы собираемся исследовать.

Практически достаточно, чтобы перечисленные условия выполнялись хотя бы приближенно. Например, падающая волна может быть сферической (от источника, находящегося достаточно далеко от границы), тогда ее небольшие участки приближенно рассматриваются как плоские и к ним применимы результаты, полученные для неограниченных плоских волн. Аналогично обстоит дело в случае неплоской границы раздела, отдельные участки которой можно приближенно рассматривать как плоские. Для этого размеры таких участков должны быть велики по сравнению с длиной волны.

Монохроматичность падающей волны предполагает установившийся характер всех процессов. Полное электромагнитное поле, включающее падающую, отраженную и преломленную волны, должно удовлетворять определенным *граничным условиям*, которые могут быть получены предельным переходом из уравнений Максвелла. Эти условия заключаются в непрерывности тангенциальных составляющих векторов **E** и **B** на границе.

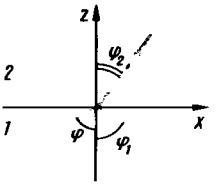
Для получения граничных условий можно взять контур в виде небольшой прямоугольной петли $ABCD$ (рис. 3.1), стороны AB и DC которой параллельны границе раздела сред и проходят по разные стороны от нее. Применим к контуру уравнение Максвелла (1.12) [или (2.9)] в интегральной форме и устремим длины сторон AD и BC к нулю, чтобы в пределе стороны AB и DC совпали на границе. Тогда циркуляция вектора **E** в левой части (1.12) сводится в пределе к $E_{1\tau}\Delta l - E_{2\tau}\Delta l$, где $E_{1\tau}$ и $E_{2\tau}$ — проекции векторов $\mathbf{E}_1$ в первой и второй средах на направление вектора $\boldsymbol{\tau}$, параллельного границе (и стороне AB), а поток вектора **B** в правой части обращается в нуль, так как площадь охватываемой контуром поверхности стремится к нулю. Отсюда и следует, что $E_{1\tau} = E_{2\tau}$. Аналогично на основе интегральной формы уравнения Максвелла (2.7) можно показать, что $B_{1\tau} = B_{2\tau}$ (при отсутствии поверхностных токов на границе). Так как вектор $\boldsymbol{\tau}$ может иметь любое направление в плоскости границы (т. е. два независимых компонента), то здесь мы имеем четыре независимых граничных условия, которые справедливы для любых непрерывных сред.



3.1
Контур интегрирования для получения граничных условий

Еще два граничных условия можно получить из уравнений Максвелла (2.6) и (2.8). Эти условия выражают непрерывность нормальных составляющих векторов **B** и **D** на границе: $B_{1n} = B_{2n}$, $D_{1n} = D_{2n}$.

Но для монохроматических полей уравнения (2.6) и (2.8) являются следствием уравнений (2.7) и (2.9). Поэтому граничные условия для нормальных составляющих не дают ничего нового: они выполняются автоматически при соблюдении условий для тангенциальных составляющих.



Направления падающей, отраженной и преломленной волн

Отметим, что наличие во второй среде только *одной* (преломленной) волны, уходящей от границы, не следует непосредственно из уравнений Максвелла, а основано на дополнительном предположении, известном как *условие излучения*. Можно обеспечить выполнение граничных условий, предполагая во второй среде наличие двух волн, одна из которых распространяется от границы, другая — к границе. Так пришлось бы поступать при исследовании волнового процесса не в полубесконечной среде, а в слое, ограниченном с двух сторон (в плоскопараллельной пластинке). Разные предположения приводят к разным результатам. Условие излучения, связанное с *принципом причинности*, дает критерий отбора имеющих физический смысл решений: возбуждаемое тело может порождать лишь уходящие от него волны (отраженные, рассеянные и т. п.). В задаче о преломлении на границе полубесконечной среды физический смысл имеет решение, основанное на предположении о наличии только *трех* волн: падающей, отраженной и преломленной. а рис. 3.2 показаны направления рассматриваемых волн. Все величины, относящиеся к распространяющимся в первой среде с показателем преломления $n_1 = \sqrt{\varepsilon_1}$ падающей и отраженной волнам, будем отмечать соответственно индексами 0 и 1, а к преломленной волне во второй среде с показателем преломления $n_2 = \sqrt{\varepsilon_2}$ — индексом 2. Первую среду считаем прозрачной, для второй среды такого предположения пока делать не будем. Для каждой из трех плоских волн используем комплексную запись. Таким образом, на границе

$$\begin{aligned} E_{0\tau}e^{i(\mathbf{k}_0\mathbf{r}-\omega_0 t)} + E_{1\tau}e^{i(\mathbf{k}_1\mathbf{r}-\omega_1 t)} &= E_{2\tau}e^{i(\mathbf{k}_2\mathbf{r}-\omega_2 t)}, \\ B_{0\tau}e^{i(\mathbf{k}_0\mathbf{r}-\omega_0 t)} + B_{1\tau}e^{i(\mathbf{k}_2\mathbf{r}-\omega_1 t)} &= B_{2\tau}e^{i(\mathbf{k}_2\mathbf{r}-\omega_2 t)}. \end{aligned} \qquad (3.1)$$

Чтобы граничные условия выполнялись в любой момент времени, коэффициенты при t в показателях экспонент для всех трех волн должны быть одинаковы. Поэтому часто́ты отраженной и преломленной волн равны частоте ω падающей волны, что, впрочем, очевидно, если эти волны рассматривать как результат сложения вторичных волн, излучаемых зарядами вещества при их вынужденном движении*.

---

* При больших амплитудах напряженности поля падающей волны, сравнимых с внутриатомными полями, вынужденные колебания атомных осцилляторов могут происходить не только на частоте падающей волны, но и на кратных частотах, что приводит к появлению гармоник в преломленном и отраженном излучении (см. гл. 10).

Направим ось z перпендикулярно границе раздела. Углы φ, φ_1 и φ_2, образуемые волнами с осью z, называют соответственно углами падения, отражения и преломления. Граничные условия должны выполняться сразу во всех точках границы раздела. Это возможно лишь тогда, когда зависимость E_τ и B_τ от координат точки в плоскости xy у всех трех волн одинакова, т. е. равны тангенциальные компоненты k_x, k_y их волновых векторов. Отсюда следует прежде всего, что *направления распространения всех трех волн лежат в одной плоскости*, проходящей через ось z *(плоскость падения)*. Выберем в качестве нее плоскость xz (рис. 3.2). Таким образом,

$$k_{1x} = k_{2x} = k_{0x} = (\omega/c)n_1 \sin\varphi. \tag{3.2}$$

Остается найти нормальные границе раздела компоненты волновых векторов отраженной и преломленной волн. Для этого воспользуемся соотношением (2.23), согласно которому

$$k_1^2 = k_{1x}^2 + k_{1z}^2 = (\omega^2/c^2)\varepsilon_1, \quad k_2^2 = k_{2x}^2 + k_{2z}^2 = (\omega^2/c^2)\varepsilon_2. \tag{3.3}$$

Учитывая (3.2), находим

$$k_{1z} = -k_{0z} = -\frac{\omega}{c} n_1 \cos\varphi, \quad k_{2z} = \frac{\omega}{c}\sqrt{\varepsilon_2 - \varepsilon_1 \sin^2\varphi}. \tag{3.4}$$

Величина k_{2z} в поглощающей среде (при комплексном значении ε_2) комплексна. Она будет комплексной (чисто мнимой) и в прозрачной среде, если $\sin^2\varphi > \varepsilon_2/\varepsilon_1$ (условие полного отражения). Эти случаи будут рассмотрены ниже.

Если вторая среда прозрачна и $\sin\varphi < \sqrt{\varepsilon_2/\varepsilon_1} = n_2/n_1$, то из (3.2) следуют известные законы отражения и преломления света, определяющие направления отраженной и преломленной волн. Учитывая, что $k_{0x} = (\omega/c)n_1 \sin\varphi$, $k_{1x} = (\omega/c)n_1 \sin\varphi_1$, $k_{2x} = (\omega/c)n_2 \sin\varphi_2$, находим

$$\varphi_1 = \varphi, \quad n_1 \sin\varphi = n_2 \sin\varphi_2. \tag{3.5}$$

Эти законы получены без каких бы то ни было предположений относительно комплексных амплитуд E_τ в (3.1) и поэтому справедливы при любом состоянии поляризации падающей волны.

Контрольные вопросы

1. Каким условиям должны удовлетворять векторы электрического и магнитного полей на границе раздела?
2. К какому условию приводит каждое из уравнений Максвелла (2.6)—(2.9)?
3. Сколько независимых условий для векторов электромагнитного поля должно выполняться на границе?
4. Поясните, как из граничных условий (3.1) вытекают геометрические законы отражения и преломления света.

При выводе законов отражения и преломления информация, содержащаяся в граничных условиях (3.1), не была использована полностью: для соблюдения (3.1) комплексные амплитуды отраженной и преломленной волн должны иметь вполне определенные значения при заданной амплитуде падающей волны. Поэтому граничные условия (3.1) позволяют определить не только направления отраженной и преломленной волн, но и их амплитуды и состояния поляризации.

Ограничимся пока случаем прозрачных сред и, кроме того, будем считать, что выполняется условие $\sin\varphi < n_2/n_1$, при котором во второй среде существует преломленная волна [т. е. значение k_{2z} в (3.4) вещественно]. Противоположный случай полного отражения $\sin\varphi > n_2/n_1$ рассмотрен в § 3.3.

Разложим каждую из трех волн в (3.1) на две составляющие: поляризованную в плоскости падения (снабдив ее индексом $\parallel$) и поляризованную в перпендикулярном плоскости падения направлении (индекс $\perp$). Для векторов **E** и **B**, лежащих в плоскости падения, условимся выбирать положительные направления в каждой из волн так, как показано на рис. 3.3. В перпендикулярной плоскости положительное направление задается единичным вектором **j**, направленным вдоль оси y. Таким образом, $\mathbf{E}_0^{\perp} = E_0^{\perp}\mathbf{j}$, $\mathbf{E}_0^{\parallel} = E_0^{\parallel}\mathbf{e}_0$ и т. д.

Вектор **B** в каждой из трех волн выражается через соответствующий вектор **E** с помощью соотношения (2.21), что дает следующую связь между амплитудами $\mathbf{E}^{\parallel}$, $\mathbf{E}^{\perp}$ и $\mathbf{B}^{\parallel}$, $\mathbf{B}^{\perp}$:

$$\mathbf{B}_0^{\perp} = n_1 E_0^{\perp}\mathbf{e}_0, \quad \mathbf{B}_1^{\perp} = n_1 E_1^{\perp}\mathbf{e}_1, \quad \mathbf{B}_2^{\perp} = n_2 E_2^{\perp}\mathbf{e}_2,$$
$$\mathbf{B}_0^{\parallel} = -n_1 E_0^{\parallel}\mathbf{j}, \quad \mathbf{B}_1^{\parallel} = -n_1 E_1^{\parallel}\mathbf{j}, \quad \mathbf{B}_2^{\parallel} = -n_2 E_2^{\parallel}\mathbf{j}. \tag{3.6}$$

Используя соотношения (3.6), четыре граничных условия (3.1) можно переписать так, чтобы в них входили амплитуды напряженностей только электрических полей каждой из волн. Учитывая, что $e_{0x} = \cos\varphi$, $e_{1x} = -\cos\varphi$, $e_{2x} = \cos\varphi_2$, получаем:

$$E_0^{\perp} + E_1^{\perp} = E_2^{\perp}, \quad n_1\cos\varphi(E_0^{\perp} - E_1^{\perp}) = n_2\cos\varphi_2 E_2^{\perp},$$
$$n_1(E_0^{\parallel} + E_1^{\parallel}) = n_2 E_2^{\parallel}, \quad \cos\varphi(E_0^{\parallel} - E_1^{\parallel}) = \cos\varphi_2 E_2^{\parallel}. \tag{3.7}$$

Заметим, что эти уравнения распадаются на две группы, одна из которых содержит только компоненты $\mathbf{E}^{\perp}$, другая — компоненты $\mathbf{E}^{\parallel}$, т. е. $\mathbf{E}_1^{\perp}$ и $\mathbf{E}_2^{\perp}$ выражаются только через $\mathbf{E}_0^{\perp}$ и не зависят от $\mathbf{E}_0^{\parallel}$, и наоборот. Это значит, что волны указанных двух типов можно рассматривать независимо друг от друга.

Уравнения (3.7) можно разрешить относительно компонент отраженной и преломленной волн, выразив их через компоненты падающей волны:

$$E_1^{\perp} = \frac{n_1\cos\varphi - n_2\cos\varphi_2}{n_1\cos\varphi + n_2\cos\varphi_2} E_0^{\perp}, \quad E_2^{\perp} = \frac{2n_1\cos\varphi}{n_1\cos\varphi + n_2\cos\varphi_2} E_0^{\perp}, \tag{3.8}$$

$$E_1^{\parallel}=\frac{n_2\cos\varphi-n_1\cos\varphi_2}{n_2\cos\varphi+n_1\cos\varphi_2}E_0^{\parallel},\quad E_2^{\parallel}=\frac{2n_1\cos\varphi}{n_2\cos\varphi+n_1\cos\varphi_2}E_0^{\parallel}. \tag{3.9}$$

Эти соотношения, называемые *формулами Френеля,* полностью определяют характеристики отраженной и преломленной волн. Их обычно пишут в несколько иной форме, которую можно получить из (3.8) и (3.9), исключив n_1/n_2 с помощью закона преломления (3.4):

$$E_1^{\perp}=-\frac{\sin(\varphi-\varphi_2)}{\sin(\varphi+\varphi_2)}E_0^{\perp},E_2^{\perp}=\frac{2\cos\varphi\sin\varphi_2}{\sin(\varphi+\varphi_2)}E_0^{\perp}, \tag{3.10}$$

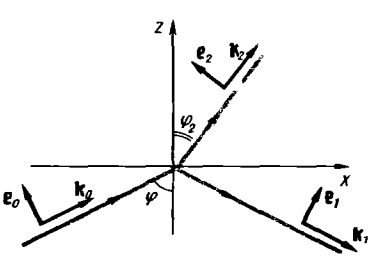
$$E_1^{\parallel}=\frac{\operatorname{tg}(\varphi-\varphi_2)}{\operatorname{tg}(\varphi+\varphi_2)}E_0^{\parallel},\quad E_2^{\parallel}=\frac{2\cos\varphi\sin\varphi_2}{\sin(\varphi+\varphi_2)\cos(\varphi-\varphi_2)}E_0^{\parallel}. \tag{3.11}$$

В случае нормального падения $\varphi=0$ и, следовательно, $\varphi_2=0$. Тогда соотношения (3.8) принимают вид

$$E_1=(n_1-n_2)E_0/(n_1+n_2),\ E_2=2n_1E_0/(n_1+n_2). \tag{3.12}$$

При этом различие между параллельной и перпендикулярной компонентами исчезает, так как теряет смысл понятие плоскости падения. Заметим, что при переходе к случаю нормального падения в формулах (3.9) для E_1 получается выражение, отличающееся знаком от (3.12). Это отличие чисто формальное и возникает из-за того, что в соответствии с принятым выше определением положительное направление для $\mathbf{E}_1^{\parallel}$ (рис. 3.3) при нормальном падении совпадает с отрицательным направлением $\mathbf{E}_0^{\parallel}$, тогда как положительные направления $\mathbf{E}_1^{\perp}$ и $\mathbf{E}_0^{\perp}$ всегда одинаковы.

При $n_2>n_1$ знак амплитуды E_1 в (3.12) противоположен знаку E_0. Это значит, что векторы напряженностей электрического поля падающей и отраженной волн на границе раздела совершают колебания в противофазе. Об этом обычно говорят как о *потере полуволны* при отражении от оптически более плотной среды (отметим, что векторы индукции магнитного поля падающей и отраженной волн при этом колеблются синфазно). Амплитуда E_2 преломленной волны в (3.12) всегда (при любом соотношении между n_1 и n_2) совпадает по знаку с E_0, т. е. во второй среде колебания напряженности электрического поля на границе происходят синфазно с падающей волной. Очевидно, что при нормальном падении линейно поляризованного света преломленная и отраженная волны также будут поляризованы линейно в той же плоскости, что и падающая волна. При нормальном падении циркулярно поляризованного света волна во второй среде будет иметь такую же



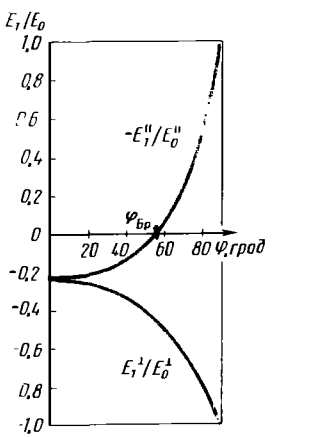
3.3

Условный выбор положительных направлений векторов **E** и **B**

круговую поляризацию, а отраженная волна — круговую поляризацию противоположного знака, что обусловлено изменением направления ее распространения на противоположное при прежнем направлении вращения вектора E в пространстве.

В общем случае наклонного падения света на границу прозрачной среды коэффициенты пропорциональности между амплитудами преломленной и отраженной волн и амплитудой падающей волны в формулах (3.8)—(3.11) вещественны. Отсюда следует, что отражение и преломление не сопровождаются изменением фаз, за исключением, быть может, изменения фазы отраженной волны на π (если соответствующий коэффициент отрицателен). Поэтому в случае линейной поляризации падающего света отраженная и преломленная волны будут тоже поляризованы линейно. С помощью формул Френеля (3.10)—(3.11) можно показать, что в отраженной волне направление поляризации отклоняется от плоскости падения на больший угол, а в преломленной — на меньший угол, чем в падающей волне (см. задачу 1).

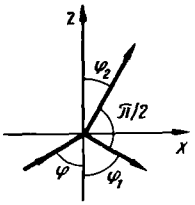
Рассчитанные по формулам Френеля зависимости отношения амплитуд напряженности электрического поля отраженного и падающего света от угла падения для границы воздух—стекло приведены на рис. 3.4. В случае линейной поляризации в направлении, перпендикулярном плоскости падения, знак отношения $E_1^{\perp}/E_0^{\perp}$ всегда отрицателен, т. е. при любых углах падения колебания напряженностей электрического поля в отраженной и падающей волнах на границе раздела происходят в противофазе. Для поляризации в плоскости падения на рис. 3.4 приведен график отношения $E_1^{\parallel}/E_0^{\parallel}$, взятого для большей наглядности с противоположным знаком. Перемена знака сделана для того, чтобы при переходе к предельному случаю нормального падения выбор положительных направлений амплитуд отраженной и падающей волн стал одинаковым (ср. с рис. 3.3).



3.4
Отношение амплитуд отраженной и падающей волн на границе воздух — стекло ($n = 1{,}5$)

Из этого графика видно, что колебания напряженностей электрического поля отраженной и падающей волн происходят в противофазе, как и в случае $E_1^{\perp}/E_0^{\perp}$, пока угол падения меньше некоторого значения $\varphi_{Бр}$, называемого углом Брюстера. При переходе через это значение фаза отраженной волны, поляризованной в плоскости падения, скачком изменяется на π.

Угол Брюстера определяется из условия $\varphi_{Бр} + \varphi_2 = \pi/2$: при падении света под таким углом направления отраженной и преломленной волн взаимно перпендикулярны (рис. 3.5). Из формулы (3.11), содержащей



3.5
К определению угла полной поляризации

в знаменателе $\mathrm{tg}(\varphi+\varphi_2)$, видно, что при $\varphi=\varphi_{Бр}$ получается $E_{\parallel}=0$: отраженной волны не будет, если падающий под углом $\varphi_{Бр}$ свет поляризован в плоскости падения. Отсюда следует, что отражение естественного света, который можно представить как некогерентную смесь двух линейно поляризованных волн с ортогональными направлениями поляризации, обладает замечательным свойством: при падении естественного света под углом $\varphi=\varphi_{Бр}$ отразится только составляющая, поляризованная перпендикулярно плоскости падения, и отраженный свет будет полностью линейно поляризованным. В этом состоит *закон Брюстера*, открытый экспериментально в 1815 г. Угол $\varphi_{Бр}$ называется еще *углом полной поляризации*. Так как в этом случае $\sin\varphi_2=\cos\varphi_1$ (рис. 3.5), то из закона преломления получаем

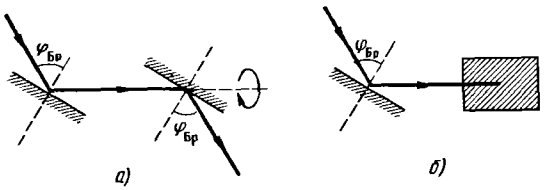
$$\mathrm{tg}\,\varphi_{Бр}=n_2/n_1. \qquad (3.13)$$

При отражении видимого света на границе воздух—стекло $n_2/n_1=1{,}5$, что соответствует углу Брюстера $\varphi_{Бр}\approx 56°$

Использование отражения естественного света под углом Брюстера дает один из простейших способов получения поляризованного света. Его недостатком является малая интенсивность отраженного света.

Описанные выше закономерности отражения света под углом Брюстера можно продемонстрировать простым опытом, схема которого представлена на рис. 3.6. Отражение света происходит от черного стекла, в котором преломленная волна полностью поглощается. Это необходимо для того, чтобы не было мешающего света, отраженного второй поверхностью стекла. Конечно, применение черного стекла не вполне соответствует отражению света от прозрачной среды. Но поглощение света в черном стекле в действительности слабое (оно происходит на расстояниях, больших по сравнению с длиной волны, т. е. мнимая часть показателя преломления много меньше вещественной), и связанные с поглощением искажения не скажутся на результатах опыта с отраженным светом.

Падающий на первое зеркало под углом Брюстера естественный свет после отражения оказывается полностью поляризованным. Если второе зеркало расположено параллельно первому (рис. 3.6, *а*), то падающий на него поляризованный свет отразится полностью, так как



3.6
Схема опыта, демонстрирующего поляризацию света, отраженного под углом Брюстера

направление поляризации перпендикулярно плоскости падения (плоскости чертежа). Будем поворачивать второе зеркало вокруг оси пучка поляризованного света, полученного при отражении от первого зеркала. При этом угол падения остается неизменным и равным углу Брюстера. Интенсивность отраженного вторым зеркалом света начнет постепенно убывать вплоть до полного исчезновения, когда плоскость падения совпадет с плоскостью поляризации (рис. 3.6, *б*). При таком расположении второе зеркало не может отражать свет той поляризации, которая возникает при отражении от первого зеркала.

В газовых лазерах торцовые окна разрядной трубки представляют собой плоскопараллельные стеклянные пластинки, расположенные под углом Брюстера к оси трубки (см. рис. 9.8). Благодаря этому излучение, распространяющееся вдоль оси трубки в образованном зеркалами открытом резонаторе и поляризованное в плоскости падения на пластинки, проходит сквозь них беспрепятственно, не испытывая отражения. В результате лазер генерирует свет, поляризованный в этой плоскости.

Введем, по определению, *коэффициент отражения* R границы (по другой терминологии, отражательную способность) как отношение среднего по времени отраженного от поверхности потока энергии к падающему потоку. Для нормального падения коэффициент отражения не зависит от состояния поляризации и, как видно из формулы (3.12), следующим образом выражается через показатели преломления n_1 и n_2 граничащих сред:

$$R=(E_1/E_0)^2=[(n_1-n_2)/(n_1+n_2)]^2. \qquad (3.14)$$

Значение R не изменится, если n_1 и n_2 поменять местами. Поэтому коэффициент отражения на границе прозрачных сред не зависит от того, в каком направлении падает свет: из первой среды во вторую или наоборот. Для границы воды ($n=1{,}33$) и воздуха $R=2\%$, стекла ($n=1{,}5$) и воздуха $R=4\%$, т. е. ни вода, ни стекло при нормальном падении не могут служить зеркалом. В обычных зеркалах используется отражение от металлической поверхности, стекло же служит только для ее защиты. Слабое отражение от передней поверхности делает такие зеркала непригодными для оптических целей. В оптических приборах используют зеркала, у которых отражающий металлический слой нанесен на переднюю поверхность, стекло здесь служит лишь удобным материалом для подложки.

В сложных оптических системах, содержащих большое число преломляющих поверхностей (до нескольких десятков, например в многолинзовых объективах), даже незначительные ($\approx 4\%$) потери энергии на отражение при каждом прохождении границы стекло — воздух существенно ослабляют полезный световой поток. Для их уменьшения на поверхности линз наносят специальные покрытия (см. § 5.7).

Коэффициент отражения тем меньше, чем ближе показатели преломления граничащих сред. При $n_1=n_2$ отражение вообще отсутствует. Можно, например, так подобрать смесь бензола с сероуглеродом, чтобы ее показатель преломления был таким же, как у стекла. Тогда отражение исчезает и граница между стеклом и жидкостью становится незаметной: жидкость с погруженными в нее кусочками стекла становится оптически однородной. Это явление используют в минералогии для измерения показателя преломления прозрачных образцов неправильной формы. Свет должен быть по возможности монохроматическим, так как жидкость и погруженное в нее тело обычно обладают различной дисперсией: если их показатели преломления для какой-то длины волны совпадают, то для других длин волн совпадения уже не будет.

Коэффициент отражения при наклонном падении зависит от состояния поляризации падающего света. Для света, поляризованного в плоскости падения и перпендикулярно ей, формулы (3.10) дают

$$R_\perp=\sin^2(\varphi-\varphi_2)/\sin^2(\varphi+\varphi_2);\ R_\parallel=\mathrm{tg}^2(\varphi-\varphi_2)/\mathrm{tg}^2(\varphi+\varphi_2). \quad (3.15)$$

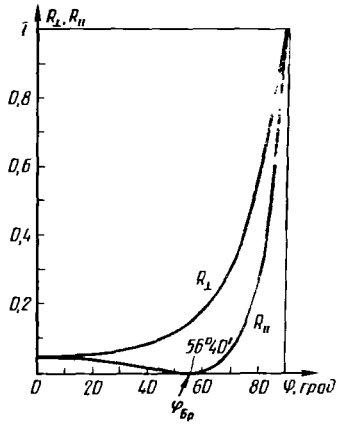
Если плоскость поляризации падающего света составляет с плоскостью падения угол γ, то $E_0^\perp=E_0\sin\gamma$, $E_0^\parallel=E_0\cos\gamma$ и коэффициент отражения R_γ можно выразить через $R_\perp$ и $R_\parallel$ из (3.15):

$$R_\gamma=R_\perp\sin^2\gamma+R_\parallel\cos^2\gamma.$$

Коэффициент отражения для неполяризованного света можно получить усреднением этого выражения по γ:

$$R=(R_\perp+R_\parallel)/2.$$

Он, конечно, равен полусумме $R_\perp$ и $R_\parallel$, так как естественный падающий свет можно рассматривать как сумму двух некогерентных волн одинаковой интенсивности, поляризованных в плоскости падения и в перпендикулярном ей направлении.



3.7
Зависимость $R_\perp$ и $R_\parallel$ от угла падения для границы воздух — стекло

Отметим, что выражения (3.15) не изменяются при взаимной замене φ и φ_2. Это значит, что коэффициент отражения для волны, падающей из первой среды под углом φ, равен коэффициенту отражения для волны, падающей из второй среды под углом φ_2.

Зависимости $R_\perp$ и $R_\parallel$ от угла падения имеют существенно различный характер. На рис. 3.7 приведены графики зависимости $R_\perp$ и $R_\parallel$ от φ для видимого света на границе воздуха со

стеклом ($n_1=1$, $n_2=1{,}52$), построенные по формулам (3.15). Полезно сравнить их с графиками $E_1^{\perp}/E_0^{\perp}$ и $E_1^{\parallel}/E_0^{\parallel}$ на рис. 3.4. При малых углах падения ($\varphi<10°$) изменения коэффициентов отражения незначительны, $R_{\perp}$ и $R_{\parallel}$ имеют практически такое же значение (3.14), что и при нормальном падении. Затем $R_{\perp}$ монотонно возрастает по мере увеличения φ, в то время как $R_{\parallel}$ сначала убывает, обращается в нуль при $\varphi=\varphi_{\text{Бр}}$ и лишь затем начинает монотонно возрастать. Если отражение происходит от оптически более плотной среды, т. е. $n_2>n_1$, то возрастание $R_{\perp}$ и $R_{\parallel}$ продолжается вплоть до $\varphi=\pi/2$ («скользящее» падение), когда они оба достигают единицы. При скользящем падении отражение будет полным. Благодаря этому можно наблюдать изображение противоположного берега, отражающегося в спокойной поверхности воды.

При отражении от оптически менее плотной среды ($n_2<n_1$) оба коэффициента обращаются в единицу уже при угле падения $\varphi=\varphi_{\text{m}}$, где $\sin\varphi_{\text{m}}=n_2/n_1$. Угол φ_{m} называется *предельным углом полного отражения*. При $\varphi=\varphi_{\text{m}}$ угол преломления $\varphi_2=\pi/2$, т. е. преломленная волна распространяется параллельно границе раздела. Отражение под углами $\varphi>\varphi_{\text{m}}$ требует особого рассмотрения (см. § 3.3), так как k_{2z} в (3.4) становится чисто мнимым, т. е. поле во второй среде затухает. Затухание волны при отсутствии поглощения (диссипации энергии) означает, что на границе происходит полное отражение падающей волны.

азличие в зависимости $R_{\perp}$ и $R_{\parallel}$ от угла падения приводит к тому, что при наклонном падении *естественного* света отраженный от поверхности свет в общем случае оказывается *частично поляризованным*. Впервые это было экспериментально обнаружено в 1808 г. Малю. Так как $R_{\parallel}<R_{\perp}$ (за исключением нормального падения $\varphi=0$ и скользящего $\varphi=\pi/2$), то в отраженном свете преимущественное направление электрического поля перпендикулярно плоскости падения. Преломленный свет будет частично поляризован с преимущественным направлением **E** в плоскости падения. Частично поляризованный свет можно рассматривать как смесь естественного с линейно поляризованным. Для его характеристики вводят *степень поляризации* Δ:

$$\Delta=\frac{I_{\perp}-I_{\parallel}}{I_{\perp}+I_{\parallel}}\,100\%.$$

Для естественного света $I_{\perp}=I_{\parallel}$ и $\Delta=0$. Свет, отраженный под углом Брюстера, имеет $I_{\parallel}=0_{\parallel}$ и для него $\Delta=100\%$, т. е. он полностью поляризован перпендикулярно плоскости падения. При принятом определении отрицательные значения Δ соответствуют преимущественной поляризации в плоскости падения. Для преломленного света из формул Френеля (3.10) и (3.11) в случае естественного падающего света находим

$$I_{\perp}/I_{\parallel}=(E_2^{\perp}/E_2^{\parallel})^2=\cos^2(\varphi-\varphi_2)\leqslant 1,$$

т. е. $\Delta<0$. При падении естественного света из воздуха на стекло под углом Брюстера преломленная волна имеет $\Delta=-8\%$ (см. задачу 3). При прохождении света через плоскопараллельную пластинку на второй поверхности опять происходит преломление под углом Брюстера и степень поляризации увеличивается еще на 8%. Если последовательно расположить несколько пластинок (стопа Столетова), то на выходе можно получить высокую степень поляризации света.

В настоящее время этот способ почти не применяется, так как существуют более удобные и совершенные способы получения поляризованного света с помощью кристаллических поляризационных призм и поляроидов (см. § 4.4).

При выводе формул Френеля граница раздела между двумя различными средами рассматривалась как математическая плоскость. В действительности граница раздела представляет собой не геометрическую поверхность, а тонкий переходный слой, на протяжении которого показатель преломления изменяется от n_1 до n_2. Для справедливости формул Френеля необходимо, чтобы толщина слоя была мала по сравнению с длиной волны. Для этого граничная поверхность должна быть свободна от посторонних примесей и хорошо отполирована. Если же показатель преломления постепенно изменяется на протяжении нескольких длин волн, преломление имеет совсем другой характер. Когда длина волны мала по сравнению с размерами неоднородностей среды, выполняются условия применимости геометрической оптики (см. § 7.1). Преломление волны можно при этом рассматривать как распространение лучей, испытывающих в переходном слое рефракцию (постепенное отклонение) без всякого отражения.

Измерения коэффициентов отражения для чистых поверхностей при различных углах падения и различных поляризациях падающего света дают значения, которые находятся в хорошем согласии с формулами Френеля. Исключение составляет лишь случай отражения света под углом Брюстера, при котором наблюдаются небольшие отступления от формул Френеля. Тщательные опыты показали, что закон Брюстера соблюдается не строго: коэффициент отражения $R_\parallel$ ни для какого угла падения в нуль не обращается, а достигается только минимум при $\varphi=\varphi_{Бр}$, хотя в минимуме $R_\parallel$ очень мал. Кроме того, свет, линейно поляризованный под углом к плоскости падения, при отражении под углом Брюстера (и в малой его окрестности) становится поляризованным эллиптически, хотя по формулам Френеля он должен оставаться линейно поляризованным.

Происхождение этих аномалий, согласно Друде, можно объяснить наличием на поверхности отражающей среды тонкого переходного слоя. Теория показывает, что толщина переходного слоя порядка нескольких межатомных расстояний уже достаточна для того, чтобы вызвать наблюдаемые отклонения от формул Френеля. Физические причины возникновения переходных слоев на поверхностях еще окончательно не выяснены, но несомненно, что для чистых веществ они обусловлены молекулярной структурой самой отражающей среды вблизи ее поверхности.

Контрольные вопросы

- □ Как связаны (в точках границы) фазы отраженной и преломленной волн с фазой падающей волны в случае нормального падения?
- □ Что такое угол Брюстера? Чем он замечателен?
- □ Как можно измерить показатель преломления прозрачного образца неправильной формы?
- □ Почему при наклонном падении естественного света ($\varphi \neq \varphi_{Бр}$) отраженный свет частично поляризован?
- □ Какие отступления от формул Френеля наблюдаются на опыте?

Задачи

1. У падающего на границу прозрачных сред света направление поляризации составляет угол γ с плоскостью падения. Какие углы γ_1 и γ_2 составляет направление поляризации с плоскостью падения у отраженного и преломленного света?

◇ Ответ: $\operatorname{tg}\gamma_1 = -\cos(\varphi-\varphi_2)\operatorname{tg}\gamma/\cos(\varphi+\varphi_2)$, $\operatorname{tg}\gamma_2 = \cos(\varphi-\varphi_2)\operatorname{tg}\gamma$. Учитывая, что $0<\varphi, \varphi_2<\pi/2$, и полагая $0<\gamma<\pi/2$, получаем неравенства $\gamma_1>\gamma$ и $\gamma_2<\gamma$, т. е. плоскость поляризации при отражении поворачивается от плоскости падения, а при преломлении — в сторону плоскости падения. ◇

2. Каким должен быть преломляющий угол α призмы из стекла с показателем преломления $n=1{,}5$, чтобы свет линейной поляризации мог пройти сквозь нее без потерь на отражение?

◇ Ответ: $\alpha = \pi - 2\varphi_{Бр} = 67°22'$. ◇

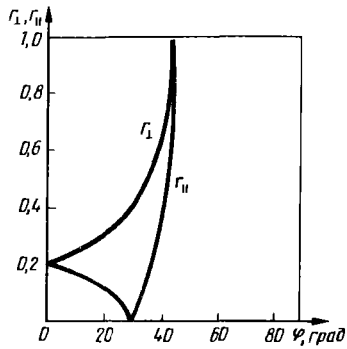
3. Найти степень поляризации преломленного света, если естественный свет падает из воздуха на поверхность стекла ($n=1{,}5$) под углом Брюстера.

◇ Ответ: $\Lambda = [4n^2-(n^2+1)^2]/[4n^2+(n^2+1)^2] = -(n^2-1)^2/[4n^2+(n^2+1)^2] = -8\%$. ◇

3.3. Полное отражение

Проведенное в § 3.2 исследование коэффициентов отражения $R_\perp$ и $R_\parallel$ показало, что для падения света из оптически более плотной среды *1* на границу менее плотной среды *2* ($n_1>n_2$) оба коэффициента обращаются в единицу, когда угол падения φ достигает значения φ_m, определяемого условием $\sin\varphi_m = n_2/n_1$ и называемого предельным углом полного отражения. На границе стекло — воздух ($n_1=1{,}5$; $n_2=1$) $\varphi_m = 41°$.

Графики зависимости $r_\perp \equiv E_1^\perp/E_0^\perp$ и $r_\parallel \equiv E_1^\parallel/E_0^\parallel$ от угла падения для $n_1>n_2$, построенные по формулам Френеля (3.10) — (3.11), приведены на рис. 3.8. Отличие от аналогичных графиков для $n_1<n_2$ (см. рис. 3.4) состоит не только в том, что $r_\perp$ и $|r_\parallel|$ достигают теперь единицы уже при угле падения $\varphi=\varphi_m<\pi/2$ (а не при скользящем падении, как для $n_1<n_2$), но и в самом характере их приближе-



3.8
Отношение амплитуд отраженной и падающей волн

ния к единице. С помощью формул Френеля легко убедиться (см. задачу 1), что в точке $\varphi=\varphi_m$ кривые имеют вертикальную касательную.

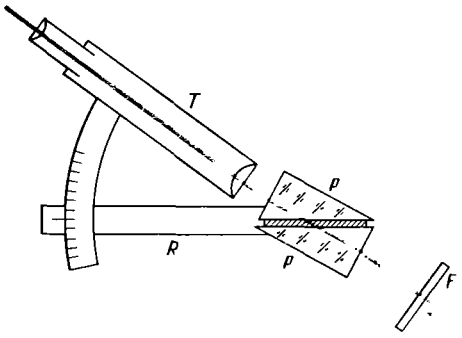


Схема рефрактометра Аббе

Это обстоятельство играет большую роль в экспериментальных методах определения показателей преломления жидкостей. На рис. 3.9 приведена схема рефрактометра Аббе, действие которого основано на явлении полного отражения. Между двумя призмами PP, изготовленными из стекла с высоким показателем преломления, помещают каплю исследуемой жидкости. Пучок света от источника S проходит через светофильтр F и испытывает полное отражение при переходе из первой призмы в жидкость. Призму вместе с рычагом R можно поворачивать относительно зрительной трубы T. Благодаря отмеченному обстоятельству граница полного отражения наблюдается с исключительной резкостью: нарастание интенсивности преломленного света кажется скачкообразным. По углу наклона зрительной трубы по отношению к призме при визировании границы полного отражения определяют показатель преломления жидкости. Рефрактометр Аббе обеспечивает измерение показателя преломления жидкости с точностью до 0,1%.

Когда угол падения φ превышает предельное значение φ_m, не существует вещественного угла преломления φ_2, так как закон преломления (3.5) дает для $\sin\varphi_2$ значение, превосходящее единицу, а для $\cos\varphi_2$ — чисто мнимое значение. Но формулы Френеля (3.8) — (3.11) останутся справедливыми и в этом случае, если закон преломления (3.5) рассматривать просто как определение входящих в них величин $\sin\varphi_2$ и $\cos\varphi_2$. Справедливость понимаемых таким образом формул Френеля следует из того, что они обеспечивают выполнение граничных условий (3.1) и в случае полного отражения.

При $\varphi>\varphi_m$ волновой вектор $\mathbf{k}_2$ волны во второй среде имеет вещественную проекцию k_{2x} на направление границы и мнимую проекцию k_{2z} на направление нормали: формулы (3.2) и (3.4) для этого случая дают

$$k_{2x}=n_1\omega\sin\varphi/c,\quad k_{2z}=\pm i\omega\sqrt{n_1^2\sin^2\varphi-n_2^2}/c. \tag{3.16}$$

Это значит, что электромагнитное поле во второй (оптически менее плотной) среде представляет собой неоднородную волну, у которой поверхности постоянной фазы — это плоскости $x=\text{const}$, перпендикулярные границе, а поверхности постоянной амплитуды — плоскости $z=\text{const}$, параллельные границе раздела*. Знак перед

* Волновой вектор $\mathbf{k}_2$ можно записать в виде $\mathbf{k}_2=\mathbf{k}_2'+i\mathbf{k}_2''$, где вещественные векторы $\mathbf{k}_2'$ и $\mathbf{k}_2''$ имеют следующие проекции: $\mathbf{k}_2'(n_1\omega\sin\varphi/c,\ 0,\ 0)$, $\mathbf{k}_2''(0,\ 0,\ \omega\sqrt{n_1^2\sin^2\varphi-n_2^2}/c)$. Векторы $\mathbf{k}_2'$ и $\mathbf{k}_2''$ ортогональны. Это соответствует (см. § 2.2) случаю, когда в прозрачной среде (с вещественной диэлектрической проницаемостью) вектор $\mathbf{k}$ может быть комплексным, если $\mathbf{k}'\mathbf{k}''=0$.

корнем в (3.16) определяется из требования, чтобы при удалении от границы раздела амплитуда этой неоднородной волны уменьшалась. Только такое решение имеет физический смысл. Подставляя k_{2x} и k_{2z} из (3.16) в выражение $\mathbf{E}_2 \exp i(\mathbf{k}_2\mathbf{r} - \omega t)$, получаем

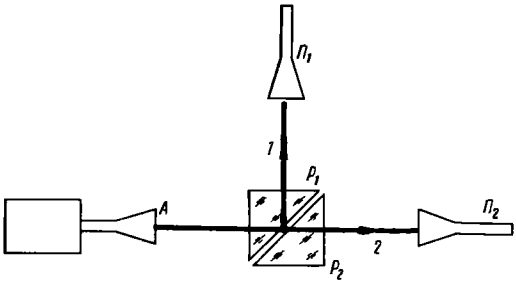
$$\mathbf{E}_2 e^{-z/l} e^{i(k_{2x}x - \omega t)}, \quad l = c/(\omega\sqrt{n_1^2 \sin^2\varphi - n_2^2}). \tag{3.17}$$

Убыванию амплитуды в направлении оси z соответствует знак «+» в (3.16). Величина l характеризует глубину проникновения волны во вторую среду: на этом расстоянии от границы амплитуда волны убывает в e раз. По существу, волну во второй среде можно считать поверхностной: ее амплитуда заметна только на расстоянии нескольких длин волн от граничной поверхности. Амплитуда $\mathbf{E}_2$ волны на границе (при $z=0$) может быть найдена по формулам Френеля.

Поверхности постоянной фазы этой неоднородной волны перемещаются вдоль границы раздела со скоростью $v = \omega/k_{2x} = c/(n_1 \sin\varphi)$. В рассматриваемом случае вторая среда прозрачна, т. е. поглощение света в ней не происходит. Поэтому затухание амплитуды волны в глубь второй среды означает, что энергия падающего света целиком возвращается в первую среду. В этом можно убедиться и непосредственным вычислением среднего потока энергии: вектор $\langle \mathbf{S} \rangle$ во второй среде имеет только параллельную границе составляющую, которая убывает в e раз на расстоянии $l/2$ от границы (см. задачу 4).

Экспериментальная проверка проникновения света во вторую среду при полном отражении, т. е. доказательство существования неоднородной волны, представляет собой довольно трудную задачу из-за малой глубины проникновения. Простой и интересный метод был предложен Мандельштамом и Зелени. Призма полного отражения своей отражающей гранью погружается в раствор флуоресцирующего вещества. Пучок света падает на эту грань под углом, большим предельного. В тонком слое раствора, прилегающем к призме, виден зеленоватый свет флуоресценции.

Другой метод обнаружения возмущения в менее плотной среде основан на помещении второй преломляющей среды на близком расстоянии от границы, где происходит полное отражение. Сделав прослойку менее плотной среды (например, воздуха между стеклами) тоньше, чем длина волны, можно получить во второй среде обычную однородную световую волну, так как неоднородная волна в прослойке достигает второй границы еще не слишком ослабленной. Меняя толщину прослойки, можно варьировать интенсивность проходящего света. На таком принципе работает один из модуляторов света. Прохождение света через зазор между средами при падении под углом, большим предельного, называют *нарушенным полным отражением*. В такой ситуации необходимо учитывать граничные условия и на второй близкой поверхности. Вно-



3.10
Схема опыта, демонстрирующего нарушенное полное отражение

симое второй границей возмущение изменяет поведение волны на первой границе, нарушая полное отражение*.

Описанные опыты легко выполнить с электромагнитными волнами сантиметрового диапазона. Две парафиновые призмы P_1 и P_2 полного отражения (для парафина $n=1{,}5$ при $\lambda \approx 3$ см) располагаются рядом с небольшим зазором между диагональными гранями (рис. 3.10). Рупор антенны A клистронного передатчика направляется на боковую грань одной из призм. Антенны приемников воспринимают волны, распространяющиеся в направлениях *1* и *2*. Если зазор между призмами больше длины волны, то в призме P_1 происходит полное отражение и только приемник $П_1$ регистрирует сигнал. По мере сближения призм можно наблюдать перераспределение энергии между приемниками. Когда призмы прижаты друг к другу, реагирует только приемник $П_2$.

При исследовании отраженной волны будем исходить из формул Френеля (3.8) — (3.9), подставив в них $\cos\varphi_2 = i\sqrt{(n_1\sin\varphi/n_2)^2 - 1}$ из закона преломления (3.4):

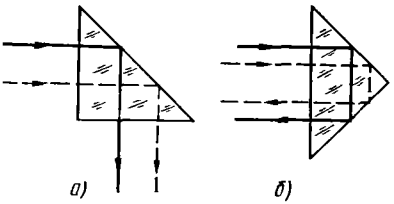
$$r_\perp = (n_1\cos\varphi - i\sqrt{n_1^2\sin^2\varphi - n_2^2}) / (n_1\cos\varphi + i\sqrt{n_1^2\sin^2\varphi - n_2^2}),$$

$$r_\parallel = (n_2^2\cos\varphi - in_1\sqrt{n_1^2\sin^2\varphi - n_2^2}) / (n_2^2\cos\varphi + in_1\sqrt{n_1^2\sin^2\varphi - n_2^2}). \quad (3.18)$$

Отсюда следует, что $|r_\perp| = |r_\parallel| = 1$ при $\varphi > \varphi_m$: при любой поляризации падающего света интенсивность отраженного света равна интенсивности падающего, т. е. отражение действительно полное, что соответствует заключению, сделанному выше на основе закона сохранения энергии. Этот эффект легко наблюдается на опыте и находит многочисленные практические применения. На рис. 3.11 показаны стеклянные призмы полного отражения. На границе стекло — воздух $\varphi_m \approx 42°$, поэтому при падении света на грань призмы под углом 45° отражение полное. Призма на рис. 3.11, *а* поворачивает лучи под прямым углом, оборотная призма на рис. 3.11, *б* переворачивает изображение. Призмы полного отражения широко используются во многих оптических приборах, в частности в зрительных трубах и полевых биноклях для получения прямого изображения.

* Аналогичное явление — проникновение частицы через потенциальный барьер — имеет место в квантовой механике и называется туннельным эффектом.

В последние годы для передачи информации с помощью генерируемого лазерами когерентного оптического излучения все большее распространение получают тонкопленочные и волоконные *диэлектрические световоды*. Их действие также основано на явлении полного отражения. Световой поток, испытывая многократные отражения на границе материала световода и оболочки из оптически менее плотного материала, распространяется вдоль волокна, несмотря на его изгибы.

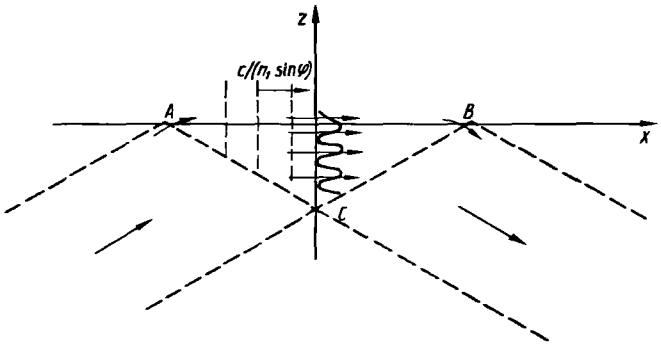


3.11
Стеклянные призмы полного отражения

Во многих отношениях оптическое волокно аналогично полым *волноводам* с внутренними поверхностями из хорошо проводящего металла, широко применяемым в технике СВЧ. Электромагнитные поля в этих системах имеют подобную структуру. Распространение света в цилиндрическом прозрачном волокне или прямоугольной диэлектрической пленке носит волноводный характер. Физические принципы действия оптических волноводов и других тонкопленочных структур составляют теоретическую базу новой бурно развивающейся области прикладной физики, получившей название *интегральной оптики*. Интерес к оптическим способам передачи и обработки информации быстро растет, что обусловлено преимуществами оптической связи в таких системах, где требуется высокая надежность, помехозащищенность, большая скорость передачи информации при малых габаритах и массе. Основные трудности реализации таких систем связаны с потерями световой энергии в диэлектрическом световоде, вызванными поглощением или рассеянием света в волокне, а также нерегулярностями границы раздела между сердцевиной и оболочкой. Эти потери предъявляют очень жесткие требования к технологии изготовления световодов. В результате интенсивной исследовательской работы в 70-х годах была разработана технология получения оптических волокон и световодных кабелей с малыми потерями из кварца и специальных стекол, что открыло путь к практической реализации оптических систем дальней связи.

Ослабление световой волны вызывается также радиационными потерями на изгибах световода, что ставит предел минимальным размерам схем с использованием оптических волокон.

Обращение модулей коэффициентов отражения $r_\perp$ и $r_\parallel$ (3.17) в единицу при $\varphi > \varphi_m$ находится в полном соответствии со сделанным выше выводом о равенстве нулю среднего потока энергии во второй среде в направлении, перпендикулярном границе. На вопрос о том, как проникла световая энергия (которой обладает неоднородная волна) во вторую среду, проведенный выше анализ ответить не может, так как он относился к стационарному состоянию и, кроме того, был основан на предположении, что граничная поверхность и волновая поверхность падающей волны бесконечны. В реальном эксперименте падающая волна ограничена и в пространстве, и во времени. В начале процесса, пока происходит установление колебаний, небольшая энергия проникает во вторую среду и создает там неоднородную волну. Когда падающий пучок света имеет поперечное сечение конечной ширины, вблизи его границ (точки A и B на рис. 3.12) полученные выше формулы неприменимы: здесь энергия падающей волны вполне может проходить из более плотной среды в менее плотную или обратно. Это и есть та энергия, которая вдали



3.12

Полное отражение пучка света конечного поперечного сечения

от краев пучка переносится неоднородной волной во второй среде в направлении, параллельном границе раздела.

Электромагнитное поле в первой среде в том месте, где происходит наложение падающей и отраженной волн (область внутри треугольника на рис. 3.12), тоже образует неоднородную плоскую волну, распространяющуюся параллельно границе раздела. Поверхности постоянной фазы этой волны, как и неоднородной волны во второй среде, представляют собой плоскости, перпендикулярные границе раздела. Они перемещаются вдоль границы с такой же скоростью $v=c/(n_1\sin\varphi)$. Амплитуда этой волны зависит от z, изменяясь периодически с пространственным периодом $\lambda_1/\cos\varphi$ (λ_1 — длина падающей волны в первой среде), в отличие от экспоненциального затухания вдоль z амплитуды неоднородной волны во второй среде. Средний поток энергии здесь тоже направлен вдоль границы и периодически зависит от координаты z (см. задачу 4), т. е. имеет слоистую структуру (рис. 3.12).

Рассмотрим изменение поляризации света при полном отражении. Как видно из формул (3.18), $r_\perp$ и $r_\parallel$ при полном отражении становятся комплексными. Это значит, что на границе фаза отраженной волны отличается от фазы падающей волны. Чтобы найти изменение фазы при отражении, представим комплексное число, стоящее в числителе выражения $r_\perp$ в (3.18), в показательной форме:

$$n_1\cos\varphi - i\sqrt{n_1^2\sin^2\varphi - n_2^2} = ae^{-i\delta_\perp/2},$$

где

$$\operatorname{tg}(\delta_\perp/2) = \sqrt{n_1^2\sin^2\varphi - n_2^2}/(n_1\cos\varphi). \qquad (3.19)$$

Знаменатель в $r_\perp$ — это число, комплексно сопряженное числителю, т. е. $a\exp(i\delta_\perp/2)$. Поэтому $r_\perp = e^{-i\delta_\perp}$ и $\delta_\perp$ в (3.19) представляет собой скачок фазы при отражении волны, поляризованной перпендикулярно плоскости падения.

Аналогичио найдем скачок фазы $\delta_\parallel$ для волны, поляризованной в плоскости падения, воспользовавшись выражением для $r_\parallel$ в (3.38):

$$\operatorname{tg}(\delta_\parallel/2) = n_1\sqrt{n_1^2\sin^2\varphi - n_2^2}/(n_2^2\cos\varphi). \qquad (3.20)$$

Скачки фаз $\delta_\perp$ и $\delta_\parallel$ для волн двух поляризаций различны. Если падающая волна поляризована *линейно*, то в отраженной волне возникает некоторая разность фаз $\delta = \delta_\perp - \delta_\parallel$ между компонентами поля, перпендикулярной и параллельной плоскости падения. Поэтому отраженная волна будет, вообще говоря, *эллиптически поляризована*. Для относительной разности фаз δ из (3.19) и (3.20) находим

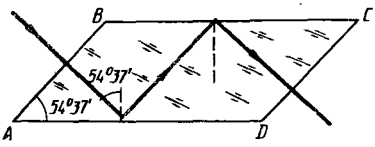
$$\operatorname{tg}(\delta/2) = \cos\varphi\sqrt{n_1^2\sin^2\varphi - n_2^2}/(n_1\sin^2\varphi). \qquad (3.21)$$

Разность фаз δ обращается в нуль при скользящем падении ($\varphi = \pi/2$) и при падении под критическом углом ($\varphi = \varphi_m$). Между этими значениями лежит угол, соответствующий максимуму относительной разности фаз. Максимум (3.21) достигается при $\cos^2\varphi = (n_1^2 - n_2^2)/(n_1^2 + n_2^2)$. Подставляя это значение в (3.21), получаем для максимального значения разности фаз δ_m выражение

$$\operatorname{tg}(\delta_m/2) = (n_1^2 - n_2^2)/(2n_1n_2). \qquad (3.22)$$

Чем больше различие в показателях преломления двух сред, тем большую разность фаз между параллельной и перпендикулярной компонентами волны можно получить при полном отражении. Для получения круговой поляризации отраженного света амплитуды этих компонент должны быть равны между собой, а разность фаз должна быть $\pi/2$. Первое условие выполняется, если падающая волна линейно поляризована в направлении, составляющем угол 45° с нормалью к плоскости падения. Удовлетворить второму условию, т. е. получить $\delta = \pi/2$, можно только при $n_1^2 - n_2^2 \geqslant 2n_1n_2$. Если $n_2 \approx 1$ (воздух), то показатель преломления n_1 оптически более плотной среды должен превосходить значение $\sqrt{2}+1 = 2{,}41$. В видимой области спектра столь большой показатель преломления имеет лишь алмаз.

Для стекла ($n_1 = 1{,}5$) максимальная разность фаз достигается при $\varphi = 51°20'$ и составляет $45°56'$. При двух значениях угла падения $\varphi = 48°37'$ и $\varphi = 54°37'$ разность фаз равна $\pi/4$. Двукратное отражение под одним из этих углов дает разность фаз $\pi/2$. Этим свойством воспользовался Френель для преобразования линейной поляризации света в круговую с помощью стеклянного параллелепипеда с углом $54°37'$ (рис. 3.13). Линейно поляризованный свет падает по нормали на грань AB и после двукратного полного отражения на гранях AD и BC выходит в прежнем



3.13
Параллелепипед Френеля

направлении через грань CD. Если плоскость поляризации падающего света составляет угол 45° с плоскостью падения, то выходящий свет имеет круговую поляризацию. Это устройство можно использовать и для превращения циркулярной поляризации света в линейную.

Контрольные вопросы

При каких условиях отражение на границе прозрачных сред будет полным?
Какими будут поверхности постоянной фазы и поверхности постоянной амплитуды волны в оптически менее плотной среде при полном отражении?
Как согласовать с законом сохранения энергии полное отражение падающей волны и наличие потока энергии, переносимого поверхностной волной во второй среде?
Как изменяется состояние поляризации света при полном отражении?

Задачи

С помощью формул Френеля показать, что при отражении от оптически менее плотной среды ($n_2 < n_1$) крутизна кривых $r_\perp(\varphi)$ и $r_\parallel(\varphi)$ при приближении φ к предельному углу полного отражения φ_m стремится к бесконечности.
При каком значении показателя преломления параллелепипед Френеля имеет угол 45°?
Ответ: $n = 1{,}55$.
Падающий на параллелепипед Френеля свет имеет круговую поляризацию. Каким будет направление линейной поляризации выходящего света?
Плоская волна падает из среды с показателем преломления n_1 на границу оптически менее плотной среды ($n_2 < n_1$) под углом φ, превышающим предельный угол φ_m. Волна поляризована перпендикулярно плоскости падения. Найти среднюю плотность потока энергии $\langle S \rangle$ в каждой среде.
В первой среде в области наложения падающей и отраженной волн вектор $\langle S \rangle$ направлен вдоль границы и периодически зависит от координаты z с периодом $\lambda_1/(2\cos\varphi)$:

$$\langle S_x^{(1)} \rangle = 2In_1 \sin\varphi \cos^2[(\omega n_1 \cos\varphi/c)z + \delta_\perp/2],\ \langle S_y^{(1)} \rangle = \langle S_z^{(1)} \rangle = 0.$$

Во второй среде вектор $\langle S \rangle$ также направлен вдоль границы, а его модуль экспоненциально убывает с увеличением z, т. е. фактически весь поток энергии сосредоточен в прилегающем к границе слое толщиной $l/2$:

$$\langle S_x^{(2)} \rangle = 2IA^2 n_1 \sin\varphi \exp(-2z/l),\ \langle S_y^{(2)} \rangle = \langle S_z^{(2)} \rangle = 0.$$

Здесь I — интенсивность падающей волны.

$$A = \frac{n_1 \cos\varphi}{\sqrt{n_1^2 - n_2^2}},\ \operatorname{tg}\frac{\delta_1}{2} = \frac{\sqrt{n_1^2 \sin^2\varphi - n_2^2}}{n_1 \cos\varphi},\ l = \frac{c}{\omega\sqrt{n_1^2 \sin^2\varphi - n_2^2}}.$$

При $z = 0$ (на границе) $\langle S_x^{(1)} \rangle = \langle S_x^{(2)} \rangle$ — плотность потока энергии в каждой из сред имеет одно и то же значение вблизи границы.

3.4. Отражение света от поверхности металлов

Особенности отражения света от металлической поверхности связаны с наличием в металле свободных электронов, ответственных за его электропроводность. Вынужденные колебания свободных электронов под действием поля падающей на границу металла электромагнитной волны, происходящие в примыкающем к этой границе тонком слое, создают сильную отраженную волну. Ее интенсивность может приближаться к интенсивности падающей волны. Вследствие большой плотности свободных электронов (около 10^{22} см$^{-3}$) даже сравнительно тонкие слои металла отражают бо́льшую часть падающего на них света и поэтому практически непрозрачны в оптическом диапазоне. Благодаря высокой отражательной способности металлы играют важную роль в оптике: поверхности некоторых металлов служат прекрасными зеркалами.

Частичное проникновение света в металл создает токи проводимости. С ними связано выделение джоулевой теплоты, т. е. *поглощение* света — необратимое превращение электромагнитной энергии в энергию хаотического теплового движения. Чем выше проводимость металла, тем меньшая доля падающего света проникает в металл и поглощается там. В *идеальном* проводнике, которому формально соответствует бесконечно большая проводимость, потери на джоулеву теплоту вообще отсутствуют, так что падающий свет полностью отражается.

Отражение монохроматического света от поверхности металла, как и его распространение в поглощающей среде, можно рассмотреть на основе макроскопических уравнений Максвелла и материальных уравнений, в которых диэлектрическая проницаемость $\varepsilon(\omega)$ комплексна. Ее мнимая часть ответственна за поглощение света, т. е. описывает джоулевы потери. При использовании комплексной диэлектрической проницаемости уравнения Максвелла и вытекающие из них граничные условия для векторов электромагнитного поля формально принимают такой же вид, как и в прозрачной среде. Поэтому полученные выше законы отражения и преломления остаются в силе и для поглощающих сред, включая металлы, если входящий в них показатель преломления n считать комплексным: $\sqrt{\varepsilon} = n + i\varkappa$ (см. § 2.2). В сильно поглощающих средах и металлах мнимая часть преобладает. В макроскопической теории величины $n(\omega)$ и $\varkappa(\omega)$ полностью характеризуют оптические свойства поглощающей среды. Экспериментальные методы их определения основаны на изучении отраженного света. Измерение характеристик отраженного света позволяет как бы «заглянуть» внутрь металла и получить сведения о значениях n и $\varkappa$ для массивного образца, несмотря на малую глубину проникновения зондирующего света.

Рассмотрим падающую (из вакуума или воздуха) на поверхность металла плоскую монохроматическую волну, волновой вектор которой $\mathbf{k}_0$ образует с нормалью угол φ (рис. 3.14). В результате

в первой среде возникают отраженная волна с волновым вектором $\mathbf{k}_1$ и неоднородная волна, прошедшая в металл. Ее волновой вектор комплексный: $\mathbf{k}_2=\mathbf{k}_2'+i\mathbf{k}_2''$. Обращаясь к формулам (3.2), находим

$$k_{1x}=k_{2x}=k_{0x}=(\omega/c)\sin\varphi. \qquad (3.23)$$

Отсюда прежде всего следует, что геометрический закон отражения от металлов такой же, как и для границы прозрачных сред. Для волны в металле из (3.23) получаем, что составляющая вектора $\mathbf{k}_2$, направленная вдоль границы, вещественна: $k_{2x}=(\omega/c)\sin\varphi$. Поэтому вектор $\mathbf{k}_2''$ (мнимая часть) перпендикулярен поверхности металла. Это значит, что плоскости равных амплитуд прошедшей волны параллельны границе. Вектор $\mathbf{k}_2'$ перпендикулярен плоскостям постоянных фаз и характеризует направление прошедшей волны. Угол ψ, который он образует с нормалью к границе, называется вещественным углом преломления. Отношение синусов угла падения и вещественного угла преломления $\sin\varphi/\sin\psi$ зависит от угла падения в отличие от преломления на границе прозрачной среды, где $\sin\varphi/\sin\varphi_2=\text{const}$.

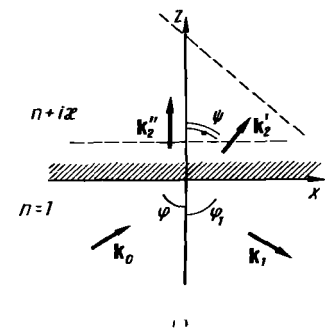
Формулы Френеля (3.8) — (3.9) остаются в силе и для волн, отраженных от поверхности металла, если в них рассматривать $\cos\varphi_2$ как комплексную величину, определяемую законом преломления (3.4):

$$\cos\varphi_2=\sqrt{1-\sin^2\varphi/(n+i\varkappa)^2},$$

причем нужно взять то значение квадратного корня, которое имеет положительную мнимую часть. Тогда неоднородная волна будет затухать в глубь металла. Явные выражения для амплитуд отраженных от поверхности металла волн в общем случае получаются чрезвычайно громоздкими, и мы их здесь не приводим.

В случае комплексного показателя преломления отношения амплитуд отраженных волн к амплитудам падающих $E_1^\perp/E_0^\perp$ и $E_1^\parallel/E_0^\parallel$, вычисляемые по формулам Френеля для каждой из двух поляризаций, также комплексные:

$$E_1^\perp/E_0^\perp=\rho_\perp e^{i\Delta_\perp},\quad E_1^\parallel/E_0^\parallel=\rho_\parallel e^{i\Delta_\parallel}.$$



Отражение и преломление света на поверхности металла

В общем случае $\Delta_\perp\neq\Delta_\parallel$. Поэтому при линейной поляризации падающего света между двумя компонентами отраженной волны появляется сдвиг фаз, приводящий к эллиптической поляризации отраженного света. Отраженный свет остается линейно поляризованным, если падающий поляризован в плоскости падения или в перпендикулярном направлении. При произвольном направлении линейной поляризации падающего света отраженный ос-

тается линейно поляризованным при нормальном ($\varphi=0$) и при скользящем ($\varphi=\pi/2$) падении. В этих случаях направление поляризации в пространстве остается неизменным.

Измерение эллиптической поляризации света, отраженного от поверхности металла при наклонном падении линейно поляризованного света, лежит в основе предложенного Друде экспериментального метода определения оптических характеристик n и $\varkappa$ металла. Теория связывает n и $\varkappa$ с эксцентриситетом и положением осей эллипса колебаний. По данным измерений этих величин можно рассчитать n и $\varkappa$. Наибольшая чувствительность метода (и одновременное упрощение расчетных формул) достигается при определенном угле падения (*главном* угле падения, играющем при отражении от поглощающих сред ту же роль, что и угол Брюстера при отражении от прозрачных сред). В большинстве случаев он лежит вблизи 70°. Для этого угла отраженный свет имеет круговую поляризацию, если соответствующим образом подобрать направление поляризации падающего света.

Информацию об оптических характеристиках металла можно получить не только из измерений состояния поляризации отраженного света, но и из сравнения интенсивностей отраженного и падающего света. Рассмотрим нормальное падение света на поверхность металла. В этом случае для амплитуды отраженной волны можно воспользоваться формулой (3.12), подставив в нее $n_1=1$, $n_2=n+i\varkappa$:

$$E_1/E_0=(1-n-i\varkappa)/(1+n+i\varkappa)=|r|e^{i\delta}. \tag{3.24}$$

Отсюда, умножая (3.24) на комплексно-сопряженную величину $|r|e^{-i\delta}$, находим коэффициент отражения при нормальном падении:

$$R=[(n-1)^2+\varkappa^2]/[(n+1)^2+\varkappa^2]. \tag{3.25}$$

У металлов слагаемое $\varkappa^2$ в числителе и знаменателе этой формулы часто значительно больше другого слагаемого. Тогда значение R близко к единице, т. е почти вся энергия падающего света отражается. В видимой области натрий отражает свыше 97%, серебро — 95% света, падающего на чистую поверхность.

Волновой вектор прошедшей в металл волны при нормальном падении, как видно из (3.2), имеет только z-составляющую, которая находится из (3.4): $k_{2z}=(\omega/c)\sqrt{\varepsilon}=(\omega/c)(n+i\varkappa)$. В этом случае поверхности равных фаз и равных амплитуд параллельны границе. Амплитуда волны на границе находится из (3.12): $E_2=2E_0/(1+n+i\varkappa)$. Таким образом, для напряженности электрического поля волны в металле получаем

$$E_2(z,t)=2E_0\frac{1+n-i\varkappa}{(1+n)^2+\varkappa^2}e^{-\omega\varkappa z/c}e^{-i\omega(t-zn/c)}. \tag{3.26}$$

Амплитуда волны уменьшается в глубь металла по закону $\exp(-z/l)$, где $l=c/(\varkappa\omega)=\lambda_0/(2\pi\varkappa)$ характеризует глубину проникновения (толщину скин-слоя); $\lambda_0=2\pi c/\omega$ — длина волны падающего излучения в вакууме. При $\varkappa=1$ в слое толщиной в одну длину волны амплитуда уменьшается в $e^{2\pi}$ раз, а интенсивность уменьшается в $e^{4\pi}\approx 3\cdot 10^5$ раз. Для большинства металлов при измерениях в видимой области значение $\varkappa$ лежит между 2 и 5. В инфракрасной области значение $\varkappa$ еще больше: у серебра $\varkappa\approx 40$ при $\lambda=6$ мкм. Эти цифры дают представление о том, насколько мала глубина проникновения света в металл.

Определяемые экспериментально значения оптических характеристик металлов не отличаются высокой точностью. Воспроизводимость измеряемых значений n и $\varkappa$ в пределах нескольких процентов считается удовлетворительной. Причина этого связана с тем, что в случае сильно поглощающих сред, таких, как металлы, все процессы происходят в тонких слоях ($\sim 10^{-4}$ мм) вблизи поверхности. Поверхностные слои не защищены от внешних воздействий, их свойства изменяются со временем и зависят от способа обработки поверхности. Образование переходных слоев на поверхности при ее обработке может внести заметные искажения в результаты измерений, когда толщина их сравнима с глубиной проникновения.

Последовательный теоретический расчет оптических характеристик металлов $n(\omega)$ и $\varkappa(\omega)$ возможен только в рамках квантовой теории дисперсии. Основанная на упрощенных модельных представлениях классическая теория дисперсии в металлах рассматривалась в § 2.6.

При достаточно высоких частотах ($\omega\gg\gamma$) роль силы «трения» в уравнении движения свободного электрона становится несущественной, и для $\varepsilon(\omega)$ теория дисперсии дает выражение (2.54), получающееся из полной формулы (2.53) при $\gamma=0$. Случай $\gamma=0$ формально соответствует «идеальному» металлу, обладающему бесконечно большой проводимостью $\sigma\to\infty$. При $\omega<\omega_p$ диэлектрическая проницаемость идеального проводника отрицательна, поэтому $n=0$, а

$$\varkappa(\omega)=\sqrt{\omega_p^2/\omega^2-1}\,.$$

В этом случае (3.25) дает $R=1$, т. е. отражение от поверхности идеального проводника полное. Наложение отраженной волны на падающую приводит к образованию стоячей волны, в то время как в металле напряженность поля убывает в e раз на расстоянии $l=\lambda_0/(2\pi\varkappa)$ от поверхности (см. задачу).

Основанное на макроскопической электродинамике описание оптических свойств металлов построено в предположении, что действующее на отдельный свободный электрон поле волны можно считать однородным. Характерный пространственный масштаб изменения напряженности поля в металле дает глубина проникновения (или толщина скин-слоя) l. Если средняя длина свободного пробега электрона мала по сравнению с толщиной скин-слоя, упомянутое выше предположение выполняется.

Но с понижением температуры проводимость металла растет и, следовательно, уменьшается глубина проникновения. Одновременно средняя длина свободного пробега электрона возрастает, так что она может стать больше глубины скин-слоя. В таких условиях скин-эффект называется аномальным. За время одного свободного пробега электрон будет двигаться через области с разной напряженностью поля. Поэтому при расчете вклада свободного электрона в поляризованность металла действующее на него электрическое поле нельзя считать однородным. В результате оказывается, что при низких температурах макроскопическая теория применима в области достаточно низких частот, когда глубина скин-слоя значительно больше длины свободного пробега, и в области высоких частот, когда столкновения электронов становятся несущественными и глубина скин-слоя должна превосходить не длину свободного пробега, а среднее расстояние, проходимое электроном в течение одного периода колебаний поля. При промежуточных частотах ни одно из этих условий не выполняется и нужно учитывать пространственную зависимость напряженности электрического поля, действующего на свободные электроны. Задача теоретического описания оптических свойств металла в таких условиях становится чрезвычайно сложной. Она может быть решена методами физической кинетики.

Терминология, используемая в случае поглощающих сред, нередко приводит к недоразумениям. Поглощение называют слабым, если глубина проникновения много больше длины волны — амплитуда постепенно уменьшается на протяжении многих длин волн. Но «слабо» поглощающие вещества, такие, как растворы красителей (разбавленные чернила), могут при достаточной толщине слоя поглотить практически *всю* энергию падающего на них излучения. Поглощение называют сильным, если глубина проникновения меньше длины волны. «Сильно» поглощающие вещества, например металлы, способны поглотить лишь *малую долю* энергии падающего света: подавляющая часть энергии отражается. Это общее правило: если материал обладает сильным поглощением на какой-то частоте, то отражение света данной частоты на его границе очень велико и лишь малая доля энергии попадает внутрь и поглощается.

Следующий простой опыт может служить иллюстрацией этого эффекта. На поверхность стеклянной пластинки наносится слой фиолетовых чернил. Пластинка имеет одинаковую фиолетовую окраску как в проходящем, так и в отраженном свете, ибо окраска слабопоглощающей среды (раствора красителя) обусловлена избирательным поглощением света в толще слоя. Когда чернила засохнут, они превратятся в среду с сильным поглощением. В проходящем свете пластинка по-прежнему имеет фиолетовую окраску, но в отраженном свете засохшие чернила приобретают золотистый металлический блеск. Он обусловлен избирательным отражением света от поверхности для тех длин волн, которые испытывают сильное поглощение в слое красителя.

Поверхностная окраска некоторых металлов обусловлена зависимостью их оптических характеристик от длины волны. Так как металл представляет собой среду с сильным поглощением, тонкие пленки металла при наблюдении на просвет будут иметь дополнительную окраску. Например, очень тонкие пленки золота, имеющего желтый цвет, в проходящем свете кажутся зелеными.

Контрольные вопросы

- ☐ В чем заключаются основные особенности отражения света от металлов (по сравнению с отражением от прозрачных сред)?
- ☐ В каких случаях отраженный металлическим зеркалом свет сохранит линейную поляризацию?
- ☐ Почему результаты измерений оптических характеристик металлов по поляризации отраженного света имеют плохую воспроизводимость?
- ☐ Почему металлы сильно поглощают свет, а обладающие высокой проводимостью электролиты прозрачны?
- ☐ В чем заключаются различия поверхностной окраски сред, обладающих сильным и слабым поглощением?

Задача

Световая волна частотой ω падает по нормали из вакуума на плоскую поверхность идеального металла. Найти напряженности электрического поля в металле и в вакууме для $\omega < \omega_p$.

◇ При $\omega < \omega_p$ диэлектрическая проницаемость идеального проводника отрицательна $(\varepsilon = 1 - \omega_p^2/\omega^2 < 0)$, поэтому $n = 0$, а $\varkappa = (\omega_p^2/\omega^2 - 1)^{1/2}$. Для отношения амплитуд отраженной и падающей волн из (3.24) находим

$$E_1/E_0 = (1 - i\varkappa)/(1 + i\varkappa) = e^{-2i\Delta}, \quad \operatorname{tg}\Delta = \varkappa = \sqrt{\omega_p^2/\omega^2 - 1}. \qquad (3.27)$$

Отражение от поверхности идеального проводника будет полным даже при нормальном падении. Напряженность электрического поля в вакууме находится сложением напряженностей полей падающей и отраженной волн:

$$E_0[e^{i(kz - \omega t)} + e^{i(-kz - \omega t - 2\Delta)}] = 2E_0 \cos(kz + \Delta)\, e^{-i(\omega t + \Delta)}.$$

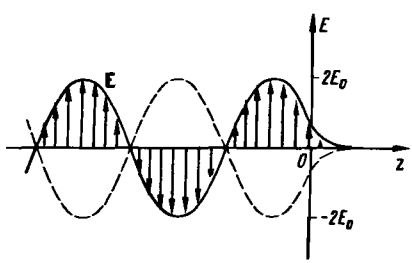
Для нахождения амплитуды напряженности электрического поля в металле воспользуемся второй из формул (3.12):

$$E_2/E_0 = 2/(1 + i\varkappa) = 2e^{-i\Delta}/\sqrt{1 + \varkappa^2} = 2(\omega/\omega_p)e^{-i\Delta}.$$

Зависимость напряженности электрического поля в металле от координат и времени получим, подставляя E_2 в выражение $E_2(z, t) = E_2 \exp i(k_{2z}z - \omega t)$ с $k_{2z} = i\omega\varkappa/c$ из (3.4):

$$E_2(z, t) = 2E_0(\omega/\omega_p)e^{-\omega\varkappa z/c} e^{-i(\omega t + \Delta)}. \qquad (3.28)$$

Напряженности поля в вакууме и в металле имеют одинаковую зависимость от времени. Значения их на границе (т. е. при $z = 0$) совпадают, так как $\cos\Delta = 1/\sqrt{1 + \operatorname{tg}^2\Delta} = 1/\sqrt{1 + \varkappa^2} = \omega/\omega_p$. Этого и следовало ожидать из граничных условий. Зависимость напряженности электрического поля от z показана на рис. 3.15. Напряженность поля в металле затухает экспоненциально, убывая в e раз на глубине $l = c/(\omega\varkappa) = \lambda_0/(2\pi\varkappa)$, где λ_0 — длина волны в вакууме. В вакууме (при $z < 0$) образуется стоячая волна, первый узел напряженности электрического поля которой расположен от поверхности металла на расстоянии, несколько меньшем половины длины волны. ◇



3.15

Напряженность электрического поля при нормальном падении волны на границу идеального металла

3.5. Световое давление. Импульс электромагнитной волны

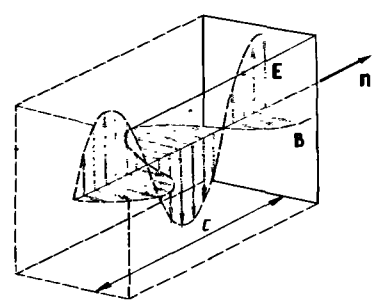
Поглощение света в веществе, а также отражение и преломление на границе сопровождаются появлением механических сил, действующих на вещество. Эти силы принято называть *световым давлением*.

Идея о световом давлении была высказана еще Кеплером для объяснения формы кометных хвостов. В рамках корпускулярных представлений о природе света такая гипотеза была естественной, так как световые частицы должны были бы передавать свой импульс поглощающим и отражающим телам, т. е. производить давление. Неудачи ранних попыток обнаружить световое давление на опыте приводились Франклином и Юнгом как один из аргументов против корпускулярной теории. Волновая теория, рассматривавшая свет как поперечные упругие волны, отрицала световое давление. Однако пришедшая ей на смену электромагнитная волновая теория света дает объяснение возникновению светового давления и позволяет его рассчитать. Экспериментально световое давление было впервые обнаружено и измерено П. Н. Лебедевым в 1900 г. в исключительно тонких опытах. Измерения Лебедева, подтвердившие рассчитанное Максвеллом световое давление, сыграли большую роль в становлении электромагнитной теории света.

Рассмотрим сначала простейший случай нормального падения света на поверхность поглощающей среды. Покажем, что среда при этом испытывает давление в направлении падающей волны. Падающая волна взаимодействует с электрическими зарядами среды. Электрическое и магнитное поля волны действуют на заряд Q с силой $\mathbf{F}=Q\mathbf{E}+Q\mathbf{v}\times\mathbf{B}$, где $\mathbf{v}$ — скорость движения заряда. В бегущей электромагнитной волне, согласно (1.31), в каждой точке в любой момент времени $B=E/c$, поэтому при $v\ll c$ второе слагаемое в $\mathbf{F}$ — обусловленная магнитным полем волны сила Лоренца — всегда много меньше первого слагаемого QE. Это позволяет считать, что вынужденное движение заряда обусловлено электрическим полем волны, и не учитывать силу Лоренца в основном уравнении электронной теории дисперсии (2.30). Но именно с силой Лоренца связано давление, оказываемое волной на среду.

Выразим в формуле для силы Лоренца $Q\mathbf{v}\times\mathbf{B}$ индукцию магнитного поля волны через напряженность электрического поля. Введем для этого единичный вектор $\mathbf{n}$ вдоль направления распространения волны: $\mathbf{n}=\mathbf{k}/k$ (рис. 3.16), тогда $\mathbf{B}=\mathbf{n}\times\mathbf{E}/c$. Таким образом, выражение для силы Лоренца принимает вид

$$\mathbf{F}_{\text{л}}=Q\mathbf{v}\times(\mathbf{n}\times\mathbf{E})/c=\mathbf{n}Q(\mathbf{vE})/c. \qquad (3.29)$$



К расчету светового давления

Здесь при преобразовании двойного векторного произведения по формуле $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=\mathbf{B}(\mathbf{AC})-\mathbf{C}(\mathbf{AB})$ мы предположили, что вектор скорости $\mathbf{v}$ направлен по одной прямой с вектором напряженности $\mathbf{E}$ электрического поля световой волны, так как именно электрическое поле волны возбуждает движение зарядов в веществе. Произведение силы $Q\mathbf{E}$ на скорость $\mathbf{v}$ равно мощности P, отдаваемой заряду электромагнитной волной. Поэтому формулу (3.29) можно записать в виде

$$\mathbf{F}_{\text{л}}=\mathbf{n}P/c. \tag{3.30}$$

Так как взаимодействие волны с зарядом Q приводит в среднем к передаче энергии волны заряду, то среднее за период значение мощности P положительно. Поэтому направление действующей на заряд средней силы, как видно из (3.30), совпадает с $\mathbf{n}$, т. е. с направлением волны. Таким образом, при усреднении по времени полной силы, действующей на заряд со стороны электромагнитного поля световой волны, вклад силы $Q\mathbf{E}$ обращается в нуль, а сила Лоренца $Q\mathbf{v}\times\mathbf{B}$ приводит к световому давлению.

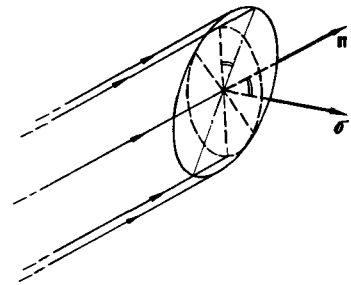
Важно отметить, что из выражения (3.30) для силы, действующей на заряженную частицу, выпала основная характеристика частицы — ее заряд Q. Это позволяет обобщить формулу (3.30) и не связывать полученный результат с модельными представлениями электронной теории, использованными при ее выводе. Пусть, например, вся энергия падающего света поглощается преградой. Тогда поглощаемая мощность P может быть выражена через вектор Пойнтинга $\mathbf{S}=\mathbf{n}wc$ падающей волны (w — объемная плотность излучения). Мощность dP, поглощаемая элементарной площадкой $d\sigma$, равна $\mathbf{S}d\sigma$, а действующая на нее сила в соответствии с (3.30) составляет $\mathbf{n}(\mathbf{S}d\sigma)/c$. Сила светового давления на всю поверхность σ поглощающей преграды определяется полным потоком излучения сквозь эту поверхность:

$$\mathbf{F}=\mathbf{n}\int_{\sigma}(\mathbf{S}d\sigma)/c. \tag{3.31}$$

На плоскую площадку, характеризуемую вектором $\boldsymbol{\sigma}$, действует сила светового давления $\mathbf{F}=\mathbf{n}(\mathbf{S}\boldsymbol{\sigma})/c=w\mathbf{n}(\mathbf{n}\boldsymbol{\sigma})$, где $(\mathbf{n}\boldsymbol{\sigma})$ — площадь перпендикулярной направлению волны проекции площадки (рис. 3.17).

Существование светового давления приводит к заключению о том, что электромагнитное поле обладает *импульсом*. Понятие импульса, как и понятие энергии, по мере развития науки значительно изменялось. Вначале эти понятия применялись только к механическим движениям. В дальнейшем они были обобщены так, что стали охватывать и другие формы движения материи. Это позволило сформулировать универсальные законы сохранения энергии и импульса, учитывающие возможность их превращения из одной формы в другую.

В частности, давление света на материальные тела заставляет приписать ему определенный импульс. Сила светового давления, как и любая другая механическая сила, изменяет импульс тела, на которое она действует. Исходя из универсальной применимости закона сохранения импульса нужно считать, что тело приобретает механический импульс за счёт соответствующего изменения импульса другой формы движения. Так как единственный процесс, сопровождающий в данных условиях изменение импульса тела, — это поглощение света, то и носителем другой формы импульса может быть только свет. За 1 с импульс тела получает приращение, численно равное действующей силе. При нормальном падении сила давления света на единичную площадку поверхности поглощающего тела равна $\mathbf{S}/c=\mathbf{n}w$. Как раз таким импульсом должен обладать поглощенный за 1 с свет. Его энергия $S=wc$ была распределена с плотностью w в опирающемся на единичную площадку объеме, вытянутом вдоль волны на расстояние c (рис. 3.17). Естественно считать, что и переданный площадке импульс был распределен в том же объеме. Так как этот объем равен c, то для импульса, приходящегося на единицу объема, т. е. для *объемной плотности импульса электромагнитного поля* $\mathbf{p}$, получаем



3.17
Световое давление на наклонную площадку

$$\mathbf{p}=\mathbf{S}/c^2=\mathbf{n}w/c. \qquad (3.32)$$

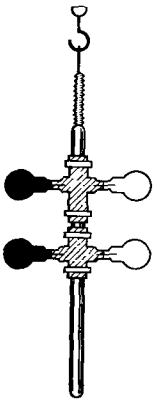
Таким образом, импульс световой волны равен ее энергии, деленной на скорость света, и направлен в сторону распространения света.

Используя представление об импульсе электромагнитного поля, легко найти силу светового давления и в тех случаях, когда падающий свет не обязательно полностью поглощается телом, а может также частично или полностью отражаться или рассеиваться. Сила давления света всегда равна передаваемому телу за 1 с импульсу света, т. е. векторной разности импульсов падающего и отраженного (или рассеянного) за 1 с света. Например, при нормальном падении на поверхность идеального зеркала отраженный свет обладает противоположно направленным импульсом. Поэтому передаваемый зеркалу импульс вдвое больше, чем в случае полного поглощения. При нормальном падении на зеркало с коэффициентом отражения R давление равно $w(1+R)$. При наклонном падении сила светового давления имеет как нормальную, так и тангенциальную (при $R\neq 1$) составляющие (см. задачу 1). Для тела, рассеивающего падающий свет равномерно по всем направлениям,

как, например, зеркальный шар, сила светового давления будет такой же, как и при полном поглощении, так как суммарный импульс изотропно рассеянного света равен нулю (см. задачу 2). Когда свет падает со всех направлений на внутреннюю зеркальную стенку полости, равномерно заполненной изотропным излучением с объемной плотностью w, его давление равно $w/3$ (см. задачу 3).

Итак, давление света на поверхность тела обусловлено передачей телу определенного импульса отражающимся светом, подобно тому как давление газа на стенки сосуда объясняется тем, что молекулы газа передают этим стенкам некоторый импульс при соударениях с ними.

В лабораторных условиях импульс света и световое давление обычно настолько малы, что их непосредственное измерение затруднительно. Например, достигающий Земли солнечный свет действует на расположенное перпендикулярно его лучам зеркало с силой 10^{-5}Н на 1 м^2 поверхности. В опытах Лебедева использовался легкий подвес на тончайшей упругой нити, по одну сторону которого были прикреплены тонкие и легкие зеркальные металлические крылышки, а по другую — зачерненные (рис. 3.18). Подвес находился в откачанном сосуде, образуя чувствительные крутильные весы. Давление света на зеркальное крылышко приблизительно вдвое больше, чем на зачерненное, что должно приводить к соответствующему закручиванию подвеса при освещении. Основные трудности таких опытов связаны с тем, что при недостаточно высоком вакууме вторичные явления могут полностью замаскировать искомый эффект светового давления. Главные из них — это конвекционные потоки остаточного газа и радиометрический эффект, возникающий в разреженном газе вследствие разности температур освещенной и неосвещенной сторон крылышка. Радиометрический эффект вызывает закручивание подвеса в направлении, противоположном действию светового давления. Лебедеву удалось почти полностью исключить радиометрическое действие благодаря тщательно продуманной методике и технике эксперимента. Его измерения дали величину, согласующуюся с теорией с точностью до 20%. Опыты Лебедева были повторены Герлахом в 1923 г., использовавшим более совершенные методы получения вакуума. Полученные результаты согласуются с теорией в пределах погрешности опыта (до 2%).



Один из подвесов, использованных П. Н. Лебедевым

После создания лазеров световое давление снова привлекло внимание исследователей. Высокие временна́я и пространственная когерентности лазерного излучения позволяют сфокусировать его в области с размерами порядка длины световой волны. При этом достигаются очень

большие значения плотности энергии и импульса световой волны. Простые оценки показывают, что сила давления лазерного излучения мощностью 1 Вт (такую мощность легко получить для зеленого света аргонового лазера), сфокусированного на частицу с размерами порядка длины световой волны, должна сообщить частице ускорение, примерно в миллион раз превышающее ускорение свободного падения. При попытках экспериментального осуществления этой идеи возникает ряд трудностей. Так, например, нельзя использовать металлические частицы, так как даже при коэффициенте отражения 98% оставшихся двух процентов падающей мощности, поглощаемых частицей, достаточно для ее плавления и испарения. Движение под действием давления сфокусированного лазерного излучения можно наблюдать, используя малые прозрачные сферические частицы (например, шарики полиметилкрилата), взвешенные в дистиллированной воде. Несмотря на малый коэффициент отражения [как видно из формулы (3.14), он пропорционален квадрату разности показателей преломления материала частицы и окружающей среды], силы светового давления оказываются достаточными для их обнаружения в наблюдаемом движении частицы на фоне вторичных процессов (конвекционных потоков жидкости, реактивных сил, возникающих, подобно радиометрическим силам в газе, при неравномерном нагреве частицы падающим излучением и т. п.).

Давление и импульс излучения проявляются в двух противоположных по масштабам областях: астрономической и субатомной. Например, притяжение верхних слоев звезд к их центру в значительной степени уравновешивается давлением излучения, идущего от центра звезды наружу. Световое давление приводит к некоторому предельному значению массы, при котором звезда еще остается устойчивой. Этот вывод согласуется с астрономическими данными, согласно которым звезды с массой, превосходящей некоторый известный предел, не наблюдаются.

Из явлений микромира отметим эффект Комптона (см. § 9.6), при котором рентгеновское излучение передает часть своего импульса электронам, на которых оно рассеивается, и тем самым сообщает этим «электронам отдачи» большие скорости. Импульс излучения обнаруживает себя также в «отдаче», которую испытывает атомное ядро при испускании гамма-лучей. Это явление вполне аналогично отдаче ружья при выстреле. Эффект отдачи в принципе существует и при испускании света атомами, но в оптической области он приводит к ничтожному сдвигу частоты испускаемого света (значительно меньшему естественной ширины линии).

В случае классической модели элементарного источника света — линейного осциллятора — эффект отдачи отсутствует ввиду симметрии непрерывно испускаемой им волны. Суммарный импульс такой волны равен нулю. Наблюдаемый на опыте эффект отдачи при испускании гамма-лучей атомными ядрами свидетельствует о неклассическом дискретном характере испускания,

когда вылетающий в определенном направлении гамма-квант уносит конечную порцию импульса.

Свет круговой поляризации наряду с энергией и импульсом обладает также *моментом импульса*. Для его определения можно рассмотреть вынужденное движение изотропного заряженного осциллятора под действием электрического поля циркулярно поляризованной волны. Это движение будет происходить по окружности с угловой скоростью ω, равной частоте падающей волны, причем смещение заряда из положения равновесия не совпадает с направлением действующей на него силы $q\mathbf{E}$. Поэтому на заряд действует некоторый момент сил $\mathbf{N}$ относительно центра, благодаря которому волна передает осциллятору мощность $P=N\omega$. Наличие момента сил означает, что осциллятор за 1 с приобретает механический момент импульса, численно равный этому моменту сил $\mathbf{N}$. Основываясь на универсальном законе сохранения момента импульса, заключаем, что за 1 с осциллятор вместе с энергией P получил механический момент импульса $\mathbf{N}$ от электромагнитного поля действующей на него световой волны круговой поляризации. Таким образом, луч света левой круговой поляризации с энергией w переносит момент импульса w/ω, направленный вдоль луча. Правополяризованный свет обладает противоположно направленным моментом импульса.

В области макроскопических явлений экспериментальное измерение момента импульса света представляет очень трудную задачу из-за ничтожной величины связанных с ним эффектов. Тем не менее в исключительно тонких экспериментах удалось обнаружить момент импульса, передаваемый светом полуволновой кристаллической пластинке, при прохождении через которую правополяризованный свет становится левополяризованным, или наоборот (см. § 4.1). Но в области атомных явлений обмен моментом импульса между светом и веществом имеет существенное значение. Например, при испускании света круговой поляризации возбужденным атомом момент импульса электронов в атоме изменяется на величину, сравнимую с моментом импульса всего атома.

Контрольные вопросы

- ☐ Какую роль играют электрическое и магнитное поля световой волны в возникновении давления света?
- ☐ Как сила светового давления, действующая на совершающий вынужденное движение заряд, зависит от передаваемой заряду мощности?
- ☐ Как импульс электромагнитного поля световой волны выражается через ее энергию?
- ☐ Чему равно давление на поверхность с коэффициентом отражения R падающей по нормали световой волны, плотность энергии в которой равна w?
- ☐ С какими трудностями сопряжено измерение светового давления в лабораторных условиях?
- ☐ В каких явлениях давление и импульс излучения играют существенную роль?

Задачи

1. Найтн нормальиую и тангенцнальиую составляющие силы светового давления при паденнн волны с объемной плотностью энергии w под углом 0 на поверхность, которая: а) поглощает все падающее на иее нзлучеиие; б) полностью отражает; в) частично отражает (коэффицнент отражения R); г) рассеивает равномерно по всем направлениям (по закону Ламберта).

◇ Ответ: а) $f_\perp = w\sigma\cos^2\theta$, $f_\parallel = w\sigma\sin\theta\cos\theta$; б) $f_\perp = 2w\sigma\cos^2\theta$, $f_\parallel = 0$; в) $f_\perp = w\sigma(1+R)\cos^2\theta$, $f_\parallel = w\sigma(1-R)\sin\theta\cos\theta$; г) $f_\perp = w\sigma\cos\theta(\cos\theta + {}^1/_2)$, $f_\parallel = w\sigma(1-R)\sin\theta\cos\theta$; здесь w — плотность энергии падающей волны. ◇

2. Найти силу давления плоской световой волны на шар радиусом R_0, когда его поверхность: а) полностью поглощает падающее излучение; б) зеркально отражает ($R=1$).

◇ Ответ: сила давления в случаях а) и б) одинакова и равна $w\pi R_0^2$. ◇

3. Найти давленне изотропного излучения с объемной плотностью w на идеальное зеркало.

◇ Ответ: $p = {}^1/_3 w$. ◇

РАСПРОСТРАНЕНИЕ СВЕТА В АНИЗОТРОПНОЙ СРЕДЕ

● Прохождение света через анизотропное вещество, оптические свойства которого в разных направлениях не одинаковы, сопровождается рядом своеобразных явлений, имеющих большое принципиальное и практическое значение. Особенности оптических явлений в анизотропных средах связаны с тем, что индуцируемый электромагнитной волной дипольный момент физически бесконечно малого элемента объема среды, вообще говоря, не совпадает по направлению с электрическим полем волны. Так происходит потому, что в анизотропном веществе под действием внешней силы элементарные заряды смещаются в одних направлениях легче, чем в других.

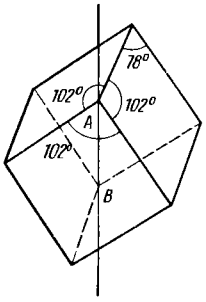
Физическая природа анизотропии вещества связана с особенностями строения его молекул или особенностями кристаллической решетки, в узлах которой находятся атомы или ионы. В таких условиях модель изотропного осциллятора для описания оптического электрона в атоме может оказаться непригодной, так как действующая на него возвращающая сила при смещении из положения равновесия обусловлена не только сферически симметричным полем ионного остова, но и полем соседних атомов или ионов, которое имеет более низкую симметрию.

● Изучение распространения света в анизотропных средах построено в этой главе не на учете их атомной структуры, а с помощью феноменологической электромагнитной теории. В рамках такой теории анизотропия учитывается тем, что в материальном уравнении восприимчивость $\varkappa(\omega)$ представляет собой тензор (а не скаляр, как в изотропной среде).

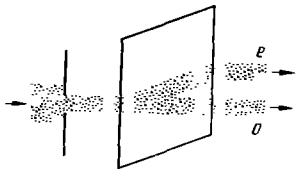
ри падении световой волны на границу изотропной среды в этой среде от границы распространяется одна волна, волновой вектор которой лежит в плоскости падения и составляет с нормалью к границе угол φ_2, определяемый известным законом преломления: $n_1 \sin \varphi_1 = n_2 \sin \varphi_2$. Если среда анизотропна, то в ней в общем случае возникают *две волны,* распространяющиеся от границы в разных направлениях и с разными скоростями. Это явление называется *двойным лучепреломлением*. Оно было открыто Бартолином в 1670 г. в кристаллах исландского шпата и впоследствии подробно исследовано Гюйгенсом. Гюйгенс дал объяснение двойного лучепреломления на основе гипотезы о том, что падающая на границу волна порождает в кристалле элементарные вторичные волны двух видов: сферические (обыкновенные) и эллипсоидальные (необыкновенные), скорость которых зависит от направления.

Исландский шпат представляет собой разновидность углекислого кальция $CaCO_3$ (кальцит), относящуюся к гексагональной системе. Он встречается в природе в виде довольно больших и оптически чистых кристаллов. Раскалыванием по определенным плоскостям (плоскостям спайности) кристаллу можно придать форму ромбоэдра — фигуры, которую легче всего представить себе как куб, несколько сжатый вдоль пространственной диагонали AB (рис. 4.1). Грани его имеют вид ромбов с углами 78 и 102°. В двух противоположных вершинах A и B сходятся стороны трех тупых углов.

Для демонстрации двойного лучепреломления узкий параллельный пучок света направляют перпендикулярно грани естественного ромбоэдра. Из противоположной грани выходят два пучка, имеющие направления, параллельные первоначальному. Один из них представляет продолжение первичного, а второй смещен в сторону, т. е. для него угол преломления отличен от нуля, несмотря на то что угол падения равен нулю. Это обстоятельство дало повод назвать второй пучок *необыкновенным* (*е*), а первый, подчиняющийся закону преломления, — *обыкновенным* (*о*). Если падающий пучок достаточно узок, а кристалл имеет достаточную толщину, то выходящие пучки пространственно разделены (рис. 4.2) и образуют два пятна на экране. Когда падающий свет естественный, оба пятна имеют одинаковую освещенность. При повороте кристалла вокруг направления падающего пучка одно пятно остается на месте, а второе обходит вокруг него. С помощью анализатора легко убедиться, что выходящие из кристалла пучки света *линейно поляризованы* во взаимно перпендикулярных направлениях. Увеличив сечение падающего пучка, можно получить частичное перекрывание пятен на экране. При повороте анализатора неперекрывающие-



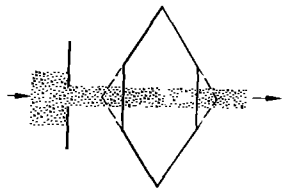
Кристалл исландского шпата



4.2
Двойное лучепреломление света

ся части пятен периодически становятся светлыми и темными: когда у одной части освещенность максимальна, у другой — обращается в нуль. Освещенность перекрывающейся части остается неизменной.

В кристалле исландского шпата пространственная диагональ AB ромбоэдра (см. рис. 4.1) представляет собой ось симметрии: при повороте на 120° вокруг этой оси кристалл совмещается сам с собой. Спилив тупые углы в вершинах ромбоэдра по плоскостям, перпендикулярным оси, и отшлифовав эти плоскости, можно исследовать распространение света в направлении оси (рис. 4.3). Оказывается, что в этом случае двойное лучепреломление отсутствует: падающий пучок света не раздваивается, и состояние его поляризации не изменяется. Обладающее таким свойством направление называется *оптической осью* кристалла. Отметим, что речь идет именно о *направлении*, а не об отдельной прямой, так как пучок не будет испытывать двойного преломления при распространении вдоль любой прямой, параллельной этому направлению.



4.3
Распространение света вдоль оптической оси

Кристаллы, имеющие лишь одно направление, вдоль которого не происходит двойное лучепреломление, называют *одноосными* (исландский шпат, кварц и др.). Существуют кристаллы, имеющие два таких направления. Они называются *двуосными*. В дальнейшем основное внимание будет уделено одноосным кристаллам, широко используемым в оптических экспериментах.

Рассмотрим распространение света в одноосном кристалле в направлении, перпендикулярном оптической оси. Возьмем плоскопараллельную кристаллическую пластинку, грани которой вырезаны вдоль оптической оси. Опыт показывает, что падающий нормально пучок света распространяется в пластинке в прежнем направлении. Однако состояние поляризации света при прохождении через пластинку изменяется: если падающий свет поляризован линейно, прошедший свет в общем случае поляризован эллиптически. Изменение поляризации легко понять, если падающую волну разложить на две составляющих, в одной из которых вектор **E** параллелен оптической оси, в другой — перпендикулярен. Оказывается, что эти составляющие распространяются с разными скоростями и при прохождении через пластинку между ними возникает сдвиг по фазе.

Различие скоростей этих волн можно объяснить на основе электронной теории дисперсии. Одинаковым смещениям оптического электрона атома вдоль оптической оси и в перпендикулярном на-

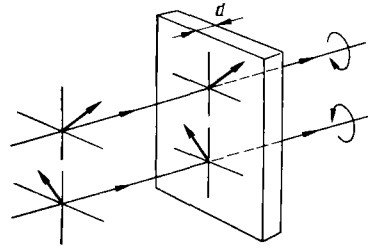
правлении соответствуют разные квазиупругие возвращающие силы. Поэтому будут различаться и собственные частоты $\omega_{\parallel}$ и $\omega_{\perp}$ колебаний электрона в этих взаимно перпендикулярных направлениях. Поляризуемость атома определяется разностью квадратов частоты ω света и частоты собственных колебаний электрона. Следовательно, различным направлениям колебаний в световой волне определенной частоты соответствуют различные значения поляризуемости, диэлектрической проницаемости и показателя преломления.

Обозначим n_o показатель преломления для волны, в которой вектор **E** перпендикулярен оптической оси, n_e — для волны, в которой вектор **E** лежит в плоскости, содержащей оптическую ось и направление распространения волны. Этим волнам с ортогональными направлениями поляризации соответствуют различные фазовые скорости $v_o = c/n_o$ и $v_e = c/n_e$. Пусть на пластинку падает линейно поляризованная волна. Тогда на входе обе волны имеют одинаковую фазу. Разность их фаз δ на выходе из пластинки зависит от ее толщины d:

$$\delta = (\omega/c)(n_o - n_e)\, d. \qquad (4.1)$$

Если $\delta = k\pi$, где $k = \pm 1, \pm 2, \ldots$, то выходящая волна тоже имеет линейную поляризацию. В остальных случаях поляризация волны эллиптическая (если только направление поляризации падающей волны не совпадает с оптической осью и не перпендикулярно оси, когда в кристалле возбуждается лишь одна из волн и выходящий свет остается линейно поляризованным).

Чтобы с помощью кристаллической пластинки можно было получить свет круговой поляризации, разность фаз δ должна быть равна $(2k+1)\pi/2$. При заданных значениях n_o и n_e это условие удовлетворяется подбором толщины d пластинки в (4.1). В результате получается так называемая *пластинка в четверть длины волны* (пластинка $\lambda/4$). Разумеется, необходимая разность фаз получится только для света определенной частоты. Это обусловлено как прямой зависимостью δ от ω в (4.1), так и дисперсией показателей преломления n_o и n_e. Свет круговой поляризации получается при условии, что плоскость поляризации падающей волны составляет угол $\pm\pi/4$ с оптической осью пластинки (рис. 4.4), так как только в этом случае равны амплитуды волн, между которыми пластинка $\lambda/4$ вносит сдвиг фаз $\pi/2$.



4.4
Получение света круговой поляризации

Пластинки $\lambda/4$ изготовляют обычно из кварца или из тонких листочков слюды. Качество пластинки $\lambda/4$ можно проверить, исследуя с помощью анализатора выходящий из нее циркулярно поляризованный

свет. При правильно подобранной толщине пластинки вращение анализатора не должно приводить к изменению интенсивности проходящего света.

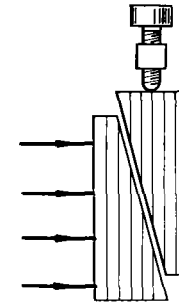
Пластинка $\lambda/4$ может быть использована и для обратного превращения циркулярно поляризованного света в линейно поляризованный. Ясно, что направление поляризации выходящего света при этом составляет угол $\pm\pi/4$ с оптической осью пластинки.

С помощью четвертьволновой пластинки можно также отличить на опыте свет круговой поляризации от естественного, а эллиптический — от частично поляризованного. Одного только поляризационного прибора (анализатора) недостаточно, чтобы различить эти типы поляризации. Как для поляризованного по кругу, так и для света естественного, интенсивность после прохождения через анализатор одинакова при любой его ориентации. Если же предварительно ввести пластинку $\lambda/4$, то поляризованный по кругу свет превратится в линейно поляризованный, который можно полностью погасить при определенной ориентации анализатора. Естественный свет можно рассматривать как наложение двух волн одинаковой интенсивности с ортогональными поляризациями, разность фаз между которыми изменяется в течение времени наблюдения случайно. Внесение четвертьволновой пластинкой дополнительной постоянной разности фаз между ними не может изменить случайного характера соотношения фаз ортогональных составляющих. Поэтому прошедший через четвертьволновую пластинку свет остается неполяризованным и его интенсивность не меняется при повороте анализатора.

Эллиптически поляризованный свет можно представить как сумму двух волн, линейно поляризованных вдоль главных осей эллипса, разность фаз между которыми $\pm\pi/2$. Пропуская исследуемый свет через пластинку $\lambda/4$, к этой разности можно добавить еще $\pm\pi/2$ и тем самым сделать ее равной 0 или π, т. е. превратить эллиптическую поляризацию в линейную, в чем можно убедиться с помощью анализатора. Для этой цели пластинка $\lambda/4$ должна быть ориентирована так, чтобы ее главные направления (т. е. направление оптической оси и перпендикулярное ей) совпадали с главными осями эллипса колебаний, определенными предварительно с помощью анализатора (напомним, что для превращения света круговой поляризации в линейную пластинка может быть ориентирована как угодно). Таким образом, по направлению оптической оси пластинки определяют ориентацию осей эллипса колебаний, а по положению анализатора, при котором гасится выходящий из пластинки пучок, — отношение этих осей.

Описанным выше методом можно легко отличить эллиптически поляризованный свет от частично поляризованного, который можно рассматривать как смесь линейно поляризованного света с естественным. И в том, и в другом случае при повороте анализатора наблюдается лишь изменение интенсивности света между некоторыми максимальным и минимальным значениями. Если же предварительно ввести в пучок пластинку $\lambda/4$, соответственным образом ее

ориентировав, то эллиптически поляризованный свет превратится в линейно поляризованный и может быть полностью погашен анализатором. В случае, частично поляризованного света пластинка $\lambda/4$ при указанной ориентации никакого влияния не оказывает, т. е. выходящий из нее свет не будет погашен анализатором ни при какой его ориентации.



4.5
Компенсатор

Для анализа поляризованного света наряду с пластинкой $\lambda/4$ используются приспособления, которые позволяют скомпенсировать до нуля (или дополнить до π) любую разность фаз между двумя волнами. Они называются *компенсаторами*. Простейший компенсатор состоит из двух слабо скошенных кварцевых клиньев (рис. 4.5). Сложенные вместе, они образуют плоскопараллельную кристаллическую пластинку с оптической осью, ориентированной вдоль ее граней. Один из клиньев можно перемещать относительно другого с помощью микрометрического винта, изменяя тем самым их общую толщину и, следовательно, вносимую компенсатором разность фаз между двумя волнами.

Контрольные вопросы

- ☐ Что называется оптической осью анизотропной среды?
- ☐ Какую поляризацию может иметь свет, распространяющийся вдоль оптической оси?
- ☐ Циркулярно поляризованный свет при прохождении через пластинку $\lambda/4$ превращается в линейно поляризованный. Каким будет направление его поляризации?
- ☐ Как с помощью пластинки $\lambda/4$ и анализатора можно отличить естественный свет от циркулярно поляризованного? Частично поляризованный от эллиптически поляризованного?
- ☐ Как отличить свет с частичной круговой поляризацией (т. е. смесь естественного с циркулярно поляризованным) от естественного?

Задачи

1. Если естественный кристалл исландского шпата положить на страницу книги и рассматривать оказавшийся под ним текст, то все буквы кажутся раздвоившимися, причем одно изображение параллельно смещено относительно другого. Маленькие предметы (например, точки) дают два отдельных изображения. Будет ли происходить раздваивание изображения, если через такой кристалл смотреть на удаленный предмет?

◇ Ответ: не будет.

2. Узкий пучок естественного света падает нормально на грань кристалла исландского шпата и преломляется, как показано на рис. 4.2. Что произойдет, если на пути выходящих из кристалла пучков поместить еще один точно такой же кристалл, ориентированный так же, как и первый? Что изменится, если второй кристалл повернуть на 90° вокруг оси падающего пучка?

3. Какую толщину может иметь четвертьволновая пластинка из кварца для желтого света натрия ($n_e = 1{,}553$, $n_o = 1{,}544$)?

◇ Ответ: $d_{\min} = \lambda/4(n_e - n_o) = 16$ мкм. ◇

4.2. Плоские монохроматические волны в анизотропной среде. Одноосные кристаллы

Рассмотрим распространение света в прозрачной анизотропной среде на основе электромагнитной теории. Фундаментальные уравнения Максвелла (2.6)—(2.9) для электромагнитного поля в веществе имеют универсальный характер и в полной мере применимы к анизотропным средам. Будем искать их решение в виде плоских монохроматических волн, где **E**, **D**, **B** зависят от координат и времени по закону $\exp i(\mathbf{kr}-\omega t)$. Введем единичный вектор волновой нормали **N**, направленный вдоль волнового вектора **k** (т. е. перпендикулярно плоскостям равных фаз):

$$\mathbf{N}=\mathbf{k}/k=(v/\omega)\mathbf{k}. \tag{4.2}$$

Тогда из уравнений Максвелла (2.7) и (2.9) получим

$$\mathbf{N}\times\mathbf{E}=v\mathbf{B},\quad \varepsilon_0 c^2\mathbf{N}\times\mathbf{B}=-v\mathbf{D}. \tag{4.3}$$

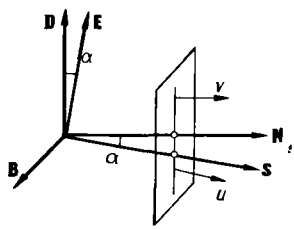
Здесь v — фазовая скорость волны, т. е. скорость, с которой поверхность равных фаз перемещается в направлении волновой нормали **N**. Прежде чем вводить материальное уравнение, связывающее векторы **E** и **D** в анизотропной среде, рассмотрим те свойства электромагнитных волн, которые следуют непосредственно из уравнений (4.3). Эти свойства отражают взаимное расположение векторов **D**, **E**, **B** и **N**.

Из второй формулы (4.3) видно, что вектор **D** перпендикулярен векторам **B** и **N**, а из первой — что векторы **B** и **N** взаимно перпендикулярны. Таким образом, в бегущей волне векторы **D**, **B** и **N** образуют правую тройку ортогональных векторов (рис. 4.6). Что касается вектора **E**, то в анизотропной среде его направление, вообще говоря, не совпадает с направлением **D**. Из первого уравнения (4.3) следует, что вектор **E** ортогонален вектору **B**, т. е. лежит в плоскости, образуемой векторами **D** и **N**. Это значит, что бегущие волны в анизотропной среде поперечны в отношении векторов **D** и **B**, но в общем случае они не поперечны в отношении вектора **E** (рис. 4.6).

Направление переноса энергии в электромагнитной волне определяется вектором Пойнтинга **S** (1.50). Для характеристики этого направления введем ориентированный вдоль **S** единичный вектор **s**:

$$\mathbf{s}=\mathbf{S}/S=(\mathbf{E}\times\mathbf{B})/(EB). \tag{4.4}$$

Его называют *лучевым вектором*, так как направление переноса энергии — это и есть направление лучей. В изотропной среде лучи параллельны волновой нормали. Однако в анизотропной среде в общем случае это не так. Из рис. 4.6 видно, что вектор **s**, ортогональный векторам **E** и **B**, лежит в одной плоскости с **D**, **E** и **N** и составляет с вектором **N** такой же угол α, что и вектор **E** с **D**.



4.6
Взаимное расположение векторов **B**, **D**, **E**, **N** и **s**

Плоскость равных фаз перемещается вдоль вектора $\mathbf{N}$ со скоростью v. Скорость перемещения этой плоскости вдоль вектора луча $\mathbf{s}$ называется *лучевой скоростью* u. Когда $\mathbf{N}$ и $\mathbf{s}$ не совпадают, лучевая и фазовая скорости не равны и, как видно из рис. 4.6, связаны соотношением

$$v = u\cos\alpha = u(\mathbf{Ns}). \tag{4.5}$$

Особенности распространения лучей (т. е. переноса энергии) в анизотропной среде обусловлены как дисперсией волн (т. е. зависимостью фазовой скорости от частоты), так и отличием направлений волновых нормалей $\mathbf{N}$ и лучей $\mathbf{s}$. Дисперсия в равной мере присуща как изотропным, так и анизотропным средам. Чтобы выделить особенности, специфичные только для анизотропной среды, будем в дальнейшем пренебрегать дисперсией, т. е. полагать $dv/d\lambda = 0$. В такой недиспергирующей среде вектор лучевой скорости $\mathbf{u} = u\mathbf{s}$ характеризует направление и скорость переноса энергии световой волны. Поэтому задача определения лучевой скорости в зависимости от направления луча представляет наибольший интерес и на ее решении будет сосредоточено основное внимание.

Преобразуем уравнения (4.3) так, чтобы вместо вектора волновой нормали $\mathbf{N}$ в них фигурировал лучевой вектор $\mathbf{s}$, а вместо фазовой скорости v — лучевая скорость u. Для этого умножим векторно левую и правую части каждого из них на $\mathbf{s}$. Двойные векторные произведения в левых частях преобразуются следующим образом:

$$\mathbf{s}\times(\mathbf{N}\times\mathbf{E}) = \mathbf{N}(\mathbf{sE}) - \mathbf{E}(\mathbf{sN}) = -(v/u)\mathbf{E},$$

$$\mathbf{s}\times(\mathbf{N}\times\mathbf{B}) = \mathbf{N}(\mathbf{sB}) - \mathbf{B}(\mathbf{sN}) = -(v/u)\mathbf{B}.$$

Здесь мы воспользовались тем, что вектор $\mathbf{s}$ в соответствии с (4.4) ортогонален векторам $\mathbf{E}$ и $\mathbf{B}$, а произведение $\mathbf{sN}$ заменили на v/u с помощью формулы (4.5). В результате вместо (4.3) получаем

$$\mathbf{E} = -u(\mathbf{s}\times\mathbf{B}),\quad \varepsilon_0 c^2\mathbf{B} = u(\mathbf{s}\times\mathbf{D}). \tag{4.6}$$

Теперь можно исключить из этих уравнений индукцию $\mathbf{B}$ магнитного поля:

$$(c/u)^2\varepsilon_0\mathbf{E} + \mathbf{s}\times(\mathbf{s}\times\mathbf{D}) = 0.$$

Раскрывая двойное векторное произведение и учитывая, что $\mathbf{s}^2 = 1$, получаем уравнение, которому должны удовлетворять $\mathbf{E}$ и $\mathbf{D}$ в бегущей плоской волне:

$$(c/u)^2\varepsilon_0\mathbf{E} + \mathbf{s}(\mathbf{sD}) - \mathbf{D} = 0. \tag{4.7}$$

Соотношения (4.6) и (4.7) получены на основе уравнений Максвелла без каких бы то ни было предположений о свойствах среды. Чтобы продвинуться дальше, необходимы *материальные уравнения*, связывающие $\mathbf{E}$ и $\mathbf{D}$ в рассматриваемой среде. В анизотропной среде поляризованность $\mathbf{P}$, как уже отмечалось, в общем случае не совпа-

дает по направлению с создающим ее электрическим полем **E**. Поэтому не совпадают и направления **E** и **D**, т. е. в материальном уравнении, связывающем **E** и **D** в монохроматической волне, диэлектрическая проницаемость $\varepsilon(\omega)$ представляет собой тензор второго ранга:

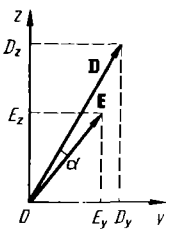
$$D_i = \varepsilon_0 \sum_k \varepsilon_{ik}(\omega) E_k \quad (i,\ k = x,\ y,\ z). \tag{4.8}$$

Компоненты тензора диэлектрической проницаемости для той или иной модели среды могут быть рассчитаны на основе электронной теории дисперсии. В рамках феноменологической теории (которая положена в основу дальнейшего рассмотрения) их можно считать параметрами, определяемыми на опыте.

В различных системах координат x, y, z компоненты тензора ε_{ik} имеют разные значения. Специальным выбором системы координат соотношение (4.8) можно упростить, приведя тензор ε_{ik} к диагональному виду:

$$D_x = \varepsilon_0 \varepsilon_x E_x,\ D_y = \varepsilon_0 \varepsilon_y E_y,\ D_z = \varepsilon_0 \varepsilon_z E_z. \tag{4.9}$$

Если вектор **E** направлен вдоль одной из этих осей, то вектор **D** совпадает с ним по направлению. Соответствующие оси координат x, y, z называются *главными осями* тензора, а величины ε_x, ε_y, ε_z — его главными значениями или *главными диэлектрическими проницаемостями*. Различие главных значений и отражает несовпадение направлений векторов **E** и **D** (рис. 4.7). Если два главных значения диэлектрического тензора ε_{ik} совпадают ($\varepsilon_x = \varepsilon_y$), то среда оптически одноосная. Ее оптические свойства полностью определяются двумя параметрами $\varepsilon_\perp \equiv \varepsilon_x = \varepsilon_y$ и $\varepsilon_\parallel = \varepsilon_z$, называемыми поперечной и продольной диэлектрическими проницаемостями. Когда вектор **E** лежит в плоскости xy, т. е. перпендикулярен оси z (направление которой параллельно оптической оси), вектор **D** совпадает с ним по направлению. Это значит, что в отношении оптических (и электрических) свойств одноосная среда обладает полной симметрией вращения относительно направления оптической оси, хотя в отношении других свойств (например, механических) симметрия может быть более низкой.

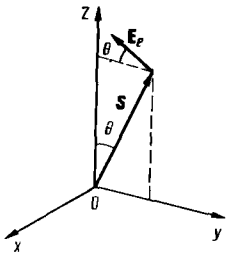


4.7 Векторы **D** и **I** в анизотропной среде

К оптически одноосным средам относятся все кристаллы тетрагональной, гексагональной и тригональной (ромбоэдрической) систем; оптическая ось совпадает здесь с осью симметрии соответственно четвертого, шестого или третьего порядка. Изотропное твердое тело (например, стекло), подверженное однородной деформации растяжения или сжатия в одном направлении, или жидкость из анизотропных молекул, помещенная в однородное электрическое поле, также будут оптически одноосными.

В кристаллах более низкой симметрии (триклинная, моноклинная, ромбическая системы) все

три главных значения тензора ε_{ik} различны. Можно показать, что в этом случае существует два направления (оптические оси), вдоль которых обе волны с ортогональными поляризациями распространяются с одной скоростью.



1 к
К выбору направления осей координат

В кристаллах кубической системы (таких, как каменная соль NaCl, флюорит CaF_2, алмаз C и т. д.) все три главных направления диэлектрического тензора физически эквивалентны, поэтому главные значения ε_x, ε_y и ε_z одинаковы. Это значит, что тензор ε_{ik} вырождается в скаляр (векторы **E** и **D** всегда совпадают по направлению) и кристаллы кубической системы в отношении оптических свойств ведут себя как изотропная среда. В отношении других свойств, выражаемых тензорами более высокого ранга (например, упругих), кубические кристаллы анизотропны. Оптическая анизотропия кубических кристаллов появляется только при учете очень слабых эффектов пространственной дисперсии, описываемых тензором четвертого ранга (см. § 2.9).

Перейдем к исследованию распространения света в оптически одноосных кристаллах. Если свет распространяется вдоль оптической оси, то при любой его поляризации векторы **E** и **D** лежат в плоскости xy и, как и в изотропной среде, совпадают по направлению, причем $\mathbf{D}=\varepsilon_0\varepsilon_\perp\mathbf{E}_\perp$. Поэтому скорость волн, распространяющихся вдоль оси, равна $c/\sqrt{\varepsilon_\perp}$, а поляризация может быть любой (линейной, круговой, эллиптической). Ниже будет показано, что в любом другом направлении могут распространяться только линейно поляризованные волны с ортогональными направлениями поляризации, причем скорости этих двух волн различны.

Из-за симметрии выбор направления осей координат в плоскости xy произволен. Воспользуемся этим для упрощения уравнений. Пусть направление луча **s** (рис. 4.8) составляет некоторый угол θ с оптической осью (осью z). Выберем ось y так, чтобы она лежала в плоскости, образуемой оптической осью и лучом (ее называют *плоскостью главного сечения*): Тогда вектор **s** имеет следующие проекции: $\mathbf{s}(0, \sin\theta, \cos\theta)$. Скалярное произведение **sD** в (4.7) имеет вид $\sin\theta D_y+\cos\theta D_z$. Выразим проекции **D** в (4.7) с помощью материальных уравнений (4.9), учитывая, что $\varepsilon_x=\varepsilon_y=\varepsilon_\perp$, $\varepsilon_z=\varepsilon_\parallel$, и запишем (4.7) в проекциях на оси координат x, y, z:

$$\begin{aligned}
&[(c/u)^2-\varepsilon_\perp]E_x=0,\\
&[(c/u)^2-\varepsilon_\perp\cos^2\theta]E_y+\varepsilon_\parallel\sin\theta\cos\theta\ \ E_z=0,\\
&\varepsilon_\perp\sin\theta\cos\theta\ \ E_y+[(c/u)^2-\varepsilon_\parallel\sin^2\theta]\ \ E_z=0.
\end{aligned}\qquad(4.10)$$

Мы получили систему однородных уравнений для нахождения проекций вектора **E** плоской волны. Система имеет ненулевое реше-

ние только тогда, когда ее определитель равен нулю. Это условие и дает уравнение для нахождения лучевой скорости $u(0)$ при данном направлении луча. Определитель распадается на произведение двух множителей, один из которых равен $(c/u)^2-\varepsilon_\perp$. Отсюда сразу находим первый корень уравнения: $u=c/\sqrt{\varepsilon_\perp}=c/n_o$. Подставив его в коэффициенты системы (4.10), получим, что у соответствующего этому корню решения проекции напряженности поля E_y и E_z тождественно равны нулю, а E_x может иметь любое значение. Это значит, что описываемая этим решением волна линейно поляризована вдоль оси x, т. е. перпендикулярно оптической оси (и плоскости главного сечения). Ее лучевая скорость $u=c/n_o$ не зависит от угла 0, т. е. от направления распространения. Такую волну называют *обыкновенной* и относящиеся к ней величины снабжают индексом o.

Приравнивая нулю второй множитель в определителе системы (4.10), находим еще один корень:

$$u(0)=c/(\varepsilon_\perp \cos^2 0+\varepsilon_\parallel \sin^2 0)^{1/2} \qquad (4.11)$$

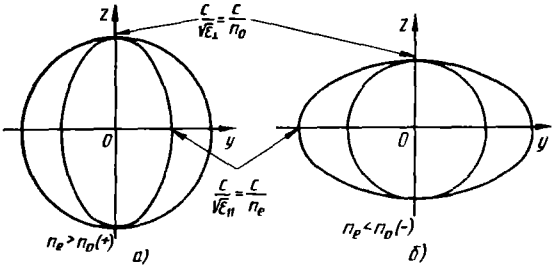
Подставляя его в коэффициенты системы (4.10), находим $E_x=0$, $E_z/E_y=-\operatorname{tg} 0$. Это значит, что распространяющаяся с зависящей от направления (т. е. от угла 0) скоростью $u(0)$ (4.11) волна поляризована в плоскости главного сечения, причем вектор **E** перпендикулярен **s** (рис. 4.8). Эту волну называют *необыкновенной* (индекс e). Наряду с главными диэлектрическими проницаемостями $\varepsilon_\perp$ и $\varepsilon_\parallel$ для характеристики одноосных сред используют также параметры $n_o\equiv\sqrt{\varepsilon_\perp}$ и $n_e\equiv\sqrt{\varepsilon_\parallel}$, называемые соответственно обыкновенным и необыкновенным показателями преломления.

Других решений система (4.10) не имеет, т. е. двумя найденными выше волнами с ортогональными линейными поляризациями, имеющими скорости $u_o=c/\sqrt{\varepsilon_\perp}=c/n_o$ и $u_e(0)$ (4.11), исчерпывается все многообразие нормальных* волн, которые могут распространяться по заданному направлению 0.

Для нахождения хода лучей в одноосных кристаллах обычно выполняют геометрические построения, в которых используют *поверхности лучевых скоростей* (лучевые, или волновые, поверхности). Для построения лучевой поверхности из произвольной точки O во всевозможных направлениях проводят лучи и откладывают на них отрезки, пропорциональные соответствующим значениям лучевой скорости. Множество концов отложенных отрезков образует замкнутую поверхность, которая для обыкновенной волны, очевидно,

---

* Напомним, что нормальными называют волны, состояние поляризации которых остается неизменным по мере распространения.



4.9
Лучевые поверхности обыкновенной и необыкновенной волн в одноосных кристаллах

представляет собой сферу радиусом $u_o=c/\sqrt{\varepsilon_\perp}$, а для необыкновенной волны — эллипсоид вращения с полуосями $c/\sqrt{\varepsilon_\perp}$ и $c/\sqrt{\varepsilon}$. Чтобы убедиться в этом, достаточно переписать соотношение (4.11) в виде

$$\frac{u^2\cos^2\theta}{c^2/\varepsilon_\perp}+\frac{u^2\sin^2\theta}{c^2/\varepsilon_\parallel}=1. \qquad (4.12)$$

Так как $u\cos\theta=u_z$, $u\sin\theta=u_y$, то ясно, что уравнение (4.12) определяет эллипсоид вращения в пространстве скоростей, соприкасающийся со сферой для обыкновенной волны в точках, соответствующих направлению оптической оси. Сечение лучевых поверхностей плоскостью yz, проходящей через оптическую ось, показано на рис. 4.9. При $n_e>n_o$ (кварц) вытянутый эллипсоид целиком лежит внутри сферы (рис. 4.9, *а*). Такие кристаллы называют *положительными.* У *отрицательных* кристаллов $n_e<n_o$ (исландский шпат) и сфера лежит внутри сплющенного эллипсоида (рис. 4.9, *б*).

Из этих рисунков видно, что при распространении вдоль оптической оси обе волны имеют одинаковую скорость $u=c/n_o$, определяемую обыкновенным показателем преломления n_o. Для этого направления любая плоскость, содержащая оптическую ось, будет плоскостью главного сечения, поэтому возможны как любое направление линейной поляризации, так и в равной мере круговая или эллиптическая поляризация. При распространении в перпендикулярном оптической оси направлении обыкновенная волна имеет по-прежнему скорость $u_o=c/n_o$, а необыкновенная, в которой вектор **E** направлен вдоль оптической оси, — скорость $u_e=c/\sqrt{\varepsilon_\parallel}=c/n_e$, определяемую необыкновенным показателем преломления. Для всех других направлений распространения векторы **N** и **s** не совпадают.

В § 4.3 лучевые поверхности будут использованы для исследования двойного лучепреломления на границе оптически одноосной анизотропной среды.

Контрольные вопросы

- Дайте характеристику взаимного расположения векторов **E**, **D**, **B**, **N** и **s** для плоской волны в анизотропной среде.
- Как связаны между собой лучевая и фазовая скорости?
- Какому условию удовлетворяют главные диэлектрические проницаемости оптически одноосного кристалла? Кристаллы каких систем (сингоний) оптически одноосны?
- Как зависит от направления луча лучевая скорость обыкновенной и необыкновенной волн? Как эти волны поляризованы?
  Кристаллы кубической системы оптически изотропны, т. е. скорость света в них по всем направлениям одинакова. Будет ли это утверждение справедливо и для звука, т. е. для механических упругих волн?

Задача

Определить максимальный угол α_{max} между направлениями луча и волновой нормали для кристалла исландского шпата ($n_o = 1{,}658$, $n_e = 1{,}486$).

◇ Угол α между **N** и **s**, как видно из рис. 4.6, равен углу между **D** и **E**. В необыкновенной волне эти векторы лежат в плоскости главного сечения. Пусть β — угол между **D** и оптической осью кристалла, а γ — между **E** и той же осью. Спроецируем векторы **D** и **E** на направление оси и перпендикулярное ей направление и воспользуемся материальными уравнениями $D_\perp = \varepsilon_0 \varepsilon_\perp E_\perp$, $D_\parallel = \varepsilon_0 \varepsilon_\parallel E_\parallel$:

$$D\sin\beta = \varepsilon_0 \varepsilon_\perp E \sin\gamma, \quad D\cos\beta = \varepsilon_0 \varepsilon_\parallel E \cos\gamma.$$

Разделив почленно эти уравнения, находим $\operatorname{tg}\beta = (\varepsilon_\perp/\varepsilon_\parallel)\operatorname{tg}\gamma = (n_o/n_e)^2 \operatorname{tg}\gamma$. Интересующий нас угол $\alpha = \beta - \gamma$ (см. рис. 4.7), поэтому

$$\operatorname{tg}\alpha = \operatorname{tg}(\beta - \gamma) = (n_o^2 - n_e^2)\operatorname{tg}\gamma / (n_e^2 + n_0^2 \operatorname{tg}^2\gamma).$$

Максимум α достигается при $\operatorname{tg}\gamma = n_e/n_o = 0{,}896$, т. е. при $\gamma = 41°52'$, и равен

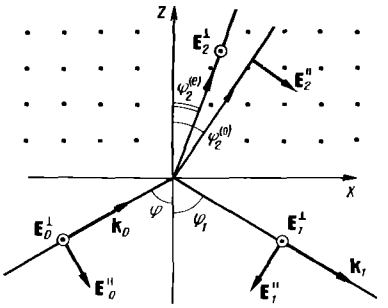
$$\alpha_{max} = \operatorname{arctg}[(n_o^2 - n_e^2)/(2n_0 n_e)] = 6°16'.$$

4.3. Преломление на границе анизотропной среды. Построение Гюйгенса

Полное количественное решение задачи о преломлении и отражении света на границе анизотропной среды может быть получено на основе электромагнитной теории. Как и в случае границы изотропных сред, электромагнитное поле должно удовлетворять тем же *граничным условиям:* тангенциальные составляющие векторов **E** и **B** по обе стороны границы должны совпадать в каждой ее точке в любой момент времени. Метод решения задачи остается прежним: в первой среде наряду с заданной падающей монохроматической плоской волной рассматривается еще одна — отраженная, а во второй среде — преломленная. Их волновые поверхности предполагаются бесконечными плоскостями (хотя на опыте, как правило, приходится иметь дело с узкоограниченными пучками). Отличие состоит в том, что для волн в анизотропной среде нужно учитывать зависимость фазовой скорости от направлений волновой нормали и поляризации. Эта зависимость может быть найдена с помощью уравнений Максвелла и материальных уравнений примерно так,

как это было сделано в § 4.2 для лучевой скорости в одноосных кристаллах.

Электромагнитная теория позволяет найти как направления отраженной и преломленных волн, так и их амплитуды. При этом не требуются никакие дополнительные допущения (как, например, принцип Гюйгенса — построение огибающей для вторичных волн). Однако в общем случае окончательные формулы оказываются чрезвычайно громоздкими. Поэтому ограничимся лишь иллюстрацией применения электромагнитной теории на наиболее простом примере.



Преломление и отражение плоской волны на границе одноосной среды

Пусть плоская волна падает из вакуума (или воздуха) на границу оптически одноосной анизотропной однородной среды, занимающей верхнее полупространство (рис. 4.10). Рассмотрим частный случай: оптическая ось параллельна границе xy и перпендикулярна плоскости падения xz (т. е. параллельна оси y). Падающую волну разложим на составляющие, поляризованные в плоскости падения и в перпендикулярном направлении. Граничные условия, как и для изотропной среды, выражаются уравнениями (3.1). Чтобы эти условия выполнялись сразу во всех точках границы, у всех трех экспонент зависимость от координат x и y должна быть одинакова. Отсюда, во-первых, следует, что у волновых векторов $\mathbf{k}_1$ и $\mathbf{k}_2$ отраженной и преломленной волн равны нулю y-составляющие, т. е. нормали к волновым поверхностям отраженной и преломленной волн лежат в плоскости падения. Во-вторых, из равенства x-составляющих векторов $\mathbf{k}_0$, $\mathbf{k}_1$ и $\mathbf{k}_2$ следуют геометрические законы отражения и преломления, определяющие направления этих волн. Так как $k_{0x}=(\omega/c)\sin\varphi$, $k_{1x}=(\omega/c)\sin\varphi_1$, то $\varphi_1=\varphi$: угол отражения φ_1 от анизотропной среды равен углу падения φ.

Что касается угла преломления φ_2, то теперь он будет иметь разные значения для двух ортогональных поляризаций. В самом деле, при рассматриваемом направлении оптической оси (вдоль оси y) в волне, поляризованной в плоскости падения, вектор $\mathbf{E}$ направлен перпендикулярно оптической оси. Скорость такой (обыкновенной) волны v_o равна $c/\sqrt{\varepsilon_\perp}=c/n_o$, а модуль ее волнового вектора $k_2^{\parallel}=\omega/v_o=(\omega/c)n_o$. Из $k_{2x}^{\parallel}=k_{ox}$, учитывая, что $k_{2x}^{\parallel}=k_2^{\parallel}\sin\varphi_2^{(o)}$, находим $n_o\sin\varphi_2^{(o)}=\sin\varphi$. Если же волна линейно поляризована в направлении, перпендикулярном плоскости падения, то для рассматриваемого расположения вектор $\mathbf{E}$ в ней направлен вдоль оптической оси. Скорость этой (необыкновенной) волны v_e равна $c/\sqrt{\varepsilon_\parallel}=c/n_e$, а модуль волнового вектора $k_2^{\perp}=\omega/v_e=(\omega/c)n_e$. Поэтому для нее угол преломления $\varphi_2^{(e)}$ определяется из условия

$n_e \sin\varphi_2^{(e)} = \sin\varphi$. Так как здесь скорость необыкновенной волны одна и та же для любого ее направления в плоскости xz, то отношение синусов угла падения и угла преломления постоянно и равно n_e.

Полученные выше выражения для $\varphi_2^{(o)}$ и $\varphi_2^{(e)}$ определяют направления волновых нормалей преломленных волн. Для сравнения с опытом важно знать ход лучей, представляющих пути распространения световой энергии. Однако при выбранном расположении преломленные волны идут перпендикулярно оптической оси, когда лучи и волновые нормали совпадают. Таким образом, в этом частном случае падающий из вакуума под углом φ луч создает два преломленных луча (обыкновенный и необыкновенный), углы преломления которых $\varphi_2^{(o)}$ и $\varphi_2^{(e)}$ (рис. 4.10) даются соотношениями

$$n_o \sin\varphi_2^{(o)} = \sin\varphi, \quad n_e \sin\varphi_2^{(e)} = \sin\varphi. \tag{4.13}$$

Граничные условия позволяют найти не только направления отраженной и преломленных волн, но и их амплитуды. Действуя так же, как и в § 3.2, мы в рассматриваемом случае придем к таким же формулам Френеля (3.8)—(3.11) с той разницей, что для поляризации, перпендикулярной плоскости падения, выражения для амплитуд отраженной и преломленной волн содержат n_e в качестве показателя преломления второй среды и угол преломления $\varphi_2^{(e)}$ из (4.13) вместо φ_2:

$$\frac{E_1^\perp}{E_0^\perp} = \frac{\cos\varphi - n_e\cos\varphi_2^{(e)}}{\cos\varphi + n_e\cos\varphi_2^{(e)}} = -\frac{\sin(\varphi-\varphi_2^{(e)})}{\sin(\varphi+\varphi_2^{(e)})},$$

$$\frac{E_2^\perp}{E_0^\perp} = \frac{2\cos\varphi}{\cos\varphi + n_e\cos\varphi_2^{(e)}} = \frac{2\cos\varphi\sin\varphi_2^{(e)}}{\sin(\varphi+\varphi_2^{(e)})}.$$

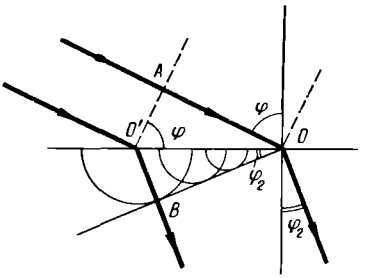
Составляющая падающей волны, поляризованная в плоскости падения, преломляется на иной угол $\varphi_2^{(o)}$, и для нее амплитуды отраженной и преломленной волн выражаются соответствующими формулами Френеля (3.9), (3.11) с заменой n_2 на n_o и φ_2 на $\varphi_2^{(o)}$:

$$\frac{E_1^\parallel}{E_0^\parallel} = \frac{n_o\cos\varphi - \cos\varphi_2^{(o)}}{n_o\cos\varphi + \cos\varphi_2^{(o)}} = \frac{\operatorname{tg}(\varphi-\varphi_2^{(o)})}{\operatorname{tg}(\varphi+\varphi_2^{(o)})},$$

$$\frac{E_2^\parallel}{E_0^\parallel} = \frac{2\cos\varphi}{n_o\cos\varphi + \cos\varphi_2^{(o)}} = \frac{2\cos\varphi\,\sin\varphi_2^{(o)}}{\sin(\varphi+\varphi_2^{(o)})\cos(\varphi-\varphi_2^{(o)})}.$$

Мы видим, что в данном случае электромагнитная теория дает исчерпывающее описание отражения и преломления света на границе анизотропной среды. При ином расположении оптической оси относительно границы принципиальные затруднения не возникают, но вычисления оказываются громоздкими. В таких случаях возможно получить частичное решение задачи — определить направления преломленных волн в одноосном кристалле — с помощью изящного геометрического построения, впервые примененного Гюйгенсом для объяснения двойного лучепреломления в исландском шпате.

Напомним, как выполняется построение Гюйгенса в случае изотропной среды (рис. 4.11). Когда волновая поверхность падающей из вакуума плоской волны достигает точки O на границе изотропной среды, вторичные волны из всех прежних точек O', распространяющиеся со свойственной им скоростью, имеют общую огибающую OB, которая и представляет собой поверхность равных фаз преломленной волны. Учитывая, что вторичная волна, распространяющаяся со скоростью $v=c/n$, проходит расстояние $O'B$ за то время, что падающая волна расстояние AO, из треугольников $OO'A$ и $OO'B$ получаем закон преломления: $\sin\varphi = n\sin\varphi_2$.



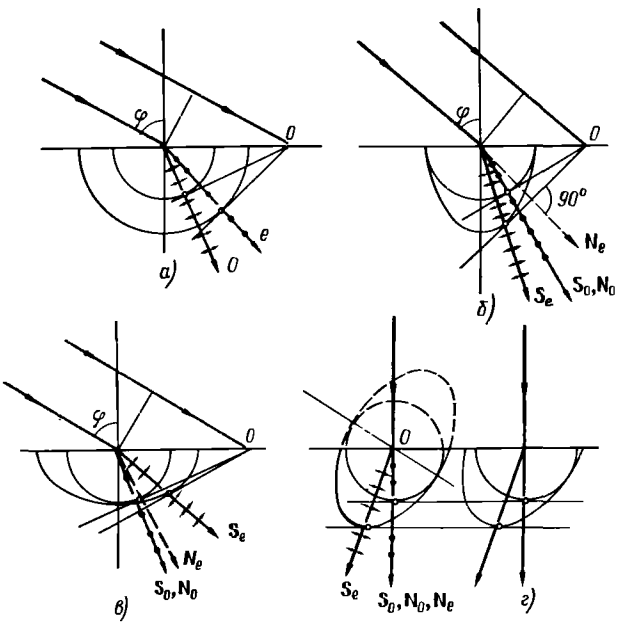
4.11
Построение преломленной волны по принципу Гюйгенса (для изотропной среды)

При обобщении построений Гюйгенса на случай анизотропной одноосной среды для вторичных волн нужно использовать найденные в § 4.2 поверхности лучевых скоростей. Касательная к ним плоскость дает положение фронта (т. е. поверхности равных фаз) преломленной волны, а прямая, проведенная из центра вторичной волны в точку касания, — направление преломленного луча. Так как лучевая поверхность состоит из сферы и эллипсоида, то построение Гюйгенса дает два луча: обыкновенный, направление которого совпадает с нормалью к фронту, как и в изотропной среде, и необыкновенный, направление которого в общем случае отклоняется от нормали к фронту необыкновенной волны. Для строгого обоснования построений Гюйгенса (которое здесь не приводится) требуется показать, что распространение света от точечного источника по некоторому направлению в анизотропной среде происходит так же, как и рассмотренных в § 4.2 плоских волн, скорости которых по разным направлениям характеризуются лучевыми поверхностями.

Для иллюстрации построений Гюйгенса рассмотрим несколько сравнительно простых частных случаев.

А. *Оптическая ось параллельна границе. Плоскость падения перпендикулярна оптической оси* (рис. 4.12, *а*). Сечения лучевых поверхностей обыкновенной и необыкновенной волн представляют собой окружности. Поэтому направления лучей и волновых нормалей совпадают как у обыкновенной, так и у необыкновенной волн. Вектор **E** в обыкновенной волне ориентирован перпендикулярно оптической оси, в необыкновенной — параллельно оси. При $n_o > n_e$ (отрицательный кристалл) обыкновенный луч преломляется сильнее, чем необыкновенный: $\sin\varphi_2^{(o)} = \sin\varphi/n_o$, $\sin\varphi_2^{(e)} = \sin\varphi/n_e$. Этот случай был рассмотрен выше на основе электромагнитной теории.

Б. *Оптическая ось параллельна границе. Плоскость падения проходит через оптическую ось* (рис. 4.12, *б*). Сечения лучевых поверх-



4.12

Построение обыкновенного и необыкновенного лучей при разном расположении оптической оси

Точками и стрелками отмечены направления колебаний вектора **E**

ностей плоскостью падения — окружность и эллипс. Направление преломленного луча задается прямой, проведенной из центра лучевой поверхности в точку ее касания с огибающей (т. е. с фронтом волны). Оба преломленных луча лежат в плоскости падения. При $n_o > n_e$ необыкновенный луч преломляется сильнее, чем обыкновенный, хотя направление волновой нормали для него (в данном случае не совпадающее с лучом) изменяется при преломлении меньше, чем для обыкновенного луча. Если свет падает по нормали ($\varphi=0$), то обе волны распространяются в прежнем направлении, но с разными скоростями. Таким образом, при $\varphi=0$ случаи А и Б соответствуют рассмотренной выше кристаллической пластинке (в частности, четвертьволновой).

В. *Оптическая ось перпендикулярна границе* (рис. 4.12, *в*). *Оба преломленных луча лежат в плоскости падения.* При $n_o > n_e$ обыкновенный луч преломляется сильнее, чем необыкновенный. Если для обыкновенного луча $\sin\varphi/\sin\varphi_2^{(o)} = n_o = \text{const}$, то для необыкновенного $\sin\varphi/\sin\varphi_2^{(e)}$ зависит от угла падения. При нормальном падении обе волны распространяются в прежнем направлении (вдоль оси) с одинаковой скоростью, т. е. нет двойного лучепреломления. Состояние поляризации волны в кристалле будет таким же, как у падающей волны.

Г. *Оптическая ось образует произвольный угол с преломляющей гранью кристалла. Свет падает по нормали* (рис. 4.12, *г*). Волновые поверхности обыкновенной и необыкновенной волн (т. е. касательные к сферам и эллипсоидам) представляют собой плоскости, параллельные границе. Из расположения точек касания этих плоскостей с эллипсоидальными лучевыми поверхностями ясно, что необыкновенные лучи при нормальном падении отклоняются от первоначального направления. Этим объясняется двойное лучепреломление при нормальном падении света на естественную грань кристалла.

При наклонном падении света преломление становится еще сложнее. Если оптическая ось не лежит в плоскости падения, то по положению точек касания огибающей плоскости с эллипсоидами вторичных волн можно установить, что необыкновенный луч при преломлении выходит из плоскости падения.

Из рассмотренных примеров видно, что построение Гюйгенса дает наглядную картину двойного лучепреломления и позволяет сравнительно просто найти направление отраженной, обыкновенной и необыкновенной преломленных волн. Однако оно оставляет открытым вопрос об их амплитудах.

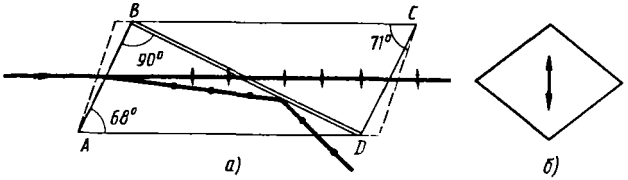
Контрольные вопросы

Каким условиям должны удовлетворять векторы **E** и **B** на границе анизотропной среды?
Как электромагнитная теория объясняет появление двух преломленных волн при падении плоской волны на поверхность одноосного кристалла?
Используя построение преломленных лучей по Гюйгенсу, выяснить, при какой ориентации оптической оси преломление необыкновенной волны на поверхности одноосного кристалла описывается обычным законом, т. е. $\sin\varphi/\sin\varphi_2 = \text{const}$.
При каких условиях преломленный луч остается в плоскости падения и при каких — выходит из нее?

4.4. Поляризационные призмы и поляроиды

Особенности преломления света на границе анизотропной среды можно использовать для получения поляризованного света. Вместо отдельного кристалла для этой цели более удобными оказываются их комбинации, называемые *поляризационными призмами.* Наиболее подходящий материал для их изготовления — исландский шпат, у которого сравнительно велико различие между обыкновенным и необыкновенным показателями преломления (n_o=1,658; n_e=1,486), реже применяют кварц (n_o=1,544; n_e=1,553).

Первая поляризационная призма была изобретена Николем в 1828 г. Она изготовляется из полученного раскалыванием по плоскостям спайности куска исландского шпата, длина которого



4.13
Ход лучей (а) и вид торцовой плоскости (б) в призме Николя

а — ход лучей, *б* — вид торцовой плоскости

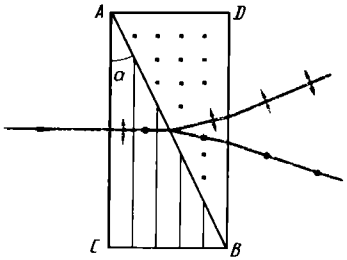
примерно в 3,5 раза больше толщины. Торцовые основания AB и CD (рис. 4.13, *а*) сошлифовывают под углом 68° к его длинным ребрам (вместо 71° у естественного кристалла). Затем кристалл разрезается по плоскости, перпендикулярной торцовым поверхностям (рис. 4.13, *б*), и отшлифованные половинки склеиваются в прежнем положении канадским бальзамом, образующим между ними тонкий прозрачный слой с показателем преломления $n=1{,}549$, имеющим промежуточное значение между n_0 и n_e. Для обыкновенного луча канадский бальзам представляет собой оптически менее плотную среду, для необыкновенного — более плотную. Поэтому обыкновенный луч может преодолеть слой бальзама только тогда, когда его угол падения на плоскость разреза меньше предельного угла полного отражения φ_m (см. § 3.3), определяемого условием $\sin\varphi_m = n/n_o = 1{,}549/1{,}658$. В данном случае $\varphi_m \approx 69°$ и обыкновенные лучи, идущие параллельно длинным ребрам призмы (и под небольшими углами к ним), устраняются за счет полного отражения. Необыкновенный луч, поляризованный в плоскости главного сечения, частично проходит через слой канадского бальзама, так как для него полное отражение исключено. В результате выходящий из призмы Николя параллельно своему направлению входа луч оказывается полностью поляризованным.

Полное отражение обыкновенного луча происходит лишь тогда, когда углы падения света на входную грань призмы лежат в определенных пределах. Угол между крайними лучами падающего пучка (29°), удовлетворяющими этому условию, определяет *апертуру полной поляризации* призмы.

Существует много модификаций призмы Николя. Для работы в ультрафиолетовой области спектра канадский бальзам непригоден из-за сильного поглощения, поэтому используют *призму Фуко* с тонкой воздушной прослойкой. Полное отражение обыкновенного луча в плоскости разреза происходит при меньших углах, и призма Фуко получается значительно короче николя, но апертура ее полной поляризации составляет всего 8°.

Иногда необходимо из падающего света получать сразу два луча, поляризованных в перпендикулярных направлениях. Служащие для этой цели поляризационные призмы называются *двухлучевыми.*

Из них наибольшим совершенством отличается *призма Волластона* (рис. 4.14). Две прямоугольные призмы склеены по гипотенузам так, что их оптические оси (показанные на рисунке линиями и точками) взаимно перпендикулярны. Обыкновенный и необыкновенный лучи, возникающие в первой призме при падении света по нормали к входной грани AC, идут по одному (прежнему) направлению, но с разными скоростями: $v_0 = c/n_o$ и $v_e = c/n_e$



Ход лучей в призме Волластона

Необыкновенный луч, в котором колебания вектора **E** происходят в плоскости чертежа, при переходе во вторую призму будет распространяться как обыкновенный, так как колебания в нем перпендикулярны оптической оси второй призмы. Его преломление на границе призм AB происходит, как при переходе между изотропными средами с показателями преломления $n_1 = n_e$ и $n_2 = n_o$. И наоборот, у обыкновенного луча первой призмы после перехода его во вторую призму колебания окажутся направленными вдоль оси; преломление на границе соответствует случаю $n_1 = n_o$ и $n_2 = n_e$. При $n_o > n_e$ (исландский шпат) первый луч отклонится в сторону ребра D, второй — в противоположную сторону. Этим обеспечивается угловое разведение выходящих из призмы пучков света, линейно поляризованных во взаимно перпендикулярных направлениях (см. задачу 1).

Кроме описанных выше поляризационных призм широкое распространение для получения поляризованного света нашли устройства, действие которых основано на явлении *дихроизма*, заключающемся в зависимости поглощения света в некоторых средах от направления колебаний. Сильным дихроизмом обладают кристаллы турмалина, в которых обыкновенный луч поглощается значительно больше необыкновенного. При достаточной толщине пластинки турмалина (≈ 1 мм) выходящий из нее свет будет практически полностью поляризован. Но для некоторых участков спектра необыкновенный луч тоже испытывает заметное поглощение, что ограничивает применение турмалина в качестве поляризатора.

Очень удобны в обращении поляризационные устройства из синтетических дихроичных материалов, известные в обиходе как *полароиды*. Поляроид представляет собой пленку целлулоида или другого прозрачного материала, в которую вкраплены определенным образом ориентированные микроскопические кристаллики сильно дихроичного вещества, обычно герапатита или других родственных ему соединений. Таким способом получают листы большой площади, обеспечивающие высокую степень поляризации проходящего света при больших апертурах практически во всей видимой области спектра и имеющие сравнительно небольшую стоимость. Недостатки

поляроидов связаны со спектральной селективностью поглощения герапатита, из-за чего фиолетовая часть спектра оказывается поляризованной лишь частично, а пленка получается неодинаково прозрачной для лучей разного цвета.

Контрольные вопросы

1. Какое явление использовано в призме Николя для удаления одного из лучей?
2. Каково направление напряженности электрического поля в выходящем из призмы Николя луче (по отношению к ее ребрам)?
3. Почему на диагональной плоскости призмы Волластона происходит расщепление падающего луча?

Задача

Призма Волластона (рис. 4.14) из исландского шпата ($n_o=1{,}658$; $n_e=1{,}486$) имеет преломляющий угол $\alpha=15°$. На какой угол θ будут разведены выходящие из призмы лучи, поляризованные в ортогональных направлениях?

Ответ: $\theta \approx 2(n_o-n_e)\operatorname{tg}\alpha = 5°17'$. ◊

4.5. Искусственная анизотропия. Эффект Керра

Многие оптически изотропные тела состоят из анизотропных молекул или других структурных элементов, хаотически ориентированных в пространстве. Микроскопическая анизотропия сглаживается лишь статистически. В результате какого-либо внешнего воздействия (механической деформации, электрического или магнитного поля), создающего физически выделенное направление в пространстве, такая среда может стать и макроскопически анизотропной. При достаточно сильном воздействии даже первоначально изотропные структурные элементы могут стать анизотропными; например, кубический кристалл каменной соли NaCl при одностороннем сжатии становится оптически анизотропным.

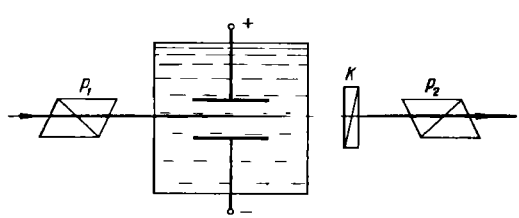
Двойное лучепреломление при механической деформации (*фотоупругость*) было открыто Брюстером в 1815 г. При одностороннем растяжении или сжатии направление деформации становится выделенным и играет роль оптической оси. Для наблюдения оптической анизотропии исследуемое тело помещают между скрещенными или параллельными поляризаторами, направления пропускания которых составляют некоторый угол (лучше всего 45°) с направлением деформации. Распространяющиеся перпендикулярно оптической оси обыкновенный и необыкновенный лучи приобретают некоторую разность фаз, и свет, прошедший через деформированное тело, становится эллиптически поляризованным. Опыт показывает, что измеряемая в таких опытах разность n_o-n_e, которую можно рассматривать как меру возникшей анизотропии, пропорциональна механиче-

скому напряжению, приложенному к деформируемому телу. В случае неоднородно деформированного тела таким методом можно получить картину распределения напряжений. Этим пользуются для исследования напряжений в сложных деталях и промышленных конструкциях: модель из подходящего прозрачного материала подвергают требуемой нагрузке и по наблюдаемой между скрещенными поляроидами картине с помощью числовых расчетов определяют внутренние напряжения. Этот метод значительно облегчает трудоемкую работу по расчету новых конструкций.

Оптический метод применяется также для исследования остаточных механических напряжений в оптическом стекле, возникающих при недостаточно медленном охлаждении после термической обработки. Разность $n_o - n_e$ зависит от длины волны (дисперсия искусственного двойного лучепреломления), и при наблюдении в белом свете картина неоднородно деформированного тела между скрещенными поляроидами оказывается разноцветной.

Искусственная анизотропия в жидкостях с сильно анизотропными элементами (например, в растворах полимеров) может возникнуть под влиянием даже очень слабых внешних динамических воздействий. Измерение двойного лучепреломления в ламинарном потоке между неподвижным и вращающимся коаксиальными цилиндрами, обусловленного ориентацией анизотропных молекул, используется для изучения свойств полимеров.

Возникновение оптической анизотропии во внешнем электрическом поле было обнаружено Керром в 1875 г. *Явление Керра* можно наблюдать в жидкостях и газах, где условия для теоретической интерпретации несравненно проще, чем в твердых телах. Благодаря этому обстоятельству теория эффекта Керра разработана значительно глубже, чем рассмотренного выше явления анизотропии при механических деформациях, трактовка которого требует исследования действия на молекулы среды межмолекулярных электрических полей, имеющих сложную структуру, в отличие от однородного внешнего поля в эффекте Керра. Большое научное значение эффекта Керра связано с тем, что он позволяет глубже проникнуть в природу анизотропии вообще и, кроме того, получить информацию о строении молекул. Вместе с тем явление Керра имеет чрезвычайно важное прикладное значение в научных исследованиях и в технике, обусловленное его способностью протекать практически безынерционно. Быстродействие керровских модуляторов света и оптических затворов может быть доведено до 10^{-12} с.



Двойное лучепреломление в электрическом поле

Схема наблюдения эф-

фекта Керра приведена на рис. 4.15. Заполненный исследуемым веществом конденсатор помещают между скрещенными поляризаторами P_1 и P_2, так что при отсутствии электрического поля свет через систему не проходит. При наложении электрического поля (направление $\mathbf{E}$ которого составляет угол 45° с направлениями пропускания поляризаторов) среда становится оптически анизотропной; выходящий из конденсатора свет поляризован эллиптически и частично проходит через анализатор P_2. Вводя компенсатор K, можно измерить разность фаз, возникшую между обыкновенным и необыкновенным лучами, и тем самым найти разность показателей преломления $n_e - n_o$ исследуемого вещества в электрическом поле. Опыт показывает, что эта разность пропорциональна квадрату напряженности внешнего электрического поля: $n_e - n_o = KE^2$. Такой характер зависимости от напряженности поля объясняется тем, что для изотропной (при отсутствии поля) среды величина $n_e - n_o$ не может зависеть от направления $\mathbf{E}$, поэтому разложение $n_e - n_o$ в ряд по степеням напряженности поля должно начинаться с квадратичного члена. В слабых полях, какие применяются на опыте, в разложении достаточно ограничиться только этим членом.

Для разности фаз, приобретаемой лучами на пути l, можно написать

$$\varphi = (2\pi l/\lambda)(n_e - n_o) = 2\pi B l E^2,$$

где $B = K/\lambda$ — *постоянная Керра,* характеризующая исследуемое вещество. Ее значения для разных веществ изменяются в широких пределах. Большим значением B обладает нитробензол, поэтому он часто используется в технических приложениях эффекта Керра. Если, например, конденсатор имеет длину $l = 5$ см и расстояние между пластинами $d = 1$ мм, то при напряжении 1500 В разность фаз достигает значения $\varphi = \pi/2$, т. е. ячейка Керра в таких условиях действует как пластинка в четверть волны. У других жидкостей и особенно у газов постоянная Керра значительно меньше, что требует использования высоких напряжений и чувствительных методов измерения разности фаз.

Для большинства веществ постоянная Керра положительна, т. е. $n_e > n_o$, что соответствует анизотропии положительного одноосного кристалла. Реже встречаются вещества, у которых $B < 0$ (этиловый эфир, спирт).

Причина оптической анизотропии вещества в электрическом поле заключается в анизотропии самих молекул. Количественная теория для газов была развита Ланжевеном в 1910 г. Поляризуемость анизотропной молекулы и, следовательно, ее вклад в показатель преломления среды зависят от ориентации молекулы относительно направления напряженности электрического поля световой волны. При отсутствии внешнего электрического поля анизотропные молекулы ориентированы хаотически, так что среда макроскопически оказы-

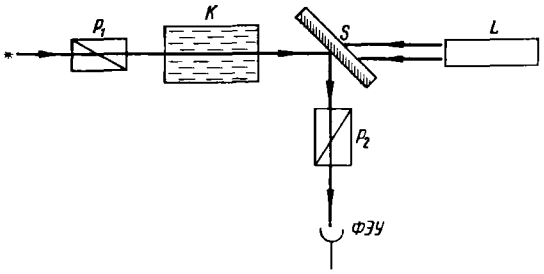
вается изотропной. Во внешнем электрическом поле молекулы преимущественно ориентируются своими осями наибольшей поляризуемости вдоль поля* и среда становится оптически анизотропной: световой волне с направлением колебаний вдоль внешнего поля (необыкновенной) соответствует большее значение показателя преломления, чем волне, поляризованной в поперечном направлении ($n_e > n_o$). Величины n_e и n_o можно, используя статистическую механику, выразить через главные значения поляризуемости отдельной молекулы.

Теория Ланжевена всегда дает $n_e > n_c$, т. е. $B > 0$, хотя на опыте для некоторых веществ получается $B < 0$. Борн в 1916 г. обобщил теорию Ланжевена, распространив ее на полярные молекулы, обладающие собственным дипольным моментом, направление которого может не совпадать с направлением наибольшей поляризуемости. Если эти направления взаимно перпендикулярны, то оси наибольшей поляризуемости молекул преимущественно ориентируются перпендикулярно внешнему полю и оптическая анизотропия среды соответствует отрицательному кристаллу ($n_o > n_e$).

Двойное лучепреломление в изотропной среде может возникнуть не только в постоянном внешнем электрическом поле, но и в переменном с частотами вплоть до оптических. Благодаря развитию лазерной техники появилась возможность получать оптическое излучение, в котором напряженность электрического поля достигает очень больших значений. Схема опыта по наблюдению эффекта Керра, вызванного электрическим полем лазерного излучения, показана на рис. 4.16. Луч зондирующего света ($\lambda \approx 500$ нм) проходит через ячейку K с исследуемой жидкостью и после отражения от полупрозрачной пластинки S направляется на фотоумножитель (ФЭУ). При скрещенных поляризаторах P_1 и P_2 свет не может попасть в ФЭУ. Когда через ячейку проходит мощный импульс инфракрасного поляризованного излучения лазера L, жидкость становится анизотропной, зондирующий свет выходит из ячейки эллиптически поляризованным и попадает в ФЭУ. Измеряя разность фаз φ между необыкновенным и обыкновенным лучами и зная среднеквадратичную напряженность поля лазерного излучения $(\langle E^2 \rangle)^{1/2}$, можно найти значение постоянной Керра в поле оптической частоты и сравнить его со значением в постоянном электрическом поле. В недипольных жидкостях эти значения практически совпадают. Однако в жидкостях с дипольными молекулами постоянная Керра уменьшается при переходе к оптическим частотам (у нитробензола приблизительно в 100 раз), так как дипольная молекула не успевает переориентироваться в такт с изменениями напряженности внешнего поля.

---

* Индуцируемый внешним полем дипольный момент анизотропной молекулы не совпадает с направлением напряженности поля. Поэтому и возникает момент сил, который стремится повернуть молекулу осью вдоль поля.

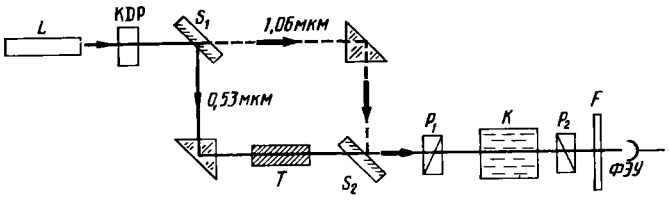


Двойное лучепреломление в электрическом поле лазерного излучения

В высокочастотных полях постоянные дипольные моменты молекул не играют роли в возникновении анизотропии. В отличие от вектора дипольного момента, имеющего определенное направление, оба направления оси наибольшей поляризуемости молекулы эквивалентны и направление действующего на молекулу момента сил не меняется при переключении направления внешнего поля на противоположное. Поэтому ориентация молекул в высокочастотных полях обусловлена только индуцированными дипольными моментами, как это предполагается в теории Ланжевена.

Большое прикладное значение эффекта Керра обусловлено его малой инерционностью. Для измерения времени установления в эффекте Керра используют мощные и короткие (до 10^{-12} с) импульсы инфракрасного лазера. Схема опыта показана на рис. 4.17. Импульс инфракрасного света ($\lambda = 1{,}06$ мкм) проходит через кристалл дигидрофосфата калия (KDP), где из-за нелинейного эффекта удвоения частоты (см. § 10.3) небольшая часть инфракрасного света превращается в зеленый ($\lambda = 0{,}53$ мкм). Зеркало S_1 пропускает инфракрасный свет и отражает зеленый, зеркало S_2, наоборот, пропускает зеленый и отражает инфракрасный. Пройдя по разным путям, зеленый и инфракрасный лучи затем снова соединяются и проходят через ячейку Керра, помещенную между скрещенными поляризаторами P_1 и P_2. Фильтр F задерживает инфракрасный свет и пропускает в ФЭУ только зеленый. С помощью стеклянных пластинок T разной толщины можно задержать импульс зеленого света относительно инфракрасного на разные промежутки времени τ. Опыт показывает, что релаксация анизотропии, созданной в жидкости инфракрасным импульсом, происходит экспоненциально по закону $\exp(-t/\tau_0)$ с постоянной времени τ_0. В сероуглероде $\tau_0 \sim 2 \cdot 10^{-12}$ с, в нитробензоле $\tau_0 \sim 5 \cdot 10^{-11}$ с.

Ячейка Керра, т. е. кювета с жидкостью, помещенная между скрещенными поляризаторами, может работать в качестве быстродействующего *оптического затвора*, управляемого кратковременными импульсами электрического поля. Если вместо электрического



417

Схема опыта для измерения времени релаксации искусственной анизотропии

импульса использовать короткий мощный лазерный импульс, то время экспозиции можно довести до 10^{-12} с. Керровский модулятор света, питаемый электрическим полем высокой частоты, позволяет осуществить до 10^9 прерываний в секунду. Модуляторы и затворы на эффекте Керра применяются, в частности, для управления режимом работы лазеров с целью получения сверхкоротких импульсов огромной мощности. Такие «гигантские» импульсы необходимы для исследований в бурно развивающейся нелинейной оптике.

Если анизотропные молекулы среды обладают постоянными *магнитными* моментами, то их преимущественная ориентация может быть создана постоянным *магнитным полем*. Поэтому в достаточно сильных магнитных полях среда становится анизотропной и в ней может наблюдаться двойное лучепреломление. В этом состоит *эффект Коттона—Мутона,* открытый в 1905 г.

Изменение оптических характеристик кристалла под действием внешнего *электрического* поля называется *электрооптическим эффектом Поккельса.* В одноосном кристалле распространение света вдоль оптической оси происходит с одной и той же фазовой скоростью $v_o = c/n_o$ независимо от направления его поляризации. Если кристалл не обладает центром симметрии, то при приложении внешнего электрического поля вдоль этой оси фазовые скорости волн с ортогональными направлениями поляризации становятся различными. В отличие от эффекта Керра, *квадратичного* по напряженности внешнего электрического поля, в электрооптическом эффекте разность фазовых скоростей таких волн пропорциональна напряженности поля (*линейный эффект Поккельса*). Безынерционность эффекта Поккельса позволяет широко использовать его для создания быстродействующих оптических затворов и высокочастотных модуляторов света. Вырезанная перпендикулярно оптической оси пластинка кристалла KDP (дигидрофосфата калия) помещается между скрещенными поляризаторами. Интенсивность света, пропускаемого такой *ячейкой Поккельса*, зависит от приложенного напряжения U по закону $I \sim \sin^2[\pi U/(2U_{\lambda/2})]$, где $U_{\lambda/2}$ — минимальное напряжение, при котором сдвиг фаз волн с ортогональными поляризациями равен π (для KDP $U_{\lambda/2} \approx 8$ кВ).

Контрольные вопросы

1. Какие практические применения имеет явление возникновения оптической анизотропии при механических деформациях?

2. Почему величина $n_e - n_o$, т. е. мера анизотропии жидкости в электрическом поле, пропорциональна квадрату напряженности поля?

   Какова физическая причина вызываемой внешним электрическим полем оптической анизотропии?

3. Почему в жидкостях с дипольными молекулами при переходе к высоким частотам внешнего электрического поля постоянная Керра уменьшается?

4. Какие практические применения находит эффект Керра?

5

ИНТЕРФЕРЕНЦИЯ СВЕТА

● Волновые свойства света наиболее отчетливо обнаруживают себя в интерференции и дифракции. Эти явления характерны для волн любой природы и сравнительно просто наблюдаются на опыте для волн на поверхности воды или для звуковых волн. Наблюдать же интерференцию и дифракцию световых волн можно лишь при определенных условиях.

● Под интерференцией света обычно понимают широкий круг явлений, в которых при наложении пучков света результирующая интенсивность не равна сумме интенсивностей отдельных пучков: в одних местах она больше, в других — меньше, т. е. возникают чередующиеся светлые и темные участки — интерференционные полосы.

● Свет, испускаемый обычными (нелазерными) источниками, не бывает строго монохроматическим. Поэтому для наблюдения интерференции свет от одного источника нужно разделить на два пучка и затем наложить их друг на друга. **Существующие экспериментальные методы получения когерентных пучков из одного светового пучка можно разделить на два класса. В методе деления волнового фронта пучок пропускается, например, через два близко расположенных отверстия в непрозрачном экране. Такой метод пригоден лишь при достаточно малых размерах источника. В другом методе пучок делится на одной или нескольких частично отражающих, частично пропускающих поверхностях. Этот метод деления амплитуды может применяться и при протяженных источниках. Он обеспечивает большую интенсивность и лежит в основе действия разнообразных интерферометров. В зависимости от числа интерферирующих пучков различают двухлучевые и многолучевые интерферометры. Они имеют важные практические применения в технике, метрологии и спектроскопии.**

5.1. Интерференция монохроматического света

Световые колебания в некоторой точке, через которую проходит строго монохроматическая волна, должны продолжаться бесконечно долго и иметь неизменную частоту и амплитуду. Свет, излучаемый любым реальным источником, этим свойством не обладает. Тем не менее монохроматическая идеализация оказывается достаточной для решения многих задач. В частности, при изучении явлений интерференции она пригодна для определения положения максимумов и минимумов интерференционной картины.

Пусть в некоторую точку приходят волны, напряженности электрического поля которых равны $\mathbf{E}_1$ и $\mathbf{E}_2$. По принципу суперпозиции*, напряженность результирующего поля равна их векторной сумме: $\mathbf{E}=\mathbf{E}_1+\mathbf{E}_2$. В результате сложения двух гармонических колебаний одинаковой частоты получается колебание той же частоты, неизменная во времени амплитуда которого зависит от соотношения фаз складываемых колебаний и поэтому в разных точках наблюдения имеет, вообще говоря, разные значения.

Из-за очень большой частоты оптических колебаний напряженность $\mathbf{E}$ невозможно измерить непосредственно. Все приемники излучения измеряют энергетические величины (интенсивность света или освещенность поверхности), усредненные за промежуток времени, очень большой по сравнению с периодом оптических колебаний. Поэтому экспериментально наблюдаемые величины пропорциональны среднему значению квадрата напряженности электрического поля $\langle E^2\rangle$ за время, определяемое инерционностью приемника излучения:

$$\langle E^2\rangle=\langle(\mathbf{E}_1+\mathbf{E}_2)^2\rangle=\langle E_1^2\rangle+\langle E_2^2\rangle+2\langle\mathbf{E}_1\mathbf{E}_2\rangle. \qquad (5.1)$$

Выражение для результирующей интенсивности помимо суммы интенсивностей каждой из волн содержит еще одно слагаемое, пропорциональное $2\langle\mathbf{E}_1\mathbf{E}_2\rangle$, называемое *интерференционным членом*. В тех случаях, когда он обращается в нуль, результирующая интенсивность равна сумме интенсивностей и интерференция отсутствует.

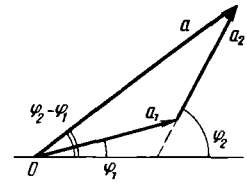
Скалярное произведение $\mathbf{E}_1\mathbf{E}_2$ равно нулю, если складываемые волны линейно поляризованы в ортогональных направлениях. Отсутствие интерференции лучей, поляризованных во взаимно перпендикулярных направлениях, было обнаружено Френелем и Араго в 1816 г. и интерпретировано в 1817 г. Юнгом как доказательство поперечности световых волн. Электромагнитная теория света полностью подтвердила это заключение.

В дальнейшем будем считать, что оба вектора $\mathbf{E}_1$ и $\mathbf{E}_2$ в точке

---

* При распространении в веществе очень интенсивных световых волн, напряженность поля в которых сравнима с напряженностью внутриатомных электрических полей (сфокусированное лазерное излучение), принцип суперпозиции нарушается, что приводит к разнообразным явлениям нелинейного взаимодействия волн, изучаемым нелинейной оптикой (см. гл. 10).

наблюдения совершают колебания вдоль одной прямой. Тогда можно отвлечься от векторного характера этих величин и записать интерференционный член в виде $2\langle E_1 E_2\rangle$. Рассмотрим сначала случай, когда в точке наблюдения, положение которой задается радиусом-вектором $\mathbf{r}$, налагаются две плоские монохроматические волны частотой ω, характеризуемые волновыми векторами $\mathbf{k}_1$ и $\mathbf{k}_2$:



Сложение гармонических колебаний с помощью векториой диаграммы

$$E_1 = a_1 \cos(\mathbf{k}_1\mathbf{r} - \omega t + \delta_1) = a_1 \cos(-\omega t + \varphi_1),$$
$$E_2 = a_2 \cos(\mathbf{k}_2\mathbf{r} - \omega t + \delta_2) = a_2 \cos(-\omega t + \varphi_2). \quad (5.2)$$

Здесь введены обозначения $\varphi_1 = \mathbf{k}_1\mathbf{r} + \delta_1$ и $\varphi_2 = \mathbf{k}_2\mathbf{r} + \delta_2$. При сложении двух гармонических колебаний одинаковой частоты получается гармоническое колебание той же частоты $E = a\cos(-\omega t + \varphi)$. Его амплитуду a проще всего найти с помощью векторной диаграммы, изображенной на рис. 5.1:

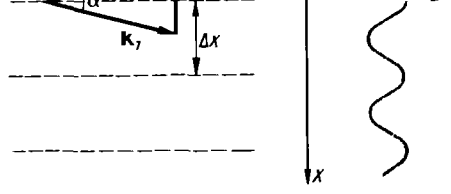
$$a^2 = a_1^2 + a_2^2 + 2a_1a_2\cos(\varphi_2 - \varphi_1) = a_1^2 + a_2^2 + 2a_1a_2\cos(\mathbf{K}\mathbf{r} + \delta_2 - \delta_1),$$
$$\mathbf{K} = \mathbf{k}_2 - \mathbf{k}_1. \quad (5.3)$$

Вводя интенсивности волн, пропорциональные квадратам их амплитуд, для интенсивности результирующего колебания получаем

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\mathbf{K}\mathbf{r} + \delta_2 - \delta_1). \quad (5.4)$$

Интенсивность зависит от положения точки наблюдения, характеризуемого радиусом-вектором $\mathbf{r}$. Поверхности равных интенсивностей определяются уравнением $\mathbf{K}\mathbf{r} = \text{const}$ и представляют собой плоскости, перпендикулярные вектору $\mathbf{K} = \mathbf{k}_2 - \mathbf{k}_1$ (рис. 5.2). Интенсивность максимальна там, где косинус в (5.4) принимает значение $+1$, и минимальна при -1. Расстояние Δx между соседними плоскостями максимальной (или минимальной) интенсивности определяется условием $K\Delta x = 2\pi$. Модули волновых векторов $\mathbf{k}_1$ и $\mathbf{k}_2$ одинаковы и равны $k = 2\pi/\lambda$, поэтому $K = 2k\sin(\alpha/2)$ (рис. 5.2), где α — угол между k_1 и k_2, т. е. между направлениями интерферирующих волн. Таким образом,

$$\Delta x = 2\pi/K = \pi/[k\sin(\alpha/2)] = \lambda/[2\sin(\alpha/2)] \approx \lambda/\alpha. \quad (5.5)$$



Плоскости равных интенсивностей и освещенность экрана при интерференции плоских волн

Последнее приближенное выражение справедливо, когда волны распространяются под малым углом друг к другу ($\alpha \ll 1$). Если на пути волн поместить плоский экран, то плоскости равной интенсивности

пересекут его по параллельным прямым, т. е. на экране будут наблюдаться чередующиеся светлые и темные *интерференционные полосы*. Когда плоскость экрана перпендикулярна биссектрисе угла между направлениями волн, ширина интерференционной полосы равна величине $\Delta x \approx \lambda/\alpha$, определяемой соотношением (5.5).

Зависимость освещенности экрана от координаты x (рис. 5.2) дается выражением

$$I(x)=I_1+I_2+2\sqrt{I_1 I_2}\cos(2\pi x/\Delta x) \qquad (5.6)$$

(начало отсчета на оси x выбрано в одном из максимумов). В частном случае интерференции волн одинаковой интенсивности $I_1=I_2$ формула (5.6) принимает вид

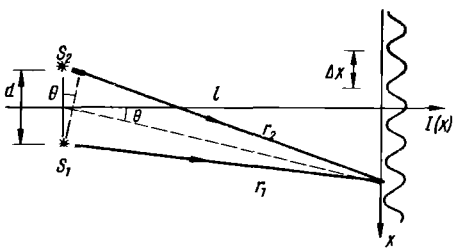
$$I(x)=2I_1[1+\cos(2\pi x/\Delta x)]=4I_1\cos^2(\pi x/\Delta x), \qquad (5.7)$$

а освещенность экрана изменяется от минимального значения, равного нулю, до максимального, равного учетверенному значению освещенности, создаваемой одной волной. Если экран наклонить на угол β, то ширина полосы увеличится и станет равной $\Delta x/\cos\beta$. Отметим, что полосы можно наблюдать при расположении экрана в любом месте, где перекрываются волны.

Рассмотрим теперь случай интерференции волн от двух одинаковых синфазных монохроматических точечных источников S_1 и S_2 (рис. 5.3), находящихся на расстоянии d друг от друга. Если расстояние l до экрана, где наблюдаются интерференционные полосы, много больше расстояния между источниками ($l \gg d$), то амплитуды обеих волн в точке наблюдения практически одинаковы и для напряженности поля в точке P можно написать

$$E=E_0\cos(kr_1-\omega t)+E_0\cos(kr_2-\omega t)=$$
$$=2E_0\cos[k(r_2-r_1)/2]\cos[k(r_2+r_1)/2-\omega t],$$

где r_1 и r_2 — расстояния от источников до точки наблюдения P (рис. 5.3). Величину $\Delta=r_2-r_1$ называют *разностью хода* интерферирующих волн. Интенсивность результирующего колебания пропорциональна квадрату амплитуды, поэтому



5.3

К интерференции волн, испускаемых точечными источниками S_1 и S_2

$$I=4I_0\cos^2(k\Delta/2)=$$

$$=2I_0(1+\cos k\Delta), \qquad (5.8)$$

где I_0 — интенсивность колебаний от одного источника. Освещенность экрана в минимумах равна нулю, а в максимумах — учетверенному значению освещенности, создаваемой одним источником. Положение максимумов определяется условием $k\Delta=\pm 2\pi m$, где целое число

$m=0, 1, 2, \ldots$ называется *порядком интерференции*. Учитывая, что $k=2\pi/\lambda$, условие максимумов можно записать в виде $\Delta=m\lambda$ — разность хода равна целому числу длин волн.

В том случае, когда волны от источников распространяются не в вакууме, а в среде с показателем преломления n, формула (5.8) остается в силе, но в ней под Δ следует понимать не геометрическую, а *оптическую разность хода* интерферирующих волн: $\Delta=n(r_2-r_1)$.

Чтобы найти зависимость освещенности экрана от координаты x (рис. 5.3), нужно разность хода Δ выразить через координату x точки наблюдения P. Для удобства введем угол θ, образуемый направлением на точку P с перпендикуляром к линии, соединяющей источники (т. е. с «оптической осью» рассматриваемой схемы). В практически важном случае малых значений θ ($\theta\ll 1$) для разности хода можно написать $\Delta\approx d\theta$. Так как $\theta\approx x/l$, то $\Delta\approx xd/l$. Подставляя Δ в (5.8), получаем

$$I(x)=2I_0[1+\cos(kdx/l)]. \qquad (5.9)$$

При $x=0$ расположен максимум, соответствующий нулевой разности хода. Для него порядок интерференции $m=0$. Это *центр интерференционной картины*. Расстояние между соседними максимумами или минимумами (пространственный период интерференционной картины) Δx определяется из условия $kd\,\Delta x/l=2\pi$, откуда $\Delta x=2\pi l/(kd)=\lambda l/d$. Если ввести угол схождения лучей $\alpha\approx d/l$, т. е. угол, под которым видны источники из точки наблюдения, то выражение для Δx можно записать следующим образом: $\Delta x\approx\lambda/\alpha$. Это совпадает с уже разобранным случаем интерференции плоских волн, распространяющихся под углом α. В самом деле, на большом расстоянии от источников сферические волны на небольших участках приближенно можно рассматривать как плоские, угол между направлениями которых при $\theta\ll 1$ приближенно равен d/l.

Отметим, что среднее значение освещенности по экрану в интерференционной картине, согласно формуле (5.9), равно удвоенной освещенности от одного источника. Это значит, что при интерференции происходит только перераспределение энергии в пространстве, а полный поток энергии остается неизменным. В действительности это условие приближенное и выполняется тем точнее, чем больше отношение d/λ. В противоположном случае $d\ll\lambda$, т. е., когда расстояние между источниками много меньше длины волны, во всех точках пространства от обоих источников происходит сложение колебаний с практически одинаковыми фазами. Это значит, что всюду интенсивность результирующих колебаний, а вместе с тем и полный поток энергии, почти в четыре раза больше, чем при одном источнике. Однако это не противоречит закону сохранения энергии, так как близко лежащие источники, взаимодействуя через создаваемое ими поле излучения, вместе излучают больше энергии, чем в том случае, когда они находятся далеко друг от друга. Для радио-

волн, излучаемых близко расположенными антеннами, увеличение полного потока энергии происходит за счет работы генератора, поддерживающего неизменные амплитуды колебаний тока в антеннах. В случае свободных колебаний в источниках увеличение излучаемой энергии приводит к более быстрому их затуханию. Уменьшение времени радиационного затухания для близко расположенных одинаковых излучающих атомов известно и в оптике (сверхизлучение).

До сих пор рассматривались только точки экрана, лежащие в плоскости чертежа на рис. 5.3. В пространстве поверхности максимальной и минимальной интенсивности представляют собой гиперболоиды вращения с фокусами в точках S_1 и S_2, так как соответствуют множеству точек, для которых разность расстояний от двух заданных точек (источников S_1 и S_2) имеет одно и то же значение. Форма интерференционных полос на экране определяется линиями пересечения этих гиперболоидов с плоскостью экрана. В небольшой центральной области экрана полосы практически можно считать равноотстоящими параллельными прямыми, ориентированными перпендикулярно плоскости чертежа на рис. 5.3.

Контрольные вопросы

В каком случае при наложении двух пучков света одинаковой частоты всегда, т. е. при любых фазовых соотношениях, происходит просто сложение интенсивностей?

По какому закону изменяется освещенность экрана, где наблюдается интерференция двух плоских монохроматических волн одинаковой интенсивности? разной интенсивности?

Какую форму имеют интерференционные полосы при падении на экран монохроматических волн от двух точечных когерентных источников?

5.2. Интерференционные опыты по методу деления волнового фронта

Особенности наблюдения явлений интерференции света от обычных (нелазерных) источников света обусловлены тем, что испускаемый ими свет никогда не бывает монохроматическим. Такой свет можно рассматривать как хаотическую последовательность отдельных цугов синусоидальных волн (см. § 1.7, 1.8). Длительность отдельного цуга не превышает 10^{-8} с даже в тех случаях, когда атомы источника не взаимодействуют (газоразрядные лампы низкого давления). Любой регистрирующий прибор имеет значительно большее время разрешения. Поэтому при наложении пучков света от разных источников фазовые соотношения между световыми колебаниями в любой точке за время наблюдения успевают многократно измениться случайным образом. В результате сложения бо́льшого числа колебаний со случайными фазами энергия результирующего колебания в любой точке будет равна сумме энергий складывающихся колебаний, т. е. не происхо-

дит характерного для интерференции перераспределения энергии в пространстве. Отсюда ясно, что для наблюдения интерференции света необходимы специальные условия: свет от одного и того же источника нужно разделить на два пучка (или несколько пучков) и затем наложить их друг на друга подходящим способом. Если разность хода этих пучков от источника до точки наблюдения не превышает длины отдельного цуга, то случайные изменения амплитуды и фазы световых колебаний в двух пучках происходят *согласованно*, т. е. эти изменения скоррелированы. О таких пучках говорят, что они полностью или частично *когерентны*, в зависимости от того, будет ли эта корреляция полной или частичной.

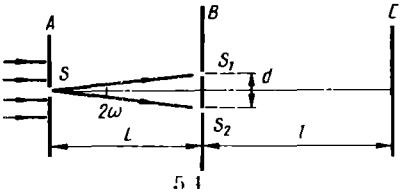
Впервые экспериментальная установка для демонстрации интерференции света была осуществлена Томасом Юнгом в начале XIX в. Яркий пучок солнечных лучей освещал экран A с малым отверстием S (рис. 5.4). Прошедший через отверстие свет вследствие дифракции образует расходящийся пучок, который падает на второй экран B с двумя малыми отверстиями S_1 и S_2, расположенными близко друг к другу на равных расстояниях от S. Эти отверстия действуют как вторичные точечные* синфазные источники, и исходящие от них волны, перекрываясь, создают интерференционную картину, наблюдаемую на удаленном экране C. Положение светлых и темных полос в ней можно находить пользуясь монохроматической идеализацией. Расстояние Δx между соседними полосами, как было показано в § 5.1, равно $\lambda l/d$.

Трудности наблюдения интерференции света в таком опыте связаны с тем, что длина волны видимого света очень мала. При $\lambda = 5 \cdot 10^{-5}$ см и расстоянии d между отверстиями S_1 и S_2, равном всего 0,5 мм, ширина интерференционных полос составляет только 1 мм при удалении экрана C на 1 м от отверстий. Измеряя ширину интерференционных полос, Юнг в 1802 г. впервые определил длины световых волн для разных цветов, хотя эти измерения и не были точными.

Введение дополнительного отверстия S (оно необходимо для когерентного возбуждения источников S_1 и S_2, см. § 5.5) резко уменьшает световой поток, что также затрудняет осуществление этого опыта. Интенсивность наблюдаемой в опыте Юнга интерференционной картины можно заметно увеличить, если вместо точечных отверстий S, S_1 и S_2 в экранах применить узкие длинные параллельные между собой щели. Вид полос вблизи центра интерференционного поля будет при этом таким же, как и при использовании точечных отверстий. Поясним это. Если точечное отверстие S перемещать перпендикулярно плоскости чертежа на рис. 5.4, то интер-

---

* Чтобы отверстия S_1 и S_2 пропускали достаточно света для возможности наблюдения интерференционных полос, они должны иметь конечные размеры. Такие источники создают направленное излучение, а не простые сферические волны (см. § 6.3). Влияние конечных размеров отверстия S обсуждается в § 5.5.



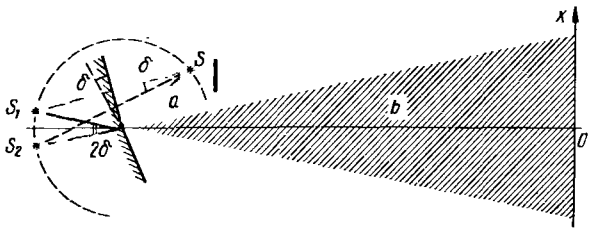
5.4

Схема опыта Юнга

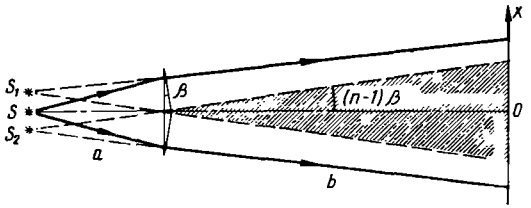
ференционные полосы на экране C, получаемые от точечных отверстий S_1 и S_2, будут просто смещаться вдоль своих направлений, т. е. также перпендикулярно плоскости чертежа. Поэтому замена отверстия S длинной щелью, т. е. непрерывной цепочкой точечных некогерентных источников, не приведет к ухудшению четкости интерференционных полос по крайней мере в той области, где их кривизна незначительна. Аналогично, не ухудшит четкости и замена отверстий S_1 и S_2 на узкие длинные щели, перпендикулярные плоскости чертежа на рис. 5.4. Увеличение же размера первого отверстия или щели S в плоскости чертежа (т. е. увеличение ширины) неизбежно приводит к уменьшению контрастности (видности) интерференционных полос (см. § 5.5).

В современной демонстрационной модификацин опыта Юнга в качестве источника света используют лазер. Прн этом для когерентного возбуждения вторичных источников S_1 и S_2 необходимость во вспомогательном отверстии S отпадает, так как в лазерном нзлучении световые колебания когерентны по всему поперечному сечению пучка (высокая пространственная когерентность лазерного нзлучення), и щели вводят непосредственно в пучок лазериого излучения.

Другой интерференционный опыт, аналогичный опыту Юнга, но в меньшей степени осложненный явлениями дифракции и более светосильный, был осуществлен Френелем в 1816 г. Две когерентные световые волны получались в результате отражения от двух зеркал, плоскости которых наклонены под небольшим углом δ друг к другу (*зеркала Френеля*, рис. 5.5). Источником служит узкая ярко освещенная щель S, параллельная ребру между зеркалами. Отраженные от зеркал пучки падают на экран, и в той области, где они перекрываются, возникает интерференционная картина. От прямого попадания лучей от источника S экран защищен ширмой. Для расчета освещенности $I(x)$ экрана можно считать, что интерферирующие волны испускаются вторичными источникам S_1 и S_2, представляющими собой мнимые изображения щели S в зеркалах. Поэтому



Зеркала Френеля



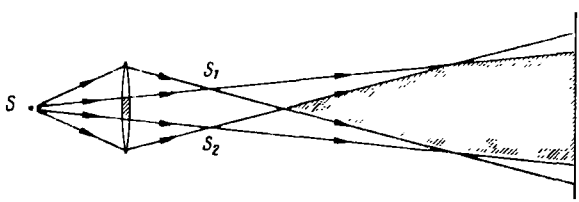
5.6
Бипризма Френеля

$I(x)$ будет определяться формулой (5.9), в которой расстояние l от источников до экрана следует заменить на $a+b$, где a — расстояние от S до ребра зеркал, b — расстояние от ребра до экрана (рис. 5.5). Расстояние d между источниками, как видно из рис. 5.5, равно $d \approx 2a\delta$. Поэтому ширина интерференционной полосы на экране равна $\Delta x \approx \lambda l/d = \lambda(a+b)/(2a\delta)$.

В другом интерференционном опыте, также предложенном Френелем, для разделения исходной световой волны на две используют призму с углом при вершине, близким к 180° (*бипризма Френеля*). Источником света служит ярко освещенная узкая щель S, параллельная преломляющему ребру бипризмы (рис. 5.6). Можно считать, что здесь образуются два близко лежащих мнимых изображения S_1 и S_2 источника S, так как каждая половина бипризмы отклоняет лучи на небольшой угол $(n-1)\beta$.

Аналогичное устройство, в котором роль когерентных источников играют действительные изображения ярко освещенной щели, получается, если собирающую линзу разрезать по диаметру и половинки немного раздвинуть (*билинза Бийе*, рис. 5.7). Прорезь закрывается непрозрачным экраном, а падающие на линзу лучи проходят через действительные изображения щели S_1 и S_2 и дальше перекрываются, образуя интерференционное поле.

Во всех перечисленных устройствах с первичным монохроматическим источником S (например, щелью, освещаемой ртутной лампой через светофильтр, выделяющий одну из узких спектральных линий) интерференционные полосы можно наблюдать в *любом месте* в области перекрытия расходящихся пучков от источников S_1 и S_2 (эти области заштрихованы на рис. 5.5—5.7).



Билинза Бийе

При описании интерференционных опытов предполагалось, что колебания векторов $\mathbf{E}$ в обоих световых пучках происходят в точке наблюдения по одному направлению. В случае ортогональной поляризации налагающихся пучков $\mathbf{E}_1\mathbf{E}_2=0$, интерференционный член в (5.1) равен нулю и происходит просто сложение интенсивностей, приводящее к равномерной освещенности в области перекрытия пучков. Это легко продемонстрировать, если в обычном интерференционном опыте на пути каждого из двух пучков поставить поляризаторы: полосы, отчетливо видные при совпадающих направлениях колебаний в обоих пучках, пропадают при повороте одного из поляризаторов на 90° (опыты Араго и Френеля).

Но для наблюдения интерференции нет необходимости использовать поляризованный свет. Неполяризованный (естественный) свет можно представить в виде суперпозиции двух некогерентных волн, поляризованных во взаимно перпендикулярных направлениях. В рассмотренных выше интерференционных опытах эти волны создают две независимые, но пространственно совпадающие системы полос, так как свет распространяется в изотропной среде, где фазовые скорости ортогонально поляризованных волн одинаковы и, следовательно, для каждой точки наблюдения обе волны имеют одну и ту же разность хода интерферирующих пучков.

В анизотропной среде волны с ортогональной поляризацией имеют различные фазовые скорости. Если на пути одного из интерферирующих пучков естественного света поместить кристаллическую пластинку в половину длины волны*, вносящую разность фаз π между волнами с ортогональной поляризацией, то светлые полосы одной из независимых интерференционных картин совпадут с темными полосами другой, что приведет к равномерной освещенности. Но в скрытом виде интерференционная картина все же существует: полосы можно наблюдать, если смотреть на экран через анализатор, направление пропускания которого соответствует поляризации одной из волн. При повороте анализатора на 90° видна вторая картина, смещенная относительной первой на полполосы (опыт С. И. Вавилова).

Иное дело, если на пути одного из пучков поместить слой оптически активного вещества (например кварцевую пластинку, грани которой перпендикулярны оптической оси), при прохождении через которое происходит поворот направления колебаний. Когда толщина пластинки подобрана так, что направление колебаний изменяется на 90°, интерференционная картина полностью пропадает: теперь в двух пучках, приходящих в точку наблюдения, одинако-

* Если опыт выполняется с немонохроматическим светом, то в другой пучок нужно ввести стеклянную пластинку, подобранную так, чтобы оптические пути обоих пучков были примерно одинаковы. Такая компенсация вносимой кристаллической пластинкой разности хода между пучками обеспечит наблюдение полос, соответствующих низким порядкам интерференции.

вое направление колебаний имеют некогерентные волны, которые в исходном пучке естественного света имели ортогональные поляризации.

Контрольные вопросы

- Почему для наблюдения интерференции света от обычных источников интерферирующие пучки должны происходить от одного и того же источника?
- Почему в центре интерференционной картины в опыте Юнга контрастность полос не ухудшается при замене точечных отверстий S, S_1 и S_2 длинными узкими параллельными щелями?

Можно ли наблюдать интерференцию неполяризованного света?

Задачи

1. Найти распределение интенсивности $I(x)$ на экране в установке с зеркалами Френеля (см. рис. 5.5), считая щель S бесконечно узкой. Сколько интерференционных полос можно наблюдать на экране?

◇ Ответ: $I(x)=I_0\cos^2\left[\dfrac{2\pi a\delta}{\lambda(a+b)}x\right]$, $N=\dfrac{4\delta^2 ab}{\lambda(a+b)}$.

2. В опыте с бипризмой Френеля (см. рис. 5.6) светлая интерференционная полоса располагается при $x=0$, т. е. в плоскости симметрии установки. Где будет эта полоса, если щель S сместить вверх на расстояние h?

◇ Ответ: при $x=-hb/a$. Указание: при наклоне призмы на небольшой угол положение мнимых изображений S_1 и S_2 щели S в первом приближении неизменно.

5.3. Деление амплитуды. Локализация интерференционных полос

Интерференцию света по методу деления амплитуды во многих отношениях наблюдать проще, чем в опытах с делением волнового фронта. В *опыте Поля* свет от источника S отражается двумя поверхностями тонкой прозрачной плоскопараллельной пластинки (рис. 5.8). В любую точку P, находящуюся с той же стороны от пластинки, что и источник, приходят два луча (при малом коэффициенте отражения повторные отражения от внутренних поверхностей пластинки можно не принимать во внимание ввиду ничтожной энергии пучков, испытавших два отражения и более). Эти лучи образуют интерференционную картину. Для определения вида полос можно представить себе, что лучи выходят из мнимых изображений S_1 и S_2

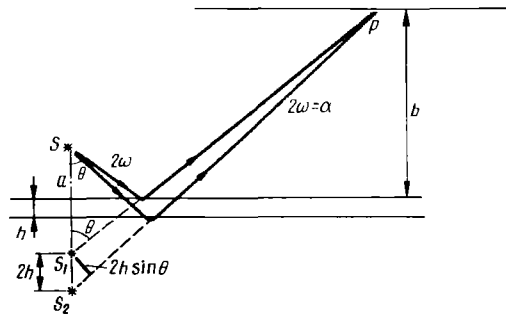


Рис. 5.8

Интерференция при отражении света двумя поверхностями тонкой пластинки

источника S, создаваемых поверхностями пластинки. На удаленном экране, расположенном параллельно пластинке, интерференционные полосы имеют вид концентрических колец с центрами на перпендикуляре к пластинке, проходящем через источник S.

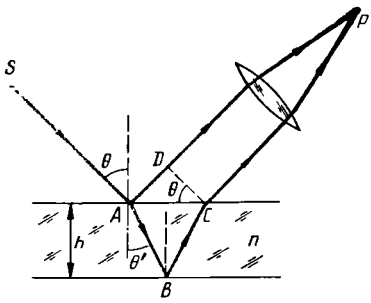
Этот опыт, как показано в § 5.5, предъявляет менее жесткие требования к размерам источника S, чем рассмотренные выше опыты. Поэтому можно в качестве S применить ртутную лампу без вспомогательного экрана с малым отверстием, что обеспечивает значительный световой поток. С помощью листочка слюды (толщиной 0,03—0,05 мм) можно получить яркую интерференционную картину прямо на потолке и стенах аудитории. Чем тоньше пластинка, тем крупнее масштаб интерференционной картины, т. е. больше расстояние между полосами.

Особенно важен частный случай интерференции света, отраженного двумя поверхностями плоскопараллельной пластинки, когда точка наблюдения P находится в бесконечности, т. е. наблюдение ведется либо глазом, аккомодированным на бесконечность, либо на экране, расположенном в фокальной плоскости собирающей линзы (рис. 5.9). В этом случае оба луча, идущие от S к P, порождены одним падающим лучом и после отражения от передней и задней поверхностей пластинки параллельны друг другу (рис. 5.9). Оптическая разность хода между ними в точке P такая же, как на линии DC:

$$\Delta = n(|AB| + |BC|) - |AD|.$$

Здесь n — показатель преломления материала пластинки. Предполагается, что над пластинкой находится воздух, т. е. $n_1 \approx 1$. Так как $|AB| = |BC| = 2h/\cos\theta'$, $|AD| = 2h \operatorname{tg}\theta' \sin\theta$ (h — толщина пластинки, θ и θ' — углы падения и преломления на верхней грани; $\sin\theta = n\sin\theta'$), то для разности хода получаем

$$\Delta = 2nh\cos\theta'. \qquad (5.10)$$



5.9
Возникновение интерференционных полос, локализованных в бесконечности

Следует также учесть, что при отражении волны от верхней поверхности пластинки в соответствии с формулами Френеля (см. § 3.2) ее фаза изменяется на π. Поэтому разность фаз δ складываемых волн в точке P равна

$$\delta = 4\pi nh\cos\theta'/\lambda_0 \pm \pi =$$

$$= 4\pi h\sqrt{n^2 - \sin^2\theta}/\lambda_0 \pm \pi, \qquad (5.11)$$

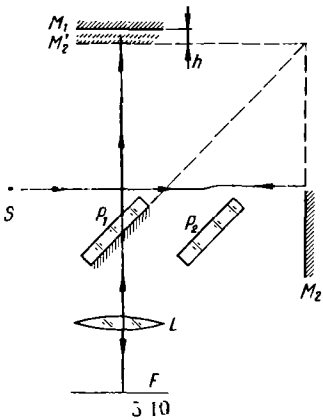
где λ_0 — длина волны в вакууме. Разность фаз δ определяется углом θ, который однозначно связан с положением точки P в фокальной

плоскости линзы (рис. 5.9). Она не зависит от положения источника S: если источник переместить, разность фаз двух волн в точке P останется прежней. Отсюда вытекает, что при использовании *протяженного* источника интерференционные полосы будут столь же отчетливыми, как и с точечным источником: каждый элемент протяженного источника создает в фокальной плоскости линзы свою интерференционную картину, причем положение полос во всех этих картинах совершенно одинаково. Так как это справедливо для одной вполне определенной плоскости наблюдения, то про такие полосы говорят, что они *локализованы*. В данном случае они локализованы в бесконечности (или в фокальной плоскости линзы).

В соответствии с формулой (5.11) светлые полосы расположены в местах, для которых $2nh\cos\theta' \pm \lambda_0/2 = m\lambda_0$, где m — целое число, называемое порядком интерференции. Полоса, соответствующая данному порядку интерференции, обусловлена светом, падающим на пластинку под вполне определенным углом θ. Поэтому такие полосы называют *интерференционными полосами равного наклона*. Если ось объектива расположена перпендикулярно пластинке, полосы имеют вид концентрических колец с центром в фокусе. В центре картины порядок интерференции максимален. Исходя из (5.11) легко показать, что угловой масштаб наблюдаемой картины пропорционален $1/\sqrt{h}$ (чем тоньше пластинка, тем шире полосы), а радиусы последовательных светлых полос пропорциональны квадратному корню из целых чисел (при условии, что в центре максимум интенсивности).

Полосы равного наклона можно получить не только в отраженном свете, но и в свете, прошедшем сквозь пластинку. В этом случае один из лучей проходит прямо, а другой — после двух отражений на внутренней стороне пластинки. Когда коэффициент отражения на поверхности пластинки мал (для стеклянной пластинки при нормальном падении примерно 0,04), интенсивности интерферирующих лучей, прошедших сквозь пластинку, сильно отличаются друг от друга. В соответствии с формулой (5.4) различие между максимальным и минимальным значениями интенсивности оказывается малым, а видность полос — низкой.

Для наблюдения полос равного наклона вместо плоскопараллельной пластинки удобно использовать *интерферометр Майкельсона*. Его упрощенная схема приведена на рис. 5.10. Свет от источника S падает на пластинку P_1, задняя сторона которой покрыта тонким полупрозрачным слоем серебра или алюминия. Здесь пучок разделяется на два взаимно перпендикулярных пучка. Отраженный пластинкой P_1 пучок падает на зеркало M_1, отражается назад, вновь попадает на пластинку P_1, где снова разделяется на две части. Одна из них идет к источнику S и не представляет интереса, а другая попадает в зрительную трубу, установленную на бесконечность, или на линзу L, в фокальной плоскости F которой расположен экран для наблюдения интерференции. Прошедший сквозь пластинку P_1 пучок от



5.10

Схема интерферометра Майкельсона

источника падает на зеркало M_2, возвращается к P_1 и частично отражается в сторону линзы L. Таким образом, от одного источника S получаются два пучка примерно одинаковой интенсивности, которые распространяются после разделения пластинкой P_1 в разных «плечах» интерферометра, затем снова встречаются и создают интерференционную картину в фокальной плоскости линзы L. Пластинка P_2, такая же, как и P_1, только без отражающего покрытия, ставится на пути второго пучка для компенсации разности хода, возникающей из-за того, что первый пучок проходит через P_1 три раза, а второй — только один раз. Зеркало M_2 неподвижно, а зеркало M_1 можно передвигать микрометрическим винтом так, что его плоскость все время остается перпендикулярной зеркалу M_2.

Построим изображение зеркала M_2, создаваемое отражающей поверхностью разделительной пластинки (M_2' на рис. 5.10). Оптическая длина пути от источника до точки наблюдения для луча, отразившегося от зеркала M_2, будет такой же, как и для воображаемого луча, отразившегося от M_2'. Поэтому можно считать, что интерференционная картина, наблюдаемая в фокальной плоскости линзы L, возникает из-за воздушного слоя между отражающей поверхностью M_1 и мнимой отражающей поверхностью M_2'. При параллельных поверхностях M_1 и M_2' полосы имеют вид концентрических окружностей с центром в фокусе линзы.

Если после разделительной пластинки P_1 пучки имеют одинаковую интенсивность, то распределение интенсивности в фокальной плоскости описывается формулой (5.8), где разность хода Δ, как и в случае плоскопараллельного воздушного слоя, в соответствии с (5.10) равна $\Delta = 2h\cos\theta$. Разность хода при заданном расстоянии h между M_1' и M_2', т. е. при фиксированном положении подвижного зеркала, зависит только от угла θ наклона луча по отношению к оптической оси. Данному значению θ соответствует кольцо радиусом $F\,\mathrm{tg}\,\theta$ в фокальной плоскости линзы. Поэтому положение и размер светлых и темных колец не зависят от положения источника S, т. е. можно использовать протяженный источник. При этом получаются интерференционные полосы, локализованные в фокальной плоскости линзы L.

Центру интерференционной картины ($\theta = 0$) соответствует максимальная разность хода $\Delta_{max} = 2h$, равная удвоенному расстоянию между M_1 и M_2'. Когда M_1 приближается к M_2', кольца стягиваются по направлению к центру. Перемещение зеркала на расстояние $\Delta m\lambda_0/2$ вызывает смещение картины на Δm порядков. Визуально смещение

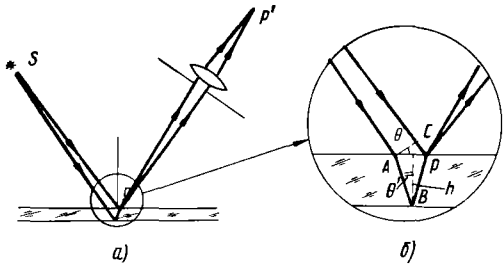
можно оценить с точностью до $^1/_{20}$ порядка, но существуют методы, позволяющие обнаружить смещения до 10^{-3} порядка. По мере приближения M_1 к M_2' угловой масштаб картины возрастает до тех пор, пока M_1 не совпадет с M_2'. При этом освещенность экрана (или поля зрения при визуальном наблюдении) становится равномерной.

Мы рассмотрели интерференционные опыты, в которых деление амплитуды световой волны от источника происходило в результате частичного отражения на поверхностях плоскопараллельной пластинки. В случае точечного источника полосы можно наблюдать всюду, т. е. они не локализованы. Но на бесконечности или в фокальной плоскости собирающей линзы полосы наблюдаются и при протяженном источнике. Локализованные полосы при протяженном источнике можно наблюдать и в других условиях. Оказывается, что для достаточно тонкой пластинки или пленки (поверхности которой не обязательно должны быть параллельными и вообще плоскими) можно наблюдать интерференционную картину, локализованную вблизи отражающей поверхности. В белом свете интерференционные полосы окрашены. Поэтому такое явление называют *цвета тонких пленок*. Его легко наблюдать на мыльных пузырях, на тонких пленках масла или бензина, плавающих на поверхности воды, на пленках окислов, возникающих на поверхности металлов при закале, и т. п.

Локализованные вблизи поверхности пленки или тонкой пластинки интерференционные полосы можно наблюдать невооруженным глазом либо с помощью лупы или сфокусированного на поверхность микроскопа. С помощью собирающей линзы интерференционную картину с поверхности пленки можно отобразить на экране (рис. 5.11, *а*). В самом деле, лучи, выходящие из точки P, вновь соберутся в сопряженной точке P' (P' — изображение точки P, создаваемое линзой). Так как оптические длины всех лучей между сопряженными точками одинаковы, интерферирующие лучи придут в точку P' с той же разностью фаз, какой они обладали в P. Поэтому линза создает не только изображение поверхности пленки, но и системы интерференционных полос, локализованных на поверхности.

Чтобы понять причину возникновения этих полос, вычислим разность хода Δ двух лучей, приходящих в точку P от источника S. Из рис. 5.11, *б* видно, что $\Delta = n(|AB| + |BP|) - |CP|$. Когда пленка мало отличается от плоскопараллельной, треугольники можно считать прямоугольными и положить $|AB| \approx |BP| \approx h/\cos\theta'$, $|CP| \approx \approx 2h\operatorname{tg}\theta'\sin\theta$. Подставляя эти величины в выражение для Δ, получаем $\Delta \approx 2nh\cos\theta'$. Учитывая изменение фазы на π для луча, отраженного передней поверхностью пленки, находим следующее выражение для разности фаз δ двух рассматриваемых лучей в точке P:

$$\delta = 4\pi nh\cos\theta'/\lambda_0 \pm \pi. \qquad (5.12)$$



5.11
Наблюдение полос, локализованных на пленке

Разность фаз δ зависит от толщины пленки h и от угла θ'*. Для данной точки P толщина h имеет определенное значение, и если использовать протяженный источник света, то различие величин δ для лучей, приходящих от разных точек источника, связано с различием для них значений $\cos\theta'$. Когда интервал возможного изменения $\cos\theta'$ достаточно мал, то разброс значений δ для пар лучей в точке P от разных точек источника много меньше 2π и полосы отчетливо видны. Практически условие малости интервала изменений $\cos\theta'$ легко выполнить при наблюдении в направлении, близком к нормальному. При наблюдении под углом для выполнения этого условия требуется ограничить входной зрачок объектива. Если полосы рассматривают невооруженным глазом, его зрачок пропускает лучи из P в пределах небольшого телесного угла, что сильно ограничивает интервал значений $\cos\theta'$. Тогда значение δ для точки P будет практически одинаковым у всех пар интерферирующих лучей, попадающих в глаз.

Заметим, что в случае точечного источника интерференционные полосы при отражении от двух поверхностей пленки (не обязательно плоскопараллельной) можно наблюдать всюду, а не только на ее поверхности. Локализация полос на поверхности пленки возникает как следствие использования протяженного источника света.

В точке P (и, следовательно, в P') будет находиться максимум интенсивности, если разность фаз (5.12) кратна 2π, или, что эквивалентно, при выполнении условия

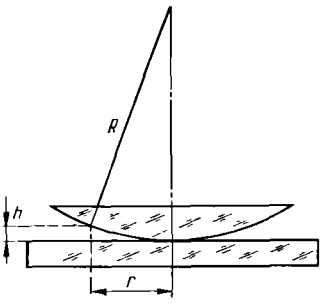
$$2nh\overline{\cos\theta'} \pm \lambda_0/2 = m\lambda_0 \quad (m=0, 1, 2, \ldots), \tag{5.13}$$

где черта над $\cos\theta'$ означает усреднение по тем точкам источника, свет от которых попадает в P'. Отметим, что соотношение (5.13) остается в силе и при неплоских поверхностях пленки при условии, что угол между ними остается малым. Поэтому интерференционные полосы соответствуют совокупности мест пленки, где ее оптическая толщина nh имеет одно и то же значение (при условии, конечно, что $\overline{\cos\theta'}$ в достаточной степени одинаков для всей области наблюдения). По этой причине такие полосы называют обычно *полосами равной толщины*.

Полосы равной толщины можно наблюдать в тонкой прослойке воздуха между поверхностями двух прозрачных пластинок. Когда

* Несмотря на внешнее сходство с формулой (5.11), физическое содержание формулы (5.12) совсем другое, так как она соответствует другим условиям наблюдения интерференции.

направление наблюдения близко к нормальному, $\cos\theta' \approx 1$ и темные полосы проходят в местах, толщина которых удовлетворяет условию $h = m\lambda/2$, где $m = 0, 1, 2, \ldots$. Переход от одной полосы к другой соответствует изменению толщины на $\lambda/2$. При постоянной толщине слоя интенсивность одинакова по всей его поверхности. Этим пользуются для испытания качества оптических поверхностей при их шлифовке путем наблюдения прослойки между исследуемой поверхностью и поверхностью эталона (контрольной пластинки).



5.12
К возникновению колец Ньютона

При клиновидной воздушной прослойке между плоскими поверхностями полосы будут проходить параллельно ребру клина на одинаковом расстоянии друг от друга, равном $\lambda/(2\alpha)$, где α — угол между плоскостями. Таким способом легко измерять углы порядка 0,1′ и меньше, а также обнаруживать дефекты поверхности с точностью, недоступной другим методам (0,1λ и менее).

Кольцевые полосы равной толщины, наблюдаемые в воздушном зазоре между соприкасающимися выпуклой сферической поверхностью линзы малой кривизны и плоской поверхностью стекла (рис. 5.12), называют *кольцами Ньютона*. Их общий центр расположен в точке касания. В отраженном свете центр темный, так как при толщине воздушной прослойки, много меньшей длины волны λ, разность фаз интерферирующих волн обусловлена различием в условиях отражения на двух поверхностях и близка к π. Толщина h воздушного зазора связана с расстоянием r до точки касания (рис. 5.12):

$$h = R - \sqrt{R^2 - r^2} = R\left(1 - \sqrt{1 - r^2/R^2}\right) \approx r^2/(2R).$$

Здесь использовано условие $(r/R)^2 \ll 1$. При наблюдении по нормали темные полосы, как уже отмечалось, соответствуют толщине $h = m\lambda/2$, поэтому для радиуса r_m m-го темного кольца получаем

$$r_m = \sqrt{mR\lambda} \quad (m = 0, 1, 2, \ldots) \tag{5.14}$$

— радиусы темных колец пропорциональны квадратному корню из натуральных чисел. Если линзу постепенно отодвигать от поверхности стекла, то интерференционные кольца будут стягиваться к центру. При увеличении расстояния на $\lambda/2$ картина принимает прежний вид, так как место каждого кольца будет занято кольцом следующего порядка.

С помощью колец Ньютона, как и в опыте Юнга, можно сравнительно простыми средствами приближенно определить длину волны света.

Полосы равной толщины можно наблюдать и с помощью интерферометра Майкельсона, если одно из зеркал M_1 или M_2 (см. рис. 5.10) отклонить на небольшой угол. Когда M_1 и M_2' расположены близко и образуют клин с малым углом, полосы имеют вид эквидистантных прямых, параллельных ребру клина и локализованных на поверхности клина либо вблизи нее. Для их наблюдения экран следует помещать не в фокальную плоскость линзы L, а в сопряженную с M_1 плоскость (где линза формирует изображение плоскости зеркала M_1), либо наблюдать в зрительную трубу, сфокусированную на поверхность зеркала M_1.

Контрольные вопросы

- □ Почему в интерференционных опытах по методу деления амплитуды с помощью тонкой прозрачной пластинки используют обычно отраженный свет, а не прошедший?
- □ Почему при наблюдении полос равного наклона можно использовать протяженный источник света? Где локализованы полосы равного наклона?
- □ Почему для колец Ньютона, получающихся между соприкасающимися поверхностями линзы и стеклянной пластинки, при наблюдении в отраженном свете центр темный?
- □ Какой вид имеют интерференционные полосы равной толщины в воздушном клине между плоскими поверхностями стеклянных пластинок?

Задача

Полосы равного наклона наблюдаются при отражении света длиной волны λ_0 от поверхностей плоскопараллельной стеклянной пластинки толщиной h (см. рис. 5.9). Найти угловой радиус θ_k k-го от центра светлого кольца. Центр колец светлый.

◇ Светлые полосы расположены в местах, для которых $2nh\cos\theta' \pm \lambda_0/2 = m\lambda_0$. Порядок интерференции m максимален в центре картины (при $\theta'=0$) и по условию равен целому числу m_0: $2nh \pm \lambda_0/2 = m_0\lambda_0$. Для k-го светлого кольца $2nh\cos\theta_k' \pm \lambda_0/2 = m_k\lambda_0$. Почленно вычитая эти равенства, получаем

$$2nh(1-\cos\theta_k') = (m_0 - m_k)\lambda_0.$$

Здесь $m_0 - m_k (k=0, 1, 2, \ldots)$ — номер светлого кольца при отсчете от центра. При $\theta_k' \ll 1$, по закону преломления, $\theta_k = n\theta_k'$ и $1-\cos\theta_k' \approx \theta_k'^2/2 \approx \theta_k^2/(2n^2)$, поэтому $h\theta_k^2/n = k\lambda_0$, откуда $\theta_k = \sqrt{kn\lambda_0/h}$. Радиус светлого кольца пропорционален корню из его номера, а масштаб наблюдаемой картины пропорционален $1/\sqrt{h}$. ◇

5.4. Интерференция квазимонохроматического света. Временна́я когерентность

В приведенном выше описании интерференционных опытов при их интерпретации предполагалось, что источник излучает монохроматический свет. Естественно, что результаты, полученные для монохроматической идеализации, имеют ограниченную применимость. В этом параграфе мы выясним, к каким изменениям в интерференционных

явлениях приводит учет спектрального состава реальных источников света.

Начнем с простейшего случая, когда источник излучает две очень узкие, близкие друг к другу спектральные линии с частотами ω_1 и ω_2. Если бы излучение на каждой из частот представляло собой бесконечную синусоидальную волну, то результирующее излучение было бы волной средней частоты с периодически изменяющейся амплитудой. Но в действительности излучение каждой из спектральных компонент представляет собой хаотическую последовательность более или менее длинных волновых цугов. Обычно за время наблюдения проходит много цугов, колебания в которых никак не связаны по фазе. Поэтому можно считать, что вместо одного имеется два расположенных в одном месте источника, независимо друг от друга излучающих волны с частотами ω_1 и ω_2. При выполнении интерференционных опытов с таким источником каждая из волн создает свою интерференционную картину и эти картины просто налагаются одна на другую.

Если часто́ты ω_1 и ω_2 мало отличаются, то интерференционные полосы в каждой картине имеют почти одинаковую ширину. В тех местах, где светлые полосы одной картины налагаются на светлые полосы другой, четкость полос в суммарной картине наибольшая. Наоборот, там, где светлые полосы одной картины приходятся на темные полосы другой, четкость полос уменьшается вплоть до их полного исчезновения.

Исследуем этот вопрос количественно. Распределение интенсивности в двухлучевой интерференционной картине для отдельной спектральной линии было получено в § 5.1. Когда обе волны имеют одинаковые амплитуды, зависимость интенсивности от разности хода Δ от вторичных источников до точки наблюдения дается формулой (5.8):

$$I(\Delta)=2I_0(1+\cos k\Delta), \qquad (5.15)$$

где $k=\omega/c$ — волновое число, а I_0 — равномерная интенсивность волны, которую создавал бы только один вторичный источник. Каждая спектральная линия первичного источника создает свою интерференционную картину, распределение интенсивности в которой описывается формулой (5.15) с соответствующим значением волнового числа k_1 или k_2.

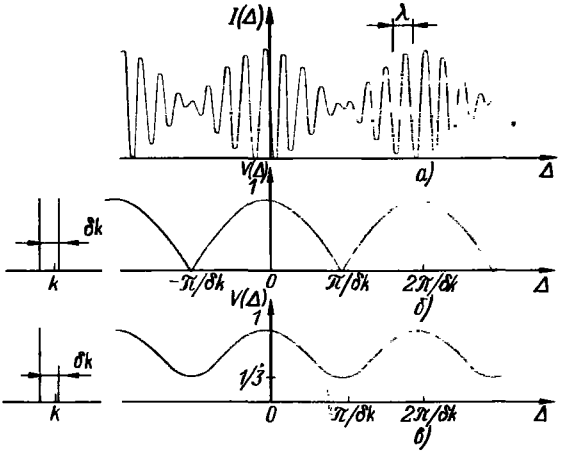
Полная интенсивность при наложении двух интерференционных картин имеет вид

$$I(\Delta)=I_1(\Delta)+I_2(\Delta)=2I_{10}(1+\cos k_1\Delta)+2I_{20}(1+\cos k_2\Delta). \qquad (5.16)$$

Будем считать, что обе спектральные линии имеют одинаковую интенсивность. Тогда $I_{10}=I_{20}=I_0$ и формулу (5.16) можно преобразовать:

$$I(\Delta)=4I_0[1+\cos(\delta k\Delta/2)\cos k\Delta], \qquad (5.17)$$

где $k=(k_1+k_2)/2$ — среднее значение волнового числа, а $\delta k=k_2-k_1$ — разность волновых чисел спектральных линий. Зависимость интенсивности от разности хода Δ при $\delta k \ll k$ показана на



Интерференционные полосы (*а*), их видность в случае источника, излучающего две близкие спектральные линии одинаковой интенсивности (*б*) и вдвое отличающиеся по интенсивности (*в*)

рис. 5.13, *а*. Расстояние между соседними полосами определяется вторым (быстро осциллирующим) множителем $\cos k\Delta$ и соответствует разности хода, равной одной (средней) длине волны $\lambda = 2\pi/k$. Первый множитель $\cos(\delta k\Delta/2)$ изменяется постепенно, оставаясь практически постоянным на протяжении многих интерференционных полос, так как $\delta k \ll k$ (разность частот спектральных линий много меньше средней частоты). В тех местах, где его значение приближается к нулю, контрастность интерференционной картины уменьшается: разность интенсивностей в соседних максимумах и минимумах стремится к нулю, полосы исчезают на фоне равномерной освещенности.

Для характеристики четкости интерференционных полос Майкельсон ввел *функцию видности* (или просто *видность*)

$$V = (I_{\max} - I_{\min})/(I_{\max} + I_{\min}), \qquad (5.18)$$

т. е. отношение разности интенсивностей в соседних максимуме и минимуме к удвоенному среднему значению интенсивности. Видность имеет максимальное значение, равное 1, когда интенсивность в минимумах $I_{\min} = 0$. Из формулы (5.17) следует, что при заданной разности хода Δ ближайшему максимуму соответствует $\cos k\Delta = \pm 1$ [в зависимости от знака $\cos(\delta k\Delta/2)$] и $I_{\max} = 4I_0(1 + |\cos(\delta k\Delta/2)|)$. Аналогично, $I_{\min} = 4I_0(1 - |\cos(\delta k\Delta/2)|)$. Подставляя $I_{\max}$ и $I_{\min}$ в (5.18), находим

$$V(\Delta) = |\cos(\delta k\Delta/2)|. \qquad (5.19)$$

Зависимость видности полос от разности хода показана на рис. 5.13, *б*. Видность принимает свое максимальное значение, равное

единице, при $\Delta=0$, т. е. вблизи полосы нулевого порядка, где освещенность в минимумах приближается к нулю. С увеличением разности хода видность уменьшается и обращается в нуль при $\delta k\,\Delta/2=\pi/2$, т. е. при $\Delta=\lambda^2/(2\delta\lambda)$. Это соответствует случаю, когда светлые полосы интерференционной картины от одной спектральной линии совмещаются с темными полосами от другой. При дальнейшем увеличении разности хода видность полос возрастает и при $\Delta=\lambda^2/\delta\lambda$ снова равна единице. Таким образом, полный период изменения видности охватывает $N=\lambda/\delta\lambda=\omega/\delta\omega$ интерференционных полос.

Кривая видности для случая, когда спектральные компоненты имеют различную интенсивность, приведена на рис. 5.13, *в* (см. задачу 1). Отличие от случая равных интенсивностей в том, что в минимумах видность в нуль не обращается, т. е. полного исчезновения полос не происходит.

Для наблюдения на опыте такой интерференционной картины с периодическими изменениями видности полос должна быть обеспечена возможность изменения разности хода в широких пределах. Можно, например, получить полосы равного наклона в интерферометре Майкельсона (см. § 5.3) с подвижным зеркалом M_1. Когда плечи интерферометра почти равны, т. е. расстояние между M_1 и M_2' мало́ (см. рис. 5.10), наблюдаемые полосы соответствуют разностям хода, равным небольшому числу длин волн. При этом (рис. 5.13) полосы имеют наибольшую видность. При перемещении зеркала M_1 разность хода Δ возрастает, а видность интерференционных полос в соответствии с (5.19) периодически изменяется (рис. 5.13, *б*).

Таким образом, из наблюдения за изменением видности интерференционных полос в зависимости от разности хода можно получить информацию о спектральном составе исследуемого света. Первые наблюдения такого рода были выполнены Физо в середине XIX в. В использованном им интерферометре наблюдались кольца Ньютона (см. § 5.3) при освещении его желтым светом натриевой лампы. При контакте линзы с пластинкой кольца были резкими. По мере отодвигания линзы от пластинки кольца стягиваются к центру, а видность полос убывает и при прохождении примерно 490-го кольца интерференционная картина пропадает. При дальнейшем увеличении расстояния кольца появляются вновь и приобретают приблизительно первоначальную видность при стягивании примерно 980-го кольца. Физо смог проследить периодическое изменение видности полос в 52 периодах из 980 колец каждый. Отсюда он сделал правильный вывод о том, что желтый свет натрия состоит из двух близких спектральных линий. Результаты этих опытов дают для отношения $\lambda/\delta\lambda$ у желтого дублета натрия значение, равное 980. Средняя длина волны желтой линии $\lambda=589{,}3$ нм, поэтому $\delta\lambda=0{,}6$ нм. Позднее более тщательные систематические измерения тонкой структуры спектральных линий были выполнены Майкельсоном. Впоследствии анализ спектров с помощью двухлучевой интерференции был вытеснен методами, основанными на многолучевой интерференции (см. § 5.8).

Рассмотренный пример с источником света, излучающим две близкие по частоте очень узкие спектральные линии, позволяет глубже проанализировать вопрос об ограниченности использованной ранее монохроматической идеализации и перейти к изучению характера интерференционных явлений в квазимонохроматическом свете. Как известно, спектр излучения разреженных газов или паров состоит из резких ярких линий, разделенных темными промежутками. Выделив свет одной из этих линий с помощью светофильтра или монохроматора, можно использовать его для наблюдения, например, полос равного наклона в интерферометре Майкельсона. Оказывается, что интерференционные полосы видны очень отчетливо, когда длины путей обоих интерферирующих пучков примерно одинаковы. Если отодвигать одно из зеркал так, чтобы разность хода Δ пучков увеличивалась, видность полос уменьшается (в общем случае немонотонно) и в конце концов они исчезают.

Введение разности хода между пучками эквивалентно задержке одного из них во времени, поэтому способность световых колебаний в одной точке исходного пучка к интерференции после его разделения на два пучка и последующего их соединения с некоторой разностью хода называется *временной когерентностью*. Максимальная разность хода, при которой возможна интерференция, называется *длиной когерентности излучения* $l_{\text{ког}}$, а соответствующее ей запаздывание — *временем когерентности* $\tau_{\text{ког}} = l_{\text{ког}}/c$.

Условие временной когерентности световых колебаний можно записать в виде

$$\Delta < l_{\text{ког}} = c\tau_{\text{ког}}.$$

Уменьшение когерентности световых колебаний с увеличением временной задержки, т. е. уменьшение видности интерференционных полос при возрастании разности хода, связано с конечной шириной спектральной линии источника квазимонохроматического света. Как было показано в § 1.6—1.8, такое излучение можно рассматривать как совокупность не скоррелированных между собой отдельных монохроматических волн, частоты которых сплошь заполняют некоторый интервал $\delta\omega$, малый по сравнению со средней частотой ω. Каждая монохроматическая волна из этой совокупности создает в интерферометре свою картину полос, и полное распределение освещенности, как и в приведенном выше примере, определяется простым наложением этих картин.

При малых разностях хода интерферирующих лучей (порядка нескольких длин волн) положение полос в картинах, создаваемых отдельными монохроматическими составляющими, практически одинаково. Поэтому полосы суммарной картины отчетливы.

По мере увеличения разности хода отдельные картины полос смещаются относительно друг друга из-за различия в длинах волн, и в конце концов суммарная картина окажется полностью размытой.

Оценить разность хода, при которой происходит исчезновение полос, можно следующим образом. Примем, что отдельные монохроматические компоненты *равномерно* заполняют спектральный интервал шириной $\delta\omega$ («прямоугольный» контур линии). Разделим мысленно этот интервал на множество пар бесконечно узких спектральных линий, отстоящих по частоте на $\delta\omega/2$, т. е. на $\delta k/2$ по шкале волновых чисел ($k=\omega/c$). Распределение освещенности от каждой пары дается формулой (5.17), в которой δk (расстояние между монохроматическими компонентами) следует теперь заменить на $\delta k/2$. Полосы в этой картине пропадают при такой разности хода $\Delta_{\max}=l_{\text{ког}}$, когда аргумент первого косинуса в (5.17) становится равным $\pi/2$. Заменяя δk на $\delta k/2$, находим

$$l_{\text{ког}}=2\pi/\delta k=\lambda^2/\delta\lambda. \qquad (5.20)$$

Условие исчезновения полос для всех пар монохроматических компонент, на которые был разделен спектральный интервал $\delta\omega$, одинаково. Поэтому при разности хода $\Delta=l_{\text{ког}}$ из (5.20) происходит размытие полос полной интерференционной картины.

Формулу (5.20) можно использовать для оценки длины когерентности и в случае более сложной формы контура спектральной линии квазимонохроматического света, понимая под δk (или $\delta\lambda$) полуширину контура (т. е. ширину на половине высоты). Длине когерентности соответствует максимально возможный порядок интерференции $m_{\max}=\lambda/\delta\lambda=\omega/\delta\omega$.

Для белого света (солнце, лампа накаливания, дуга с угольными электродами) и визуального наблюдения эффективный диапазон длин волн $\delta\lambda$ простирается приблизительно от 400 до 700 нм, т. е $\delta\lambda\sim\lambda$. В этом случае $m_{\max}\sim 1$ и интерференционные полосы, казалось бы, наблюдаться не должны. Действительно, приемник излучения, обладающий примерно одинаковой чувствительностью в различных участках спектра, например термоэлемент, покажет при перемещении в поле зрения поперек полос почти равномерное распределение освещенности. Но глаз представляет собой селективный приемник с сильно изменяющейся чувствительностью в зависимости от длины волны (см. кривую видности на форзаце), что дает некоторым длинам волн преимущество перед другими. Визуальное наблюдение полос в белом свете облегчается и благодаря способности нашего зрения различать цвет, а не только интенсивность света. Поэтому в белом свете глаз различает около десятка окрашенных интерференционных полос. При равной нулю разности хода, т. е. в тех местах, куда обе интерферирующие волны приходят в одинаковых фазах, условие максимума выполняется для всех длин волн. В этом месте получается ахроматическая (т. е. неокрашенная) светлая полоса. По обе стороны от нее находятся окрашенные максимумы и минимумы, а за ними поле зрения кажется глазу равномерно освещенным белым светом. Таково происхождение красивых интерференционных цветов в тонких пленках масла или бензина на поверхности воды.

Легко оценить максимальную толщину пленки, при которой возможно визуальное наблюдение интерференции в белом свете. Принимая $m_{max} \approx 10$, для максимально достижимой разности хода получаем $\sim 10\bar{\lambda}$ (где $\bar{\lambda} \approx 0{,}5$ мкм). При этом максимальная толщина примерно вдвое меньше: $h_{max} \approx 5\bar{\lambda} \approx 2$ мкм.

Картина интерференционных полос в белом свете иногда может оказаться полезной при интерферометрических измерениях, так как с ее помощью можно обнаружить для монохроматического света полосу нулевого порядка, соответствующую равной нулю разности хода.

Используя излучение узких спектральных линий газоразрядных источников низкого давления, можно наблюдать интерференцию при разностях хода, достигающих нескольких десятков сантиметров. Основная причина, ограничивающая в этом случае длину когерентности, — это хаотическое тепловое движение излучающих атомов, приводящее к доплеровскому уширению спектральных линий (см. § 1.8). Так, например, для излучения красной линии кадмия $\lambda = 643{,}85$ нм, впервые исследованного Майкельсоном, длина когерентности около 20 см. Для оранжевой линии $\lambda = 605{,}78$ нм стандартной криптоновой лампы, используемой в современном эталоне длины (по определению, 1 м равен 1 650 763,73 длин волн этой линии), длина когерентности достигает 0,8 м.

Исключительно высокая монохроматичность излучения стабилизированных по частоте газовых лазеров позволяет получить интерференцию при разности хода в несколько километров. Предельная разность хода, при которой возможно наблюдение интерференции, на практике ограничивается не длиной когерентности лазерного излучения, а трудностями создания стабильной интерференционной установки подобных размеров и неоднородностью земной атмосферы.

При количественном исследовании значения немонохроматичности источника света в интерференционных явлениях нужно найти, как кривая видности полос связана со спектральным распределением интенсивности источника. Пусть это распределение характеризуется некоторой функцией $I_1(k)$, так что $I_1(k)\mathrm{d}k$ — интенсивность одного из интерферирующих пучков в интервале волновых чисел $(k,\ k+\mathrm{d}k)$. Распределение интенсивности в элементарной интерференционной картине, созданной излучением из этого спектрального интервала, получается из (5.15) заменой I_0 на $I_1(k)\mathrm{d}k$, а полная интенсивность результирующей картины находится суммированием по всему спектру:

$$I(\Delta) = 2\int I_1(k)\,(1+\cos k\Delta)\mathrm{d}k. \qquad (5.21)$$

Для узкой спектральной линии $I_1(k)$ быстро убывает при удалении от некоторого среднего волнового числа k_0 — центра спектральной линии. В (5.21) удобно перенести начало отсчета для k в центр линии, т. е. ввести переменную $x = k - k_0$. Подставляя $k = k_0 + x$ в (5.21), получаем

$$I(\Delta) = 2\int f(x)[1 + \cos(k_0 + x)\Delta]dx = 2\int f(x)\,dx + \tag{5.22}$$
$$+ 2\int f(x)\cos(x\Delta)dx \cdot \cos k_0\Delta - 2\int f(x)\sin(x\Delta)\,dx \cdot \sin k_0\Delta,$$

где $f(x) = I_1(k_0 + x)$.

Будем считать, что контур спектральной линии симметричен относительно центра, т. е. $f(x)$ — четная функция от x. Тогда последний интеграл в (5.22) равен нулю. Второй интеграл, если его рассматривать как функцию от Δ, изменяется медленно по сравнению с быстро осциллирующим множителем $\cos k_0\Delta$, так как для узкой спектральной линии $f(x)$ равно нулю всюду, за исключением малой окрестности центра (т. е. вклад в интеграл дают только значения x, для которых $|x| \ll k_0$). Вводя обозначения

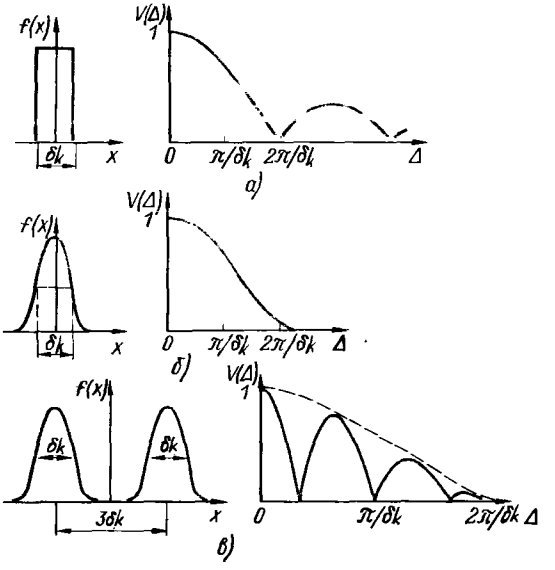
$$I_0 = \int f(x)dx,\ C(\Delta) = \int f(x)\cos(x\Delta)\,dx/I_0 \tag{5.23}$$

(I_0 — полная интенсивность одного из интерферирующих пучков), для полной интенсивности из (5.22) получаем

$$I(\Delta) = 2I_0[1 + C(\Delta)\cos k_0\Delta]. \tag{5.24}$$

Учитывая медленность изменения $C(\Delta)$, можно считать, что экстремумы $I(\Delta)$ (5.24) будут при тех значениях Δ, где $\cos k_0\Delta = \pm 1$, поэтому $I_{экстр} = 2I_0[1 \pm C(\Delta)]$. Следовательно, зависимость видности от разности хода Δ описывается выражением

$$V = (I_{max} - I_{min})/(I_{max} + I_{min}) = |C(\Delta)|. \tag{5.25}$$



5.14

Кривые видности, соответствующие различным спектральным распределениям интенсивности источника

В случае прямоугольного спектрального контура, т. е. равномерного распределения в интервале δk вблизи k_0, функция $f(x)=\text{const}$ при $-\delta k/2<x<\delta k/2$ и $f(x)=0$ вне этого интервала. Тогда

$$C(\Delta)=\frac{1}{\delta k}\int\limits_{-\delta k/2}^{\delta k/2}\cos(x\Delta)\mathrm{d}x=\frac{\sin(\delta k\cdot\Delta/2)}{\delta k\cdot\Delta/2}, \tag{5.26}$$

а видность интерференционных полос в соответствии с (5.25) определяется модулем этой функции. Кривая видности, соответствующая равномерному распределению в спектральном интервале шириной δk, приведена на рис. 5.14, *а*. При $\Delta=0$ видность максимальна: $V=1$. С увеличением разности хода видность полос убывает и при $\Delta=2\pi/\delta k=\lambda^2/\delta\lambda$ обращается в нуль. Этот результат уже был получен ранее [см. (5.20)]. С дальнейшим увеличением Δ интерференционные полосы появляются вновь, но их видность незначительна (рис. 5.14, *а*).

В случае отдельной спектральной линии газоразрядного источника, уширенной вследствие эффекта Доплера, форма контура описывается функцией Гаусса $f(x)\sim\exp(-\alpha^2x^2)$. Для нахождения видности (5.25) нужно рассчитать значение $C(\Delta)$, определяемое формулой (5.23). Вычисляя соответствующий интеграл (см. задачу 2), получаем $V(\Delta)=\exp\{-[\Delta/(2\alpha)]^2\}$. С увеличением разности хода видность полос монотонно убывает (рис. 5.14, *б*) и полосы практически исчезают при $\Delta\approx 2\pi/\delta k$, где $\delta k=\sqrt{\ln 2}/\alpha$ — ширина спектрального контура на половине высоты. Именно такую кривую видности получил Майкельсон при исследовании красной линии кадмия.

Когда спектр состоит из двух одинаковых компонент гауссовской формы с шириной δk, отделенных друг от друга интервалом в несколько δk (рис. 5.14, *в*), периодические изменения видности полос, обусловленные наложением интерференционных картин от разных компонент (ср. с рис. 5.13), оказываются промодулированными монотонно убывающей огибающей, обусловленной наложением картин от разных монохроматических составляющих каждой компоненты (ср. с рис. 5.14, *б*). Сравнение теоретически рассчитанных кривых видности с экспериментальными позволяет установить спектральное распределение интенсивности исследуемого источника света.

В § 1.8 была рассмотрена статистическая модель излучения макроскопического источника света, содержащего большое число атомов — независимых элементарных излучателей. Свет такого источника представляет собой хаотическую последовательность отдельных волновых цугов конечной длительности. Когда цуги волн, испускаемые разными атомами в случайные моменты времени, одинаковы, спектральное распределение интенсивности излучения будет таким же, как и у отдельного цуга (однородное уширение спектральной линии). Связь между длительностью τ волнового цуга и шириной $\delta\omega$ соответствующего ему спектрального распределения обсуж-

далась в § 1.7. Разным формам огибающей волнового цуга соответствуют и разные спектральные контуры, но соотношение между $\delta\omega$ и τ имеет универсальный характер [см. (1.89)]:

$$\delta\omega \cdot \tau \approx 2\pi \text{ или } \delta k \cdot l \approx 2\pi, \tag{5.27}$$

где $\delta k = \delta\omega/c$, а $l = c\tau$ — протяженность волнового цуга в пространстве. Чем больше протяженность l цуга, тем уже соответствующий ему спектральный интервал $\delta\omega$.

Исчезновение полос в интерференционных опытах при увеличении разности хода легко объяснить на основе этой модели. Каждый волновой цуг в интерференционном опыте делится на два цуга одинаковой протяженности, которые затем по разным путям приходят в точку наблюдения. Если оптическая разность длин этих путей превышает протяженность цуга, то один из цугов минует точку наблюдения раньше, чем другой дойдет до нее. Тем самым интерференция двух цугов, образовавшихся из одного, становится невозможной. В точке наблюдения идет наложение цугов, порожденных разными цугами в излучении источника. Результат будет таким же, как при наложении волновых цугов от разных источников: за время наблюдения проходит большое число цугов, фазы колебаний в которых никак не связаны друг с другом, поэтому интерференционный член в среднем обращается в нуль и происходит просто сложение интенсивностей.

Отсюда вытекает, что максимальная разность хода, при которой возможна интерференция, т. е. длина когерентности, — это и есть длина l волнового цуга. Выражая длину волнового цуга через ширину δk спектрального распределения интенсивности с помощью (5.27), получаем для длины когерентности $l_{\text{ког}} = 2\pi/\delta k$, что совпадает с выражением (5.20), полученным как результат наложения интерференционных картин, создаваемых отдельными монохроматическими компонентами в спектральном распределении излучения источника.

Конечно, рассмотренный пример, в котором все волновые цуги одинаковы, дает лишь идеализированное представление об излучении реальных источников. Тепловое движение излучающих атомов приводит вследствие эффекта Доплера к различию средних частот, сопоставляемых отдельным цугам. Во многих случаях такое неоднородное уширение определяет форму и ширину спектральных линий. Поэтому нельзя ожидать, что для излучения реальных источников квазимонохроматического света понятие длины когерентности можно столь просто и наглядно трактовать в буквальном смысле как протяженность волновых цугов. Однако для любого излучения, занимающего спектральный диапазон δk, длину когерентности $l = 2\pi/\delta k$ всегда можно принять за некоторую *эффективную* протяженность волновых цугов. При таком подходе две возможные интерпретации исчезновения интерференционных полос при больших разностях хода — в рамках представлений о хаотической последовательности волновых цугов конечной протяженности или представ-

лений о наложении интерференционных картин, создаваемых отдельными монохроматическими компонентами в спектре излучения, — оказываются эквивалентными.

Количественная теория когерентных свойств излучения, т. е. его способности к интерференции, в рамках спектральных представлений была рассмотрена выше. Теория, в которой используется временно́е описание излучения, базируется на математической теории *случайных функций*. Колебания напряженности электрического поля в световой волне, представляющей собой наложенне волновых цугов, можно записать в виде

$$E(t)=E_0(t)e^{-i\omega t}=a(t)e^{-i[\omega t-\varphi(t)]}, \tag{5.28}$$

где ω — некоторая средняя частота, а $E_0(t)=a(t)\exp[i\varphi(t)]$ — зависящая от времени комплексная амплитуда. В случае квазимонохроматического излучения, частотный спектр которого сосредоточен в узком интервале $\delta\omega$ вблизи средней частоты ω, т. е. когда $\delta\omega\ll\omega$, амплитуда $a(t)$ и фаза $\varphi(t)$ в (5.28) представляют собой случайные функции, которые изменяются *медленно* по сравнению с функцией $\exp(-i\omega t)$. Другими словами, можно считать, что (5.28) описывает *хаотическую модуляцию* амплитуды и фазы колебания с частотой ω. Следует, однако, заметить, что хотя изменения $a(t)$ и $\varphi(t)$ происходят медленно в масштабе периода колебаний, они происходят чрезвычайно быстро в масштабе времени, требуемого для наблюдения. Измеряемая на опыте интенсивиость пропорциональиа среднему за время иаблюдения зиачению квадрата вещественной амплитуды $a(t)$:

$$I\sim\langle a^2(t)\rangle=\langle E_0(t)E_0^*(t)\rangle. \tag{5.29}$$

Хаотические изменения $a^2(t)$ сглаживаются при усреднении за время наблюдения. В случае стациоиарных оптичесиих полей интенсивиость в (5.29) не зависит от времени (точнее, от положеиия иитервала усреднения на шкале времени).

В иитерференционных экспериментах пучок квазимонохроматического света расщепляется на два, которые затем вновь встречаются в иекоторой точке наблюдения P. Будем для простоты считать, что иитеисивности этих пучков одинаковы. Поскольку оптические пути пучков от места разделения до точки P различаются на Δ, колебания в одном из них происходят с запаздыванием на время $\tau=\Delta/c$. Поэтому результирующее колебание в P описывается функцией

$$E(t)+E(t-\tau)=[E_0(t)+E_0(t-\tau)e^{i\omega\tau}]e^{-i\omega t}.$$

В соответствии с (5.29) интенсивность этого колебания

$$I\sim\langle[E_0(t)+E_0(t-\tau)e^{i\omega\tau}][E_0^*(t)+E_0^*(t-\tau)e^{-i\omega\tau}]\rangle=\langle E_0(t)E_0^*(t)\rangle+$$
$$+\langle E_0(t-\tau)E_0^*(t-\tau)\rangle+2\mathrm{Re}[\langle E_0(t)E_0^*(t-\tau)\rangle e^{-i\omega\tau}]. \tag{5.30}$$

Первые два слагаемых в (5.30) одинаковы и пропорциональны интенсивностям I_0 интерферирующих пучков. Выражение $\langle E_0(t)E_0^*(t-\tau)\rangle$ называют в теории случайных функций *автокорреляционной функцией* для комплексной амплитуды $E_0(t)$. Нормированную автокорреляциониую функцию называют *комплексной степенью временно́й когерентности* колебаний и обозначают $\gamma(\tau)$:

$$\gamma(\tau)=\langle E_0(t)E_0^*(t-\tau)\rangle/\langle E_0(t)E_0^*(t)\rangle. \tag{5.31}$$

Эта функция характеризует корреляцию световых колебаний в моменты времеии, разделеиные интервалом τ^*. Используя определение $\gamma(\tau)$ (5.31), из (5.30) получаем выражеиие для интенсивиости в точке наблюдеиия:

$$I=2I_0[1+\mathrm{Re}\gamma(\tau)e^{-i\omega\tau}]. \tag{5.32}$$

---

* Так как функция $\gamma(\tau)$ характеризует в то же время корреляцию колебаний в один момеит времени в двух точках, находящихся на расстоянии $\Delta=c\tau$ *вдоль* пучка, то *временну́ю* когерентиость иногда называют *продольной*.

Если комплексную функцию $\gamma(\tau)$ представить как $\gamma(\tau)=|\gamma(\tau)|e^{-i\delta(\tau)}$, то формуле (5.32) можно придать вид

$$I=2I_0[1+|\gamma(\tau)|\cos(\omega\tau+\delta)]=2I_0[1+|\gamma(\tau)|\cos(k\Delta+\delta)]. \quad (5.33)$$

Это выражение отличается от формулы (5.8) для интенсивности при интерференции монохроматических волн наличием множителя $|\gamma(\tau)|$ в интерференционном члене и добавочным слагаемым $\delta(\tau)$ в аргументе косинуса. Зависящий от положения точки наблюдения P множитель $\cos(k\Delta+\delta)$ в интерференционном члене описывает быстрые осцилляции интенсивности в пространстве при переходе от одной полосы к другой. Изменение плавной функции $|\gamma(\tau)|$ при переходе от одной полосы к соседней незначительно, т. е. она имеет приблизительно одно и то же значение для целой области интерференционного поля, содержащей много полос.

Когда $|\gamma(\tau)|=1$, интерференция квазимонохроматического света с хаотически изменяющимся амплитудой и фазой осуществляется так же, как и в случае регулярных строго монохроматических волн. Поэтому при $|\gamma(\tau)|=1$ говорят о *полной когерентности* интерферирующих пучков. При $\gamma(\tau)=0$ происходит простое сложение интенсивностей пучков: $I=2I_0$. В этом случае интерференции нет и колебания называют *некогерентными*. Если $0<|\gamma(\tau)|<1$, то говорят о *частичной когерентности* интерферирующих пучков. Можно представить себе частично когерентный свет как бы состоящим из полностью когерентной и некогерентной частей, причем доля когерентного света в этой смеси равна $|\gamma(\tau)|$. В самом деле, формулу (5.33) можно записать в виде

$$I=2I_0|\gamma(\tau)|[1+\cos(k\Delta+\delta)]+2I_0(1-|\gamma(\tau)|).$$

Здесь первое слагаемое описывает интерференцию полностью когерентных волн с равными интенсивностями $I_0|\gamma(\tau)|$, а второе — наложение некогерентных волн с интенсивностями $I_0(1-|\gamma(\tau)|)$. Этим объясняется, почему величина $|\gamma(\tau)|$ называна степенью когерентности.

Экспериментальное определение функции $|\gamma(\tau)|$ для исследуемого излучения может быть основано на измерении видности $V(\Delta)$ интерференционных полос в зависимости от разности хода $\Delta=c\tau$. Ввиду медленности изменения $|\gamma(\tau)|$ в (5.33) максимумы интенсивности соответствуют значениям $\cos(k\Delta+\delta)=1$, а минимумы — значениям $\cos(k\Delta+\delta)=-1$. Поэтому

$$I_{max}=2I_0(1+|\gamma(\tau)|),\quad I_{min}=2I_0(1-|\gamma(\tau)|).$$

Таким образом, видность интерференционных полос (при равных интенсивностях интерферирующих пучков)

$$V(\Delta)=(I_{max}-I_{min})/(I_{max}+I_{min})=|\gamma(\tau)|, \quad (5.34)$$

т. е. равна модулю комплексной степени когерентности. Кривые видности $V(\Delta)$, примеры которых даны на рис. 5.14, можно рассматривать и как графики функции $|\gamma(\tau)|$ для соответствующих излучений, заменив Δ на $c\tau$. Значительно труднее определить на опыте аргумент $\gamma(\tau)$, т. е. функцию $\delta(\tau)$. Для этого нужно, как видно из (5.33), при каждом значении разности хода Δ сравнить наблюдаемое положение полос от исследуемого источника света с положением полос от монохроматического источника с частотой ω, которое определяется формулой (5.8). Таким способом в принципе можно экспериментально полностью определить комплексную степень временной когерентности, характеризующую исследуемое излучение.

Привлекательная черта теории частичной когерентности состоит в том, что она оперирует величинами, которые в принципе можно определить из эксперимента (корреляционные функции и интенсивности). В этом она существенно отличается от элементарной оптической волновой теории, где основную величину, т. е. $E(t)$, из-за большой частоты оптических колебаний невозможно измерить реальными инерционными приемниками излучения.

Комплексную степень когерентности $\gamma(\tau)$ можно найти, используя ту или иную статистическую модель рассматриваемого излучения. Например, в простейшем примере, когда квазимонохроматическое излучение представляется в виде хаотического наложения оди-

наковых волновых цугов длительностью τ_0, прямой расчет на основе определения $\gamma(\tau)$ (5.31) дает

$$\gamma(\tau)=\begin{cases}1-\tau/\tau_0 & \text{при } \tau<\tau_0,\\ 0 & \text{при } \tau>\tau_0.\end{cases} \tag{5.35}$$

Степень когерентности линейно убывает от 1 до 0 с увеличением τ от 0 до τ_0. Формула (5.35) отражает физически очевидный факт: колебания в моменты времени t и $t+\tau$ когерентны, если промежуток времени τ меньше длительности τ_0 отдельного цуга. Другими словами, в такой модели излучения длительность τ_0 цуга совпадает с временем когерентности колебаний.

Найдем связь комплексной степени когерентности $\gamma(\tau)$ со спектральным распределением интенсивности излучения $I(\omega)$. Для этого воспользуемся разложением квазимонохроматического колебания $E(t)=E_0(t)e^{-i\omega t}$ (5.28) в интеграл Фурье [см. (1.83) и 1.84)]:

$$E(t)=\int\limits_{-\infty}^{\infty}E_\omega e^{-i\omega t}\frac{d\omega}{2\pi},\quad E_\omega=\int\limits_{-\infty}^{\infty}E(t)\,e^{i\omega t}dt. \tag{5.36}$$

Выражая в определении $\gamma(\tau)$ [см. (5.31)] комплексную амплитуду $E_0(t)$ через $E(t)$ с помощью (5.28), приводим корреляционную функцию для амплитуд к виду

$$\langle E_0(t)E_0^*(t-\tau)\rangle=\langle E(t)E^*(t-\tau)\rangle e^{i\omega\tau}=\int\limits_{-\infty}^{\infty}E(t)E^*(t-\tau)dt\,e^{i\omega\tau}.$$

При усреднении по времени наблюдения, большому по сравнению с характерным временным масштабом изменения $E_0(t)$, пределы интегрирования по t можно распространить до $\pm\infty$. Затем вместо $E^*(t-\tau)$ подставляем его разложение в интеграл Фурье согласно (5.36) и изменяем порядок интегрирования по t и ω':

$$\langle E(t)E^*(t-\tau)\rangle=\int\limits_{-\infty}^{\infty}E(t)\left[\int\limits_{-\infty}^{\infty}E_{\omega'}^*e^{i\omega'(t-\tau)}\frac{d\omega'}{2\pi}\right]dt=$$

$$=\int\limits_{-\infty}^{\infty}E_{\omega'}^*\left(\int\limits_{-\infty}^{\infty}E(t)e^{i\omega' t}dt\right)e^{-i\omega'\tau}\frac{d\omega'}{2\pi}.$$

Стоящий в скобках интеграл согласно второй из формул (5.36) есть фурье-компонента $E_{\omega'}$ функции $E(t)$. Поэтому

$$\langle E_0(t)E_0^*(t-\tau)\rangle=2\int\limits_0^{\infty}E_{\omega'}^*E_{\omega'}\,e^{-i\omega'\tau}\frac{d\omega'}{2\pi}e^{i\omega\tau}. \tag{5.37}$$

Знаменатель в определении $\gamma(\tau)$ (5.31) получается из (5.37) при $\tau=0$. Поэтому

$$\gamma(\tau)=e^{i\omega\tau}\int\limits_0^{\infty}|E_{\omega'}|^2e^{-i\omega'\tau}d\omega'\Big/\int\limits_0^{\infty}|E_{\omega'}|^2d\omega'. \tag{5.38}$$

Понимая под $I(\omega)$ нормированную на единицу функцию спектрального распределения интенсивности

$$I(\omega)=|E_\omega|^2\Big/\int\limits_0^{\infty}|E_\omega|^2d\omega,\quad\int\limits_0^{\infty}I(\omega)d\omega=1,$$

для $\gamma(\tau)$ получаем окончательное выражение:

$$\gamma(\tau)=e^{i\omega\tau}\int\limits_0^{\infty}I(\omega')\,e^{-i\omega'\tau}d\omega'. \tag{5.39}$$

Проиллюстрируем применение формулы (5.39) на простых примерах. Пусть излучение равномерно заполняет узкий спектральный интервал $\delta\omega$ со средней частотой ω (прямоугольный спектральный контур). Тогда $I(\omega')$ равно $1/\delta\omega$ в пределах этого интервала $(\omega-\delta\omega/2,\ \omega+\delta\omega/2)$ и нулю — вне его. При вычислении интеграла в (5.39) удобно перейти к переменной $x=\omega'-\omega$:

$$\gamma(\tau)=e^{i\omega\tau}\frac{1}{\delta\omega}\int_{-\delta\omega/2}^{\delta\omega/2}e^{-i(\omega+x)\tau}dx=\frac{\sin(\tau\delta\omega/2)}{\tau\delta\omega/2}. \qquad (5.40)$$

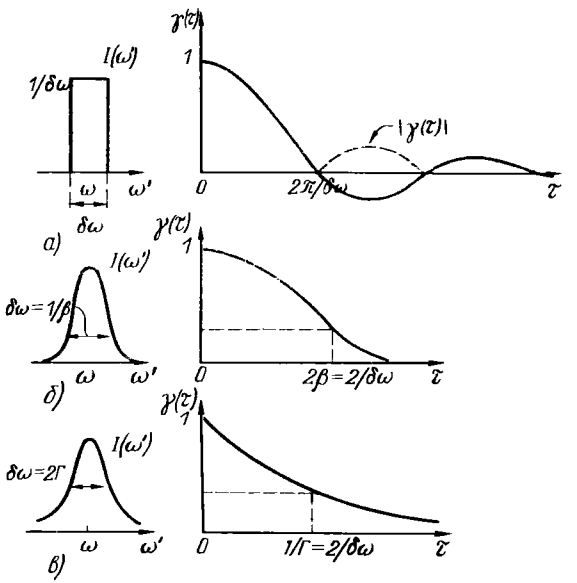
График этой функции дан на рис. 5.15, *а*. Модуль $\gamma(\tau)$, равный в соответствии с (5.34) видности интерференционных полос, показан штриховой линией. Сравните этот график с соответствующей кривой видности на рис. 5.14, *а*.

Для излучения с гауссовым спектральным контуром (доплеровское уширение) нормированная функция распределения интенсивности имеет вид $I(\omega')=(\beta/\sqrt{\pi})\times\times\exp[-\beta^2(\omega'-\omega)^2]$. Перейдем в (5.39) к переменной $x=\omega'-\omega$. Так как $I(\omega')$ быстро убывает при удалении от ω', то пределы интегрирования по x можно распространить до $\pm\infty$:

$$\gamma(\tau)=e^{i\omega\tau}\frac{\beta}{\sqrt{\pi}}\int_{\infty}^{\infty}e^{-\beta^2x^2-i(\omega+x)\tau}dx=e^{-[\tau/(2\beta)]^2} \qquad (5.41)$$

(вычисление этого интеграла см. в задаче 2). График $\gamma(\tau)$ также представляет собой гауссову кривую (рис. 5.15, *б*). Степень когерентности монотонно убывает с увеличением промежутка времени τ. Сравните график $\gamma(\tau)$ с соответствующей кривой видности на рис. 5.14, *б*.

В случае спектральной линии с лоренцевским контуром полушириной $\delta\omega=2\Gamma$ степень когерентности $\gamma(\tau)=\exp(-\Gamma|\tau|)$ (см. задачу 3). Ее график приведен на рис. 5.15, *в*. Во всех случаях время когерентности $\tau_{\text{ког}}$ связано с шириной спектра излучения соотношением $\tau_{\text{ког}}\delta\omega\approx 2\pi$.



5.15

Степень когерентности $|\gamma(\tau)|$, соответствующая различным спектральным распределениям интенсивности излучения

Интегральное соотношение (5.39), выражающее $\gamma(\tau)$ через $I(\omega)$, можно обратить, чтобы по известной степени когерентности $\gamma(\tau)$ (например, измеренной экспериментально) рассчитывать спектральное распределение интенсивности $I(\omega)$ исследуемого излучения.

Таким образом, экспериментальное определение времени когерентности τ (или длины когерентности $l=c\tau$) может быть основано либо на измерении видности интерференционных полос, либо на измерении спектрального интервала частот $\Delta\nu \sim 1/\tau$, занимаемого исследуемым излучением.

Принципиально иной метод нахождения времени когерентности был предложен Брауном и Твиссом в 1956 г. Его упрощенная схема показана на рис. 5.16, *а*. Излучение от исследуемого источника S делится полупрозрачным зеркалом на два пучка, которые после прохождения разных расстояний падают на два малоинерционных приемника (фотоумножители ФЭУ-1 и ФЭУ-2). В отличие от обычных интерференционных опытов здесь независимо регистрируется интенсивность *каждого* из пучков, а не интенсивность результата их наложения, т. е. интерференционной картины. Для полного анализа необходимо учитывать квантовую природу фотоэффекта (статистику фотоотсчетов), но понять принцип метода можно, считая, что в идеальных условиях эксперимента сила тока на выходе каждого ФЭУ пропорциональна мгновенной интенсивности* $I(t)$ падающего света, а флуктуация этого тока пропорциональна $\Delta I(t)=I(t)-\langle I(t)\rangle$. Силы фототоков $I_1(t-\tau)$ и $I_2(t)$ перемножаются специальной электронной схемой в корреляторе C, и произведение усредняется за большой промежуток времени. Перемещением одного из приемников можно изменять временную задержку τ между пучками и получить в результате измерений корреляционную функцию интенсивностей:

$$G(\tau)=\langle I_1(t-\tau)I_2(t)\rangle/I^2=1+\langle\Delta I_1(t-\tau)\Delta I_2(t)\rangle/I^2.$$

(Предполагается, что пучки имеют одинаковые средние интенсивности: $\langle I_1(t)\rangle=\langle I_2(t)\rangle\equiv I$.)

Для пучков строго постоянной интенсивности $\Delta I_1=\Delta I_2=0$ и $G(\tau)=1$. Но для квазимонохроматического света «мгновенная» интенсивность, пропорциональная квадрату медленно изменяющейся амплитуды $a(t)$ в (5.28), представляет собой случайную функцию времени. При $\tau=0$ (детекторы на одинаковом расстоянии от разделительного зеркала) изменения интенсивности света на обоих ФЭУ происходят одинаково: $I_1(t)=I_2(t)\equiv I(t)$. Поэтому $G(0)=\langle I^2(t)\rangle/I^2=1+\langle(\Delta I)^2\rangle/I^2>1$. С увеличением времени задержки τ корреляция между изменениями интенсивностей на приемниках уменьшается и исчезает совсем, когда τ превосходит время когерентности, т. е. эффективную длительность волновых цугов исследуемого излучения: $\langle\Delta I_1(t-\tau)\Delta I_2(t)\rangle\to 0$ при $\tau>\tau_{\text{ког}}$ и $G(\tau)\to 1$. Характерный вид зависи-

* Точнее, усредненной за время, большое по сравнению с периодом оптических колебаний, но малое по сравнению с характерным временным масштабом изменения амплитуды $a(t)$ квазимонохроматического света [см. (5.18)].

мости $G(\tau)$ показан на рис. 5.16, *б*.

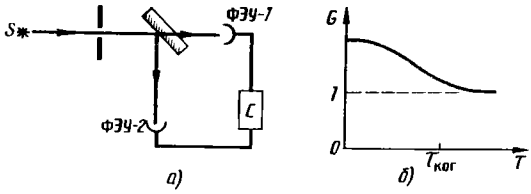


Схема опыта для измерения корреляционной функции интенсивностей $G(\tau)$ (*а*) и график $G(\tau)$ (*б*)

Функцию $G(\tau)$ можно рассчитать, если выбрать определенную статистическую модель излучения. Для типичных источников квазимонохроматического света, атомы которых излучают независимо друг от друга, флуктуации интенсивности $\langle(\Delta I)^2\rangle^{1/2}$ порядка самой интенсивности I и $G(0)\approx 2$. Это легко понять в рамках модели излучения в виде хаотической последовательности одинаковых волновых цугов. Пусть n — число атомов источника, волновые цуги от которых достигают в данный момент приемника. Сложение n колебаний одинаковой амплитуды a_1 со случайными фазами дает колебание, квадрат амплитуды которого $a^2 = na_1^2$, поэтому $I \sim a^2 = na_1^2$ и $\langle I\rangle \sim \langle n\rangle a_1^2$. Для полностью независимых атомов флуктуация числа n атомов, излучающих в данных момент, равна $\langle(\Delta n)^2\rangle^{1/2} = \langle(n - \langle n\rangle)^2\rangle^{1/2} = \langle n\rangle$, откуда и $\langle(\Delta I)^2\rangle^{1/2} = \langle(\Delta n)^2\rangle^{1/2} a_1^2 = \langle n\rangle a_1^2 = I$.

Минимальное время когерентности, которое удается измерить данным способом (он называется *интерферометрией интенсивностей*), определяется быстродействием элементов установки и может быть порядка 1—10 нс.

Контрольные вопросы

Как определена функция видности интерференционных полос?
Дайте качественное объяснение кривой видности в случае источника, спектр излучения которого состоит из двух близких линий (см. рис. 5.13).
Какую величину называют длиной когерентности? Чему равна длина когерентности для квазимонохроматического излучения, занимающего спектральный интервал шириной $\delta\lambda$ со средним значением длины волны λ?
Почему для наблюдения полос равной толщины в белом свете пленка (или пластинка) должна быть очень тонкой?
Каким образом из наблюдения полос двухлучевой интерференционной картины можно получить информацию о спектральном составе излучения?
Сопоставьте спектральный и временной подходы к объяснению исчезновения полос в квазимонохроматическом свете при большой разности хода.
Что называется степенью временной когерентности колебаний? В каком случае говорят о частичной когерентности интерферирующих пучков? Как степень когерентности связана с видностью интерференционных полос?
Как найти степень когерентности, если известен спектральный состав излучения?
Как измеряется время когерентности исследуемого излучения в методе интерферометрии интенсивности?

Задачи

1. Построить кривую видности $V(\Delta)$ интерференционных полос, когда спектр излучения источника состоит из двух очень узких близких спектральных линий ($\delta k \ll k$), интенсивность одной из которых вдвое больше интенсивности другой.

◇ Ответ: $V(\Delta)=\sqrt{8\cos^2(\delta k\Delta/2)+1}/3$. В минимумах видность принимает значение $V=1/3$. Кривая $V(\Delta)$ приведена на рис. 5.13, *в*. ◇

2. Найти видность $V(\Delta)$ интерференционных полос в случае источника, контур спектральной линии излучения которого имеет гауссову форму $I_1(k)=\text{const}\cdot\exp[-\alpha^2(k-k_0)^2]$ (неоднородно уширенная спектральная линия).

◇ Воспользуемся формулами (5.25) и (5.23), где в качестве $f(x)=I_1(k_0+x)$ нужно взять $\text{const}\cdot e^{-\alpha^2x^2}$. Тогда

$$V(\Delta)=\int\limits_{-\infty}^{\infty} e^{-\alpha^2x^2}\cos(x\Delta)\,dx \Big/ \int\limits_{-\infty}^{\infty} e^{-\alpha^2x^2}dx = \operatorname{Re}\int\limits_{-\infty}^{\infty} e^{-\alpha^2x^2+ix\Delta}dx \Big/ \int\limits_{-\infty}^{\infty} e^{-\alpha^2x^2}dx.$$

Показатель экспоненты в интеграле, стоящем в числителе, преобразуем к виду $-[\alpha x-i\Delta/(2\alpha)]^2-[\Delta/(2\alpha)]^2$ и множитель $e^{-[\Delta/(2\alpha)]^2}$ вынесем за знак интеграла. Тогда интегралы в числителе и знаменателе сокращаются. В результате получаем $V(\Delta)=e^{-[\Delta/(2\alpha)]^2}$. Кривая $V(\Delta)$ приведена на рис. 5.14, *б*. ◇

3. Найти степень когерентности $\gamma(\tau)$ для излучения, спектральный контур которого имеет лоренцевскую форму с полушириной 2Γ:

$$I(\omega')=\frac{\Gamma/\pi}{(\omega'-\omega)^2+\Gamma^2}$$

(однородно уширенная спектральная линия).

◇ Вычисление $\gamma(\tau)$ производится по формуле (5.39). При интегрировании по ω' удобно перейти к новой переменной $x=\omega'-\omega$:

$$\gamma(\tau)=\frac{\Gamma}{\pi}\int\limits_{-\infty}^{\infty}\frac{e^{-i\tau x}}{x^2+\Gamma^2}dx=e^{-\Gamma|\tau|}.$$

График $\gamma(\tau)$ приведен на рис. 5.15, *в*. ◇

5.5. Роль конечных размеров источника света. Пространственная когерентность

При анализе интерференционных опытов в § 5.4 первичный источник предполагался точечным. Однако все реальные источники света имеют конечные размеры. Увеличение размеров источника, как и расширение спектра излучаемого им света, приводит к ухудшению контрастности (уменьшению видности) интерференционных полос и даже к полному их исчезновению. Чтобы выяснить роль только первого из этих факторов, будем считать здесь излучение монохроматическим.

Протяженный самосветящийся источник состоит из большого числа точечных взаимно некогерентных элементов. Поэтому интенсивность в любом месте равна сумме интенсивностей в интерференционных картинах, создаваемых отдельными точечными элементами источника.

В интерференционных опытах по методу деления волнового фронта (см. § 5.2) полосы на экране перпендикулярны плоскости, в кото-

рой находятся первичный точечный источник S и вторичные источники S_1 и S_2. Использование вместо S линейного источника, т. е. достаточно узкой щели, вытянутой перпендикулярно этой плоскости, увеличивает интенсивность и не приводит к ухудшению четкости интерференционных полос. Это уже отмечалось в § 5.2. Однако увеличение ширины щели приводит к тому, что полосы становятся менее четкими или пропадают совсем.

Если размеры источника (т. е. ширина щели S) много меньше длины световой волны, то интерференционная картина будет резкой, так как разность хода интерферирующих лучей от любой точки источника до некоторой точки наблюдения P будет практически одна и та же. Но обычно размеры источника значительно больше длины волны, поэтому одинаковые интерференционные картины от разных его элементов сдвинуты одна относительно другой. В результате наложения этих картин интерференционные полосы оказываются более или менее размытыми. Их можно наблюдать лишь при выполнении определенных условий, налагаемых не геометрию эксперимента. Эти условия подробно обсуждаются ниже.

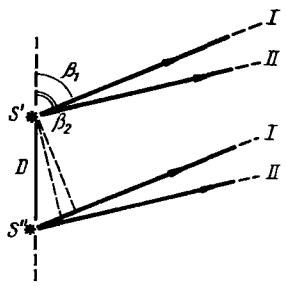
Рассмотрим сначала случай, когда источник состоит всего из двух одинаковых некогерентных светящихся точек S' и S'', находящихся на небольшом расстоянии D друг от друга (рис. 5.17). В интерференционных опытах свет от каждого источника попадает в некоторую точку наблюдения по двум различным путям. Пусть луч, идущий от S' по одному из этих путей, образует с соединяющей источники линией $S'S''$ угол β_1 (луч I), а по другому пути — угол β_2 (луч II). В опыте Юнга, например, лучи I и II идут в направлении отверстий S_1 и S_2. Если расстояние D между источниками S' и S'' достаточно мало (много меньше расстояния от них до вторичных источников S_1 и S_2), то можно считать, что аналогичные лучи, исходящие из второго источника S'', образуют с линией $S'S''$ такие же углы β_1 и β_2. Оптическая разность хода лучей $S'P(\mathrm{I})$ и $S''P(\mathrm{I})$, приходящих в точку наблюдения P от S' и S'' по первому пути, как видно из рис. 5.17, равна $|S''P(\mathrm{I})| - |S'P(\mathrm{I})| = D\cos\beta_1$. Аналогично, для лучей, приходящих в P по второму пути, можно написать $|S''P(\mathrm{II})| - |S'P(\mathrm{II})| = D\cos\beta_2$. Вычтем почленно эти равенства и перегруппируем слагаемые в левой части:

$$[|S''P(\mathrm{I})| - |S''P(\mathrm{II})|] - [|S'P(\mathrm{I})| - |S'P(\mathrm{II})|] = D(\cos\beta_1 - \cos\beta_2). \tag{5.42}$$

Здесь в первой скобке стоит разность хода Δ'' лучей, приходящих в P по двум путям от источника S''. Ее значение определяет, будет ли в P максимальная, минимальная или промежуточная интенсивность в интерференционной картине, создаваемой источником S''. Во второй скобке стоит разность хода Δ' лучей, выходящих из S'. Разность этих величин

$$\Delta'' - \Delta' = D(\cos\beta_1 - \cos\beta_2) \tag{5.43}$$

определяет сдвиг двух интерференционных картин, создаваемых



5.17
К нахождению видиости полос

источниками S' и S''. Если $\Delta''-\Delta'=0$ (или $|\Delta''-\Delta'|\ll\lambda$), то максимумы одной картины совпадают с максимумами второй. При таком точном совмещении интерференционных картин видность полос максимальна ($V=1$), так как интенсивность в минимумах равна нулю. С увеличением $|\Delta''-\Delta'|$ видность полос начинает уменьшаться и при $|\Delta''-\Delta'|=\lambda/2$ обращается в нуль, так как светлые полосы одной картины совмещаются с темными полосами другой. В соответствии с (5.43) это происходит при

$$D|\cos\beta_1-\cos\beta_2|=\lambda/2. \qquad (5.44)$$

При дальнейшем возрастании $|\Delta''-\Delta'|$ интерференционная картина появляется вновь, причем видность полос периодически изменяется. Когда $|\Delta''-\Delta'|=m\lambda$ ($m=1, 2, \ldots$), видность полос $V=1$.

Найдем зависимость видности полос от расстояния D между источниками S' и S''. В каждой из налагающихся интерференционных картин распределение интенсивности дается формулой (5.8). В результате их наложения получаем

$$I=2I_0(1+\cos k\Delta')+2I_0(1+\cos k\Delta'')=4I_0\{1+\cos[k(\Delta''-\Delta')/2]\times\cos[k(\Delta''+\Delta')/2]\}.$$

При перемещении точки наблюдения P второй сомножитель быстро осциллирует, описывая изменения интенсивности при переходе от одной полосы к другой. Светлые и темные полосы расположены в тех местах, где $\cos[k(\Delta''+\Delta')/2]=\pm1$, а интенсивность в максимумах и минимумах равна $I_{\text{экстр}}=4I_0\{1\pm|\cos[k(\Delta''-\Delta')/2]|\}$. Отсюда для видности полос суммарной картины получаем

$$V=\frac{I_{\max}-I_{\min}}{I_{\max}+I_{\min}}=\left|\cos\frac{k(\Delta''-\Delta')}{2}\right|. \qquad (5.45)$$

Пусть, например, два источника S' и S'' освещают экран с отверстиями S_1 и S_2 в установке Юнга (см. рис. 5.4). Очевидно, что $\beta_2-\beta_1=\alpha$ — угол, под которым видны отверстия S_1 и S_2 с места расположения источников. Так как $\alpha\ll1$, для $\Delta''-\Delta'$ приближенно можно написать

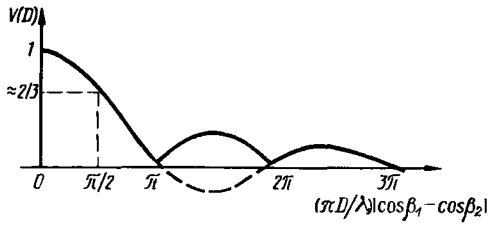
$$\Delta''-\Delta'=D(\cos\beta_1-\cos\beta_2)\approx D\sin\beta\cdot(\beta_2-\beta_1)=D\sin\beta\cdot\alpha, \qquad (5.46)$$

где $\beta=(\beta_1+\beta_2)/2$. Если линия $S'S''$ перпендикулярна оптической оси установки, то $\beta=\pi/2$ и $\Delta''-\Delta'=D\alpha$. В общем случае (при $\beta\neq\pi/2$) произведение $D\sin\beta=D_\perp$ есть проекция отрезка $|S'S''|$ на перпендикулярное оптической оси направление и $\Delta''-\Delta'=D_\perp\alpha$. Так как $\alpha=d/l$, где l — расстояние от источников до экрана с отверстиями S_1 и S_2, то $\Delta''-\Delta'=D_\perp d/l$ и из (5.45) получаем

$$V=\left|\cos\frac{kD_\perp d}{2l}\right|=\left|\cos\frac{\pi D_\perp d}{\lambda l}\right|. \qquad (5.47)$$

Зависимость видности интерференционных полос от расстояния $D_\perp$ между источниками нашла интересное применение в астрономии. У многих двойных звезд компоненты находятся на столь малом угловом расстоянии θ друг от друга, что при наблюдении в телескоп их уширенные дифракцией (см. § 7.6) изображения сливаются. Для измерения угла θ Физо в 1868 г. предложил следующий метод. Перед объективом телескопа помещается экран с двумя параллельными щелями, находящимися на расстоянии d друг от друга, которые играют ту же роль, что и щели S_1 и S_2 в опыте Юнга. Свет каждой из компонент двойной звезды создает свою картину интерференционных полос в фокальной плоскости объектива телескопа. Видность суммарных полос, возникающих при наложении этих картин, определяется формулой (5.47), в которой отношение $D_\perp/l$ и представляет собой искомое угловое расстояние θ между компонентами двойной звезды. Расстояние d между щелями можно изменять. При малом d, пока $\theta d/\lambda \ll 1$, полосы видны отчетливо, так как в (5.47) $V \approx 1$ (при условии, что звезды имеют одинаковую яркость). Если увеличивать расстояние d, налагающиеся интерференционные картины будут смещаться относительно друг друга и при некотором значении $d = d_0$ светлые полосы одной совпадут с темными полосами другой. Как видно из (5.47), первый раз такое исчезновение интерференционных полос произойдет при $\pi\theta d_0/\lambda = \pi/2$. Отсюда по измеренному значению расстояния d_0 можно определить угол θ: $\theta = \lambda/(2d_0)$. В 1920 г. Майкельсон измерил по методу Физо угловое расстояние между компонентами двойной звезды Капеллы, оказавшееся равным 0,042″. Этим методом можно даже проследить орбитальное движение звезд, так как щели на объективе должны быть ориентированы соответствующим образом в зависимости от расположения компонент двойной звезды.

Рассмотрим теперь более общий случай протяженного источника света в виде равномерно светящейся полоски (или щели) шириной D. Можно представить себе, что такой источник состоит из элементарных взаимно некогерентно излучающих полосок, расположенных перпендикулярно линии, соединяющей точки S' и S'' на рис. 5.17. Условие, при котором происходит первое исчезновение интерференционных полос по мере увеличения ширины D щели, можно получить следующим образом. Разделим мысленно всю светящуюся полоску на множество пар одинаковых элементов так, чтобы расстояние между элементами любой пары было равно половине ширины полоски, т. е. $D/2$. Если положение светлых полос интерференционной картины от одного элемента пары совпадает с положением темных полос картины от другого элемента этой пары, то интерференционные полосы от всего протяженного источника наблюдаться не будут, так как условия совпадения одинаковы для всех пар элементов. Это условие дается формулой (5.44), в которой теперь расстояние D между источниками S' и S'' нужно заменить на $D/2$ — расстояние между



5.18

Зависимость видности интерференционных полос от ширины D протяженного источника

двумя элементами одной пары протяженного источника:

$$D|\cos\beta_1 - \cos\beta_2| = \lambda. \quad (5.48)$$

Последующие исчезновения интерференционных полос произойдут, если ширину источника увеличить в 2, 3, ... раз. При промежуточных значениях D полосы появляются, однако видность их незначительна, так как они наблюдаются на равномерном светлом фоне, создаваемом участком щели, на ширине которого укладывается целое число значений D из (5.48), и лишь оставшаяся меньшая часть создает интерференционные полосы.

Количественное исследование (см. задачу 1) показывает, как видность интерференционных полос зависит от ширины D протяженного источника (рис. 5.18). При $D|\cos\beta_1 - \cos\beta_2| = m\lambda$ ($m = 1, 2, ...$) видность, как уже отмечалось, обращается в нуль. Ориентировочно в качестве условия хорошего наблюдения интерференционной картины от протяженного источника можно принять неравенство

$$D|\cos\beta_1 - \cos\beta_2| \leqslant \lambda/2. \quad (5.49)$$

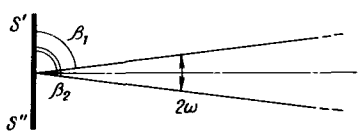
При его выполнении видность полос $V \geqslant 2/\pi \approx {}^2/_3$.

В тех случаях, когда интерферирующие лучи выходят из какой-либо точки протяженного источника *симметрично* относительно перпендикуляра к линии $S'S''$, проведенного из этой точки (т. е. когда протяженный источник $S'S''$ ориентирован перпендикулярно оси симметрии установки, рис. 5.19), угол $\beta_2 = \pi - \beta_1$ и $\cos\beta_2 = -\cos\beta_1$. Тогда условие (5.49) можно записать в виде

$$D\sin\omega \leqslant \lambda/4, \quad (5.50)$$

где 2ω — угол между выходящими из источника интерферирующими лучами, называемый *апертурой интерференции*.

При больших апертурах наблюдать интерференцию можно только от источников, размеры которых меньше длины световой волны. Если $\omega \approx \pi/2$, т. е. интерферирующие лучи выходят из источника почти в противоположных направлениях, то из (5.50) следует, что его протяженность D должна быть меньше $\lambda/4$. Для наблюдения интерференции с использованием источника, размеры которого много больше длины волны света, геометрия эксперимента должна быть такой, чтобы интерферирующие лучи выходили из источника под малым углом друг к другу. Когда $\beta_2 \approx \beta_1 \equiv \beta$, приближенно



5.19

Симметричное расположение протяженного источника

можно написать $\cos\beta_1 - \cos\beta_2 \approx \sin\beta \cdot (\beta_2 - \beta_1) = 2\omega\sin\beta$ и критерий (5.49) принимает вид

$$D \sin\beta \cdot 2\omega \leqslant \lambda/2, \text{ т. е. } D_\perp \omega \leqslant \lambda/4 \qquad (5.51)$$

— при малых апретурах интерференции ($\omega \ll 1$) играет роль протяженность источника только в направлении, перпендикулярном направлению выходящих из него лучей: $D_\perp = D\sin\beta$.

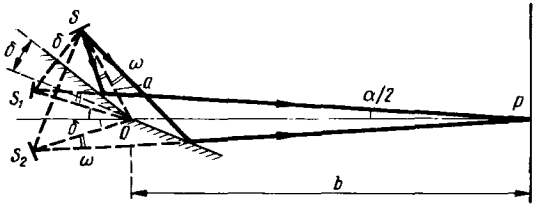
Применим критерий (5.51) к рассмотренным выше интерференционным опытам. В опыте Юнга (см. рис. 5.4) $2\omega \approx d/L$, поэтому ширина дополнительной щели S и ее угловой размер $\theta = D_\perp/L$ должны удовлетворять условию

$$D_\perp \leqslant \lambda L/(2d), \quad \theta \leqslant \lambda/(2d). \qquad (5.52)$$

При $\lambda = 5\cdot 10^{-5}$ см, $L = 1$ м и $d = 0{,}5$ мм из (5.52) находим, что ширина щели $D_\perp$ должна быть меньше 0,5 мм.

Рассмотрим опыт с зеркалами Френеля (см. рис. 5.5). Для точки наблюдения P, лежащей в центре интерференционного поля, угол 2ω между выходящими из источника S интерферирующими лучами легко найти из построения, приведенного на рис. 5.20. Угол δ (равный углу между зеркалами) является внешним для треугольника S_1OP и поэтому $\delta = \omega + \alpha/2$. Для половины угла схождения лучей можно написать $\alpha/2 = a\delta/(a+b)$. Исключая из этих уравнений $\alpha/2$, находим $\omega = \delta b/(a+b)$. Подставляя ω в (5.51), получаем следующее ограничение на ширину $D_\perp$ щели источника S: $D_\perp \leqslant \lambda(a+b)/(4b\delta)$. Если $b \gg a$, то это условие принимает вид $D_\perp \leqslant \lambda/(4\delta)$. Чтобы можно было наблюдать полосы с источником, для которого $D_\perp \gg \lambda$, угол между зеркалами должен быть очень мал ($\delta \ll 1$). Можно показать, что в опыте с зеркалами Френеля апертура интерференции 2ω имеет практически одно и то же значение при любом положении точки наблюдения P на экране в области, где перекрываются интерферирующие пучки (см. задачу 2). Поэтому видность полос одинакова по всему интерференционному полю.

В опыте с бипризмой Френеля (см. рис. 5.6 и задачу 3) апертура интерференции также практически одинакова по всему полю и равна $2\omega = \beta(n-1)b/(a+b) \approx \beta(n-1)$ (последнее справедливо при $a \ll b$).



5.20
Апертура интерференции в опыте с зеркалами Френеля

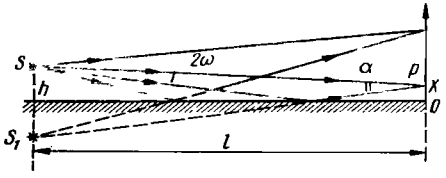
Подставляя ω в (5.51), находим $D_\perp \leqslant \lambda/[2\beta(n-1)]$. При использовании протяженного источника с $D_\perp \gg \lambda$ угол β бипризмы должен быть мал.

При анализе интерференционного опыта Поля (см. рис. 5.8) для упрощения будем пренебрегать преломлением в слюде, т. е. заменим пластинку двумя отражающими параллельными плоскостями, расстояние между которыми равно толщине h пластинки. Тогда расстояние между вторичными источниками S_1 и S_2 (мнимыми изображениями источника S) равно $2h$. Легко видеть, что угол схождения лучей в точку наблюдения в этом опыте равен апертуре интерференции 2ω. Расстояние $|S_1P|$ равно $(a+b)/\cos\theta$, поэтому $2\omega = 2h\sin\theta/|S_1P| = h\sin 2\theta/(a+b)$. Толщина h листочка слюды очень мала ($\approx 0{,}05$ мм) по сравнению с $a+b$ (≈ 5 м), поэтому мала и апертура интерференции 2ω (при любом положении точки наблюдения P, включая $\theta = 45°$). Следовательно, размер источника S может быть достаточно большим: из (5.51) находим $D_\perp \leqslant \lambda(a+b)/(2h\sin 2\theta)$. При $\lambda = 5\cdot 10^{-5}$ см, $h = 0{,}05$ мм, $a+b = 5$ м и $\theta = 45°$ должно быть $D_\perp \leqslant 2{,}5$ см. Для демонстрации опыта Поля можно использовать небольшую ртутную лампу без всяких дополнительных щелей, что обеспечивает большую светосилу. С помощью листочка слюды площадью несколько квадратных сантиметров можно получить яркую интерференционную картину больших размеров, покрывающую потолок и стены аудитории. Размер источника (≈ 10 мм) гораздо больше расстояния $|S_1S_2|$ ($\approx 0{,}1$ м), так что мнимые изображения источника почти полностью перекрываются.

При наблюдении локализованных в бесконечности полос равного наклона (см. рис. 5.9) оба интерферирующих луча выходят из источника в одном направлении, т. е. апертура интерференции равна нулю.

Поэтому здесь нет никаких ограничений на размер источника.

Зависимость видности полос от апертуры интерференции можно наглядно продемонстрировать в опыте с зеркалом Ллойда (рис. 5.21). Здесь прямой пучок света от источника S интерферирует с пучком, отраженным от зеркала при почти скользящем падении. В отличие от опытов Юнга или Френеля в опыте Ллойда апертура интерференции сильно зависит от положения точки наблюдения P на экране, установленном перпендикулярно плоскости зеркала. Из рис. 5.21 видно, что $2\omega \approx 2x/l$. Апертура тем меньше, чем ближе точка P к плоскости зеркала. При использовании протяженного источника S видность полос заметно убывает при удалении от точки $x=0$. Критерий (5.51) позволяет определить размер области, в пределах которой видность полос $V \geqslant 2/3$ при за-



5.21
Интерференционный опыт с зеркалом Ллойда

данной ширине $D_\perp$ щели источника S: $x \leqslant \lambda l/(4D_\perp)$. На вдвое большем расстоянии $\lambda l/(2D_\perp)$ полосы пропадут. Угол схождения лучей, как видно из рис. 5.21, равен $\alpha \approx 2h/l$. Ширина интерференционной полосы $\Delta x \approx \lambda/\alpha \approx \lambda l/(2h)$. Поэтому полное число полос на этом расстоянии составит $N = h/D_\perp$. Это дает оценку наивысшего порядка интерференции в монохроматическом свете от протяженного источника в опыте Ллойда.

При исследовании видности полос в опытах с протяженным источником рассматривалось наложение интерференционных картин, создаваемых отдельными его элементами. Но возможна и другая постановка этого вопроса, основанная на понятии *пространственной когерентности колебаний* в пучке света от протяженного источника. Обратимся для определенности к схеме опыта Юнга (см. рис. 5.4). Будем считать, что первый экран A с отверстием S отсутствует, а пучок света от монохроматического источника падает непосредственно на экран B с точечными отверстиями S_1 и S_2. Эти отверстия можно принять за вторичные источники, посылающие световые волны на экран C, где наблюдается интерференционная картина.

Когда первичный источник точечный, световые колебания в отверстиях S_1 и S_2 когерентны и видность полос на экране C максимальна: $V = 1$. В случае протяженного источника видность полос меньше единицы. При заданном расстоянии d между отверстиями S_1 и S_2 она зависит от отношения поперечного размера источника $D_\perp$ к расстоянию L между источником и экраном B, т. е. от углового размера источника $\theta = D_\perp/L$. Если $\theta \leqslant \lambda/(2d)$, то из (5.52) следует, что видность $V \geqslant {}^2/_3$, т. е. полосы видны отчетливо. С увеличением θ видность уменьшается, и при $\theta = \lambda/d$ полосы пропадают совсем. Уменьшение видности полос можно объяснять *частичной когерентностью* световых колебаний в точках S_1 и S_2, возбуждаемых протяженным источником. Для количественной характеристики этой когерентности колебаний в разных точках поперечного сечения светового пучка вводится понятие *степени пространственной когерентности* γ_{12}. Она характеризует способность световых колебаний в пространственно удаленных точках S_1 и S_2, взятых в некотором поперечном сечении пучка, к созданию стационарной интерференционной картины, если свет из точек S_1 и S_2 будет каким-либо способом сведен в одну точку (в опыте Юнга это происходит в результате дифракции на отверстиях в экране B, совпадающих с точками S_1 и S_2).

В пучке света от точечного источника колебания в S_1 и S_2 полностью когерентны, поэтому $\gamma_{12} = 1$ и видность полос при интерференции волн из S_1 и S_2 максимальна. В пучке света от протяженного источника степень пространственной когерентности γ_{12} зависит от расстояния d между точками S_1 и S_2 и от угловых размеров источника $\theta = D_\perp/L$. При $\theta \cdot d \gtrsim \lambda$ степень когерентности обращается в нуль: колебания в S_1 и S_2 *некогерентны* и при наложении волн из S_1 и S_2 наблюдается просто сложение интенсивностей, т. е. интерференция отсутствует.

Световое колебание в какой-либо точке, скажем S_1, в пучке света от протяженного источника возникает в результате сложения колебаний в волнах, приходящих в эту точку от разных элементарных излучателей (атомов), содержащихся в источнике. Амплитуда и фаза результирующего колебания в S_1 представляют собой случайные функции времени:

$$E_1(t)=E_{10}(t)e^{-i\omega t}=a_1(t)e^{-i[\omega t-\varphi_1(t)]}.$$

Изменения $a_1(t)$ и $\varphi_1(t)$, а следовательно, и комплексной амплитуды $E_{10}(t)$ происходят медленно в масштабе периода световых колебаний $2\pi/\omega$, но очень быстро в масштабе времени, необходимого для наблюдения. Если сместиться из точки S_1 в другую (близкую) точку S_2, то фазы суммируемых колебаний от отдельных атомов несколько изменятся из-за того, что расстояния от них до S_2 будут иными. При малом расстоянии между S_1 и S_2, пока длины оптических путей от отдельных элементов источника до S_1 и S_2 отличаются на величину, малую по сравнению с длиной волны, случайные изменения амплитуды и фазы результирующих колебаний в S_1 и S_2 происходят согласованно. С увеличением расстояния между S_1 и S_2 эта корреляция ослабевает и пропадает совсем, когда оптические длины до S_1 и S_2 различаются больше чем на λ.

Пусть оптические пути от S_1 и S_2 до точки наблюдения P различаются на Δ, т. е. колебание из S_2 происходит в точке P с запаздыванием на время $\tau=\Delta/c$. Тогда результирующее колебание в P описывается функцией

$$E_1(t)+E_2(t-\tau)=[E_{10}(t)+E_{20}(t-\tau)e^{i\omega\tau}]e^{-i\omega t}$$

и его интенсивность пропорциональна выражению

$$I\sim\langle E_{10}(t)E_{10}^*(t)\rangle+\langle E_{20}(t-\tau)E_{20}^*(t-\tau)\rangle+2\mathrm{Re}\langle E_{10}(t)E_{20}^*(t-\tau)\rangle e^{-i\omega\tau}. \tag{5.53}$$

Первые два слагаемых здесь пропорциональны интенсивностям волн из S_1 и S_2. Эти интенсивности мы будем считать одинаковыми. Выражение $\langle E_{10}(t)E_{20}^*(t-\tau)\rangle$ представляет собой *корреляционную функцию* для комплексных амплитуд $E_{10}(t)$ и $E_{20}(t)$. В стационарном случае она зависит от τ, но так же, как и средние интенсивности, не зависит от времени. Корреляционную функцию, деленную на интенсивность одной волны $I_0\sim\langle E_{10}(t)E_{10}^*(t)\rangle=\langle E_{20}(t)E_{20}^*(t)\rangle$, т. е. нормированную корреляционную функцию

$$\gamma_{12}(\tau)=\langle E_{10}(t)E_{20}^*(t-\tau)\rangle/\langle E_{10}(t)E_{10}^*(t)\rangle, \tag{5.54}$$

называют *комплексной степенью когерентности* колебаний в точках S_1 и S_2. Учитывая это определение, перепишем формулу (5.53):

$$I=2I_0[1+\mathrm{Re}\gamma_{12}(\tau)e^{-i\omega\tau}]=2I_0[1+|\gamma_{12}(\tau)|\cos(\omega\tau+\delta_{12})]. \tag{5.55}$$

В частном случае, когда точки S_1 и S_2 совпадают, $\gamma_{11}(\tau)$ характеризует корреляцию между колебаниями в одной точке в разные

моменты времени t и $t-\tau$ и представляет собой рассмотренную в § 5.4 комплексную *степень временно́й когерентности* (5.31): $\gamma_{11}(\tau)=\gamma(\tau)$. Для света от квазимонохроматического источника она описывает зависимость видности полос от разности хода $\Delta=c\tau$ двух интерферирующих волн, возбуждаемых колебанием в точке S_1.

В случае несовпадающих точек S_1 и S_2, но при $\tau=0$, величина $\gamma_{12}(0)$ характеризует корреляцию колебаний в разных пространственных точках в один момент времени. Это *степень пространственной когерентности* световых колебаний в точках S_1 и S_2. Ее модуль в соответствии с (5.55) равен видности интерференционных полос в том месте, куда колебания из S_1 и S_2 приходят с разностью хода $\Delta\approx 0$ (т. е. полос низкого порядка). Когда $|\gamma_{12}(0)|=1$, говорят о полной пространственной когерентности, когда $0<|\gamma_{12}(0)|<1$ — о частичной когерентности колебаний в точках S_1 и S_2. Свет от протяженного источника, возбуждающий колебания в отверстиях S_1 и S_2, можно рассматривать как смесь когерентного и некогерентного, причем модуль степени пространственной когерентности $|\gamma_{12}(0)|$, как видно из (5.55), определяет долю полностью когерентного света в этой смеси.

Уменьшение степени пространственной когерентности колебаний в световом пучке обусловлено конечным угловым размером источника. Второй подход к описанию уменьшения видности полос при увеличении размеров источника, основанный на понятии пространственной когерентности, отличается от разобранного ранее тем, на каком этапе производится суммирование действий различных участков источника. В первом подходе это суммирование производилось на последнем этапе, т. е. в интерференционной картине, во втором — на промежуточном этапе, в той плоскости, где расположены отверстия S_1 и S_2.

В выбранном поперечном сечении пучка света от протяженного источника степень пространственной когерентности γ_{12} убывает с увеличением расстояния d между точками S_1 и S_2. В частном случае, когда источник представляет собой равномерно светящуюся полоску с угловой шириной θ, зависимость γ_{12} от d имеет вид (см. задачу 1)

$$\gamma_{12}=\frac{\sin(\pi\theta d/\lambda)}{\pi\theta d/\lambda}. \tag{5.56}$$

Этот результат получен при суммировании интерференционных картин, создаваемых отдельными элементами протяженного источника (т. е. с использованием первого подхода). Но степень пространственной когерентности можно рассчитать и непосредственно на основе определения (5.54), выбрав ту или иную статистическую модель протяженного источника света. Знание γ_{12} может быть использовано при определении геометрии эксперимента, обеспечивающей возможность наблюдения интерференции (т. е. допустимых угловых размеров источника, апертуры интерферирующих лучей и т. д.).

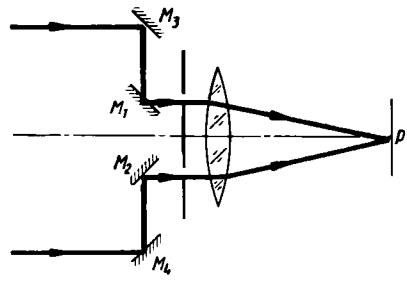
Пока расстояние d между точками S_1 и S_2 мало ($d \ll \lambda/\theta$), степень пространственной когерентности $\gamma_{12} \approx 1$. При возрастании d она уменьшается и, как видно из (5.56), при $d = \lambda/\theta$ обращается в нуль. С дальнейшим ростом d $|\gamma_{12}|$ испытывает осцилляции постепенно убывающей амплитуды (см. рис. 5.18)*, но не превышает значения $\approx 0{,}2$. Поэтому в качестве *размера области когерентности* (т. е. части поперечного сечения пучка, в пределах которой световые колебания в любой паре точек частично когерентны) можно принять $d_{\text{ког}} \approx \lambda/\theta$. Так как $\theta = D_\perp/L$, то размер области когерентности возрастает пропорционально расстоянию L от источника.

Если источник имеет равномерно светящуюся поверхность в форме прямоугольника, размеры области когерентности обратно пропорциональны соответствующим сторонам источника: $d_1 = \lambda/\theta_1$, $d_2 = \lambda/\theta_2$. В практически важном случае источника в форме равномерно светящегося диска с угловым диаметром θ область когерентности представляет собой круг диаметром $d \approx 1{,}22\lambda/\theta$ (для точек S_1 и S_2, находящихся на таком расстоянии, γ_{12} первый раз обращается в нуль). Солнечный диск виден с Земли под углом $\theta \approx 30' \approx 10^{-2}$ рад. Пренебрегая изменением яркости по его поверхности, для диаметра области когерентности при прямом солнечном освещении получим $d \approx 0{,}06$ мм (при $\lambda = 0{,}5 \cdot 10^{-4}$ см). Если щели S_1 и S_2 в опыте Юнга расположить на расстоянии, меньшем 0,06 мм, то интерференцию можно наблюдать без вспомогательного экрана A с отверстием S при непосредственном освещении щелей S_1 и S_2 лучами Солнца.

Измеряя зависимость степени пространственной когерентности $|\gamma_{12}|$ (видности полос) от расстояния между точками S_1 и S_2, можно определить угловой размер протяженного источника света. Такой метод был предложен Физо и осуществлен Майкельсоном для определения угловых размеров астрономических объектов. Об измерении угловых расстояний между компонентами двойных звезд уже было сказано выше. Попытки измерить угловые диаметры одиночных звезд, помещая экран с щелями перед объективом телескопа, оказались неудачными, так как полосы оставались четкими даже при наибольших расстояниях между S_1 и S_2, допустимых размерами телескопов. Майкельсон преодолел эту трудность с помощью *звездного интерферометра,* принцип действия которого понятен из рис. 5.22. Расположенные против щелей зеркала M_1 и M_2 неподвижны, а зеркала M и M_4 можно одновременно раздвигать. Очевидно, что видность полос зависит от степени когерентности световых колебаний на зеркалах M_3 и M_4, хотя период интерференционной картины (ширина полос) определяется расстоянием между зеркалами M_1 и M_2. При максимальном расстоянии между внешними зеркалами ~ 6 м наименьший поддающийся измерению угловой

* В тех интервалах изменения d, где $\gamma_{12} < 0$, в центре интерференционного поля расположена темная полоса.

диаметр составляет 0,02″. Несмотря на простоту идеи, практическое осуществление звездного интерферометра сопряжено с преодолением многочисленных технических трудностей, связанных с жесткими требованиями к его механической конструкции.



5.22
Схема звездного интерферометра Майкельсона

Первой звездой, у которой удалось определить угловой диаметр, была Бетельгейзе (α Ориона), относящаяся к красным гигантам. Он оказался равным 0,047″. Зная расстояние до Бетельгейзе, рассчитанное по параллаксу, можно найти линейный диаметр звезды. Он равен примерно $4 \cdot 10^8$ км, что почти в 300 раз больше диаметра Солнца и превышает диаметр земной орбиты ($3 \cdot 10^8$ км). Таким способом были измерены угловые диаметры нескольких звезд. Все они, подобно Бетельгейзе, гиганты, во много раз превосходящие Солнце. Подавляющее большинство звезд мало отличается по своему диаметру от Солнца. На расстоянии до ближайшей звезды солнечный диск был бы виден под углом лишь 0,007″, что соответствует области когерентности размером ≈ 20 м. Постройка интерферометра с такой базой (расстоянием между внешними зеркалами) представляет собой крайне сложную техническую задачу. Кроме того, при большой базе наблюдения осложняются турбулентностью атмосферы, хотя на работе интерферометра это сказывается меньше, чем при наблюдении в телескоп. Изменения показателя преломления воздуха перед зеркалами влияют на разность фаз лучей и лишь смещают интерференционную картину, не сказываясь на ее видности, так что полосы остаются различимыми, если эти изменения происходят медленно.

По принципу звездного интерферометра Майкельсона работают *радиоинтерферометры:* сигналы с двух радиотелескопов, установленных в разных местах, подают на общий детектор. Большое угловое разрешение достигается за счет значительного увеличения расстояния между антеннами. Переход к большой длине волны ($\lambda \approx 10$ см) радиодиапазона по сравнению с оптическим ведет к снижению разрешающей способности, но это компенсируется возможностью осуществления радиоинтерферометров с очень большими базами (вплоть до межконтинентальных расстояний). Таким путем было достигнуто почти в сто раз большее разрешение, чем у звездного интерферометра Майкельсона.

Определение размеров области когерентности и угловых размеров удаленного источника возможно не только из измерений видности интерференционных полос, но и с помощью предложенного Брауном и Твиссом метода *интерферометрии интенсивностей.* (В § 5.4 было

рассмотрено применение этого метода для нахождения времени когерентности.) Свет от звезды фокусируется двумя вогнутыми зеркалами на два фотоумножителя, сигиалы которых после усиления перемножаются электронной схемой. Корреляция флуктуаций регистрируется в зависимости от расстояния между зеркалами. С увеличением расстояния корреляция уменьшается и пропадает совсем, когда это расстояние превысит размер области когерентности. Таким образом корреляционные измерения интенсивности также позволяют определить степень пространственной когерентности $|\gamma_{12}(0)|$ исследуемого излучения. Интерферометр интенсивностей имеет некоторые преимущества по сравнению со звездным интерферометром Майкельсона. Здесь регистрируется непосредственно интенсивность, зависящая только от амплитуды, поэтому искажения фазы световой волны, возникающие вследствие нерегулярных изменений показателя преломления атмосферы, не влияют на результаты корреляционных измерений. Этот метод значительно менее чувствителен к неточностям в перемещении зеркал, что позволяет использовать гораздо бо́льшие базы и находить очень малые угловые размеры звезд. Был создан прибор с базой до 180 м для измерения угловых диаметров звезд вплоть до 0,0005″.

Контрольные вопросы

- □ Дайте качественное объяснение периодическому изменению видности полос в опыте Юнга при увеличении расстояния d между отверстиями S_1 и S_2, если на них падает свет от двух точечных источников, находящихся на небольшом угловом расстоянии.
- □ Какому условию удовлетворяет ширина протяженного источника в виде полоски равномерной яркости при первом исчезновении интерференционных полос в опыте Юнга?
- □ Какую величину называют степенью пространственной когерентности? Как она связана с видностью интерференционных полос?
- □ Как размеры области когерентности в пучке света от протяженного источника зависят от расстояния и от размеров источника? Оцените размер области когерентности при прямом солнечном освещении.
- □ Объясните принцип действия звездного интерферометра Майкельсона.

Задачи

1. Найти зависимость видности интерференционных полос от ширины протяженного источника света.

◇ Воспользуемся рис. 5.17, где под S' и S'' будем понимать крайние точки равномерно светящейся полоски (или щели), вытянутой в направлении, перпендикулярном плоскости чертежа. Направим ось x вдоль $S'S''$, выбрав начало отсчета в середине этого отрезка. Разделим весь источник на элементарные полоски шириной dx. Распределение интенсивности в интерференционной картине, создаваемой одним элементом dx, определяется формулой (5.8), в которой I_0 следует заменить на $I_0 dx/D$. Разность

хода $\Delta(x)$ от данного элемента до точки наблюдения P для двух интерферирующих лучей может быть записана в виде

$$\Delta(x)=\Delta_0+(\cos\beta_1-\cos\beta_2)x,$$

где Δ_0 — разность хода для лучей, выходящих из середины источника (точки $x=0$). Вывод этого выражения аналогичен выводу формулы (5.42). Распределение интенсивности в интерференционной картине от всего протяженного источника находится суммированием картин от отдельных элементов:

$$I=\frac{2I_0}{D}\int\limits_{-D/2}^{D/2}[1+\cos k\Delta(x)]\,dx=2I_0\left\{1+\frac{1}{D}\int\limits_{-D/2}^{D/2}\cos(k\Delta_0+\gamma x)\,dx\right\},$$

где $\gamma=k(\cos\beta_1-\cos\beta_2)$. Выполняя интегрирование, получаем

$$I=2I_0\{1+[\sin(k\Delta_0+\gamma D/2)-\sin(k\Delta_0-\gamma D/2)]/(\gamma D)\}=$$

$$=2I_0\left\{1+\frac{\sin(\gamma D/2)}{\gamma D/2}\cos k\Delta_0\right\}.$$

Значения Δ_0 и $\gamma=k(\cos\beta_1-\cos\beta_2)$ зависят от положения точки наблюдения P. При перемещении P сомножитель $\cos k\Delta_0$ быстро осциллирует, описывая изменения интенсивности при переходе от одной интерференционной полосы к другой, в то время как медленно изменяющийся сомножитель $\sin(\gamma D/2)/(\gamma D/2)$ дает плавную огибающую этих осцилляций. Поэтому можно считать, что экстремумы интенсивности имеются в тех точках, где $\cos k\Delta_0=\pm 1$, т. е. $k\Delta_0=m\pi$ (m — целое число). Таким образом, интерференционные полосы от протяженного источника расположены там же, где и полосы от точечного источника, находящегося при $x=0$, т. е. в середине протяженного источника.

Для видности полос получаем следующее выражение:

$$V=\left|\frac{\sin(\gamma D/2)}{\gamma D/2}\right|=\left|\frac{\sin[\pi D(\cos\beta_1-\cos\beta_2)/\lambda]}{\pi D(\cos\beta_1-\cos\beta_2)/\lambda}\right|.$$

В общем случае β_1 и β_2 зависят от положения точки наблюдения P, поэтому видность полос может быть неодинаковой в разных местах интерференционного поля. Зависимость видности от ширины D источника показана на рис. 5.18. В опыте Юнга (см. рис. 5.4) $D(\cos\beta_1-\cos\beta_2)\approx\approx D_\perp d/L=\theta\cdot d$ и видность полос равна

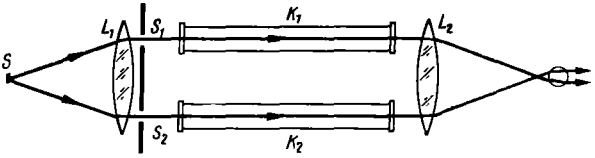
$$V=\left|\frac{\sin[\pi D_\perp d/(\lambda L)]}{\pi D_\perp d/(\lambda L)}\right|=\left|\frac{\sin(\pi\theta d/\lambda)}{\pi\theta d/\lambda}\right|. \quad \Diamond$$

2. Показать, что в опыте с зеркалами Френеля апертура интерферирующих лучей для точки наблюдения P, лежащей на краю интерференционного поля, равна $2\omega\approx 2\delta b/(a+b)\approx 2\delta$ (последнее при $a\ll b$), т. е. практически такая же, как и для центральной области.
3. Найти апертуру интерференции для опыта с бипризмой Френеля (см. рис. 5.6).

$\Diamond$ Ответ: $2\omega\approx 2\beta(n-1)b/(a+b)$. $\Diamond$

5.6. Двухлучевые интерферометры

Интерферометрами называют оптические приборы, действие которых основано на явлении интерференции света. Они предназначены для точных измерений длин, углов, характеристик оптических поверхностей, показателей преломления сред или их изменений, спектрального состава исследуемого излучения и т. п. Наблюдение интерфе-



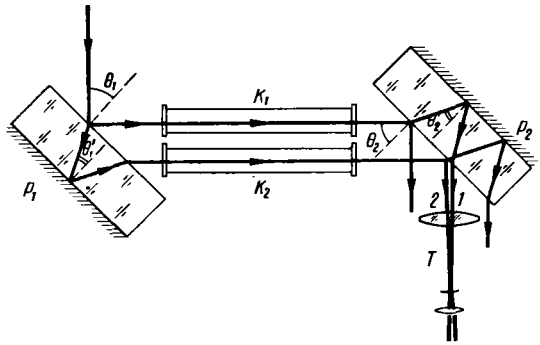
5.23
Схема интерферометра Рэлея (вид сверху)

ренционных полос при этом становится не целью исследования, а средством проведения измерений. В зависимости от характера решаемой задачи к оптической схеме интерферометра и его конструкции предъявляются различные требования.

В *интерферометре Рэлея,* предназначенном для измерения показателей преломления газов и жидкостей, использован, как и в опыте Юнга, метод деления волнового фронта. Источник в виде узкой щели S расположен в фокальной плоскости линзы L_1 (рис. 5.23). Выходящий из нее параллельный пучок идет через диафрагму с двумя щелями S_1 и S_2, параллельными щели S. Пучки света от S_1 и S_2 проходят через кюветы K_1 и K_2 и образуют интерференционные полосы в фокальной плоскости линзы L_2. Введение кювет, содержащих исследуемые газы или жидкости, требует значительного расстояния между S_1 и S_2, вследствие чего интерференционные полосы располагаются тесно и для их наблюдения требуется большое увеличение. Для этой цели удобен цилиндрический окуляр в виде тонкой стеклянной палочки, ось которой параллельна полосам. Кюветы занимают только верхнюю половину пространства между линзами L_1 и L_2, а внизу свет идет вне кювет. Благодаря этому возникает вторая система интерференционных полос с таким же расстоянием между полосами, которая может служить шкалой для отсчета. Верхняя система полос сдвинута относительно нижней, так как при прохождении света через кюветы появляется добавочная разность хода $\Delta=(n_2-n_1)l$, где n_1 и n_2 — коэффициенты преломления веществ, заполняющих кюветы. По этому смещению определяют n_2-n_1. В один из пучков ставится компенсатор, с помощью которого можно добиваться, чтобы плавно изменялась оптическая разность хода, противоположная по знаку той, которая обусловлена прохождением света через кюветы.

Совпадение двух систем полос используется для установления полной компенсации разности хода. Визуально можно установить совпадение с точностью до $^1/_{40}$ порядка, что при $l=0{,}1$ м, $\lambda=$ $=550$ нм позволяет обнаружить изменение n_2-n_1 около 10^{-7}.

Интерферометр Жамена состоит из двух одинаковых толстых пластин P_1 и P_2 (рис. 5.24), изготовленных из весьма однородного стекла (или кварца для работы в ультрафиолетовой области спектра). Задние поверхности пластин посеребрены. Пучок света от протяженного источника падает под углом, близким к 45°, на одну из пластин. В результате отражения от передней и задней поверхностей пластины P_1 возникают два параллельных пучка, разделен-



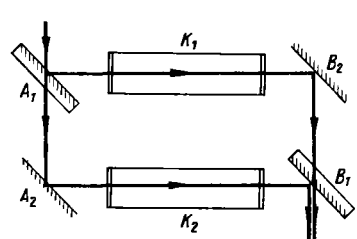
5.24
Схема интерферометра Жамена (вид сверху)

ных тем больше, чем толще пластина. Каждый из них в свою очередь раздваивается при отражении от двух поверхностей пластины P_2. Средние пучки *1* и *2* налагаются и образуют интерференционную картину в фокальной плоскости зрительной трубы T. Для разности хода между ними с помощью формулы (5.10) находим

$$\Delta = 2nh(\cos\theta_2' - \cos\theta_1') \approx 2nh\sin\theta' \cdot \delta\theta',$$

где h — толщина пластин; n — показатель преломления их материала; θ_1 и θ_2 — углы падения на поверхности пластин P_1 и P_2; θ_1' и θ_2' — соответствующие углы преломления. Если пластины строго параллельны, то $\theta_1' = \theta_2'$ и $\Delta = 0$. Поле зрения будет равномерно освещенным. При юстировке одну из пластин слегка наклоняют, поворачивая вокруг горизонтальной оси. При этом интерференционные полосы, наблюдаемые в установленную на бесконечность зрительную трубу, горизонтальны и эквидистантны. Они соответствуют низким порядкам интерференции и потому могут наблюдаться в белом свете. Значительное разведение пучков между пластинами позволяет поместить на их пути кюветы K_1 и K_2 с исследуемыми веществами. При этом оптическая разность хода изменится на $(n_2 - n_1)l$, что вызовет смещение интерференционной картины.

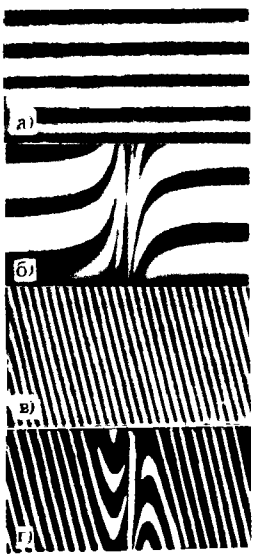
По такому же принципу устроен *интерферометр Рождественского* (рис. 5.25). Роль делителей пучков — внутренних граней пластин в интерферометре Жамена — играют здесь полуотражающие плоскопараллельные пластины A_1 и B_1, а посеребренных наружных граней пластин — зеркала A_2 и B_2. Это позволяет без использования толстых пластин значительно раздвинуть пучки света и ввести кюветы K_1 и K_2,



5.25
Схема интерферометра Рождественского

одна из которых окружена печью (для исследования паров металлов). Пластины A_1, A_2 и B_1, B_2 установлены попарно на общих основаниях строго параллельно. Блоки из A_1, A_2 и B_1, B_2 могут быть разнесены на значительное расстояние (≈ 1 м). Один из них наклоняется на небольшой угол поворотом вокруг горизонтальной оси. Поэтому, как и в интерферометре Жамена, наблюдаются горизонтальные полосы равного наклона, соответствующие (при отсутствии кювет) низким порядкам интерференции.

На основе такого прибора Д. С. Рождественским в 1912 г. были выполнены классические исследования зависимости показателя преломления от длины волны вблизи линий поглощения (т. е. аномальной дисперсии, см. § 2.5) в парах металлов. Горизонтальные интерференционные полосы в белом свете проецируются на вертикально расположенную входную щель спектрографа. Положение светлой полосы нулевого порядка не зависит от длины волны, и поэтому в сплошном спектре, даваемом спектрографом, ей соответствует горизонтальная светлая полоса, тянущаяся вдоль всего спектра. Максимумы, соответствующие порядкам интерференции $m = \pm 1$, ± 2, ..., находятся для разных длин волн падающего света на разных высотах щели спектрографа. Им отвечают в сплошном спектре светлые полосы, лежащие выше и ниже нулевой полосы и расходящиеся веером от синей к красной области спектра, так как расстояние между максимумами растет с увеличением длины волны (рис. 5.26, *а*).



5.26
Интерференционные полосы, развернутые в спектр

Направим ось x вдоль нулевой полосы в спектре. Точки этой оси соответствуют разным значениям длины волны λ. Ось y направим параллельно щели спектрографа. Расстояние между интерференционными максимумами пропорционально длине волны, поэтому для ординаты полосы m-го порядка можно написать $y_m = \alpha m \lambda$, где α — постоянная, определяемая геометрией установки. Наклон полос $dy_m/d\lambda = \alpha m$ возрастает с увеличением порядка интерференции m. Если кювету на пути одного из пучков заполнить веществом с показателем преломления $n = n(\lambda)$, а вторую кювету откачать, то оптическая разность хода изменится на $(n-1)l$ и полосы сместятся вверх или вниз на расстояние $\alpha(n-1)l$. Тогда ордината полосы m-го порядка определяется выражением

$$y_m = \alpha[m\lambda \pm (n-1)l], \qquad (5.57)$$

причем знак зависит от того, в какое из плеч интерферометра введена кювета с исследуемым веществом. Полоса нулевого

порядка ($m=0$), ранее удовлетворявшая условию $y=0$ (т. е. совпадавшая с осью абсцисс), теперь будет описываться уравнением $y=\pm\alpha(n-1)l$. Это значит, что нулевая полоса в определенном масштабе, зависящем от настройки интерферометра, вычерчивает зависимость $n-1$ от λ, т. е. дает непосредственно кривую дисперсии. Полосы с $m\neq 0$ прочертят почти подобные кривые, так как слагаемое $\alpha m\lambda$ в (5.57) обычно невелико и изменяется с длиной волны очень медленно. На рис. 5.26, *б* показаны полосы, воспроизводящие ход дисперсии с ее характерными особенностями вблизи линии поглощения.

Описанный метод измерения $n(\lambda)$, предложенный Пуччианти в 1901 г., нагляден, но мало пригоден для количественного исследования дисперсии, так как изменение положения точек на круто изменяющей свое направление кривой сопряжено с большими погрешностями. Рождественский разработал новый метод исследования дисперсии вблизи линии поглощения (метод «крюков»), позволяющий проводить измерения с большой точностью. В одно из плеч интерферометра вводится тонкая плоскопараллельная стеклянная пластинка определенной толщины l'. Это ведет к большой добавочной разности хода $(n'-1)l'$, где n' — показатель преломления пластинки. Пока в кювете, расположенной в другом плече, исследуемого вещества нет, будут наблюдаться наклонные интерференционные полосы высоких порядков $m\gg 1$ (рис. 5.26, *в*). При одновременном действии исследуемого вещества (паров металла) и стеклянной пластинки вызываемые ими противоположные смещения полос суммируются для каждого значения λ. Вдали от линии поглощения показатель преломления n разреженных паров близок к единице, поэтому наклон полос обусловлен только стеклянной пластинкой. Вблизи линии поглощения показатель преломления паров изменяется очень сильно и найдется такая длина волны, для которой действия паров и пластинки будут точно скомпенсированы, так что наклон интерференционной кривой пройдет через нуль. В результате полосы вблизи линии поглощения своеобразно изгибаются, образуя крюки, положения вершин которых на шкале длин волн можно точно измерить (рис. 5.26, *г*).

В электронной теории дисперсии зависимость показателя преломления от частоты в разреженных газах и парах описывается выражением (2.51). В окрестности одиночной линии поглощения на частоте ω_{0k} можно оставить в (2.51) только слагаемое с ω_{0k}. Пренебрегая γ_k по сравнению с $|\omega_{0k}-\omega|$, получаем для $n(\omega)$ следующее выражение (формула Зельмейера):

$$n^2-1=\frac{Ne^2}{\varepsilon_0 m_0}\frac{f_k}{\omega_{0k}^2-\omega^2}. \qquad (5.58)$$

Здесь N — концентрация атомов, находящихся в основном состоянии; m_0 — масса электрона; f_k — сила осциллятора для перехода из основного состояния E_0 в возбужденное с энергией $E_k=E_0+$

$+\hbar\omega_{0k}$. Перейдем в (5.58) от частот к длинам волн, подставив $\omega=2\pi c/\lambda$, и учтем, что $n^2-1=(n-1)(n+1)\approx 2(n-1)$:

$$n-1\approx -\frac{Ne^2f_k}{8\pi^2\varepsilon_0 m_0 c^2}\frac{\lambda_{0k}^2\lambda^2}{\lambda_{0k}^2-\lambda^2}\approx \frac{Ne^2f_k\lambda_{0k}^3}{(4\pi)^2\varepsilon_0 m_0 c^2(\lambda-\lambda_{0k})}. \tag{5.59}$$

В последнем равенстве учтено, что в окрестности линии поглощения $\lambda\approx\lambda_{0k}$ и $\lambda^2-\lambda_{0k}^2\approx 2\lambda_{0k}(\lambda-\lambda_{0k})$.

Теоретическая зависимость показателя преломления от длины волны, описываемая вблизи линии поглощения формулой Зельмейера (5.59), была подтверждена экспериментально в работах Рождественского и его учеников. Основное достоинство метода крюков заключается в том, что он позволяет по расстоянию между вершинами крюков найти произведение Nf_k. Если число атомов в единице объема N известно, то можно найти одну из важных атомных констант — силу осциллятора f_k.

Покажем, как Nf_k можно вычислить по измеренному расстоянию между вершинами крюков. При наличии кюветы с парами исследуемого вещества в одном плече интерферометра и стеклянной пластинки в другом ордината y_m полосы m-го порядка будет вместо (5.57) определяться выражением

$$y_m=\alpha[m\lambda\pm(n-1)l\mp(n'-1)l']. \tag{5.60}$$

Для нахождения положения вершин крюков m-й полосы, т. е. экстремумов зависимости y_m от λ, составим уравнение $dy_m/d\lambda=0$:

$$m\pm l\frac{dn}{d\lambda}\mp l'\frac{dn'}{d\lambda}=0. \tag{5.61}$$

Показатель преломления паров вблизи линии поглощения изменяется очень быстро, в то время как для стекла во всей видимой области, где оно прозрачно, n' изменяется очень медленно. Поэтому последним членом в уравнении (5.61) можно пренебречь и для длин волн, соответствующих вершинам крюков, приближенно получаем

$$dn/d\lambda=\mp m/l=-|m|/l.$$

С другой стороны, формула Зельмейера (5.59) позволяет связать $dn/d\lambda$ с Nf_k:

$$\frac{dn}{d\lambda}=-\frac{Ne^2f_k\lambda_{0k}^3}{(4\pi)^2\varepsilon_0 m_0 c^2(\lambda-\lambda_{0k})^2}.$$

Приравнивая правые части последних двух выражений, получаем

$$Nf_k=|m|\frac{(4\pi)^2\varepsilon_0 m_0 c^2}{le^2\lambda_{0k}^3}(\Delta\lambda)^2, \tag{5.62}$$

где $\Delta\lambda=\lambda-\lambda_{0k}$ — расстояние от вершины крюка до центра линии поглощения. У одиночной линии поглощения крюки расположены симметрично, так что расстояние между их вершинами равно $2\Delta\lambda$. Для определения порядка интерференции $|m|$, соответствующего полосе, на которой измеряется расстояние $2\Delta\lambda$ между крюками,

можно воспользоваться картиной наклонных полос, пока кювета с исследуемым металлом не нагрета. В этом случае $n-1=0$ и из (5.60) $y_m=\alpha[m\lambda\mp(n'-1)l']$. Измеряя длину волны λ_m в точке пересечения некоторой полосы с осью x (когда $y_m=0$), находим соответствующий этой полосе порядок интерференции $|m|=(n'-1)l'/\lambda_m$. Можно обойтись и без знания величины $(n'-1)l'$, если измерить кроме λ_m длину волны λ_{m+p} в точке пересечения оси x полосой, отстоящей от рассматриваемой на p порядков. Тогда $m\lambda_m=(m+p)\lambda_{m+p}$ и $|m|=p\lambda_{m+p}/(\lambda_m-\lambda_{m+p})$. Определив порядок интерференции $|m|$ некоторой полосы, по измеренному на ней расстоянию $2\Delta\lambda$ между крюками находим с помощью (5.62) силу осциллятора f_k.

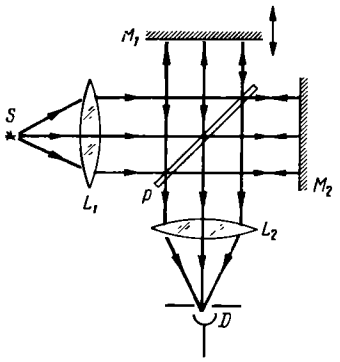
Среди двухлучевых интерферометров широкое распространение получили различные модификации *интерферометра Майкельсона* (рис. 5.10), принцип действия которого был рассмотрен в § 5.3. О его применении для исследования тонкой структуры спектральных линий говорилось в § 5.4.

Главная особенность интерферометра Майкельсона по сравнению с интерферометрами других типов заключается в том, что с его помощью можно непрерывно изменять разность хода между пучками в широких пределах путем перемещения одного из зеркал и наблюдать при этом интерференционные полосы высоких порядков. Это необходимо как для измерения длины когерентности излучения узких спектральных линий, так и для выполнения метрологических работ по прямому сравнению длины световой волны (т. е. первичного эталона длины) с концевой мерой, представляющей собой металлический стержень с параллельными зеркально отполированными торцовыми плоскостями.

Первые опыты по сравнению стержневого эталона метра с длиной световой волны были выполнены Майкельсоном и Бенуа в 1892—1893 гг. Использовался свет красной линии кадмия с $\lambda\approx 644$ нм, характеризующийся наибольшей известной тогда длиной когерентности. При этом пришлось использовать ряд промежуточных эталонов, длина которых определялась расстоянием между двумя зеркалами, укрепленными строго параллельно на массивном металлическом основании.

В результате дальнейших поисков более подходящих источников света была создана стандартная криптоновая лампа, длина когерентности оранжевого излучения которой достигает 0,8 м. Это излучение и положено в настоящее время в основу первичного эталона длины: по определению, 1 м равен 1650763,73 длин волн в вакууме излучения, соответствующего переходу между уровнями $5d_5\rightarrow 2p_{10}$ атома криптона-86. Такому определению соответствует длина волны $\lambda(\mathrm{Kr})=605{,}780211$ нм.

Оптическая схема интерферометра Майкельсона применяется в современных спектральных приборах нового типа — *фурье-спектрометрах*, особенно эффективных для исследования инфракрасной части спектра слабых источников, когда приходится применять сравнительно малочувствительные приемники излучения с относительно большими собственными шумами. Рассмотрим кратко принцип их действия.



5.27
К объяснению принципа действия фурье-спектрометра

Пусть подвижное зеркало M_1 интерферометра Майкельсона перемещается с постоянной скоростью v (рис. 5.27). Тогда разность хода между двумя интерферирующими пучками будет изменяться по закону $\Delta = 2vt$. При монохроматическом освещении интерферометра интенсивность света, попадающего в приемник D, в соответствии с (5.8) изменяется по закону

$$I(t) = I_0(1 + \cos k\Delta) = I_0(1 + \cos 2kvt). \tag{5.63}$$

Интенсивность, а вместе с ней и сигнал приемника оказываются промодулированными с частотой $\Omega = 2kv = 2\omega v/c$*. Частота модуляции Ω зависит от оптической частоты ω монохроматического излучения. Измеряя Ω, можно найти ω, т. е. получить информацию о спектре источника.

При освещении интерферометра немонохроматическим светом со спектральной плотностью $I(\omega)$ каждая монохроматическая составляющая дает в сигнал вклад, определяемый выражением (5.63). Полный сигнал находится интегрированием по всем частотам:

$$I(t) = \int_0^\infty I(\omega)[1 + \cos(2\omega v/c)t]\,d\omega = \int_0^\infty I(\omega)\,d\omega + \int_0^\infty I(\omega)\cos(2wv/c)t\,d\omega.$$

Фурье-компонента тока приемника, соответствующая частоте $\Omega = 2\omega v/c$, пропорциональна интенсивности спектральной компоненты излучения источника на частоте ω. Другими словами, спектр источника трансформируется в спектр тока приемника — от оптических частот на низкие частоты модуляции интенсивности. Сигнал поступает с приемника на широкополосный усилитель переменного тока. Первое, не зависящее от времени слагаемое (постоянная составляющая) не пропускается усилителем. Поэтому сигнал на выходе $F(t)$ пропорционален второму слагаемому — фурье-образу функции $I(\omega)$:

$$F(t) \sim \int_0^\infty I(\omega)\cos\left(\frac{2\omega v}{c}t\right)d\omega.$$

Искомая функция $I(\omega)$ распределения энергии по частотам выра-

* Биения интенсивности попадающего в приемник света [см. (5.63)] можно интерпретировать как результат нестационарной интерференции двух когерентных волн с близкими частотами: волны частотой ω, отраженной неподвижным зеркалом M_2, и волны, отраженной движущимся зеркалом M_1, частота которой вследствие эффекта Доплера сдвинута на $\Delta\omega = (2v/c)\omega$.

жается через наблюдаемую функцию $F(t)$ с помощью обратного преобразования Фурье:

$$I(\omega) \sim \int_0^\infty F(t) \cos\left(\frac{2\omega v}{c} t\right) dt. \tag{5.64}$$

Таким образом, для получения искомого спектрального распределения исследуемого излучения нужно провести соответствующее преобразование Фурье (5.64) наблюдаемого сигнала, чем и обусловлено название метода. Для выполнения этой операции используются быстродействующие электронно-вычислительные машины.

Преимущества фурье-спектрометра перед обычными спектральными приборами (например, с дифракционной решеткой, см. § 6.6) обусловлены увеличением проходящего через прибор светового потока (для достижения высокого разрешения здесь не требуется уменьшать ширину входной щели) и одновременной регистрацией всего спектра при использовании фотоэлектрических приемников. Для коротковолновой части спектра эти преимущества значительно снижаются в связи с тем, что на приемник попадает одновременно с модулированной и немодулированная часть излучения, из-за чего возрастает уровень шумов на выходе. В коротковолновой части спектра шум определяется полным световым потоком, в то время как для малочувствительных детекторов инфракрасного излучения основной шум обусловлен собственными шумами детектора и в широких пределах не зависит от падающего на приемник потока излучения. Поэтому преимущества метода реализуются лишь в инфракрасной области, особенно в далекой.

Рассмотренные примеры использования двухлучевых интерферометров дают представление о широких возможностях интерференционных методов исследования, хотя многие их приложения остались, конечно, вне нашего поля зрения.

5.7. Многолучевая интерференция

При наложении двух когерентных световых пучков образуются интерференционные полосы, в которых распределение интенсивности описывается функцией $I \sim \cos^2(k\Delta/2)$ (Δ — разность хода пучков). Максимумы и минимумы интенсивности, т. е. светлые и темные полосы, в двухлучевой интерференционной картине имеют одинаковую ширину. При наложении большого числа пучков распределение интенсивности в интерференционной картине существенно иное. Изменение характера интерференционных полос при увеличении числа n пучков качественно можно предсказать на основе закона сохранения энергии. Амплитуда световых колебаний в максимумах интенсивности, где сложение колебаний происходит в одинаковой фазе, в n раз больше, а интенсивность в n^2 раз больше, чем от одного пучка (при условии, что когерентные пучки имеют одинаковую или почти одинаковую интенсивность). Но полная энергия, приходящаяся на одну интерференционную полосу, лишь в n раз больше, чем в одном пучке. Увеличение интенсивности в максимумах в n^2 раз возможно только в случае существенного перераспределения потока энергии в пространстве: при прежнем расстоянии между светлыми полосами их ширина должна быть примерно в n раз меньше этого расстояния. Благодаря образованию узких максимумов, т. е. резких светлых полос, разделенных широкими темными

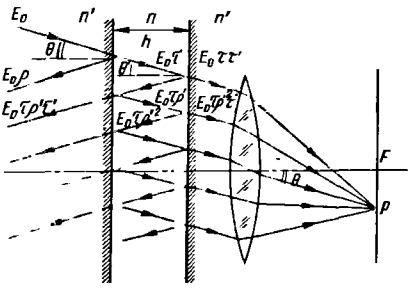
промежутками, многолучевая интерференция получила важное практическое применение.

Большое число когерентных световых пучков может возникнуть в результате дифракции при прохождении плоской волны через экран с одинаковыми регулярно расположенными отверстиями (метод деления волнового фронта). Распределение интенсивности в такой многолучевой интерференционной картине будет рассмотрено в § 6.5 на примере дифракционной решетки. Здесь мы изучим интерференцию при многократных отражениях света от двух параллельных поверхностей (метод деления амплитуды). На этом принципе действует *интерферометр Фабри—Перо,* широко используемый в спектроскопии высокого разрешения и в метрологии. Он может быть выполнен в виде плоскопараллельной стеклянной или кварцевой пластины, на обе поверхности которой нанесены отражающие слои, либо в виде двух пластин, у которых покрытые отражающими слоями плоскости установлены строго параллельно друг другу и разделены воздушным промежутком.

Отражение света от двух параллельных плоскостей приводит (см. § 5.3) к образованию локализованных в бесконечности (или фокальной плоскости линзы) интерференционных полос равного наклона. В некоторую точку P фокальной плоскости линзы собираются лучи, которые до линзы образуют с ее оптической осью один и тот же угол θ (рис. 5.28). Разность хода Δ двух соседних интерферирующих лучей определяется формулой (5.10): $\Delta = 2nh\cos\theta'$. Максимумы интенсивности в проходящем свете расположатся там, где Δ составляет целое число длин волн:

$$2nh\cos\theta' = m\lambda_0. \qquad (5.65)$$

Линиям равных интенсивностей соответствует одно и то же значение угла θ, поэтому интерференционные полосы в фокальной плоскости линзы имеют вид концентрических колец с центром на оси линзы. Центру картины соответствует наибольший порядок интерференции. При этом расположение максимумов интенсивности будет таким же, как в полосах равного наклона при двухлучевой интерференции. Однако для определения структуры максимумов в случае высокого коэффициента отражения светоделительных поверхностей необходимо учесть интерференцию всех приходящих в точку P волн, образующихся при многократных отражениях.



5.28
Полосы равного наклона при многолучевой интерференции

При каждом пересечении отражающей поверхности падающая световая волна порождает две волны: прошедшую и отражен-

ную. Отношения амплитуд этих волн к амплитуде падающей волны (коэффициенты пропускания τ и отражения ρ), вообще говоря, зависят от угла падения и состояния поляризации света. Для границы двух прозрачных сред (без нанесенных на нее отражающих покрытий) это было показано в § 3.2 с помощью формул Френеля. Мы здесь ограничимся исследованием распределения интенсивности в интерференционных полосах равного наклона, которым соответствуют малые углы θ. В этом случае τ и ρ практически не зависят от угла падения и от поляризации падающего света (см. рис. 3.5 и 3.10) и для них можно принять значения, соответствующие нормальному падению.

Пусть τ и ρ — амплитудные коэффициенты пропускания и отражения при переходе волны из окружающей среды в плоскопараллельную пластинку (которая, в частности, может быть воздушным промежутком между зеркалами), а τ' и ρ' — при переходе из пластинки в среду. Согласно формулам (3.12),

$$\tau=\frac{2n'}{n+n'},\ \tau'=\frac{2n}{n+n'};\ \rho=\frac{n'-n}{n+n'},\ \rho'=\frac{n-n'}{n+n'}, \tag{5.66}$$

где n и n' — показатели преломления пластинки и окружающей среды. На границе прозрачных сред τ и ρ вещественны, а отрицательное значение ρ при $n>n'$ учитывает изменение фазы волны на π при отражении от оптически более плотной среды. Из формул (5.66) легко видеть, что отражательная способность границы, или энергетический коэффициент отражения R (см. § 3.2), не зависит от того, идет свет из окружающей среды в пластинку или наоборот:

$$R=\rho^2=\rho'^2, \tag{5.67}$$

и что

$$\tau\tau'=1-\rho^2=1-R. \tag{5.68}$$

В случае светоделительных поверхностей с нанесенными на них многослойными непоглощающими диэлектрическими покрытиями (см. ниже) коэффициенты τ, τ' и ρ, ρ' будут иными (в частности, отражательная способность $R=\rho^2$ для определенной длины волны может иметь значения, очень близкие к единице). Но можно показать, что соотношение $\rho'=-\rho$ остается в силе и связь τ, τ', ρ, ρ' с отражательной способностью R по-прежнему выражается формулами (5.67), (5.68).

Предположим, что на пластинку падает под углом θ плоская монохроматическая волна с амплитудой E_0 (рис. 5.28). При нахождении комплексной амплитуды суммарной волны, прошедшей через пластинку, нужно учесть, что фаза каждой последующей волны больше фазы предыдущей на

$$\delta=k\Delta=2nkh\cos\theta', \tag{5.69}$$

где $k=2\pi/\lambda_0$ — волновое число. Поэтому комплексная амплитуда последующей волны отличается дополнительным множителем $\exp(i\delta)$ от амплитуды предыдущей. В результате комплексная амплитуда всей прошедшей волны представится геометрической прогрессией:

$$E_2=E_0\tau\tau'(1+\rho'^2e^{i\delta}+\rho'^4e^{i2\delta}+\ldots)=\frac{E_0\tau\tau'}{1-\rho'^2e^{i\delta}}. \qquad (5.70)$$

Мы считаем здесь размеры пластинки и линзы достаточно большими, чтобы можно было не учитывать дифракцию на их краях (см. § 6.3) и виньетирование наклонных пучков (т. е. ограничение их поперечного сечения краями пластинки и линзы). Для нахождения интенсивности прошедшей волны умножим E_2 в (5.70) на комплексно-сопряженную величину и воспользуемся формулами (5.67) и (5.68):

$$I_{\text{пр}}=I_{\text{пад}}\frac{(1-R)^2}{1+R^2-2R\cos\delta}=I_{\text{пад}}\frac{1}{1+[4R/(1-R)^2]\sin^2(\delta/2)}. \qquad (5.71)$$

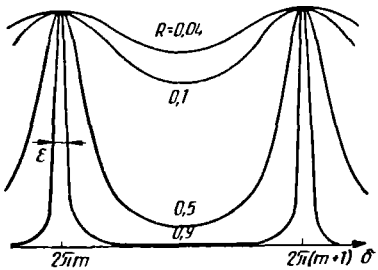
Таким же способом легко получить выражение для амплитуды отраженной волны:

$$E_1=E_0\rho+E_0\tau\tau'\rho'e^{i\delta}(1+\rho'^2e^{i\delta}+\ldots)=E_0\rho\left(1-\frac{\tau\tau'e^{i\delta}}{1-\rho'^2e^{i\delta}}\right).$$

Здесь учтено, что $\rho'=-\rho$. Для интенсивности отраженной волны находим

$$I_{\text{отр}}=I_{\text{пад}}\frac{2R(1-\cos\delta)}{1+R^2-2R\cos\delta}=I_{\text{пад}}\frac{[4R/(1-R)^2]\sin^2(\delta/2)}{1+[4R/(1-R)^2]\sin^2(\delta/2)}. \qquad (5.72)$$

Соотношения (5.71) и (5.72) известны как *формулы Эйри*. Из них видно, что $I_{\text{пр}}+I_{\text{отр}}=I_{\text{пад}}$, как и должно быть при отсутствии поглощения. Очевидно, что интенсивность в точке P фокальной плоскости линзы (рис. 5.28) относится к интенсивности в P' при отсутствии пластинки как $I_{\text{пр}}/I_{\text{пад}}$. Поэтому формула (5.71) дает распределение интенсивности в многолучевой интерференционной картине для тех мест фокальной плоскости линзы, где при отсутствии пластинки была бы равномерная освещенность. Максимумы $I_{\text{пр}}/I_{\text{пад}}=1$ получаются при $\delta/2=m\pi$ (m — целое число). Подставляя сюда δ из (5.69), получаем то же условие (5.65), что и при двухлучевой интерференции. При



5.29
Кривые распределения интенсивности в проходящем свете

малом коэффициенте отражения $R \ll 1$ формулы (5.71) и (5.72) дают то же самое и для распределения интенсивности. Это распределение, однако, существенно изменяется при увеличении R, в особенности когда R приближается к единице. На рис. 5.29 приведены построенные по формуле (5.71) кривые интенсивности прошедшего света в зависимости от δ. При $R \ll 1$ они имеют вид, типичный для двух интерферирующих пучков. С увеличением коэффициента отражения R максимумы сужаются, а в промежутках между ними интенсивность становится очень мала. Отношение интенсивности в максимумах и минимумах, характеризующее *контрастность* интерференционных полос, как видно из (5.71), определяется только коэффициентом отражения:

$$I_{\max}/I_{\min} = (1+R)^2/(1-R)^2. \qquad (5.73)$$

Многолучевая интерференционная картина в прошедшем свете при значениях R, близких к единице, имеет вид узких светлых полос на почти совершенно темном фоне. В отраженном свете наблюдается дополнительная картина в виде узких темных полос на почти равномерном светлом фоне. Подобное пространственное перераспределение потока энергии с концентрацией его в некоторых преимущественных направлениях всегда возникает при интерференции многих пучков.

Острота интерференционных максимумов характеризуется шириной контура интенсивности на половине его высоты, т. е. расстоянием между точками, лежащими по обе стороны максимума в тех местах, где интенсивность уменьшается до половины максимального значения. Около максимума порядка m эти точки находятся при $\delta = 2\pi m \pm \varepsilon/2$, где ε в соответствии с (5.71) определяется соотношением $\dfrac{1}{1+[4R/(1-R)^2]\sin^2(\varepsilon/4)} = \dfrac{1}{2}$, т. е. $\dfrac{2\sqrt{R}}{1-R}\sin\dfrac{\varepsilon}{4} = 1$.

При большом коэффициенте отражения R значение ε настолько мало, что можно положить $\sin(\varepsilon/4) \approx \varepsilon/4$. Тогда для изменения разности фаз δ, соответствующего ширине контура интерференционной полосы, получаем

$$\varepsilon = 2(1-R)/\sqrt{R}. \qquad (5.74)$$

В случае $\varepsilon \ll 1$ описываемый формулой Эйри (5.71) контур полосы в монохроматическом проходящем свете в окрестности каждого максимума $\delta = 2\pi m$ принимает лоренцевскую форму:

$$I_{\text{пр}}/I_{\text{пад}} = 1/[1+(2\Delta\delta/\varepsilon)^2], \qquad (5.75)$$

где $\Delta\delta = \delta - 2\pi m$ — отклонение δ от его значения в максимуме.

Отношение расстояния между соседними полосами (которому соответствует изменение δ на 2π) к ширине максимума называется *резкостью полос* F:

$$F = 2\pi/\varepsilon = \pi\sqrt{R}/(1-R). \qquad (5.76)$$

С увеличением резкости полос распределение интенсивности становится более благоприятным для определения на опыте положения максимумов. Чем выше резкость полос, тем более четко разделяются полосы, принадлежащие различным спектральным компонентам падающего излучения.

Из формулы (5.76) видно, что резкость полностью определяется коэффициентом отражения R. Она растет с увеличением R, стремясь в пределе к бесконечности при $R \to 1$. На границе воздух—стекло при нормальном падении коэффициент отражения составляет всего 0,04. Необходимое для достижения хорошей резкости большое значение R может быть получено нанесением тонкого металлического покрытия (серебро, алюминий) путем термического испарения в вакууме. Но металлические пленки поглощают свет (см. § 3.4), поэтому использованное при выводе формул Эйри условие (5.68) должно быть для них заменено следующим:

$$\tau\tau' = 1 - R - A,$$

где A — доля светового потока, поглощаемая слоем металла. Тогда из (5.70) получим вместо (5.71) выражение

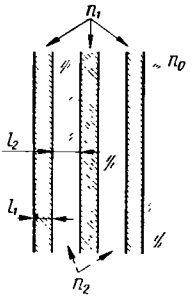
$$\frac{I_{пр}}{I_{пад}} = \left(1 - \frac{A}{1-R}\right)^2 \frac{1}{1 + [4R/(1-R)^2]\sin^2(\delta/2)}. \qquad (5.77)$$

При прежнем характере пространственного распределения интенсивности поглощение приводит к ее общему уменьшению.

Значительное расширение возможных применений интерферометра Фабри—Перо связано с использованием вместо металлических зеркал *многослойных диэлектрических покрытий*, которые обеспечивают высокий коэффициент отражения (и, следовательно, большую резкость полос) и в то же время не поглощают свет. Распределение интенсивности в интерференционной картине и в этом случае описывается формулой (5.77), но пропускание в максимумах может быть значительно больше, чем у интерферометра с металлическими зеркалами. Уменьшение интенсивности по сравнению с идеализированным случаем, выражаемым формулой (5.71), обусловлено здесь в основном рассеянием света на практически неизбежных неоднородностях покрытий.

Рассмотрим идею получения диэлектрических зеркал. Пусть на поверхность стекла (или кварца), показатель преломления которого n_0, нанесен слой прозрачного диэлектрика с показателем преломления $n > n_0$. Толщина слоя l выбрана так, чтобы его оптическая толщина nl была равна $\lambda_0/4$, т. е. четверти длины волны в вакууме. При этом отражательная способность поверхности возрастает. В самом деле, волны, отраженные (при нормальном падении) передней и задней границами слоя, находятся в фазе, так как отставание фазы второй волны на π, накопившееся при ее распространении внутри пленки туда и обратно, компенсируется

изменением фазы первой волны на π при ее отражении от оптически более плотной среды. Покрытия из TiO_2 ($n=2{,}45$) или сернистого цинка ($n=2{,}3$) дают $R\approx 0{,}3$. Но добиться более высоких коэффициентов отражения таким способом практически невозможно. Эффект можно значительно усилить, используя последовательность чередующихся диэлектрических слоев с высоким n_1 и низким n_2 показателями преломления (рис. 5.30). Если оптическая толщина всех слоев одинакова и равна $\lambda_0/4\,(n_1l_1=n_2l_2=\lambda_0/4)$, то отраженные их границами волны находятся, как легко заметить, в одинаковой фазе и в результате интерференции усиливают друг друга.



5.30
Схема многослойного диэлектрического покрытия ($n_1>n_0$, $n_1>n_2$, $n_1l_1=n_2l_2=\lambda_0/4$)

Такие многослойные диэлектрические покрытия дают высокую отражательную способность только в ограниченной области длин волн вблизи значения λ_0, для которого оптическая толщина слоев равна $\lambda_0/4$. Обычно наносят от 5 до 15 слоев сульфида цинка ($n_1=2{,}3$) и криолита ($n_2=1{,}35$). С семью слоями легко добиться $R=0{,}9$ в спектральной области шириной порядка 50 нм. Для получения коэффициента отражения $R=0{,}99$ (такие зеркала используют в лазерных резонаторах) надо нанести 11—13 слоев.

Нанесением на поверхность стекла диэлектрических покрытий можно решить и противоположную задачу уменьшения коэффициента отражения (*просветление оптики*). Это очень важно для сложных оптических систем с большим числом преломляющих поверхностей, где даже при малом коэффициенте отражения ($R\approx 0{,}04$) в итоге накапливаются значительные потери света. Очевидно, что нанесенный на стекло слой диэлектрика оптической толщины $nl=\lambda_0/4$ при $n<n_0$ приведет к уменьшению R, так как отраженные от его передней и задней границ волны находятся в противофазе. При $n=\sqrt{n_0}$ весь падающий по нормали свет с длиной волны λ_0 проходит через границу (см. задачу). Использование более сложных трехслойных покрытий позволяет получить сравнительно высокое пропускание в широкой области спектра.

При близких к единице значениях коэффициента отражения R зеркал интерферометра резкость полос, как видно из (5.76), может быть очень большой. Однако следует иметь в виду, что формула (5.76) справедлива для идеального интерферометра, зеркальные поверхности которого абсолютно плоские и строго параллельны друг другу. Дефекты поверхностей приводят к уширению максимумов и уменьшению резкости полос. Если в двух участках интерферометра расстояние между зеркальными поверхностями различается на Δh, то разность хода (при малых углах падения) разли-

чается на $2\Delta h$ и кольца от этих участков смещены друг относительно друга на величину, составляющую $2\Delta h/\lambda$ от расстояния между соседними максимумами. Наложение колец от разных участков зеркал интерферометра уширяет максимумы и искажает их форму. Для достижения теоретической резкости (5.76) при высоких коэффициентах отражения зеркала должны быть изготовлены с большой точностью.

Именно дефекты поверхностей остаются главной причиной, ограничивающей достижимые значения резкости полос и разрешающей способности (см. § 6.6) интерферометра Фабри—Перо.

Остановимся на применении интерферометра Фабри—Перо в спектроскопии. Положение максимумов интерференционной картины в монохроматическом свете зависит от длины волны λ. Изменение их положения при изменении λ служит важной характеристикой спектрального прибора, которая называется *дисперсией**. В данном случае максимумы имеют форму колец в фокальной плоскости линзы (рис. 5.28) и их положение определяется углом θ, а величина $d\theta/d\lambda$ характеризует угловую дисперсию прибора. В соответствии с (5.65) для максимума порядка m

$$\cos\theta = m\lambda/(2h), \tag{5.78}$$

откуда $d\theta/d\lambda = -m/(2h\sin\theta)$. Исключая отсюда m с помощью (5.78), получаем

$$d\theta/d\lambda = -1/(\lambda \operatorname{tg}\theta),\ d\theta/d\lambda \approx -1/(\lambda\theta) \tag{5.79}$$

(последнее справедливо при малых θ, когда $\operatorname{tg}\theta \approx \theta$). Из (5.79) следует, что угловая дисперсия не зависит от расстояния h между зеркалами и других параметров интерферометра. Она неограниченно возрастает при приближении к направлению нормали $\theta = 0$. Знак минус в (5.79) показывает, что с ростом λ угол θ для фиксированного максимума убывает.

Найдем угловое расстояние $\Delta\theta$ между соседними кольцами интерференционной картины в монохроматическом свете с данной длиной волны λ. Из (5.78) видно, что $\cos(\theta + \Delta\theta) - \cos\theta = \lambda/(2h)$. Так как $\cos(\theta + \Delta\theta) \approx \cos\theta - \sin\theta\Delta\theta$, то

$$\Delta\theta = -\lambda/(2h\sin\theta). \tag{5.80}$$

С увеличением угла θ кóльца располагаются все ближе друг к другу. При большом расстоянии h между отражающими плоскостями кольца расположены теснее, чем при малом. По известному угловому расстоянию между кольцами (5.80) и угловой дисперсии (5.79) легко найти интервал длин волн, соответствующий расстоя-

* Термин «дисперсия» имеет здесь совсем другой смысл, чем в гл. 2, где дисперсией называлась зависимость показателя преломления от длины волны.

нию между кольцами, т. е. между максимумами соседних порядков:

$$\Delta\lambda = \frac{\Delta\theta}{d\theta/d\lambda} = \frac{\lambda^2}{2h\cos\theta}. \qquad (5.81)$$

Для малых углов падения $\cos\theta \approx 1$ и $\Delta\lambda = \lambda^2/(2h)$. Спектральный интервал, занимаемый исследуемым излучением, не должен превышать этой величины, чтобы максимумы соседних порядков от отдельных монохроматических компонент излучения не перекрывались. По этой причине интервал $\Delta\lambda$ называют *свободной областью дисперсии* или *постоянной интерферометра*. В § 6.6 показано, что с увеличением расстояния h между пластинами возрастает разрешающая сила прибора, характеризующая способность разделять две близкие по длине волны монохроматические спектральные линии. Однако из (5.81) видно, что увеличение h сопровождается уменьшением области дисперсии $\Delta\lambda = \lambda^2/(2h)$. При типичных значениях ($h = 5$ мм; $\lambda = 0{,}5$ мкм) $\Delta\lambda$ составляет менее 0,03 нм. Это значит, что при работе с интерферометром Фабри—Перо требуется (за очень редким исключением) дополнительный более грубый спектральный прибор для выделения в излучении источника спектрального интервала, не превосходящего дисперсионной области интерферометра. В простейшем случае может быть применен фильтр, но чаще интерферометр «скрещивают» с призменным или дифракционным (см. § 6.6) спектральным прибором. Можно, например, спроецировать интерференционные кольца на плоскость щели спектрографа так, чтобы центр картины совпал с серединой щели. Когда исследуемый спектр состоит из отдельных линий, изображения щели в свете этих линий, получающиеся в соответствующих местах фокальной плоскости спектрографа, оказываются пересеченными поперечными дугами, представляющими участки колец (рис. 5.31). Таким образом можно изучать структуру спектральных линий, состоящих из нескольких близко расположенных компонент, так как каждая из компонент образует свою систему интерференционных колец. Измеряя на спектрограмме, какую долю от расстояния $\Delta\lambda$ между дугами колец соседних порядков составляет расстояние между дугами расщепившихся колец, можно определить спектральные интервалы между компонентами линии, структура которой не разрешается спектрографом. Измерения обычно производят на втором или третьем от центра кольце, где дисперсия еще достаточно велика, но изменяется не столь быстро, как в центре интерференционной картины.

На рис. 5.31 приведен образец такой спектрограммы для излучения изотопа урана-233. Каждое интерференционное кольцо расщеплено на шесть компонент. Такое расщепление спектральной линии возникает, как показывает теория атомных спектров, в результате

5.31
Структура линии $\lambda = 591{,}5$ нм изотопа урана-233

взаимодействия магнитного момента атомного ядра с электронной оболочкой атома (сверхтонкая структура). По характеру и величине сверхтонкого расщепления спектральных линий можно определить основные ядерные константы. В данном случае расщепление на шесть компонент свидетельствует о том, что момент импульса (спин) I ядра исследуемого изотопа урана равен $^5/_2$ (линия расщепляется на $2I+1=6$ компонент).

Интерферометр Фабри—Перо применяется в метрологии для сравнения длины волны излучения эталонного источника с длинами волн других спектральных линий.

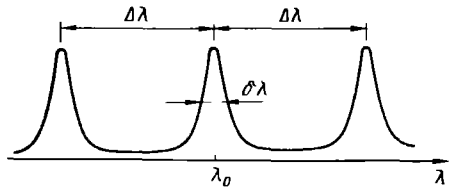
Наиболее точные сравнения концевых мер с первичным эталоном длины (излучением оранжевой линии криптона-86) также производятся с помощью этого интерферометра.

При исследовании формы контуров спектральных линий используют фотоэлектрическую регистрацию. Фотоумножитель (в отличие от фотопластинки) не обладает способностью пространственного разрешения, поэтому для измерения распределения интенсивности в центр интерференционной картины помещают круглую диафрагму и каким-либо способом изменяют оптическую толщину nh интерферометра. Тогда через центр последовательно проходят максимумы разных порядков всех компонент исследуемой линии, и фотоумножитель регистрирует изменения проходящего через отверстие диафрагмы потока излучения.

Размер диафрагмы выбирается так, чтобы вырезаемый ею спектральный интервал не превосходил существенно ширины отдельного максимума в монохроматическом свете — в противном случае уменьшится разрешающая способность (см. § 6.6). Сканирование обычно осуществляется изменением давления и, следовательно, показателя преломления газа, находящегося между отражающими поверхностями. Оптическую толщину можно изменять и перемещением зеркал. Это достигается изменением электрического напряжения, приложенного к изготовленному из пьезоэлектрического материала распорному кольцу между пластинами.

В 1968 г. А. М. Прохоров предложил использовать интерферометр Фабри—Перо в качестве высокодобротного *открытого резонатора* для электромагнитных колебаний оптического диапазона. Такой резонатор был необходим для создания оптического квантового генератора. Первый газовый лазер, осуществленный в 1961 г. А. Джаваном и др., представлял длинную газоразрядную трубку со смесью гелия и неона, помещенную внутрь резонатора, образованного плоскими зеркалами с очень высоким коэффициентом отражения (более 99%). В открытом резонаторе лазера размеры зеркал много меньше расстояния между ними. Теория такого резонатора должна строиться с учетом потерь света в результате дифракции (см. § 6.3) при ограничении площади поперечного сечения пучка света из-за конечных размеров зеркал. Этим она сущест-

венно отличается от элементарной теории интерферометра Фабри—Перо, в которой поверхности зеркал предполагались бесконечными. (Впрочем, в обычных спектроскопических применениях интерферометра Фабри—Перо дифракция несущественна и предположения элементарной теории вполне оправданны.) Впоследствии по ряду причин в газовых лазерах стали использовать открытые резонаторы со сферическими зеркалами, дифракционные потери в которых могут быть значительно меньше (см. § 6.4).



Полосы пропускания интерференционного фильтра

Из других применений многолучевой интерференции отметим *узкополосные оптические фильтры,* пропускающие свет лишь в узком спектральном интервале вблизи заданного значения длины волны. Принцип действия интерференционного фильтра легко понять, представив себе интерферометр Фабри—Перо с очень малым расстоянием между отражающими слоями (от $\lambda/2$ до нескольких длин волн). При падении по нормали света с широким спектральным составом в проходящем свете возникает система максимумов (рис. 5.32), расстояние $\Delta\lambda$ между которыми определяется в соответствии с (5.81) оптической толщиной nh промежутка между отражающими слоями: $\Delta\lambda=\lambda^2/(2nh)$. Подбором h можно совместить один из максимумов с требуемым значением длины волны λ_0. Например, при оптической толщине $nh={}^5/_2\lambda_0$ получим $\Delta\lambda=\lambda_0/5$. Если $\lambda_0=500$ нм, то соседние максимумы, лежащие при $\lambda_0\pm\Delta\lambda$, соответствуют $\lambda_1=400$ нм и $\lambda_2=600$ нм. Они могут быть отделены от нужного максимума λ_0 с помощью обычного стеклянного цветного фильтра. Оставшийся максимум при достаточно высокой отражательной способности R зеркальных слоев может быть очень узок. Его ширина $\delta\lambda$, т. е. полоса пропускания интерференционного фильтра, меньше расстояния $\Delta\lambda$ между соседними максимумами в F раз, где F — резкость полос многолучевой интерференции [см. (5.76)]. При $R\approx0{,}9$ резкость $F\approx30$, и при $\Delta\lambda=100$ нм имеем $\delta\lambda\approx3$ нм. Из падающего по нормали белого света такой фильтр выделит вблизи $\lambda_0=500$ нм узкий спектральный интервал шириной порядка 3 нм. Чтобы фильтр не давал заметного ослабления света в этой полосе, в качестве отражающих поверхностей используют многослойные диэлектрические покрытия.

Контрольные вопросы

- Чем отличаются полосы равного наклона в двухлучевой и многолучевой интерференционных картинах?
- Каким условием определяется положение максимумов и форма полос в

наблюдаемой с интерферометром Фабри — Перо интерференционной картине в проходящем свете?

От каких параметров, характеризующих интерферометр, зависит резкость полос?

Какими преимуществами обладают зеркала с многослойными диэлектрическими покрытиями по сравнению с металлическими? В чем заключается принцип их действия?

Какие факторы ограничивают практически достижимую резкость полос в интерферометре Фабри — Перо?

Что такое дисперсионная область интерферометра? Как она зависит от его толщины? Почему в спектроскопических исследованиях интерферометр используют совместно с более грубым спектральным прибором?

Как устроены интерференционные оптические фильтры?

Задача

Каким показателем преломления n должен обладать материал для просветляющего покрытия на поверхности стекла с показателем преломления $n_0 = 1{,}52$?

◇ Коэффициент отражения на границе слоя диэлектрика с $n \leqslant 1.5$ невелик ($R < 0{,}04$), поэтому в первом приближении можно не учитывать многократные отражения. При оптической толщине покрытия $nl = \lambda_0/4$ отраженные от передней и задней границ слоя волны будут в противофазе. Чтобы они полностью погасили друг друга в результате интерференции, их амплитуды должны быть одинаковы (по модулю). Используя формулы (5.66) для амплитудных коэффициентов отражения, получаем

$$(1-n)/(1+n) \approx (n-n_0)/(n+n_0),$$

откуда $n = \sqrt{n_0} = 1{,}23$. Прозрачные диэлектрики со столь малым показателем преломления неизвестны. На практике для таких покрытий обычно применяют криолит ($n = 1{,}35$) и фтористый магний ($n = 1{,}38$). При однослойном покрытии этими диэлектриками полное просветление не достигается.

Учет многократных отражений на границах слоя не меняет полученного результата ($n = \sqrt{n_0}$). Это можно показать либо суммированием комплексных амплитуд всех отраженных волн [они, как и в (5.70), образуют геометрическую прогрессию], либо непосредственным применением граничных условий на поверхностях слоя (непрерывность тангенциальных компонент векторов **E** и **B**). При оптической толщине слоя в четверть волны это дает следующее выражение для амплитуды E_1 отраженной волны:

$$E_1 = (\rho_1 - \rho_2) E_0 / (1 - \rho_1 \rho_2),$$

где ρ_1 и ρ_2 — амплитудные коэффициенты отражения на передней и задней границах слоя. Отсюда $E_1 = 0$ (просветление) при $\rho_1 = \rho_2$, что опять приводит к $n = \sqrt{n_0}$. Таким образом, этот результат точный

5.

ДИФРАКЦИЯ СВЕТА

● В явлениях дифракции, как и в интерференции, на первый план выступают волновые свойства света. Под дифракцией света обычно понимают отклонения от простых законов распространения света, описываемых геометрической оптикой. Дифракцию можно наблюдать, например, когда на пути распространения света находятся препятствия, т. е. непрозрачные тела произвольной формы (обычно их называют в таких случаях экранами), или когда свет проходит сквозь отверстия в экранах. Геометрическая оптика (закон прямолинейного распространения света в однородной среде) предсказывает существование за экраном области тени, резко отграниченной от тех мест, куда свет попадает. Но тщательный опыт показывает, что вместо резкой границы между светом и тенью получается довольно сложная картина распределения освещенности, состоящая из темных и светлых участков — дифракционных полос. Дифракция выражена тем сильнее, чем меньше размеры экранов (или отверстий в них) и чем больше длина волны. Характерные особенности дифракционных явлений в оптике обусловлены тем, что здесь, как правило, размеры экранов много больше длины волны. Поэтому наблюдать дифракцию света можно только на достаточно больших расстояниях от преграды.

● Теория дифракции света дает строгое обоснование геометрической оптике и определяет условия ее применимости. В теории дифракции получают свое решение и многие вопросы инструментальной оптики, требующие выхода за рамки геометрической оптики, такие, как предельное разрешение оптических систем и спектральных приборов или структура оптического изображения.

6.1. Принцип Гюйгенса — Френеля. Зоны Френеля

Первое объяснение дифракции света было дано Френелем в 1818 г. В своем мемуаре он показал, что количественное описание дифракционных явлений возможно на основе построения Гюйгенса, если его дополнить принципом интерференции вторичных волн. Кирхгоф в 1882 г. дал строгое математическое обоснование принципу Гюйгенса—Френеля.

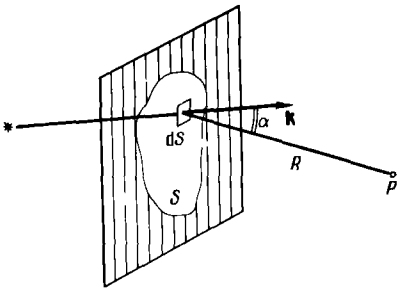
В рамках электромагнитной теории света для описания дифракционных явлений не требуется вводить какие-либо новые принципы. Но точное решение задачи о распространении света на основе уравнений Максвелла с соответствующими граничными условиями представляет большие математические трудности. В большинстве случаев, представляющих практический интерес, вполне достаточным оказывается приближенный метод решения задачи о распределении света вблизи границы между светом и тенью, основанный на принципе Гюйгенса—Френеля.

Согласно *принципу Гюйгенса*, каждую точку, в которую пришла волна от источника, можно принять за центр вторичных волн, распространяющихся во все стороны. Результирующая волна рассматривается как наложение вторичных волн. Гюйгенс считал, что отдельные вторичные волны не обладают периодичностью, что они очень слабы и заметное действие производят только на их огибающей. При таком произвольном допущении принцип Гюйгенса дает лишь некоторый рецепт построения волновых фронтов, т. е. поверхностей, до которых дошло световое возмущение. Построения Гюйгенса наглядно объясняют законы прямолинейного распространения, отражения и преломления света. Но в этих построениях не используется понятие длины волны, поэтому они не позволяют определить условия применимости упомянутых законов.

Френель вложил в принцип Гюйгенса ясное физическое содержание, отказавшись от искусственного предположения об огибающей вторичных волн и рассматривая полное световое поле как результат *интерференции вторичных волн*. При этом не только получает физическое объяснение рецепт Гюйгенса (к точкам на огибающей все вторичные волны приходят в одинаковых фазах), но и появляется возможность расчета распределения светового поля в пространстве. Изучая распределение света вблизи границы между светом и тенью на основе *принципа Гюйгенса—Френеля,* можно получить количественное описание дифракционных явлений.

Рассмотрим какой-либо экран с отверстием, через которое проходит свет от данного источника (рис. 6.1). Источник пока будем считать точечным и монохроматическим. Размеры отверстия много больше длины волны света. Будем под E понимать любую из компонент векторов **E** или **B** электромагнитного поля световой волны, опуская при этом множитель $e^{-i\omega t}$, определяющий зависимость от времени. Задача состоит в определении E_P в любой точке P за экраном. При приближенном решении этой задачи по методу Френеля делается предположение, что напряженность E в точках отверстия такова, какой она была бы в случае свободного распространения волны от источника вообще при отсутствии какого бы то ни было экрана, и что в точках, находящихся непосред-

ственно за экраном, напряженность поля равна нулю. Очевидно, что в этом предположении совершенно не учитываются специфические свойства экрана, т. е. материала, из которого он сделан, — считается, что это не играет никакой роли. Существенна только форма края отверстия (или края экрана) и совершенно несущественна форма удаленной от краев части экрана. Опыт показывает, что обсуждаемое предположение справедливо, когда размеры отверстия и расстояния до источника и точки наблюдения много больше длины волны, т. е. когда отклонения от геометрической оптики малы. Оно нарушается, например, для узкой щели, ширина которой значительно меньше длины световой волны.



6.1
К формулировке принципа Гюйгенса—Френеля

Проведем мысленно произвольную поверхность S, закрывающую отверстие в экране и ограниченную краями отверстия (рис. 6.1). Разделим эту поверхность на элементарные участки площадью dS, малые по сравнению с размерами отверстия, но большие по сравнению с длиной волны. Можно считать, что каждый из этих участков сам становится источником световой волны, распространяющейся во все стороны. Напряженность dE_P, создаваемая элементарным участком dS в точке P, пропорциональна напряженности E в самом участке dS (какой она была бы при отсутствии экрана) и проекции dS_n площади dS этого участка на плоскость, перпендикулярную вектору $\mathbf{k}$ луча, пришедшего из источника света в dS. Последнее связано с тем, что при любой форме участка dS через него будут проходить одни и те же лучи от источника, если только проекция dS_n будет неизменной, а потому и вклад его в напряженность в точке P будет тем же самым.

При вычислении вклада участка dS в E_P нужно учесть изменения амплитуды и фазы вторичной волны при ее распространении от dS к P. Это приводит к появлению в выражении для dE_P множителя e^{ikR}/R, где R — расстояние от dS до P, а $k=2\pi/\lambda$ — волновое число. Таким образом,

$$dE_P = K(\alpha) E \frac{e^{ikR}}{R} dS_n, \qquad (6.1)$$

где $K(\alpha)$ — некоторый коэффициент наклона, учитывающий зависимость амплитуды вторичных волн от угла α между вектором $\mathbf{k}$ и направлением на точку наблюдения. Естественно предположить, что модуль $K(\alpha)$ максимален в первоначальном направлении распространения света, т. е. при $\alpha=0$, и плавно убывает с увеличением α. Многие практически важные задачи дифракции можно решить при этих весьма общих предположениях относительно $K(\alpha)$, не уточняя конкретного вида его зависимости от α.

Из теории Кирхгофа, основанной на том, что напряженность поля световой волны удовлетворяет волновому уравнению, следует

$$K(\alpha)=k(1+\cos\alpha)/(4\pi i)=(1+\cos\alpha)/(2\lambda i). \tag{6.2}$$

При малых углах дифракции ($\alpha \ll 1$) можно положить $\cos\alpha \approx 1$ и $K(\alpha)\approx k(2\pi i)=-i/\lambda$.

Полное поле в точке P представляет собой суперпозицию полей (6.1) вторичных волн от всех элементов dS поверхности, закрывающей отверстие в экране:

$$E_p=\int\limits_S K(\alpha)E\frac{e^{ikR}}{R}\,dS_n. \tag{6.3}$$

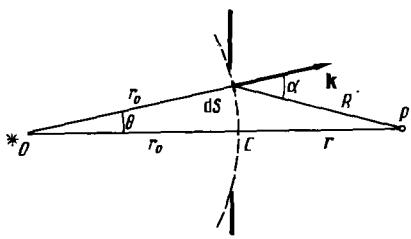
В рассматриваемом приближении интеграл (6.3) по поверхности S не зависит от формы этой поверхности.

Формула (6.3) дает математическое выражение принципа Гюйгенса—Френеля. Применим ее для определения напряженности поля в точке P за круглым отверстием в экране. Будем считать, что точечный источник O и точка P лежат на прямой, проходящей через центр отверстия и перпендикулярной его плоскости (рис. 6.2). В качестве вспомогательной поверхности S выберем часть сферы радиусом r_0 с центром в источнике, проходящей через края отверстия. В соответствии с основным предположением напряженность поля на ней будет такой же, как при отсутствии экрана. Она одинакова на всех ее элементах dS и равна

$$E=E_0e^{ikr_0}. \tag{6.4}$$

В качестве элемента dS удобно взять на сфере кольцо, все точки которого лежат на одинаковом расстоянии R от P. Площадь такого кольца $dS=dS_n=2\pi r_0\sin\theta d\theta$. Выберем R за переменную интегрирования в (6.3). Из рис. 6.2 видно, что $R^2=r_0^2+(r_0+r)^2-2r_0(r_0+r)\cos\theta$, где r — расстояние от сферы до точки P. Дифференцируя это уравнение (при постоянных r_0 и r), получаем $RdR=r_0(r_0+r)\sin\theta d\theta$, откуда

$$dS_n=2\pi\frac{r_0R}{r_0+r}\,dR. \tag{6.5}$$



6.2
К вычислению напряженности поля в точке P

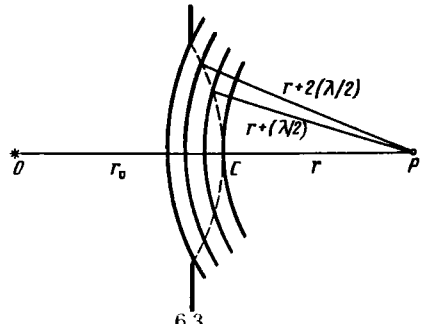
Подставляя (6.4) и (6.5) в (6.3), получаем

$$E_p=2\pi\frac{r_0}{r_0+r}E_0e^{ikr_0}\int\limits_{r_0}^{R_{max}}K(R)e^{ikR}\,dR. \tag{6.6}$$

Для приближенного вычисления интеграла в (6.6) воспользуемся следующим приемом. Построим сферы с центром в точке P, радиусы которых равны r, $r+\lambda/2$,

$r+2(\lambda/2)$, ..., $r+n(\lambda/2)$, ... (рис. 6.3). Они разобьют поверхность S на кольцевые области, называемые *зонами Френеля*. Из построения ясно, что вторичные волны от границ двух соседних зон приходят в точку наблюдения в противофазе.

Легко показать, что радиус окружности, отделяющей n-ю зону Френеля от $(n+1)$-й, приближенно равен



6.3
Построение зон Френеля

$$R_n = \sqrt{n\lambda r r_0/(r+r_0)}. \tag{6.7}$$

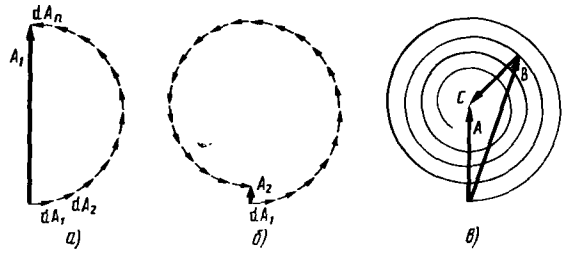
Отсюда ясно, что площади зон Френеля приблизительно одинаковы. Так как $\lambda \ll r$, то изменение R в пределах одной зоны незначительно и плавную функцию $K(R)$ в подынтегральном выражении (6.6) при интегрировании по n-й зоне можно считать постоянной. В этом приближении вклад n-й зоны в интеграл (6.6) легко вычисляется:

$$K_n \int\limits_{r+(n-1)\lambda/2}^{r+n\lambda/2} e^{ikR}\,dR = (-1)^n \frac{2K_n}{ik} e^{ikr}, \tag{6.8}$$

где K_n — значение функции $K(R)$ в пределах n-й зоны. Вычисление напряженности поля в точке P сводится к суммированию знакопеременного ряда:

$$E_p = (4\pi i r_0 E_0/k) e^{ik(r+r_0)} (K_1 - K_2 + K_3 - K_4 + \ldots). \tag{6.9}$$

Прежде чем анализировать полученные результаты, приведем наглядную геометрическую интерпретацию вычисления напряженности поля в точке P на основе принципа Гюйгенса—Френеля. Изобразим колебание напряженности поля в точке P, вызванное вторичной волной от элементарного участка dS волновой поверхности, лежащего в центре C отверстия (т. е. на линии OP), с помощью векторной диаграммы (рис. 6.4). Этому колебанию на ней сопоставляется элементарный вектор dA_1, вращающийся по часовой стрелке с угловой скоростью ω, равной частоте излучения источника. Колебание, вызванное вторичной волной от следующего (такого же по площади) элементарного кольцевого участка, изображается таким же по модулю вектором dA_2, но повернутым относительно dA_1 на небольшой угол, так как оно несколько отстает по фазе. Колебанию, приходящему в точку P от участка, прилегающего к границе первой зоны Френеля, будет соответствовать вектор dA_n, повернутый относительно dA_1 на π, так как по самому определению зон Френеля разность хода соответствующих им вторичных волн равна $\lambda/2$.



6.4
Векторные диаграммы для результирующего колебания в точке P

Результирующее колебание в P, вызываемое волнами от всей 1-й зоны Френеля, изображается на диаграмме рис. 6.4, *а* вектором A_1, замыкающим ломаную линию, образованную векторами dA_1, $dA_2,\ldots,dA_n$. Ему соответствует первое слагаемое в (6.9). В пределе, когда все $dS \to 0$, он проходит по диаметру полуокружности. Продолжим построение дальше. Векторная диаграмма результирующего колебания в P от двух первых зон Френеля показана на рис. 6.4, *б*. При строгом равенстве амплитуд складываемых колебаний dA_i от элементарных участков амплитуда результирующего колебания от двух открытых зон была бы равна нулю, т. е. вторичные волны в результате интерференции полностью гасили бы друг друга. Но коэффициент наклона $K(\alpha)$, убывающий по мере увеличения α, характеризует постепенное уменьшение амплитуд вторичных волн, т. е. модулей элементарных векторов dA_i. Поэтому амплитуда A_2 колебания от двух зон имеет конечное, хотя и очень малое, значение. Этому соответствуют два первых слагаемых в (6.9).

Таким образом, освещенность в точке P по мере увеличения диаметра отверстия в экране изменяется немонотонно. Пока открывается первая зона Френеля, освещенность в P увеличивается и достигает максимума при полностью открытой зоне. По мере открывания второй зоны Френеля освещенность убывает и при полностью открытых двух зонах уменьшается почти до нуля. Затем освещенность увеличивается снова, и т. д. К таким же выводам мы придем, если вместо увеличения диаметра отверстия будем приближать к нему точку наблюдения P вдоль прямой PO (см. рис. 6.3). Так как радиусы зон Френеля в соответствии с (6.7) зависят от расстояния r от P до экрана, то при этом будут последовательно открываться одна, две зоны и т. д. Эти на первый взгляд парадоксальные результаты, предсказываемые на основе принципа Гюйгенса—Френеля, хорошо подтверждаются экспериментом. Заметим, что они находятся в противоречии с предсказаниями геометрической оптики, согласно которой освещенность в точке P, лежащей на одной линии с источником и центром круглого отверстия, не зависит от диаметра отверстия.

Перейдем к предельному случаю, когда радиус отверстия в экране неограниченно возрастает. Это равносильно отсутствию экрана вообще, т. е. свободному распространению из O в P. На векторной диаграмме этому предельному случаю соответствует спираль, витки которой постепенно сжимаются, закручиваясь вокруг фокуса, находящегося в центре окружности (рис. 6.4, *в*). Колебание в P, вызываемое вторичными волнами от полностью открытого волнового фронта, изображается вектором A. Ему соответствует весь знакопеременный ряд в (6.9). Сравнивая рис. 6.4 *а* и *в*, видим, что $A = A_1/2$: амплитуда колебания в P при отсутствии экрана вдвое меньше, а интенсивность в четыре раза меньше, чем при наличии экрана с круглым отверстием, открывающим только одну первую зону Френеля.

Таким образом, сумма всего ряда (6.9), описывающая амплитуду колебания напряженности в точке P при полностью открытом волновом фронте, равна половине первого члена этого ряда:

$$E_p = (2\pi i r_0 E_0/k)\, e^{ik(r + r_0)} K_1. \tag{6.10}$$

Этим результатом можно воспользоваться для того, чтобы определить коэффициент K_1. В самом деле, свободно распространяющаяся из точечного источника O сферическая волна, напряженность поля которой на поверхности S, т. е. на расстоянии r_0 от источника, равна $E_0 e^{ikr_0}$ в соответствии с (6.4), в точке P, т. е. на расстоянии $r_0 + r$, имеет напряженность

$$E_p = E_0 \frac{r_0}{r_0 + r} e^{ik(r_0 + r)}. \tag{6.11}$$

Сравнивая правые части выражений (6.10) и (6.11), находим

$$K_1 = k/(2\pi i) = -i/\lambda = (1/\lambda)\, e^{-i\pi/2}. \tag{6.12}$$

что совпадает с (6.2) при $\alpha \ll 1$ в теории Кирхгофа. Множитель $\exp(-i\pi/2)$ означает, что вторичные волны опережают исходную волну по фазе на $\pi/2$ [напомним, что в выражениях (6.10) и (6.11) для E_p опущен множитель $\exp(-i\omega t)$, описывающий зависимость напряженности поля от времени].

Наиболее неожиданным в полученных выше результатах представляется, пожалуй, то, что при отверстии в экране, открывающем две зоны Френеля (или вообще небольшое четное число зон), освещенность в точке P равна нулю. Не менее неожиданно и то, что за круглым экраном как раз в центре его геометрической тени освещенность не равна нулю. Если экран перекрывает лишь несколько первых зон Френеля, то освещенность в центре тени почти такая же, как и при отсутствии экрана. В этом легко убедиться с помощью векторной диаграммы (рис. 6.4, *в*), если представить

вектор A, изображающий колебание напряженности поля в точке P при полностью открытой волновой поверхности, как сумму двух векторов, один из которых (B на рис. 6.4, *в*) изображает колебание от тех зон Френеля, которые перекрыты экраном, а другой (C) — колебание от открытого участка волновой поверхности, лежащего за пределами экрана. Вектор C лишь немного уступает по модулю вектору A.

Это предсказание теории Френеля произвело сильное впечатление на его современников. В 1818 г. член конкурсного комитета Французской академии Пуассон, рассматривавший представленный на премию мемуар Френеля, пришел к выводу о том, что в центре тени маленького диска должно находиться светлое пятно, но счел этот вывод столь абсурдным, что выдвинул его как возражение против волновой теории света, развивавшейся Френелем. Однако другой член того же комитета Араго выполнил эксперимент, показавший, что это удивительное предсказание правильно. Долгий спор между приверженцами корпускулярной и волновой теорий света был решен в пользу волновой теории.

Как уже отмечалось, когда круглое отверстие в экране открывает одну первую зону Френеля, интенсивность в точке наблюдения P в четыре раза больше, чем при полностью открытой волновой поверхности. Интенсивность в P можно во много раз усилить, если изготовить экран, который кроме 1-й зоны открывает третью, пятую и т. д., т. е. все нечетные, зоны Френеля. Вторичные волны от всех открытых зон Френеля будут приходить в точку наблюдения в одинаковой фазе и в результате интерференции усилят друг друга. Такой экран называют *зонной пластинкой* (рис. 6.5). Можно изготовить аналогичный экран, который открывает все четные зоны Френеля. Действие зонной пластинки описывается формулой (6.9), если во входящем в нее ряду оставить члены только одного знака. Зонная пластинка, содержащая n открытых зон; создает в P освещенность приблизительно в n^2 раз большую, чем отверстие в одну зону Френеля.

6.5
Зонная пластинка Френеля

Усиление интенсивности света зонной пластинкой аналогично фокусирующему действию линзы. Расстояния от пластинки до источника O и до его «изображения» P связаны таким же соотношением, что и соответствую-

щие величины для линзы. В этом легко убедиться, переписав формулу (6.7) в виде

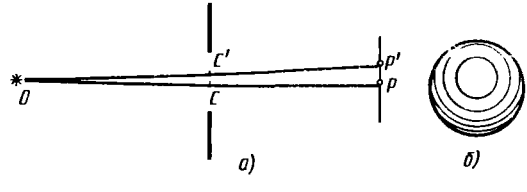
$$\frac{1}{r_0}+\frac{1}{r}=\frac{1}{f}, \tag{6.13}$$

$$\frac{1}{f}=\frac{n\lambda}{R_n^2}=\frac{\lambda}{R_1^2}, \tag{6.14}$$

где «фокусное расстояние» f — постоянная для данной пластинки величина (так как $R_n \sim \sqrt{n}$, правая часть в (6.14) не зависит от n). Но в отличие от линзы зонная пластинка имеет несколько фокусов. В самом деле, приблизив точку наблюдения P к пластинке, можно найти такое ее положение, когда в центральном светлом круге пластинки уместятся три первые зоны Френеля (а не одна). Следующие три зоны будут перекрыты темным кольцом. Затем три зоны совместятся со светлым кольцом и т. д. Действия соседних зон Френеля практически уничтожают друг друга, поэтому результат будет таким, как в случае открытых, 1, 7, 13-й зонах и т. д. Поэтому в P получится максимум, хотя и более слабый, чем в основном фокусе. Соответствующее ему фокусное расстояние $f_1 = f/3$. Аналогично для фокусов высших порядков можно написать $f_n = f/(2n+1)$, где n — целые числа. Отрицательным значениям n соответствуют дифрагировавшие волны, расходящиеся из мнимых фокусов, расположенных перед пластинкой (т. е. со стороны источника).

Интенсивность света в главном фокусе можно увеличить еще в четыре раза (по сравнению с зонной пластинкой), если изменить на π фазы вторичных волн, исходящих от всех зон с четными (или, наоборот, с нечетными) номерами. Такая пластинка была изготовлена Вудом: химическим травлением в нужных местах толщина стеклянной пластинки уменьшалась на $(n-1)\lambda/2$. В этом случае вторичные волны от всех точек волновой поверхности приходят в P в одинаковых фазах.

Метод кольцевых зон Френеля позволяет сравнительно просто найти интенсивность при дифракции света на круглом отверстии для точки наблюдения P, лежащей на оси симметрии (см. рис. 6.2 или 6.3). Расчет распределения интенсивности для всей дифракционной картины оказывается значительно сложнее. Чтобы найти напряженность $E_{P'}$ поля в какой-либо точке P', не лежащей на оси OC (рис. 6.6, a), можно построить кольцевые зоны, центр которых находится в точке C' на прямой OP'. Отверстие экрана расположится не концентрически по отношению к этим зо-



Зоны Френеля для точки наблюдения P'

нам: открытая отверстием часть зон выглядит так, как показано на рис. 6.6, *б*. Действие вторичных волн в точке P' зависит от того, какая часть каждой из зон открыта. Поэтому точный подсчет $E_{P'}$ сложен, но ясно, что при удалении P' от P периодически будут встречаться места с большей и меньшей интенсивностью. Так как вся картина должна обладать круговой симметрией, то вокруг точки P образуются чередующиеся более светлые и менее светлые кольца. Число этих колец и их положение зависят от числа зон Френеля (для точки P), умещающихся на площади отверстия. Пока на площади отверстия укладывается лишь одна центральная зона или ее часть, интенсивность максимальна в центре (т. е. в точке P) и монотонно убывает при удалении от P. Когда отверстие открывает две зоны Френеля, в точке P получается темный кружок, а вокруг него — светлое кольцо, к которому и перемещается максимум интенсивности. С увеличением числа открытых зон увеличивается и число максимумов и минимумов в радиальном распределении интенсивности. Когда на площади отверстия укладывается большое число зон Френеля, то интенсивность вблизи точки P получается почти равномерной и лишь у краев геометрической тени отверстия наблюдается чередование весьма узких светлых и темных кольцевых полос. Радиальное распределение интенсивности при этом практически такое же, как и вблизи границы тени от прямолинейного края экрана (см. рис. 6.9).

Совершенно аналогично в дифракционной картине от круглого диска центральную светлую точку (пятно Пуассона) окружает система чередующихся темных и светлых колец.

Заметим, что характер дифракционных явлений (т. е. распределение интенсивности в дифракционной картине) определяется числом зон Френеля, перекрываемых экраном (или отверстием в экране), а не абсолютными размерами экранов или отверстий. Радиусы зон Френеля определяются формулой (6.7). Для первой зоны $R_1 = \sqrt{\lambda r r_0/(r+r_0)} = \sqrt{\lambda f}$, где $1/f = 1/r_0 + 1/r$. Если $r = r_0$, т. е. экран расположен посередине между источником и точкой наблюдения, то $f = r/2$. При дифракции плоской волны, что соответствует бесконечно удаленному от экрана точечному источнику, $r_0 \to \infty$ и $f = r$. Отношение радиуса R_1 первой зоны Френеля к линейному размеру D экрана или отверстия

$$p = R_1/D = \sqrt{\lambda f}/D \qquad (6.15)$$

полностью определяет условия наблюдения, при которых дифракционные явления становятся существенными и распределение интенсивности заметно отличается от предсказаний геометрической оптики. Параметр p лежит в основе классификации дифракционных явлений. Когда $p \ll 1$, число зон Френеля, перекрываемых экраном или отверстием, велико, дифракционные эффекты незначительны и распределение интенсивности приближенно описывается законами геометрической оптики (прямолинейным распростране-

нием света). При $p \approx 1$ (заметная часть одной зоны или небольшое число зон) наблюдается сложное распределение интенсивности, называемое *дифракцией Френеля*. При $p \gg 1$ отверстие перекрывает малую часть первой зоны Френеля. В этом случае, называемом *дифракцией Фраунгофера*, дифракционная картина упрощается. Случай дифракции Фраунгофера имеет большое практическое значение для решения многих вопросов инструментальной оптики и подробно рассмотрен в § 6.3.

При изменении расстояния f в q раз и размера экрана D в $\sqrt{q}$ раз значение параметра $p=\sqrt{\lambda f}/D$ останется прежним. Поэтому дифракционные картины от подобных экранов, наблюдаемые в таких условиях, также будут подобными. В эффектных опытах В. К. Аркадьева была получена фотография дифракционной картины на расстоянии 11 км от руки, держащей тарелку, с отчетливо различимым пятном Пуассона в центре тени. В действительности фотопластинка располагалась на расстоянии 40 м, что сравнительно легко осуществить, а преграда была заменена геометрически подобной моделью из жести, уменьшенной в $\sqrt{11000/40} \approx 16{,}5$ раза.

Изучение дифракции Френеля в этом параграфе было проведено в предположении, что источник света точечный, а излучаемый им свет монохроматический. В случае протяженного источника свет от каждого его элемента дает свою дифракционную картину. Вследствие полной независимости (некогерентности) света отдельных элементов происходит просто сложение интенсивностей в каждой точке и результат дифракции определяется наложением таких несколько смещенных друг относительно друга дифракционных картин. Для наблюдения дифракции на опыте размеры источника должны быть достаточно малы, чтобы темные и светлые полосы картин от его отдельных элементов не перекрывались. Аналогично, в случае немонохроматического источника различные спектральные компоненты его излучения создают несовпадающие дифракционные картины, так как размеры зон Френеля зависят от длины волны. Наблюдаемое распределение интенсивности соответствует наложению этих дифракционных картин.

Контрольные вопросы

- В чем заключается принцип Гюйгенса — Френеля? Приведите его математическую формулировку.
- Как выбирается вспомогательная поверхность S при нахождении дифракционной картины от отверстия в непрозрачном экране? Зависит ли результат от выбора этой поверхности?
- Освещенность в точке P за круглым отверстием, открывающим одну зону Френеля, примерно в четыре раза больше, чем при полностью открытом волновом фронте. Если увеличить вдвое площадь отверстия, освещенность в P уменьшится почти до нуля, несмотря на удвоение светового потока. Как эти факты согласовать с законом сохранения энергии?

- □ Покажите с помощью векторной диаграммы, что освещенность в центре геометрической тени круглого диска, перекрывающего небольшое число зон Френеля, почти такая же, как и в освещенной области.
- □ Во сколько раз интенсивность в главном фокусе зонной пластинки больше, чем в фокусе n-го порядка?
- □ В чем заключается принцип подобия дифракционных картин?
- □ При каких условиях происходит дифракция Френеля? дифракция Фраунгофера?
- □ Объясните, почему дифракционные полосы нельзя наблюдать при протяженном или при немонохроматическом источнике света.

Задачи

1. Точечный источник монохроматического света расположен на расстоянии r_0 от круглого отверстия, а экран — с противоположной стороны на расстоянии r. При каких значениях радиуса R отверстия центр наблюдаемых на экране дифракционных колец будет светлым и при каких — темным?

◇ Ответ: при $R=\sqrt{2n\lambda f}$ центр темный, при $R=\sqrt{(2n-1)\lambda f}$ — светлый. Здесь $f=rr_0/(r+r_0)$, $n=1, 2, \ldots$. ◇

2. Монохроматический точечный источник расположен на оси зонной пластинки на расстоянии a от нее. Наиболее яркое изображение источника получается на расстоянии b от пластинки. На каких расстояниях получаются другие изображения источника?

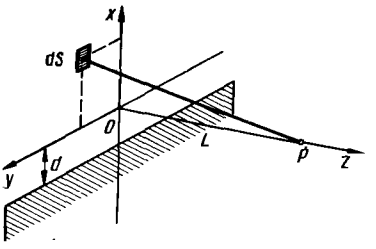
◇ Ответ: положения всех изображений b_k определяются формулой $1/a+1/b_k=1/f_k$, где $f_k=f/(2k+1)$ $(k=0, 1, 2, \ldots)$ — фокус k-го порядка, $f=ab/(a+b)$ — главный фокус пластинки. ◇

6.2. Дифракция Френеля на прямолинейном крае экрана

Принцип Гюйгенса — Френеля можно применить для нахождения распределения интенсивности вблизи границы тени, отбрасываемой краем большого экрана. Когда точка наблюдения P (рис. 6.7) находится на конечном расстоянии от экрана, задерживающего свет, для определения E_p в интеграле (6.3) играет роль сравнительно небольшой участок волновой поверхности, лежащий вблизи края экрана. В таких условиях край практически любого экрана можно считать прямолинейным.

Ограничимся случаем плоской волновой поверхности падающей волны, что соответствует бесконечно удаленному точечному источнику (или точечному источнику в фокусе линзы), и введем в этой плоскости оси x и y прямоугольной системы координат (рис. 6.7). Пусть ось z проходит через точку наблюдения P, находящуюся на расстоянии L от плоскости xy, а прямолинейный край экрана проходит в плоскости xy параллельно оси y на расстоянии d от нее (при $x=-d$). Основной интерес представляет распределение интенсивности вблизи края геометрической тени, поэтому можно считать, что $d \ll L$. Роль поверхности S при применении принципа Гюйгенса — Френеля (6.3) будет играть не закрытая экраном часть плоскости xy. Во всех ее точках поле E

падающей плоской волны одинаково. Рассматриваемому здесь случаю дифракции Френеля соответствуют малые отклонения от геометрической оптики (т. е. от прямолинейного распространения света), поэтому основной вклад в интеграл (6.3) при малых углах дифракции дают участки плоскости xy, близкие к началу координат. Тогда для коэффициента наклона $K(\alpha)$ можно принять значение $K(\alpha)\approx K_1=1/(\lambda i)=$ $=\text{const}$ из (6.12), а расстояние R от dS до P — приближенно записать в следующем виде (рис. 6.7):



6.7
Дифракция Френеля на прямолинейном крае экрана

$$R=\sqrt{L^2+x^2+y^2}=L\sqrt{1+(x^2+y^2)/L^2}\approx L+(x^2+y^2)/(2L).$$

В множителе $1/R$ [см. (6.3)], характеризующем уменьшение амплитуды вторичных волн с увеличением расстояния, можно положить $R\approx L=\text{const}$. Тогда вклад участка волновой поверхности в виде параллельной оси y полосы, простирающейся от $x=0$ до $x=x_1$, в напряженность E_P поля принимает вид

$$E_P=-\frac{iE}{\lambda L}\iint e^{ik[L+(x^2+y^2)/(2L)]}\,dx\,dy=$$

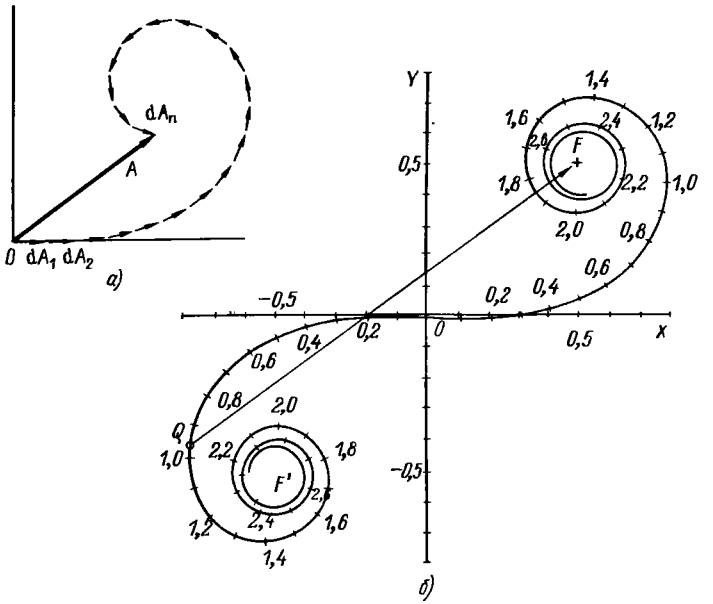
$$=-\frac{iE}{\lambda L}e^{ikL}\int\limits_{-\infty}^{\infty}e^{iky^2/(2L)}\,dy\int\limits_0^{x_1}e^{ikx^2/(2L)}\,dx.$$

Интегрирование по y даст постоянный (т. е. не зависящий от ширины x_1 выделенной полосы волновой поверхности) множитель, не представляющий интереса. Опуская его и другие постоянные множители, можно написать

$$E_P\sim\int\limits_0^{x_1}e^{ikx^2/(2L)}\,dx\sim\int\limits_0^{\eta_1}e^{i\pi\eta^2/2}\,d\eta, \qquad (6.16)$$

где вместо x введена новая безразмерная переменная интегрирования по формуле $kx^2/L=\pi\eta^2$, т. е. $\eta=x\sqrt{2/(\lambda L)}$.

Вычисление E_P по формуле (6.16) удобно проиллюстрировать с помощью векторной диаграммы аналогично тому, как это было сделано в задаче о дифракции на круглом отверстии (см. § 6.1). Колебание в P, вызываемое вторичной волной от элементарной полоски волновой поверхности шириной dx, расположенной вдоль оси y, т. е. при $x=0$, изобразим вектором dA_1 (рис. 6.8, a). Колебание от следующей полоски изобразится таким же по модулю вектором dA_2, повернутым относительно dA_1 на небольшой угол, так как



6.8
Векторная диаграмма для колебания в точке P от полосы волновой поверхности (*а*) и спираль Корню (*б*)

эта вторичная волна проходит до P большее расстояние и несколько отстает по фазе. В дальнейшем угол между соседними векторами элементарных колебаний dA_i и dA_{i+1} становится все больше, так как запаздывание по фазе вторичной волны от элементарной полоски, находящейся на расстоянии x от оси y, пропорционально *квадрату* этого расстояния x^2 [см. (6.16)]. Этим рассматриваемая векторная диаграмма отличается от диаграммы на рис. 6.4 для дифракции на круглом отверстии, где углы между любыми соседними векторами dA_i и dA_{i+1} одинаковы, так как там фаза вторичных волн растет линейно с увеличением R [см. (6.6)].

Колебание в P от широкой полосы волновой поверхности изобразится суммой векторов dA_i от всех укладывающихся на ней элементарных полосок dx (вектор A на рис. 6.8, *а*). В пределе, когда ширина dx каждой элементарной полоски стремится к нулю, цепочка векторов dA_1, dA_2, ... превращается в плавную кривую, называемую *спиралью Корню* (рис. 6.8, *б*). Она состоит из двух симметричных ветвей, закручивающихся вокруг фокусов F и F'. Ее левая половина описывает действие вторичных волн от участков волновой поверхности, лежащих ниже оси y (при $x<0$). Колебание в P от всей волновой поверхности, лежащей выше оси y на рис. 6.7 (т. е. при $0<x<\infty$), изображается вектором, проведен-

ным из O в правый фокус F спирали Корню. Колебание в P от полной волновой поверхности ($-\infty < x < \infty$) изображается вектором, соединяющим фокусы F и F'. Для нахождения колебания в P от какой-либо полосы волновой поверхности, лежащей между $x=x_1$ и $x=x_2$, нужно построить вектор, который замыкает соответствующий этой полосе участок спирали.

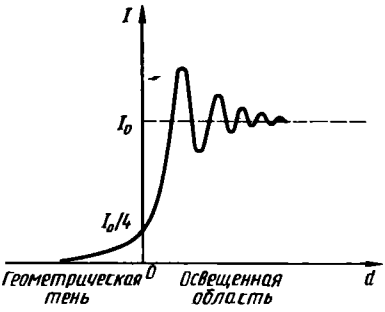
Выражение (6.16) представляет собой параметрическое уравнение спирали Корню в комплексной форме. Если в плоскости векторной диаграммы ввести прямоугольные координаты X и Y так, как показано на рис. 6.8, *б*, то уравнение спирали Корню примет вид

$$X(\eta_1)=\int_0^{\eta_1}\cos(\pi\eta^2/2)\,d\eta, \quad Y(\eta_1)=\int_0^{\eta_1}\sin(\pi\eta^2/2)\,d\eta. \tag{6.17}$$

Функции $X(\eta_1)$ и $Y(\eta_1)$ называются *интегралами Френеля*. Их значения при $\eta_1 \to \pm\infty$ дают координаты фокусов спирали Корню: $X_F=Y_F={}^1/_2$; $X_{F'}=Y_{F'}=-{}^1/_2$. Из (6.17) легко найти, что дифференциал длины дуги спирали Корню $\sqrt{(dX)^2+(dY)^2}$ равен $d\eta_1$. Это значит, что параметр η_1 есть длина дуги спирали, отсчитываемая от точки O на рис. 6.8, *б*. Напомним, что он пропорционален ширине соответствующей полосы волновой поверхности: $\eta_1=x_1\sqrt{2/(\lambda L)}$.

Угловой коэффициент касательной к спирали Корню $\operatorname{tg}\alpha=dY/dX=\operatorname{tg}(\pi\eta_1^2/2)$, откуда $\alpha=\pi\eta_1^2/2$. При $\eta_1=0$ угол $\alpha=0$, т. е. в точке O спираль касается оси X. При $\eta_1=1$ угол $\alpha=\pi/2$, т. е. касательная вертикальна. При $\eta_1=\sqrt{2}$ угол $\alpha=\pi$, т. е. касательная снова горизонтальна, но направлена влево. При $\eta_1=\sqrt{3}$ угол $\alpha=3\pi/2$, т. е. касательная вертикальна и направлена вниз. Таким образом легко проследить, как спираль Корню обвивается вокруг фокусов, совершая при этом бесконечное число витков. Соотношение $\alpha=\pi\eta_1^2/2$ позволяет по заданному значению параметра η_1 найти соответствующую точку на спирали Корню.

Теперь с помощью спирали Корню легко получить распределение интенсивности вблизи края геометрической тени при дифракции плоской волны на прямолинейном крае экрана. При любом расположении точки наблюдения P относительно края экрана верхняя часть волновой поверхности полностью открыта (см. рис. 6.7). Поэтому на векторной диаграмме колебанию в P сопоставляется вектор QF, конец которого всегда находится в верхнем фокусе F (рис. 6.8, *б*). Положение начала этого вектора (точки Q) на спирали Корню зависит от положения точки наблюдения P. Когда P находится на границе геометрической тени (т. е. край экрана на рис. 6.7 совпадает с осью y и $d=0$), точка Q совпадает с O и колебание изображается вектором OF, равным половине вектора FF', сопоставляемого колебанию при полностью открытой волновой поверхности. Поэтому интенсивность при $d=0$ в четыре раза меньше интенсивности I_0 в отсутствие экрана. При перемещении точки наблюдения P в освещенную область, т. е. вверх на рис. 6.7, точка Q на векторной диаграмме (рис. 6.8, *б*) будет перемещаться по нижней ветви спирали Корню. При этом интенсивность будет последовательно проходить через максимумы и минимумы (рис. 6.9, $d>0$). В первом, наибольшем из максимумов $I=1{,}37\,I_0$, а в первом минимуме $I=0{,}78\,I_0$. С увеличением расстояния d от края геометриче-



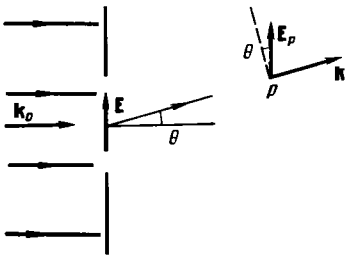
6.9
Распределение интенсивности вблизи края геометрической тени

ской тени размах колебаний интенсивности уменьшается и она приближается к значению I_0, а места максимумов и минимумов постепенно сближаются друг с другом. При погружении точки наблюдения P в область геометрической тени экрана (вниз на рис. 6.7) точка Q перемещается по верхней ветви спирали Корню. Длина вектора QF, а следовательно, и интенсивность монотонно уменьшаются с увеличением расстояния, асимптотически приближаясь к нулю (рис. 6.9, область $d<0$). Зависимость интенсивности от d можно выразить аналитически через интегралы Френеля (6.17).

Проведенное исследование показывает, что между светом и тенью от края экрана нет резкой границы: в области геометрической тени интенсивность спадает постепенно и монотонно, а край освещенной области расщепляется в дифракционные полосы. Полученные закономерности хорошо подтверждаются на опыте.

Сделаем несколько замечаний по поводу использованного выше приближенного метода решения дифракционных задач.

Чтобы с помощью принципа Гюйгенса — Френеля определить поле световой волны за экраном, нужно знать поле на поверхности экрана и в отверстии. Предполагается, что это поле в точках отверстия такое же, как и при свободном распространении падающей волны при отсутствии каких бы то ни было экранов, а в точках, лежащих непосредственно за непрозрачным экраном, поля нет. Это предположение позволяет решить задачу дифракции, но вместе с тем оно влечет за собой целый ряд принципиальных трудностей. Во-первых, оно математически противоречиво: если вычислить по принципу Гюйгенса — Френеля напряженность поля во всем пространстве, то на вспомогательной поверхности S она не совпадает с исходной напряженностью поля падающей волны, а на задней стороне экрана не обратится в нуль. Во-вторых, это предположение допускает разрыв напряженности поля на краю отверстия, что противоречит граничным условиям (непрерывности тангенциальных составляющих). В-третьих, предположение приводит к противоречию со свойством поперечности световых волн. В самом деле, допустим, что на экран с отверстием падает по нормали линейно поляризованная плоская волна (рис. 6.10). Тогда на вспомогательной поверхности S в соответствии с обсуждаемым предположением вектор $\mathbf{E}$ имеет во всех точках одно и то же направление, параллельное плоскости экрана. Применяя принцип Гюйгенса — Френеля для вычисления $\mathbf{E}_P$ в точке P, в результате сложения происходящих по этому направлению колебаний во вторичных волнах получим, что в дифрагировавшей волне вектор $\mathbf{E}_P$ всюду параллелен плоскости экрана. Но в действительности на большом расстоянии отклоненная волна поперечна и вектор $\mathbf{E}$ в



6.10
Вектор $\mathbf{E}_P$ в дифрагировавшей волне

ней образует с плоскостью экрана некоторый угол (рис. 6.10). Отмеченные трудности характерны и для теории Кирхгофа, в которой принцип Гюйгенса — Френеля получил математическое обоснование на основе (скалярного) волнового уравнения для компонент напряженности поля.

Подробный анализ показывает, что лежащие в основе метода Френеля допущения могут быть оправданы, когда размеры препятствий (или отверстий в экранах) велики по сравнению с длиной световой волны. В этом случае отклонения от геометрической оптики малы, т. е. заметная интенсивность наблюдается только при малых углах дифракции. Различие в истинном и вычисленном направлениях E при этом несущественно. Влияние края экрана на поле в отверстии простирается лишь на расстояние порядка длины волны, и при больших размерах отверстия замена истинных значений E в подынтегральном выражении формулы (6.3) на напряженность E падающей волны не приводит к заметным ошибкам, так как на большей части поверхности S эти значения совпадают. При больших по сравнению с длиной волны размерах отверстий расчеты дифракционной картины по методу Френеля хорошо подтверждаются на опыте и согласуются с результатами точного электродинамического решения (в тех случаях, когда такое решение возможно). Подтверждается и предположение о независимости дифракционной картины в этих условиях от материала экрана.

Строгое электродинамическое решение задачи дифракции плоской электромагнитной волны на прямолинейном крае идеально проводящего полубесконечного экрана было получено в 1894 г. А. Зоммерфельдом. С тех пор было найдено строгое решение лишь нескольких дифракционных задач (Л. И. Мандельштам, В. А. Фок и др.). Для большинства задач метод Френеля дает единственный путь решения и приводит к практически удовлетворительным результатам. Несмотря на отмеченные выше принципиальные трудности и ограниченную применимость, он оказался чрезвычайно плодотворным.

Контрольные вопросы

- ☐ Чем обусловлено различие векторных диаграмм для дифракции на прямолинейном крае экрана (см. рис. 6.8) и на круглом отверстии (см. рис. 6.4)?
- ☐ Как с помощью спирали Корню найти вектор, изображающий световое колебание в точке наблюдения P, лежащей в области геометрической тени? в освещенной области? Как при этом объясняется основное различие в распределении интенсивности для этих двух случаев?
- ☐ Какие трудности принципиального характера присущи приближенному методу решения дифракционных задач на основе принципа Гюйгенса — Френеля?
- ☐ При каких условиях присущие методу Френеля трудности становятся несущественными и он приводит к правильным согласующимся с опытом результатам?

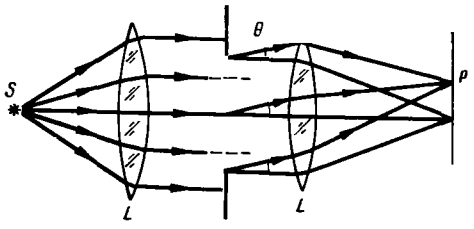
Задача

Оценить ширину дифракционных полос вблизи границы тени от прямолинейного края экрана. Дифракционная картина наблюдается в плоскости, перпендикулярной направлению падающей плоской волны ($\lambda = 500$ нм) и расположенной на расстоянии 1 м от отбрасывающего тень экрана.

◇ Максимумам и минимумам интенсивности на рис. 6.9 соответствуют точки нижней ветви спирали Корню; расстояние до которых от верхнего фокуса F соответственно максимально или минимально (см. рис. 6.8, б). Приближенно можно считать, что эти точки находятся на пересечении спирали с продолжением прямой FF', проходящей через фокусы, так как в таких точках прямая FF' практически перпендикулярна спирали Корню (см. рис. 6.8, б). Для нахождения положения максимумов воспользуемся тем, что угол α наклона касательной к спирали Корню связан с параметром η_1 соотношением $\alpha = \pi\eta_1^2/2$, а η_1 пропорционален расстоянию d до точки наблюдения от края геометрической тени: $\eta_1 = d\sqrt{2/(\lambda L)}$. Поэтому $\alpha = \pi d^2/(\lambda L)$. В первом максимуме $\alpha_1 \approx 3\pi/4$, в n-м — $\alpha_n = 3\pi/4 + 2\pi(n-1)$. Таким образом, светлые полосы находятся на расстояниях $d_n = \sqrt{(2n - {}^5/_4)\lambda L}$ ($n = 1, 2, \ldots$) от края геометрической тени. Для первой полосы $d_1 = \sqrt{3\lambda L/4} \approx 0{,}6$ мм. Расстояния между последующими максимумами постепенно уменьшаются. ◇

6.3. Дифракция Фраунгофера

Наибольший практический интерес представляют дифракционные явления, наблюдаемые при падении на экран (или на отверстие в экране) параллельного пучка света. В результате дифракции пучок утрачивает параллельность, т. е. появляется свет, распространяющийся в направлениях, отличных от первоначального. Распределение его интенсивности на очень большом (в пределе — бесконечно большом) расстоянии от экрана соответствует *дифракции Фраунгофера*. Волны, возникающие в результате ограничения фронта падающей плоской волны при прохождении сквозь отверстие в экране, называют *дифрагировавшими*, а нормали к их волновым поверхностям — дифрагировавшими лучами. Они не существуют в рамках геометрической оптики. Возникновение дифрагировавших волн при прохождении через отверстие означает, что волна с ограниченной площадью поперечного сечения не может быть строго плоской. Разложение волны с ограниченным фронтом на сумму плоских волн (т. е. пространственное разложение Фурье) содержит слагаемые с волновыми векторами различных направлений. Эти слагаемые и соответствуют дифрагировавшим волнам.



6.11
Схема наблюдения дифракции Фраунгофера

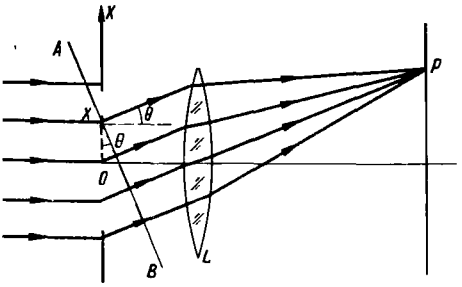
Угловой разброс в направлениях распространения для пучка шириной a из-за дифракции, как показано ниже, не может быть меньше $\Delta\theta \approx \approx \lambda/a$.

Практически дифракцию Фраунгофера наблюдают не «в бесконечности», а в фокальной плоскости объектива (собирающей линзы с

исправленными аберрациями) или с помощью зрительной трубы, установленной на бесконечность. Схема опыта показана на рис. 6.11. Падающий на экран параллельный пучок можно получить, если точечный источник S поместить в фокус линзы L' (формирующая параллельный пучок линза L' называется коллиматором). Размеры линз должны быть много больше размеров отверстий в экране, чтобы наблюдаемое распределение интенсивности было обусловлено дифракцией при ограничении фронта волны именно отверстиями в экране, а не оправами линз. Если отверстие в экране представляет собой узкую щель, изображение точечного источника S в фокальной плоскости объектива растянется в полоску, перпендикулярную щели. Освещенность полоски от середины к краям уменьшается немонотонно, проходя через ряд минимумов и максимумов. При повороте щели вся картина также поворачивается. Вместо точечного источника S можно взять светящуюся нить, параллельную щели. В этом случае наблюдаемое в фокальной плоскости объектива изображение нити оказывается растянутым в перпендикулярном щели направлении. При отсутствии экрана между линзами ограничение поперечного сечения светового пучка осуществляется апертурной диафрагмой (см. § 7.3) оптической системы (роль которой в схеме на рис. 6.11 играет оправа меньшей из линз L и L'). Создаваемое идеальной оптической системой изображение точечного источника S всегда представляет собой не точку в сопряженной с S плоскости (как это следует из геометрической оптики), а фраунгоферову дифракционную картину, возникающую вследствие ограничения сечения светового пучка.

В демонстрационных опытах с использованием лазерного излучения необходимость в коллиматоре L' и объективе L отпадает. Щель вводят непосредственно в пучок. Световые колебания когерентны по всему поперечному сечению лазерного пучка. Это значит, что в отношении когерентных свойств излучения лазер можно рассматривать как удаленный точечный источник. На экране, удаленном от щели на расстояние порядка 10 м, наблюдается фраунгоферова дифракционная картина: пятно размывается в перпендикулярную щели длинную полоску с постепенно спадающей к краям освещенностью, прорезанную эквидистантными темными минимумами. Ширина центрального максимума вдвое больше, чем боковых.

Распределение интенсивности света при дифракции Фраунгофера можно найти с помощью принципа Гюйгенса—Френеля (6.3). Интенсивность в определенной точке P фокальной плоскости объектива L (рис. 6.11) обусловлена интерференцией вторичных волн, исходящих от всех элементарных участков отверстия экрана, причем эти волны распространяются в одном и том же направлении θ, соответствующем точке P. Когда размеры отверстий в экране, ограничивающих сечение пучка, велики по сравнению с длиной световой волны, напряженность поля на вспомогательной (плоской)



Вторичная волна из точки с координатой x

поверхности S, совпадающей с отверстием в экране, можно считать такой же, какой она была бы при отсутствии экрана. Заметную интенсивность при этом имеют лишь волны, дифрагировавшие на малые углы θ.

Рассмотрим сначала простой, но практически важный случай, когда отверстие в экране имеет вид узкой длинной щели с параллельными краями. Размер волновой поверхности в направлении вдоль щели ограничен только объективом, и если вносимую им дополнительную дифракцию во внимание не принимать, то волны дифрагируют только в направлениях, перпендикулярных щели. Поэтому можно считать, что элементарные участки волнового фронта, имеющие вид узких длинных полосок, параллельных краям щели, становятся источниками цилиндрических вторичных волн. Амплитуды этих волн, приходящих в точку P от разных полосок, одинаковы, так как все элементы имеют одинаковую площадь и одинаковый наклон к направлению вторичных волн. Соотношение фаз вторичных волн в точке P будет таким же, как и в любой плоскости, перпендикулярной их направлению до линзы, например, в плоскости AB (рис. 6.12). Так как при нормальном падении света на щель фазы вторичных источников одинаковы, то исходящая под углом θ волна из элемента с координатой x (рис. 6.12) опережает по фазе волну того же направления из середины щели на $kx\sin\theta$. Амплитуда результирующего колебания в точке P, обусловленного вторичными волнами от всей щели шириной a, пропорциональна выражению

$$E_P \sim K(\theta) \int_{-a/2}^{a/2} e^{-ikx\sin\theta}\, dx = K(\theta) a \frac{\sin u}{u}, \qquad (6.18)$$

где

$$u = (ka/2)\sin\theta = (\pi a/\lambda)\sin\theta. \qquad (6.19)$$

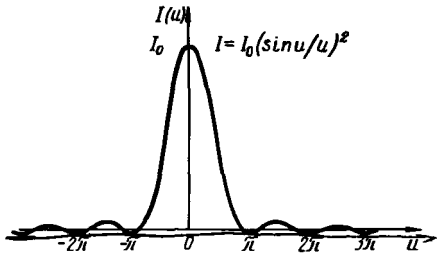
При малых углах дифракции θ коэффициент наклона $K(\theta)$ в (6.18) практически не зависит от θ и его можно заменить значением при $\theta = 0$. Тогда зависимость интенсивности дифрагировавшего света от θ определяется выражением

$$I(\theta) = I_0 (\sin u/u)^2, \qquad (6.20)$$

где I_0 — интенсивность света при $\theta = 0$, т. е. по направлению падающей волны. График распределения интенсивности по направлениям приведен на рис. 6.13. В центре дифракционной картины интенсивность максимальна и равна I_0. При $u = m\pi$, где $m = \pm 1$,

$\pm 2, \ldots$, интенсивность равна нулю. Направления θ_m на эти минимумы, как видно из (6.19), определяются условием

$$a \sin\theta_m = m\lambda. \qquad (6.21)$$



6.13
Угловое распределение интенсивности при дифракции Фраунгофера на щели

Первый минимум дифракционной картины ($m = 1$) соответствует направлению θ_1, для которого $\sin\theta_1 = \lambda/a$, а при $\lambda/a \ll 1$ $\theta_1 \approx \lambda/a$. Это условие легко получить без всяких вычислений. Рассмотрим две одинаковые элементарные полоски, находящиеся на расстоянии $a/2$. Вторичные волны от них, распространяющиеся под углом θ, имеют разность хода $(a/2)\sin\theta$. Если эта разность хода равна $\lambda/2$, т. е. $\sin\theta = \lambda/a$, то вторичные волны гасят одна другую в результате интерференции. Вся щель состоит из таких пар элементарных полосок, поэтому при $\sin\theta = \lambda/a$ интенсивность дифрагировавшего света обращается в нуль.

Между минимумами интенсивности, определяемыми условием (6.21), находятся максимумы различных порядков. Их положение определяется уравнением $\mathrm{tg}\,u = u$, имеющим корни $u_0 = 0$; $u_1 = 1{,}43\pi$; $u_2 = 2{,}46\pi$; $u_3 = 3{,}47\pi$ и т. д. Практически можно считать, что максимумы находятся посередине между соседними минимумами. Значения интенсивности в максимумах быстро убывают с увеличением порядка. Их отношения приближенно можно выразить в виде

$$1 : 0{,}047 : 0{,}017 : \ldots \approx 1 : [2/(3\pi)]^2 : [2/(5\pi)]^2 : \ldots . \qquad (6.22)$$

Таким образом, основная часть светового потока сосредоточена в центральной дифракционной полосе между минимумами порядков $m = \pm 1$, т. е. в пределах углов $-\theta_1 < \theta < \theta_1$, где $\sin\theta_1 = \lambda/a$. Угловая ширина максимумов уменьшается при увеличении ширины a щели: если $\theta \ll 1$, то $\theta_1 = \lambda/a$. Центральный максимум становится резче, первые минимумы придвигаются ближе к центру картины. Высота максимума интенсивности пропорциональна *квадрату* ширины щели, так как возрастающий пропорционально a световой поток распространяется в пределах убывающего угла ($\theta_1 \sim 1/a$). Относительная интенсивность остается неизменной: распределение света по максимумам разных порядков (6.22) не зависит от ширины щели. При сужении щели картина расширяется, а ее яркость уменьшается. Когда a приближается к λ, центральный максимум охватывает все поле зрения; освещенность экрана уменьшается от центра к краям монотонно.

Полученные результаты можно использовать для оценки *дифракционной расходимости пучков света,* ширина которых ограничена, например, в результате прохождения через диафрагму. Основная часть светового потока приходится на центральный дифракционный

максимум, поэтому его ширину можно принять в качестве оценки угловой расходимости $\Delta\theta$ пучка с поперечным сечением a:

$$\Delta\theta \sim \lambda/a. \tag{6.23}$$

Это угловое уширение пучка обусловлено волновой природой света и его в принципе нельзя устранить при заданной ширине сечения пучка. Строго параллельных световых пучков не существует. На пути длиной l пучок претерпевает дифракционное уширение порядка $l\Delta\theta = \lambda l/a$. Этим уширением можно пренебрегать лишь тогда, когда оно мало по сравнению с исходной шириной пучка, т. е. при $\lambda l \ll a^2$. В таких условиях пучок приближенно можно считать параллельным и использовать для его описания геометрическую оптику.

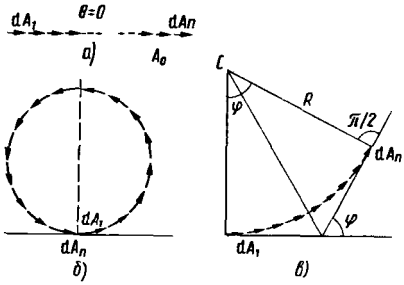
Рассмотрим, например, «параллельный» лазерный пучок диаметром $a=2$ мм, световые колебания в котором когерентны по всему поперечному сечению. Его расширение по мере распространения обусловлено дифракцией. При $\lambda=600$ нм диаметр пучка на расстоянии $l=150$ м составит приблизительно $2\lambda l/a = 10$ см. Для пучков света от нелазерных источников расширение обусловлено обычно не дифракцией, а *конечными размерами источника*. Если источник размером D (светящаяся нить) находится в фокусе линзы с фокусным расстоянием F, то выходящие из линзы пучки света от краев источника в соответствии с геометрической оптикой образуют угол D/F. Чтобы этот угол был меньше дифракционной расходимости пучка диаметром a, размер источника должен удовлетворять условию $D < \lambda F/a$. При $a=2$ мм, $F=5$ см, $\lambda=600$ нм это дает $D \lesssim 10^{-3}$ см. Размеры реальных источников гораздо больше.

Пучок света с минимально возможной при данном диаметре a угловой расходимостью $\Delta\theta \sim \lambda/a$ формируется в результате интерференции вторичных волн от всего поперечного сечения. Такая интерференция возможна только тогда, когда световые колебания когерентны по всему поперечному сечению пучка. Высокая пространственная когерентность лазерного пучка обусловлена самой природой процесса испускания света (вынужденное излучение). Когда направленный пучок («плоская» волна) формируется от обычного источника света, помещенного в фокус собирающей линзы или вогнутого зеркала, для достижения дифракционного предела расходимости необходимо, чтобы освещение всей поверхности линзы или зеркала было когерентным. Как было показано в § 5.5, размер области когерентности от протяженного источника равен $d \approx \lambda/\theta$, где $\theta = D/L$ — угловой размер источника. В данном случае расстояние L от источника равно фокусному расстоянию F и $d = \lambda F/D$. Из требования $d > a$ получаем прежнее ограничение на размер источника: $D < \lambda F/a$. Для увеличения допустимого размера источника можно увеличивать F, но при этом уменьшается та часть светового потока источника, которая попадает в формируемый пучок.

Наглядное представление о суммировании вторичных волн для нахождения распределения интенсивности при дифракции Фраунгофера на щели можно получить с помощью метода векторных диаграмм. Колебания в точке P, вызываемые вторичными волнами от одинаковых элементарных полосок вспомогательной поверхности S, изображаются векторами dA_i одинаковой длины (рис. 6.14). Центру дифракционной картины ($\theta=0$), куда все вторичные волны приходят в одинаковой фазе, соответствует диаграмма на рис. 6.14, *а*. Результирующая амплитуда равна A_0. Диаграмма на рис. 6.14, *б* соответствует такому направлению θ_1 вторичных волн, когда разность фаз от краев щели равна 2π, т. е. $a\sin\theta_1=\lambda$. Цепочка векторов dA_i, повернутых друг относительно друга на один и тот же угол, в этом случае оказывается замкнутой и амплитуда результирующего колебания равна нулю. Это первый минимум дифракционной картины. Легко видеть, что результирующая амплитуда равна нулю и тогда, когда разность фаз от краев щели равна $2\pi m$, где m — целое число. Цепочка векторов dA_i при этом замыкается после m оборотов. Общий случай представлен на диаграмме (рис. 6.14, *в*). Векторы dA_1 и dA_n, изображающие колебания от краев щели, образуют угол $\varphi=ka\sin\theta$. Результирующее колебание в P изображается вектором $A(\theta)$ — хордой окружности с центром в точке C. Длина дуги, стягиваемой этой хордой, приближенно равна A_0 — амплитуде в центре дифракционной картины (при $\theta=0$). Это заключение справедливо, если при изменении θ изменяется только направление векторов dA_i (фазы колебаний), но не их модули (амплитуды колебаний), т. е. коэффициент наклона $K(\theta)$ в (6.3) считается постоянным: $K(\theta)\approx K(0)$. Из рис. 6.14, *в* легко видеть, что $A(\theta)=2R\sin(\varphi/2)=(2A_0/\varphi)\sin(\varphi/2)$. Отсюда для распределения интенсивности $I(\theta)\sim A^2(\theta)$ снова получаем формулу (6.20).

Если плоская волна падает на щель *наклонно* под углом θ' к нормали, то разность хода между вторичными волнами, распространяющимися от краев щели в направлении θ, составляет $a(\sin\theta-\sin\theta')$. Условие дифракционных минимумов вместо (6.21) принимает вид $a(\sin\theta_m-\sin\theta')=m\lambda$. Угловое распределение интенсивности попрежнему определяется формулой (6.20), в которой теперь $\sin\theta$ следует заменить на $\sin\theta-\sin\theta'$ (см. задачу 1). Центральный максимум дифракционной картины расположен при $\theta=\theta'$, т. е. в направлении падающей волны.

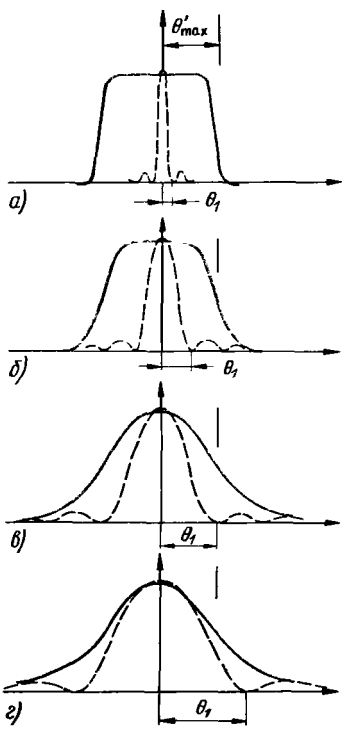
До сих пор предполагалось, что в фокусе коллиматорной линзы L' находится точечный источник S. Если в качестве S взять светящуюся линию, параллельную щели, то изображение каждой ее точки в фокальной плоскости объектива L вытянется в линию, перпендикулярно щели. Распределение интенсивности вдоль всех этих линий будет одинаковым, поэтому образуется система параллельных щели



6.14
Векторные диаграммы для дифракции Фраунгофера на щели

дифракционных полос с описываемым формулой (6.20) поперечным распределением интенсивности.

В том случае, когда расположенный в фокальной плоскости коллиматора L' источник имеет конечный размер в перпендикулярном щели направлении (светящаяся полоска шириной D), распределение интенсивности в фокальной плоскости объектива L можно рассматривать как наложение независимых дифракционных картин, создаваемых взаимно некогерентными световыми пучками от отдельных элементов протяженного источника. Отстоящий на расстоянии x от оси элемент источника посылает на щель плоскую волну, наклоненную на угол $\theta'(x) \approx x/F$, где F — фокусное расстояние коллиматора. На такой же угол будет смещен центральный максимум соответствующей дифракционной картины. Центральные максимумы от краев источника смещены на углы $\theta'_{max} \approx D/(2F)$. Если щель широкая, так что ширина максимумов $\theta_1 = \lambda/a$ значительно меньше θ'_{max}, то наложение узких максимумов от отдельных элементов создает в фокальной плоскости объектива изображение источника, которое геометрически почти подобно источнику и лишь слегка размыто по краям из-за боковых максимумов от близких к краям элементов источника (рис. 6.15, *а, б*). По мере уменьшения ширины щели угол $\theta_1 \approx \lambda/a$ увеличивается, т. е дифракционные максимумы от отдельных элементов источника расширяются, а его изображение становится все более расплывчатым: дифракционное уширение будет составлять значительную часть геометрической ширины изображения (рис. 6.15, *в, г*). При очень узкой щели, когда $\theta_1 = \lambda/a \gg \theta'_{max} = D/(2F)$, наблюдаемая в фокальной плоскости картина полностью утрачивает подобие источнику, так как «изображение» источника в этом случае почти не отличается от дифракционной картины, создаваемой линейным (бесконечно узким) источником.



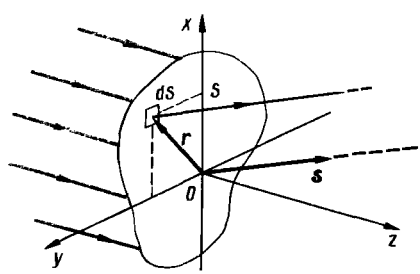
6.15
Дифракция на щели при протяженном источнике света

Характер дифракционной картины в свете от протяженного источника можно рассмотреть и на основе введенной в § 5.5 степени пространственной когерентности излучения. Размер области когерентности на поверхности коллиматорной линзы и, следовательно, на щели для источника в виде светящейся полоски шириной D в перпендикулярном полоске направлении равен $d=\lambda F/D$. Если ширина a щели много меньше этого размера, т. е. $a \ll \lambda F/D$, то световые колебания во всех точках щели (в поперечном направлении) почти полностью когерентны и распределение интенсивности в фокальной плоскости объектива практически такое же, как в дифракционной картине от линейного источника. В противоположном предельном случае широкой щели, когда $a \gg \lambda F/D$, когерентность колебаний поперек щели простирается лишь на расстояния $d \approx \lambda F/D$, малые по сравнению с ее шириной a. Для оценки ширины изображения источника здесь можно считать, что дифракция происходит как бы на щели с эффективной шириной d, т. е. свет отклоняется на углы порядка $\theta' \approx \lambda/d \approx D/F$. Это по порядку величины совпадает с угловой шириной изображения светящейся полоски. Таким образом, применение понятия частичной пространственной когерентности приводит к тем же результатам, что и суммирование независимых дифракционных картин от отдельных элементов протяженного источника.

Рассмотрим теперь дифракцию Фраунгофера при падении плоской волны на *отверстие в экране*. В отличие от длинной щели здесь волны дифрагируют во всех направлениях. Каждой точке наблюдения P соответствует определенное направление дифрагировавших волн, характеризуемое единичным вектором $\mathbf{s}$ (рис. 6.16). В качестве вспомогательной поверхности S выберем плоскость экрана xy. Разность хода идущих по направлению $\mathbf{s}$ вторичных волн из элемента dS этой поверхности и из начала координат O равна проекции вектора $\mathbf{r}$, определяющего положение dS в плоскости xy, на направление $\mathbf{s}$, т. е. $\mathbf{rs}$. В соответствии с принципом Гюйгенса — Френеля напряженность поля в точке P пропорциональна интегралу по всей площади отверстия в экране:

$$E_P \sim \int\limits_S E(\mathbf{r}) e^{-ik(\mathbf{rs})} dS = \int\limits_S E(\mathbf{r}) e^{-i\mathbf{kr}} dS, \tag{6.24}$$

где $\mathbf{k}=k\mathbf{s}$ — волновой вектор света, дифрагировавшего в направлении $\mathbf{s}$. Опущенный в (6.24) коэффициент наклона $K(\alpha)$ можно считать постоянным, когда размеры отверстия много больше длины волны, и заметную интенсивность имеют лишь волны, дифрагировавшие на малые углы α. Напряженность $E(\mathbf{r})$ в плоскости xy



6.16

К расчету дифракции Фраунгофера от отверстия

считается равной напряженности поля падающей волны в пределах отверстия экрана и равной нулю за его пределами. Понимая функцию $E(x, y)$ имеино так, можно распространить иитегрирование в (6.24) на всю плоскость xy:

$$E_p \sim \int\int E(x, y)e^{-i(k_x x + k_y y)}dxdy \equiv E(k_x, k_y). \qquad (6.25)$$

Отсюда видно, что поле в фраунгоферовой дифракционной картине, т. е. в фокальной плоскости объектива, представляет собой (с точностью до постоянного множителя) двухмерное *преобразование Фурье* функции $E(x, y)$, описывающей поле в плоскости xy. Функция $E(k_x, k_y)$, т. е. фурье-образ искаженного препятствием волнового поля $E(x, y)$ в плоскости xy, пропорциональна комплексной амплитуде плоской волны, дифрагировавшей в определенном направлении k_x, k_y. Пространственное разделение волн, дифрагировавших в разных направлениях, позволяет наблюдать отдельные фурье-компоненты функции $E(x, y)$. Поэтому можно считать, что в дифракции Фраунгофера *физически* осуществляется разложение функции $E(x, y)$ в двухмерный интеграл Фурье.

При нормальном падении плоской волны на *прямоугольное отверстие* со сторонами a и b, параллельными осям x и y, из (6.25) находим

$$E_p \sim E \int\limits_{-a/2}^{a/2} \int\limits_{-b/2}^{b/2} e^{-i(k_x x + k_y y)}dxdy = Eab\,\frac{\sin u_1}{u_1}\,\frac{\sin u_2}{u_2},$$

где $u_1 = k_x a/2$, $u_2 = k_y b/2$. Распределение интенсивности в дифракционной картине определяется формулой

$$I = I_0(\sin u_1/u_1)^2(\sin u_2/u_2)^2. \qquad (6.26)$$

Когда длина одной из сторон много больше длины другой, мы приходим к выражению (6.20) для дифракции на длинной щели. В дифракционной картине от прямоугольного отверстия (рис. 6.17, *а*) распределение интенсивности в соответствии с (6.26) дается произведением распределений от взаимно перпендикулярных щелей. Интенсивность равна нулю вдоль двух рядов линий, параллельных сторонам прямоугольника. Заметную интенсивность имеют лишь средние цепочки максимумов, образующие «крест» на рис. 6.17, *а*. Относительная высота максимумов интенсивности, расположенных вдоль этих линий, характеризуется соотношением (6.22). Величина остальных максимумов столь мала (0,2% для ближайших к центру), что они не видны на приведенной фотографии. Бо́льшая часть светового потока приходится на центральный максимум, и именно его можно рассматривать как изображение находящегося в фокусе коллиматора точечного источника, получающееся в фокальной плоскости объектива при ограничеиии сечеиия, формирующего изображение пучка света прямоугольной диафрагмой. Это изображение шире в направлеиии более короткой стороны прямоугольника.

Дифракция Фраунгофера от *круглого отверстия* представляет большой практический интерес, так как в оптических приборах оправы линз и объективов, а также диафрагмы имеют обычно круглую форму. При вычислении интеграла (6.25) целесообразно перейти к полярным координатам ρ и φ в плоскости отверстия: $x=\rho\cos\varphi$, $y=\rho\sin\varphi$. Направление s дифрагировавшей волны, соответствующее точке P, удобно характеризовать углом θ с осью z и азимутальным углом ψ: $k_x = k\sin\theta\cos\psi$, $k_y = k\sin\theta\sin\psi$. Тогда $k_x x + k_y y = k\rho\sin\theta\cos(\varphi - \psi)$ и интеграл (6.25) принимает вид

$$E_P \sim E \int_0^a \int_0^{2\pi} e^{-ik\rho\sin\theta\cos(\varphi-\psi)} \rho d\rho d\varphi.$$

Здесь a — радиус отверстия. Используя интегральное представление для бесселевых функций $J_n(z)$ при $n=0$

$$J_0(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{iz\cos\varphi} d\varphi,$$

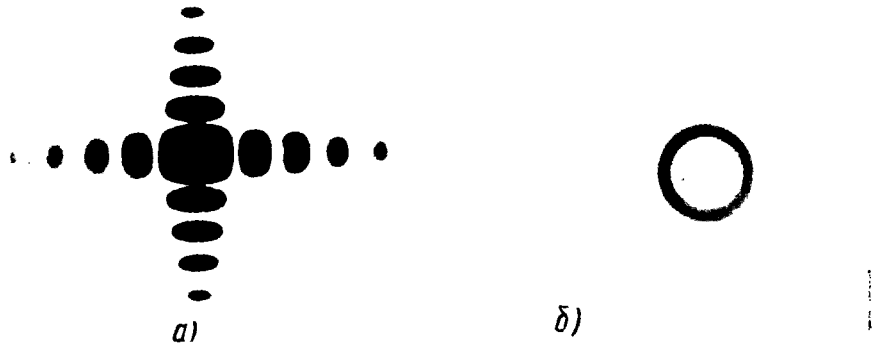
выразим E_P через интеграл от $J_0(k\rho\sin\theta)$, который вычисляется с помощью соотношения

$$\int zJ_0(z)dz = zJ_1(z).$$

Таким образом, дифракционная картина от круглого отверстия имеет вид концентрических светлых и темных колец (рис. 6.17, *б*) со следующим радиальным распределением интенсивности:

$$I(\theta) = I_0[2J_1(u)/u]^2, \quad u = ka\sin\theta = 2\pi a\sin\theta/\lambda \approx 2\pi a\theta/\lambda. \qquad (6.27)$$

График этой функции приведен на рис. 6.18. Она имеет главный максимум при $u=0$ и с ростом u осциллирует с быстрым уменьше-



Дифракция Фраунгофера от прямоугольного и круглого отверстий
Центральная область переэкспонирована (для выделения слабых побочных максимумов)

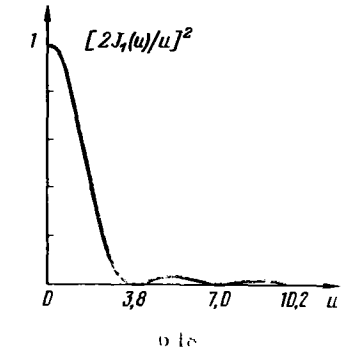


График функции $[2J_1(u)/u]^2$

нием амплитуды, подобно функции $(\sin u/u)^2$, описывающей дифракцию на щели. Угловые радиусы θ_n темных колец равны $0{,}61\lambda/a$; $1{,}12\lambda/a$; $1{,}62\lambda/a$; ... . Расстояние между соседними кольцами с увеличением их номера приближается к $\lambda/(2a)$. Эффективный размер дифракционной картины и здесь обратно пропорционален размеру отверстия. Интенсивность максимумов быстро уменьшается: уже в ближайшем максимуме она составляет менее 2% от интенсивности центрального максимума, на который приходится 84% проходящего через отверстие светового потока. Поэтому центральный максимум (*диск Эйри*), имеющий угловой радиус

$$\theta_1 = 0{,}61\lambda/a, \tag{6.28}$$

можно рассматривать как изображение точечного источника, уширенное дифракцией на круговой диафрагме радиусом a. Соотношение (6.28) играет важную роль в вопросе о разрешающей силе оптических инструментов (см. § 7.6).

Важно отметить, что распределение интенсивности в фраунгоферовой дифракционной картине не изменится, если отверстие сместить в плоскости экрана в сторону, не изменяя его ориентации. Картина в фокальной плоскости объектива всегда симметрична по отношению к его оси независимо от положения отверстия. Особый интерес представляет случай, когда в экране имеется *большое число N одинаковых отверстий*. При правильном, *регулярном,* расположении отверстий, когда их ориентация и расстояния между ними одинаковы, разность фаз между волнами, дифрагировавшими от соседних отверстий, имеет определенное значение. Интерференция этих волн существенно влияет на дифракционную картину. В направлениях, для которых разность фаз кратна 2π, амплитуда дифрагировавших волн в N раз больше, а интенсивность в N^2 раз больше, чем от одного отверстия. Такое резкое увеличение интенсивности для некоторых направлений имеет большое практическое значение. Случай регулярного расположения отверстий подробно рассмотрен на примере дифракционной решетки в § 6.5.

При *хаотическом,* беспорядочном, расположении отверстий фазовые соотношения между волнами от отдельных отверстий имеют случайный характер. Поэтому для каждого направления наблюдения происходит простое сложение интенсивностей волн, дифрагировавших от всех отверстий. Распределение интенсивности в дифракционной картине от одного отверстия не зависит от его положения. От большого числа N отверстий получается такая же картина, усиленная по интенсивности в N раз.

Контрольные вопросы

- □ Объясните назначение линз L' и L в схеме для наблюдения дифракции Фраунгофера на рис. 6.11.
- □ Как получить без вычислений соотношение $\sin\theta_1=\lambda/a$, определяющее направление на первый минимум при дифракции на щели шириной a?
- □ Можно ли какими-либо способами получить узкий параллельный пучок света?
- □ Какому условию должен удовлетворять размер источника, чтобы с помощью линзы или зеркала можно было получить пучок света с дифракционной расходимостью?
- □ Какой вид имеет фраунгоферова дифракционная картина при наклонном падении плоской волны на щель?
- □ Как изменится распределение интенсивности в дифракционной картине Фраунгофера от отверстия, если отверстие сместить в поперечном направлении?
- □ Чем отличаются дифракционные картины от большого числа одинаковых препятствий при хаотическом и упорядоченном расположениях?

Задачи

1. Найти угловое распределение интенсивности $I(\theta)$ при дифракции Фраунгофера на щели шириной a в случае наклонного падения параллельного пучка света на плоскость щели (под углом θ' к нормали).

◇ Ответ: $I(\theta)=I_0(\sin u/u)^2$, где $u=(\pi a/\lambda)(\sin\theta-\sin\theta')$, а I_0 — интенсивность при $\theta=\theta'$, т. е. в направлении падающей волны. Для малых углов дифракции, когда $\Delta\theta=\theta-\theta'\ll 1$, $\sin\theta-\sin\theta'\approx\cos\theta'\Delta\theta$ и $u\approx(\pi a\cos\theta'/\lambda)\Delta\theta$. Распределение интенсивности $I(\Delta\theta)$ при наклонном падении на щель определяется проекцией $a\cos\theta'$ ширины щели на плоскость, перпендикулярную направлению падающей волны. ◇

2. Оптическая система создает действительное изображение источника света, размером D, находящегося на расстоянии L от нее. До какой величины $a_{\min}$ можно уменьшить диаметр диафрагмы, чтобы размер изображения определялся размером источника, а не дифракцией?

◇ Ответ: $a_{\min}\approx\lambda L/D$. ◇

3. Описание картины полос в интерферометре Рэлея (см. § 5.6) приведено в предположении, что щели S_1 и S_2 бесконечно узкие. В действительности щели должны пропускать достаточно света и потому имеют конечную ширину a. Какой вид имеют при этом наблюдаемые через окуляр интерференционные полосы? Другими словами, требуется найти распределение интенсивности $I(x)$ в фокальной плоскости объектива.

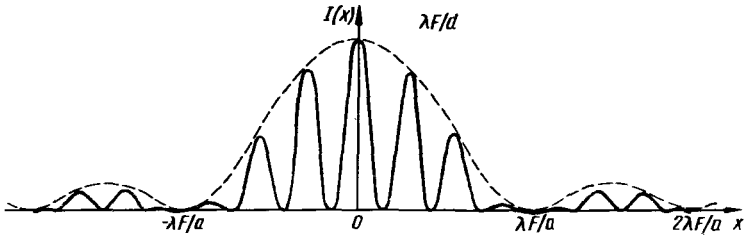
◇ В каждую точку x фокальной плоскости объектива L_2 приходят две когерентные волны от обеих щелей. Разность их хода $\Delta(x)\approx d\theta(x)\approx xd/F_2$, где d — расстояние между серединами щелей, F_2 — фокусное расстояние объектива. Предполагается, что $x_{\max}\ll F_2$, т. е. углы дифракции малы: $\theta(x)\approx x/F_2\ll 1$. Интенсивности $I_1(x)$ и $I_2(x)$ обеих волн в каждой точке x одинаковы, поэтому в соответствии с (5.8)

$$I(x)=2I_1(x)[1+\cos k\Delta(x)]=2I_1(x)[1+\cos(kd/F_2)x].$$

Распределение интенсивности $I_1(x)$ для одной волны определяется дифракцией на щели и дается формулой (6.20): $I_1(x)=I_0(\sin u/u)^2$, где $u=ka\theta(x)/2=[ka/(2F_2)]x$. Таким образом,

$$I(x)=2I_0\left\{\frac{\sin[ka/(2F_2)]x}{[ka/(2F_2)]x}\right\}^2\left[1+\cos\left(\frac{kd}{F_2}x\right)\right] \qquad (6.29)$$

— интерференционные полосы шириной $\lambda F_2/d$ промодулированы кривой дифракции на отдельной щели (рис. 6.19). Полная ширина главного максимума огибающей равна $2\lambda F_2/a$. В его пределах расположено $N=2d/a$ интерференционных полос. ◇



6.19

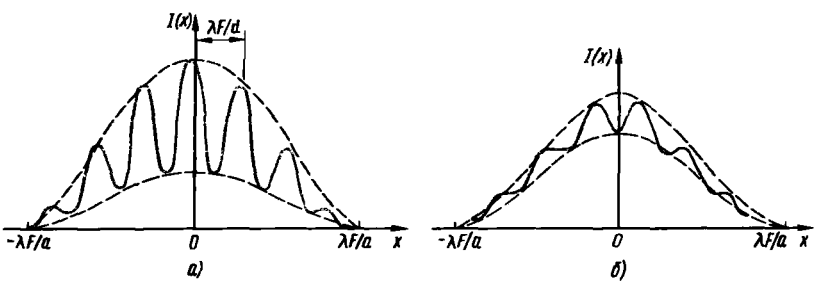
Распределение интенсивности при дифракции Фраунгофера от двух одинаковых щелей

4 Исследовать влияние ширины первичного источника S на интерференционную картину в интерферометре Рэлея.

Если ширина D щели S, находящейся в фокальной плоскости коллиматорной линзы L_1 (см. рис. 5.23), достаточно мала, то освещение обеих щелей S_1 и S_2 в диафрагме полностью когерентное и полосы в фокальной плоскости объектива L_2 описываются полученным в предыдущей задаче распределением интенсивности (6.29), график которого приведен на рис. 6.19. Для этого размер области когерентности $\lambda F_1/D$ (F_1 — фокусное расстояние коллиматора L_1) должен быть много больше расстояния d между щелями S_1 и S_2, откуда $D \ll \lambda F_1/d$. При увеличении D освещение щелей становится частично когерентным. Распределение интенсивности в интерференционных полосах описывается в этом случае выражением (5.55), где степень пространственной когерентности γ_{12} колебаний в щелях S_1 и S_2 дается формулой (5.56) с $\theta = D/F_1$. Если ширина a щелей S_1 и S_2 мала по сравнению с расстоянием d между ними, то освещение каждой из щелей можно считать полностью когерентным и принять для распределения интенсивности $I_1(x)$ одной из интерферирующих волн, как и в предыдущей задаче, выражение (6.20): $I_1(x) = I_0(\sin u/u)^2$ с $u = [ka/(2F_2)]x$ (для этого размер области когерентности $\lambda F_1/D$ должен быть больше ширины a щелей S_1 и S_2, т. е. $D < \lambda F_1/a$). Таким образом,

$$I(x) = 2I_0 \left\{ \frac{\sin [ka/(2F_2)]x}{[ka/(2F_2)]x} \right\}^2 \left\{ 1 + \frac{\sin [kDd/(2F_1)]}{kDd/(2F_1)} \cos\left(\frac{kd}{F_2}x\right) \right\}.$$

Графики $I(x)$ приведены на рис. 6.20, *а*. С увеличением D (т. е. ширины источника S) видность интерференционных полос уменьшается, и при $D = \lambda F_1/d$ они исчезают. При дальнейшем увеличении D полосы появляются вновь, однако их видность невелика, а в центре (при $x = 0$) рас-



6.20

Интерференционные полосы от двух щелей при освещении протяженным источником

положен относительный минимум этой малоконтрастной интерференционной картины (рис. 6.20, *б*). При $D=2\lambda F_1/d$ полосы снова исчезают, и т. д. Когда ширина D источника S настолько велика, что размер области когерентности $\lambda F_1/D$ становится меньше ширины a отдельной щели ($D>\lambda F_1/a$), в фокальной плоскости объектива будет наблюдаться изображение источника S, из-за дифракции несколько размытое по краям.

6.4. Гауссовы пучки. Оптические резонаторы

В § 6.3 были рассмотрены световые пучки, возникающие при прохождении когерентной плоской волны через отверстие в непрозрачном экране. В приближении Френеля—Кирхгофа амплитуда напряженности поля в плоскости экрана имеет в этом случае одно и то же значение на протяжении всего отверстия и равна нулю за его пределами. Другими словами, пучок имеет в плоскости экрана резко ограниченное краями отверстия поперечное сечение. По мере распространения из-за дифракции пучок расширяется и его границы расплываются.

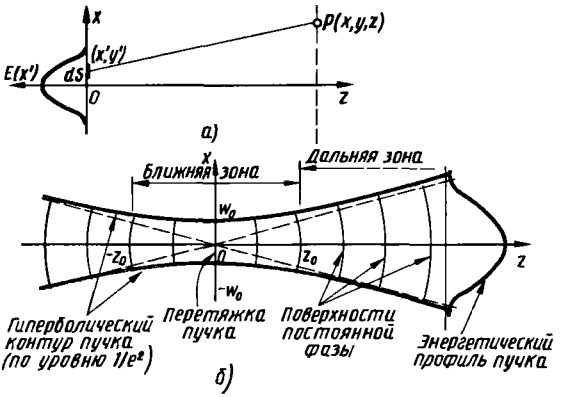
Но во многих практически важных случаях амплитуда напряженности поля плавно изменяется по поперечному сечению пучка. Такой пучок можно получить, например, в результате прохождения плоской волны через пластинку с изменяющимся от точки к точке коэффициентом пропускания. Особенно важен случай *гауссовых пучков*, для которых амплитуда напряженности поля максимальна на оси пучка и экспоненциально уменьшается к краям в соответствии с гладкой функцией Гаусса. Если выбрать ось z по направлению пучка, то в плоскости волнового фронта напряженность такого поля описывается выражением

$$E(x,\ y)=E_0 e^{-(x^2+y^2)/w_0^2}. \qquad (6.30)$$

Здесь, как и прежде, опущен множитель $\exp(-i\omega t)$, выражающий зависимость напряженности поля от времени. Величина w_0 определяет расстояние от оси пучка, на котором напряженность поля уменьшается в e раз, а интенсивность — в e^2 раз по сравнению с максимальным значением. Несколько условно w_0 можно принять за меру ширины распределения интенсивности (радиальная ширина, или *радиус пучка*) в плоскости $z=0$. В пределах этой ширины переносится 86% всей энергии.

Пучки когерентного излучения с гауссовым профилем распределения интенсивности обладают самой высокой направленностью, совместимой с волновой природой излучения. Гауссов пучок представляет собой наиболее близкое приближение, которое допускает дифракция, к параллельному пучку света с ограниченным поперечным сечением. Описываемое выражением (6.30) поперечное распределение интенсивности характерно для света, излучаемого газовыми лазерами.

При распространении в свободном пространстве (или однородной прозрачной среде) гауссов пучок постепенно расширяется из-за дифракции. Представление о его структуре можно получить на основе



6.21

Гауссов пучок:

а — к расчету напряженности поля в точке *P*; *б* — контур пучка и волновые поверхности

принципа Гюйгенса — Френеля. Пусть при $z=0$ волновая поверхность плоская и напряженность поля в разных ее точках дается формулой (6.30). Поле в некоторой точке P с координатами x, y, z (рис. 6.21, *а*) определяется суперпозицией вторичных волн от элементарных участков этой плоскости [см. (6.3)]. Из-за быстрого убывания $E(x', y')$ при удалении от оси z вклад в интеграл фактически дает только часть плоскости $z=0$ в пределах круга радиусом w_0. При малой расходимости пучка можно считать, что точка наблюдения P удалена от лежащего в этих пределах элемента dS вдоль оси пучка на расстояние z, большое по сравнению с шириной пучка (параксиальное приближение). Тогда R в показателе экспоненты [см. (6.3)] можно приближенно записать в виде

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

а множитель $1/R$ заменить на $1/z$ и вынести из-под знака интеграла. По той же причине коэффициент наклона $K(\alpha)$ можно заменить его значением $k/(2\pi i)$ [см. (6.12)] при $\alpha=0$. Таким образом, для $E(x, y, z)$ из (6.3) получаем

$$E(x,\ y,\ z)=-\frac{ikE_0}{2\pi}\frac{e^{ikz}}{z}\int\limits_{-\infty}^{\infty}\int e^{-\frac{x'^2+y'^2}{w_0^2}+ik\frac{(x-x')^2+(y-y')^2}{2z}}dx'dy'. \qquad (6.31)$$

Двойной интеграл здесь распадается на произведение интегралов по x' и y', каждый из которых приводится к интегралу Пуассона:

$$\int\limits_{-\infty}^{\infty}e^{-x^2}dx=\sqrt{\pi}\,.$$

В результате из (6.31) после громоздких, но простых преобразований находим

$$E(x,\ y,\ z)=E_0\frac{w_0}{w}\exp\left\{-\frac{x^2+y^2}{w^2}\right\}\exp\left\{ik\left(z+\frac{x^2+y^2}{2R}\right)+i\varphi\right\}, \qquad (6.32)$$

где

$$w^2(z)=w_0^2\left[1+\left(\frac{2z}{kw_0^2}\right)^2\right];\quad R(z)=z\left[1+\left(\frac{kw_0^2}{2z}\right)^2\right],$$

$$\operatorname{tg}\varphi=\frac{kw_0^2}{2z}. \qquad (6.33)$$

Структура такого пучка показана схематически на рис. 6.21, *б*. Из формулы (6.32) следует, что волна с плоским фронтом и гауссовым распределением (6.30) при $z=0$ имеет при $z\neq 0$ также гауссово распределение амплитуд $\exp[-(x^2+y^2)/w^2]$ с шириной $w(z)$, постепенно возрастающей с расстоянием вдоль оси распространения пучка согласно формуле (6.33).

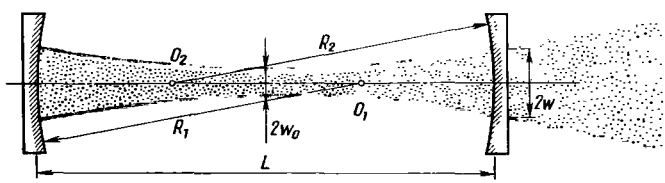
Зависимость ширины пучка от z характеризуется гиперболами $(w/w_0)^2-(z/z_0)^2=1$, где $z_0\equiv kw_0^2/2=\pi w_0^2/\lambda$ — *радиус дифракционной расходимости*. При $z=0$ радиальная ширина имеет наименьшее значение $w=w_0$ (*перетяжка*, или *шейка пучка*). В области шейки, или в ближней зоне, пока $|z|\ll z_0$, площадь сечения пучка практически постоянна. При $z=\pm z_0$ она удваивается, а на больших расстояниях $|z|\gg z_0$ (дальняя зона, или область дифракции Фраунгофера) ширина пучка возрастает линейно с увеличением z: $w(z)\approx 2z/(kw_0)=\lambda z/(\pi w_0)$. Это показано штриховыми линиями (асимптоты гипербол) на рис. 6.21, *б*. Соответствующий угол дифракционной расходимости $\theta=\lambda/(\pi w_0)$ несколько меньше, чем при прохождении плоской волны через круглое отверстие [см. (6.28)]. Важное отличие от дифракции на отверстиях, выделяющих участок волновой поверхности с примерно равными амплитудами, заключается в том, что интенсивность дифракционной картины в гауссовом пучке монотонно и быстро уменьшается с ростом угла дифракции без характерных осцилляций (т. е. чередующихся темных и светлых колец). Это качество очень полезно в оптических приборах, и иногда для подавления дифракционной структуры вместо диафрагм с резкими краями вводят искусственно постепенное ослабление пучка от оси к периферии. Такой прием называется *аподизацией*.

Поверхности постоянной фазы волны (6.32) описываются уравнением $z+(x^2+y^2)/(2R)=\text{const}$. При $x^2+y^2\ll R^2$ это уравнение сферы радиусом R с центром на оси z, т. е. волну приближенно можно рассматривать как сферическую. Однако в отличие от обычной сферической волны радиус $R(z)$ сферы и положение ее центра зависят от z, т. е. от выбранного сечения пучка (рис. 6.21, *б*). Если $z\gg z_0=kw_0/2$, то из (6.33) $R\approx z$ и центр сферы находится при $z=0$. В ближней зоне $R\approx(kw_0^2)^2/(4z)$ и в перетяжке пучка при $z=0$ волновая поверхность становится плоской. Наибольшую кривизну волиовая поверхность имеет при $z=z_0$, т. е. на границе

между ближней и дальней зонами: здесь $R=2z_0$ и центр кривизны расположен при $z=-z_0$.

Выше исследовалось распространение гауссова пучка в свободном пространстве или в однородной среде. Рассмотрим теперь структуру электромагнитного поля в *оптическом резонаторе*, образованном двумя сферическими зеркалами, обращенными друг к другу своими отражающими поверхностями (рис. 6.22). Резонатор, подобно маятнику в часах или колебательному контуру в генераторе незатухающих колебаний, составляет важную часть автоколебательной оптической системы — оптического квантового генератора (ОКГ), или лазера. Стационарные состояния поля, как и в случае резонатора с плоскими параллельными зеркалами, представляют собой стоячие электромагнитные волны, т. е. нормальные колебания напряженности поля (*моды*) с определенными дискретными частотами (см. § 1.3). Частоты мод определяются из *условия цикличности*: при двойном прохождении резонатора (туда и обратно) фаза волны изменяется на величину, кратную 2π; другими словами, на двойной длине резонатора укладывается целое число q длин волн: $2L=q\lambda_q$, откуда $\omega_q=q\pi c/L$.

Условие цикличности требует, чтобы соответствующий рассматриваемой моде световой пучок полностью воспроизводил самого себя на протяжении одного цикла, т. е. при двойном прохождении резонатора. В случае сферических зеркал этому условию удовлетворяет гауссов пучок с определенными параметрами, зависящими от геометрии резонатора. В самом деле, пусть в некоторых сечениях z_1 и z_2 (рис. 6.22) имеются сферические зеркала, отражающие поверхности которых совпадают с волновыми поверхностями гауссова пучка. Тогда исходный гауссов пучок после отражения будет преобразован в такой же пучок, распространяющийся в противоположном направлении, а после отражения от второго зеркала он полностью совпадает с исходным. При этом мы предполагаем, что диаметр $2w(z)$ пучка в месте расположения зеркал много меньше их диаметров. Практически достаточно, чтобы диаметр d зеркала в несколько раз превосходил диаметр $2w$ пучка: интенсивность настолько быстро уменьшается при $x^2+y^2>w^2$, что при $d=3\cdot 2w$ мимо зеркала проходит лишь 0,01% от полного светового потока. Эта величина характеризует *дифракционные потери* резонатора. Потери иного происхождения (например, из-за пропускания и по-



Открытый оптический резонатор

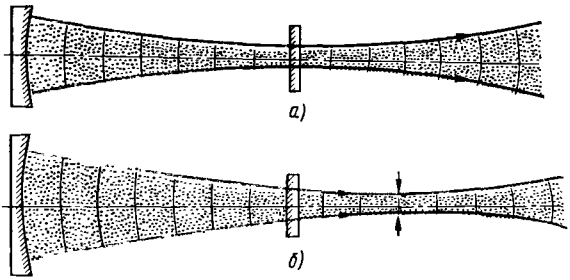
глощения зеркал), как правило, значительно выше. Обычно одно из зеркал делается частично прозрачным, что позволяет выводить наружу и использовать часть накопленной в резонаторе энергии пучка. Все виды потерь приводят к затуханию собственных колебаний напряженности поля в таком *открытом резонаторе*, что эквивалентно уменьшению добротности и уширению соответствующих его отдельным модам резонансных линий. В условиях генерации лазера потери компенсируются за счет усиления света при вынужденном излучении (см. § 9.3) в активной среде, помещенной между зеркалами резонатора.

Найдем связь параметров гауссова пучка с геометрией оптического резонатора. Будем сначала для простоты рассматривать только симметричные резонаторы, у которых оба зеркала имеют одинаковые радиусы кривизны $R_1=R_2\equiv R$. В этом случае перетяжка пучка ($z=0$) находится посередине между зеркалами, т. е. зеркала расположены при $z=\pm L/2$, где L — длина резонатора. Подставляя $z=L/2$ во вторую из формул (6.33), находим связь между радиусом кривизны R зеркал, длиной L резонатора и радиусом w_0 сечения пучка в перетяжке:

$$(kw_0^2/2)^2=(R-L/2)L/2. \tag{6.34}$$

Правая часть (6.34) должна быть положительна, поэтому гауссов пучок в симметричном резонаторе может сформироваться лишь при выполнении условия $R>L/2$. Предельное значение $R=L/2$ соответствует случаю, когда сферические поверхности зеркал имеют общий центр кривизны (*концентрический резонатор*). При $R\to L/2$ радиус перетяжки $w_0\to 0$, а радиус сечения пучка на зеркалах $w(L/2)$, как видно из первой формулы (6.33), неограниченно возрастает, т. е. при зеркалах конечных размеров значительная часть светового потока проходит мимо зеркал. Поэтому в таких условиях воспроизводящий самого себя после каждого цикла световой пучок образоваться не может. Это тем более невозможно при $R<L/2$ (неустойчивый резонатор).

Для зеркал определенной кривизны R наибольший радиус перетяжки w_0 получается, как видно из (6.34), при $L=R$, т. е. когда центр кривизны каждого зеркала находится на противоположном зеркале. В этом случае фокусы зеркал совпадают и резонатор называется *софокусным* или *конфокальным*. Для гауссова пучка в конфокальном резонаторе границы между дальней и ближней зонами $z_0=\pm kw_0^2/2$ совпадают с положениями зеркал $z=\pm L/2$ [это следует из (6.34) при $R=L$]. Радиус перетяжки w_0 равен при этом $\sqrt{L/k}=\sqrt{\lambda L/(2\pi)}$, а радиус пучка на зеркалах $w=\sqrt{2}\,w_0=\sqrt{\lambda L/\pi}$. У гелий-неонового лазера, генерирующего на длине волны $\lambda=0{,}63$ мкм, при длине симметричного конфокального резонатора $L=1$ м радиус пучка на зеркалах составляет 0,45 мм, т. е. зеркала могут иметь диаметр всего несколько миллиметров. С уве-



Оптические резонаторы с плоским и выпуклым выходными зеркалами

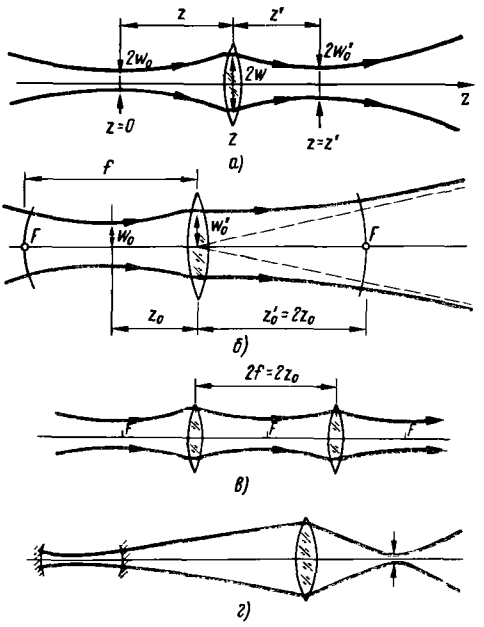
личением фокусного расстояния зеркал (при заданной длине резонатора) диаметр пучка возрастает. При этом возрастают и дифракционные потери, однако из-за малой длины волны даже очень длиннофокусные зеркала оказываются вполне приемлемыми. Так, например, из (6.34) находим, что при прежних значениях $\lambda = 0{,}63$ мкм и $L = 1$ м пучок в перетяжке будет иметь радиус $w_0 = 1$ мм, если взять зеркала с радиусом кривизны $R = 49{,}2$ м. Радиус дифракционной расходимости $z = \pi w_0^2/\lambda$ при $w_0 = 1$ мм составляет 5,0 м, поэтому весь резонатор длиной $L = 1$ м находится в пределах ближней зоны и пучок в резонаторе имеет практически постоянное сечение: $w \approx w_0 = 1$ мм.

Зеркала оптического резонатора могут иметь и разную кривизну. В самом деле, любую из сферических поверхностей равных фаз гауссова пучка (рис. 6.22) можно заменить зеркалом того же радиуса кривизны, и это не приведет к изменению структуры поля в резонаторе. В частности, одно из зеркал может быть плоским (рис. 6.23, *а*). В этом случае перетяжка гауссова пучка расположена непосредственно в плоскости зеркала, и если оно полупрозрачное, то лазерный пучок на выходе из резонатора имеет плоский волновой фронт. Если выходное зеркало сделать выпуклым (рис. 6.23, б), то перетяжка пучка расположится вне резонатора, т. е. выходящий из лазера пучок будет сходящимся.

Выше предполагалось, что находящаяся в резонаторе активная среда (стеклянный или кристаллический стержень, газоразрядная трубка и т. п.) не оказывает влияния на форму волновых поверхностей. Эта идеализация допустима для многих газовых лазеров низкой мощности, для некоторых лазеров на неодимовом стекле и на красителях.

Таким образом, параметры выходящего через частично прозрачное зеркало лазерного пучка — радиус кривизны волновой поверхности и ширина — полностью определяются геометрией резонатора. Если пренебречь небольшим преломлением в материале под-

ложки*, то кривизна волновой поверхности на выходе совпадает с кривизной зеркала. При использовании лазерного излучения нередко возникает необходимость преобразовать параметры пучка в зависимости от характера решаемой задачи. Рассмотрим преобразование гауссова пучка при прохождении через тонкую линзу. Пусть центр линзы лежит на оси пучка (рис. 6.24, *а*). Оптический путь луча, пересекающего линзу на расстоянии $\sqrt{x^2+y^2}$ от оси, меньше, чем у распространяющегося вдоль оси луча, на $(x^2+y^2)/(2F)$, где F — фокусное расстояние линзы. Поэтому при прохождении пучка через расположенную в точке z линзу величина $z+(x^2+y^2)/(2R)$, входящая в выражение для фазы волны в формуле (6.32), заменится на другую величину



6.24

Преобразование гауссова пучка при прохождении через линзу

$$z+(x^2+y^2)[1/(2R)-1/(2F)]+\text{const},$$

а распределение амплитуды $\exp[-(x^2+y^2)/w(z)]$ не изменится. Это значит, что после прохождения через линзу пучок остается гауссовым, но радиус кривизны его волновой поверхности становится другим:

$$1/R'=1/R-1/F. \qquad (6.35)$$

Если $F<R$, то $R'<0$, т. е. знак кривизны изменяется: расходящийся ($R>0$) гауссов пучок после прохождения через линзу будет сходящимся (рис. 6.24, *а*). Зная положение z линзы и ее фокусное расстояние F, можно определить положение z' и радиус перетяжки w_0' преобразованного пучка.

---

* Лазерные зеркала обычно представляют собой многослойную диэлектрическую структуру (см. §5.7) с высоким коэффициентом отражения для длины волны генерируемого излучения, нанесенную на подложку из стекла или плавленого кварца. Подложка зеркала не влияет на форму волнового фронта выходящего пучка, если обе ее поверхности имеют одинаковую кривизну. Если прозрачная поверхность плоская, то подложка действует как слабая рассеивающая линза.

Рассмотрим несколько простых случаев преобразования пучков. Введем в исходный пучок с радиусом перетяжки w_0 на расстоянии $z_0=\pi w_0^2/\lambda$ от нее, где радиус кривизны волнового фронта $R=2z_0$, линзу с фокусным расстоянием $F=2z_0$ (рис. 6.24, *б*). В соответствии с (6.35) у выходящего из линзы пучка $1/R'=0$, т. е. в плоскости линзы он имеет плоский волновой фронт. Это значит, что здесь расположена перетяжка преобразованного пучка. Ее радиус w_0' равен, очевидно, радиусу исходного пучка при $z=z_0$: $w_0'=w(z_0)=\sqrt{2}\,w_0$. Радиус дифракционной расходимости $z_0'=\pi w_0'^2/\lambda$ у нового пучка равен $2z_0$, т. е. в два раза больше, а угол расходимости в дальней зоне $\theta'=\lambda/(\pi w_0')=\theta/\sqrt{2}$ в $\sqrt{2}$ раз меньше, чем у исходного пучка.

Дифракционные потери энергии при преобразовании пучка будут тем меньше, чем больше диаметр линзы по сравнению с диаметром пучка.

Если ввести в пучок тонкую линзу с фокусным расстоянием F, равным половине радиуса кривизны волновой поверхности в этом месте пучка ($F=R/2$), то после прохождения через линзу получим, как видно из (6.35), пучок с $R'=-R$, в котором полностью воспроизводится геометрия исходного пучка. Можно, например, выбрать F равным радиусу дифракционной расходимости $z_0=\pi w_0^2/\lambda$ исходного пучка и поместить линзу на расстоянии z_0 от перетяжки, так как $R(z_0)=2z_0$. Для гелий-неонового лазера ($\lambda=0{,}63$ мкм) при радиусе перетяжки $w_0=1$ мм, $z_0=5$ м, т. е. требуется линза с $F=5$ м (оптическая сила 0,2 дптр). Последовательность таких линз, установленных на расстояниях $2z_0=10$ м, образует линзовый волновод (рис. 6.24, *в*).

По этому волноводу пучок когерентного лазерного излучения может распространяться на большое расстояние, не испытывая даже дифракционного расширения.

В некоторых случаях требуется сконцентрировать энергию лазерного пучка на возможно меньшей площади, т. е. преобразовать пучок так, чтобы получить перетяжку минимального радиуса. Для этой цели выбирают короткофокусную линзу и помещают ее, как это ни парадоксально, далеко от лазера, на расстоянии, большом по сравнению с радиусом дифракционной расходимости лазерного пучка, так, чтобы пятно лазерного излучения заполняло возможно большую часть поверхности линзы (рис. 6.24, *г*). Радиус перетяжки преобразованного пучка получается при этом приблизительно равным $2\lambda F/(\pi d)$ (d — диаметр линзы). Если $d\sim F$, то вся энергия пучка концентрируется на площадке, линейный размер которой порядка длины волны λ.

Заметим, что иллюстрируемая этими примерами возможность концентрации световой энергии в пространстве и концентрации по направлениям распространения связана с высокой пространственной когерентностью лазерного излучения.

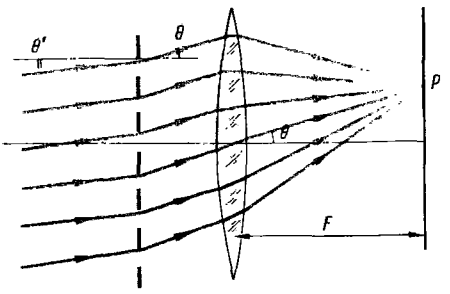
Контрольные вопросы

- Чем гауссов пучок отличается от пучка света, получающегося при прохождении плоской волны сквозь круглое отверстие? Что понимают под шириной гауссова пучка? Какую форму имеют его границы?
  Как зависит кривизна волновых поверхностей гауссова пучка от расстояния до его перетяжки? Где эта кривизна максимальна?
  В чем заключается аподизация световых пучков? С какой целью она применяется?
- Какими должны быть радиусы кривизны сферических зеркал открытого резонатора определенной длины, чтобы его дифракционные потери были минимальны?
  Как изменится раднус кривизны волнового фронта гауссова пучка при прохождении через линзу с фокусным расстоянием F?
  Где следует расположить линзу, чтобы сфокусировать лазерный пучок на возможно меньшей площади?

6.5. Дифракционные решетки

Важное практическое применение в спектроскопии имеет явление дифракции Фраунгофера на системе из большого числа одинаковых параллельных щелей, находящихся на равных расстояниях друг от друга. Так устроена простейшая *дифракционная решетка*, впервые изготовленная в 1786 г. астрономом Риттенгаузом в виде натянутых на рамку параллельных тонких проволок. Начиная с Фраунгофера, выполнившего в 1821 г. первые исследования с помощью дифракционного спектроскопа, решетки изготовляют нанесением штрихов на поверхность стеклянной или зеркальной металлической пластинки. Дифракционной решеткой можно считать любое устройство, обеспечивающее пространственную периодическую модуляцию падающей световой волны по амплитуде или фазе.

Рассмотрим распределение интенсивности света в фраунгоферовой дифракционной картине, наблюдаемой на бесконечности либо в фокальной плоскости линзы при падении на решетку плоской монохроматической волны. В каждой точке наблюдения происходит многолучевая интерференция когерентных пучков света одинаковой интенсивности, дифрагировавших на отдельных одинаковых структурных элементах решетки, например на щелях (рис. 6.25). В направлении вдоль щелей (или штрихов) размер волновой поверхности ограничен лишь размерами линзы, поэтому можно считать, что волны дифрагируют только в перпендикулярных штрихам направлениях. Обозначим $E_1(\theta)$ напряженность поля в точке наблюдения P в пучке света от



6.25
Дифракция Фраунгофера на системе одинаковых параллельных щелей

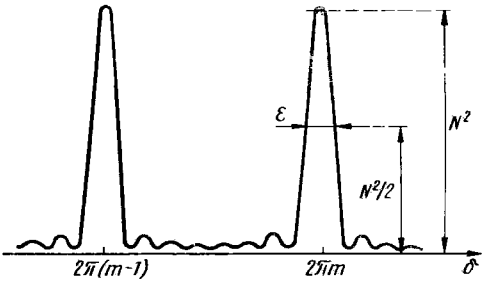
первого структурного элемента решетки. В случае решетки из щелей зависимость E_1 от угла дифракции θ выражается формулой (6.18). Напряженность E_2 в точке P от второго элемента решетки отличается от E_1 только множителем $\exp(i\delta)$, выражающим запаздывание этой волны по фазе. Напряженность от третьего элемента имеет вид $E_3=E_1\exp(2i\delta)$ и т. д. Полная напряженность в P от всех N периодических элементов решетки представляется суммой геометрической прогрессии:

$$E_P=E_1(\theta)[1+e^{i\delta}+e^{i2\delta}+...+e^{i(N-1)\delta}]=E_1(\theta)\frac{1-e^{iN\delta}}{1-e^{i\delta}}. \qquad (6.36)$$

Умножая правую часть на комплексно-сопряженное выражение, для интенсивности в точке P получаем

$$I(\theta)=I_1(\theta)\frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)}. \qquad (6.37)$$

Здесь $I_1(\theta)$ — интенсивность в P от одной щели. В результате интерференции всех N когерентных вторичных волн происходит перераспределение светового потока по направлениям и получается существенно отличающееся от $I_1(\theta)$ распределение интенсивности в фраунгоферовой дифракционной картине. В тех направлениях θ_m, для которых $\delta/2=m\pi$ ($m=0, \pm1, \pm2, ...$) и второй сомножитель в (6.37) принимает значение N^2, интенсивность в N^2 раз больше, чем от одной щели в том же направлении. Так происходит потому, что разность хода Δ вторичных волн от соседних щелей для этих направлений θ_m равна целому числу m длин волн ($\Delta=m\lambda$) и все они приходят в точку наблюдения в одинаковой фазе. Первый сомножитель $I_1(\theta)$ в (6.37), описывающий дифракцию от отдельной щели, сравнительно плавно зависит от θ, поэтому можно считать, что в направлении θ_m будет наблюдаться максимум интенсивности, если только $I_1(\theta_m)\neq0$ (если данное направление не совпадает с минимумом распределения интенсивности от одной щели). Такие максимумы называются *главными*, а целое число m — порядком главного максимума или *порядком спектра*.



6.26

Распределение интенсивности при многолучевой интерференции одинаковых пучков (график функции $\sin^2(N\delta/2)/\sin^2(\delta/2)$ при $N=8$)

Выражение (6.37) обращается в нуль каждый раз, когда $\sin(N\delta/2)=0$, но $\sin(\delta/2)\neq0$. В соответствующих таким значениям δ направлениях лежат минимумы, интенсивность света в которых равна нулю. Между ними находятся побочные (или дополнительные) максимумы. Между двумя соседними главными максимумами расположены $N-1$

минимумов и $N-2$ побочных максимумов. График второго сомножителя в (6.37), т. е. функции $\sin^2(N\delta/2)/\sin^2(\delta/2)$, приведен на рис. 6.26. При большом числе интерферирующих пучков (в современных дифракционных решетках N достигает 200 тыс.) получаются очень узкие и резкие главные максимумы. Ширина ε главного максимума на половине высоты, т. е. на уровне $N^2/2$, определяется из условия $\sin^2(N\varepsilon/4)/(\varepsilon/4)^2=N^2/2$. Обозначив $x=N\varepsilon/4$, получим для нахождения x трансцендентное уравнение $x=\sqrt{2}\sin x$. Его корень $x=1{,}39$, откуда $\varepsilon=4x/N=5{,}56/N$.

Разность фаз δ между соседними интерферирующими пучками, приходящими в некоторую точку наблюдения P, зависит от направления θ и от *постоянной решетки* d, т. е. от пространственного периода решетки. В случае простой щелевой решетки (рис. 6.26) d — это суммарная ширина прозрачного и непрозрачного участков. При нормальном падении монохроматической волны на плоскость решетки (рис. 6.27, *а*) разность хода соседних пучков $\Delta=d\sin\theta$, откуда

$$\delta=k\Delta=2\pi\Delta/\lambda=(2\pi d/\lambda)\sin\theta, \tag{6.38}$$

а направление θ_m на главный максимум m-го порядка определяется условием

$$d\sin\theta_m=m\lambda \quad (m=0,\ \pm1,\ \pm2,\ \ldots). \tag{6.39}$$

При наклонном падении плоской волны под углом θ' (рис. 6.27, *б*) разность хода соседних пучков $\Delta=d(\sin\theta-\sin\theta')$ и положение главных максимумов определяется условием

$$d(\sin\theta_m-\sin\theta')=m\lambda. \tag{6.40}$$

Эта же формула справедлива и для отражательной решетки, если углы падения θ' и дифракции θ отсчитывать в противоположных направлениях от нормали (рис. 6.27, *в*).

Структура отдельного периодического элемента (штриха) решетки (например, отношение ширины прозрачного и непрозрачного участков в случае щелевой решетки) сказывается лишь на виде функции $I_1(\theta)$ в (6.37), которая меняется при изменении угла θ значительно медленнее, чем второй (интерференционный) сомножи-

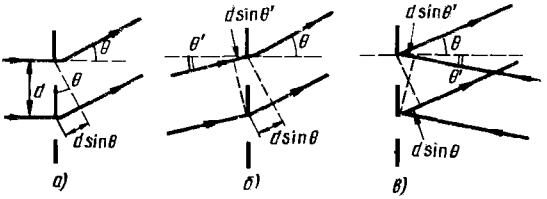


Рис. 6.27,
К вычислению разности хода волн от соседних штрихов

тель. Поэтому при большом числе штрихов огибающая функция $I_1(\theta)$ модулирует многолучевую интерференционную картину и определяет относительную интенсивность главных максимумов разных порядков, но практически не влияет на положение и ширину главных максимумов. Их положение θ_m в соответствии с (6.39)—(6.40) зависит от периода решетки d, а ширина ε еще и от полного числа штрихов N. Резкость F многолучевой интерференционной картины в монохроматическом свете (т. е. отношение расстояния между главными максимумами соседних порядков к их ширине) определяется полным числом штрихов: $F=2\pi/\varepsilon=(\pi/2{,}78)N=1{,}13N$.

Направления θ_m на главные максимумы зависят от длины волны (за исключением случая $m=0$). Поэтому решетка разлагает падающий немонохроматический свет в спектры нескольких порядков и может использоваться в качестве диспергирующего элемента в спектральных приборах. В спектрах каждого порядка наибольшее отклонение испытывают красные лучи, наименьшее — фиолетовые, в отличие от спектра, возникающего при преломлении в призме, где порядок расположения цветов обратный, так как в стекле фиолетовые лучи преломляются сильнее красных. Максимальный порядок спектра при нормальном падении ограничивается условием $|\sin|\theta\leqslant 1$, поэтому $m_{\max}\leqslant d/\lambda$. Чтобы решетка давала спектр хотя бы первого порядка, ее период должен быть больше длины световой волны: $d>\lambda$ (или $d>\lambda/2$ в случае скользящего падения).

Угловая дисперсия решетки $D_\theta=\mathrm{d}\theta/\mathrm{d}\lambda$, характеризующая изменение положения главных максимумов при изменении длины волны, может быть найдена дифференцированием обеих частей формулы (6.40) по λ при фиксированном θ':

$$\mathrm{d}\theta/\mathrm{d}\lambda=m/(d\cos\theta)=(\sin\theta-\sin\theta')/(\lambda\cos\theta). \qquad (6.41)$$

При нормальном падении ($\theta'=0$) $\mathrm{d}\theta/\mathrm{d}\lambda=(\mathrm{tg}\,\theta)/\lambda$. Таким образом, угловая дисперсия при данном значении θ не зависит от периода и других параметров решетки. Она растет с увеличением угла дифракции θ. Для спектра определенного порядка m дисперсия тем больше, чем меньше период решетки d.

Формула (6.40) позволяет также определить *свободную область дисперсии* решетки, т. е. максимальную ширину $\Delta\lambda$ спектрального интервала исследуемого излучения, при которой спектры соседних порядков еще не перекрываются. Длинноволновый конец спектра m-го порядка совпадает с коротковолновым концом спектра $(m+1)$-го порядка при выполнении условия $m(\lambda+\Delta\lambda)=(m+1)\lambda$, откуда

$$\Delta\lambda=\lambda/m \qquad (6.42)$$

— свободная область дисперсии решетки обратно пропорциональна порядку спектра. При использовании спектров низких порядков (обычно второго или третьего) решетка пригодна для исследования излучения, занимающего широкий спектральный интервал. В этом

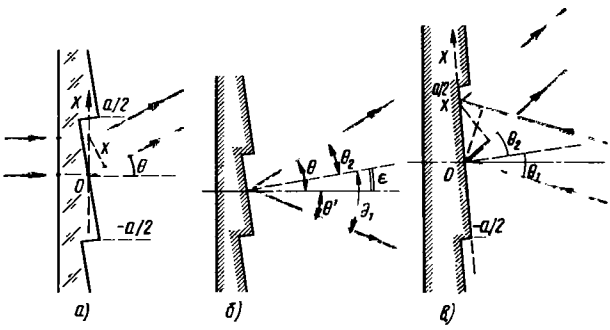
состоит главное преимущество дифракционных решеток перед интерференционными спектральными приборами (например, интерферометром Фабри—Перо, см. § 5.7), у которых из-за высоких порядков m свободные области дисперсии очень узкие. Для получения большой дисперсии в спектрах низких порядков требуется уменьшать период решетки d. Поэтому современные решетки для видимой области спектра имеют до 2400 штрихов на 1 мм.

При очень малом периоде решетки (порядка длины волны) перестают выполняться условия применимости приближенного метода Френеля, с помощью которого получена основная формула (6.37). В этом случае для нахождения распределения интенсивности $I_1(\theta)$ от одной щели решетки требуется более тонкое исследование. Оказывается, что даже при щелях с шириной $a \gtrsim \lambda$ приближение Френеля, в котором поле в пределах щели предполагается таким же, как и при отсутствии решетки, приводит к удовлетворительным результатам. Второй сомножитель в (6.37), обусловленный интерференцией вторичных волн от всех щелей, остается без изменения и при сколь угодно узких щелях. Поэтому сохраняют силу как формулы (6.39)—(6.40), определяющие положение главных максимумов, так и все рассмотренные выше следствия из них.

Структура отдельного периодического элемента решетки, т. е. форма ее штрихов, как уже отмечалось, не влияет на положение главных максимумов и, следовательно, дисперсию решетки. Не влияет она и на резкость интерференционной картины, так как резкость определяется только полным числом штрихов. От формы штрихов зависит лишь плавная функция $I_1(\theta)$, модулирующая многолучевую интерференционную картину. Вид этой функции определяет распределение энергии дифрагировавшего света между главными максимумами разных порядков. Для простой решетки с щелями шириной a функция $I_1(\theta)$ дается формулой (6.20) с $u=(\pi a/\lambda)(\sin\theta-\sin\theta')$. Ее график приведен на рис. 6.13. Значение u для направления на главный максимум порядка m в соответствии с (6.40) равно $u_m=(\pi a/d)m$. Так как интерференционный сомножитель в (6.37) для этого направления принимает значение N^2, интенсивность m-го главного максимума определяется выражением

$$I_m = I_0 N^2 \frac{\sin^2(m\pi a/d)}{(m\pi a/d)^2}. \qquad (6.43)$$

Помимо общей тенденции к уменьшению интенсивности с увеличением m формула (6.43) показывает зависимость I_m от отношения ширины a прозрачного участка к периоду d: при $a/d=m'/m$, где целое число $m'<m$, $I_m=0$. Например, при $a=d/3$ исчезает каждый третий главный максимум, при $a=d/4$ — каждый четвертый. Если $a=d/2$, все четные главные максимумы выпадают, так как они совпадают с минимумами огибающей функции $I_1(\theta)$, а интенсивность нечетных убывает с увеличением m: $I_m \sim 1/m^2$, так как в числителе формулы (6.43) $\sin^2(m\pi/2)=1$. Можно показать, что для обычной щелевой решетки интегральная интенсивность дифрагировавшего света, приходящегося на все главные максимумы, кроме $m=0$, будет наибольшей при $a=d/2$ и составит $^1/_4$ интенсивности



Дифракционные решетки с профилированным штрихом

падающего света. Только малая часть падающего света попадает в какой-нибудь один порядок с $m \neq 0$. При исследовании излучения слабых источников и при работе в инфракрасной области спектра это серьезный недостаток. У современных решеток он устраняется тем, что штрихам придают определенный профиль (рис. 6.28, *а*). Это позволяет сконцентрировать большую часть энергии дифрагировавшего света в одном или двух главных максимумах, лежащих по одну сторону от центрального (см. задачу 1). При гравировке таких решеток применяют алмазные резцы, острие которых имеет нужную форму.

Прозрачная или отражательная решетка с профилированным штрихом практически не влияет на амплитуду световой волны, а вносит периодические изменения в ее фазу. По этой причине такие решетки называют *фазовыми*. В случае отражательной решетки со ступенчатым профилем (рис. 6.28, *б*) максимальная интенсивность дифрагировавшего света наблюдается в направлении зеркального отражения от плоскости штриха, т. е. под «углом блеска» $\theta = \theta' + 2\varepsilon$, где ε — угол наклона грани штриха к поверхности решетки (см. задачу 2). Когда ширина рабочей грани занимает практически целый период ($a \approx d$), такие *блестящие*, или *концентрирующие*, решетки со ступенчатым профилем могут сконцентрировать почти весь дифрагировавший свет в спектр одного порядка.

Распределение дифрагировавшего света по главным максимумам разных порядков и в случае амплитудной решетки может существенно отличаться от закономерности (6.43), справедливой для простой щелевой решетки. Если пропускание решетки приводит к синусоидальной модуляции в пространстве амплитуды падающей волны, то возникают главные максимумы только с $m = 0$ и $m = \pm 1$ (см. задачу 3). Так получается потому, что всем остальным главным максимумам интерференционного сомножителя в общей формуле (6.37) соответствуют направления, для которых модулирующая функция $I_1(\theta)$ при синусоидальном пропускании обращается в нуль. Этот

результат становится особенно понятным, если вспомнить, что в фраунгоферовой дифракционной картине распределение напряженности поля представляет собой преобразование Фурье функции $E(x)$, описывающей пространственное распределение напряженности поля падающей волны сразу за решеткой:

$$E(k_x)\sim\int\limits_{-D/2}^{D/2}E(x)e^{-ik_xx}dx\sim\int\limits_{-D/2}^{D/2}[1+\alpha\cos(2\pi x/d)]e^{-ik_xx}dx, \qquad (6.44)$$

где $D=Nd$ — размер решетки в перпендикулярном штрихам направлении, $k_x=k\sin\theta$ характеризует направление дифрагировавшей волны. Вычисляя интеграл в (6.44), находим

$$E(k_x)\sim\frac{\sin(k_xD/2)}{k_xD/2}+\frac{\alpha}{2}\frac{\sin[(k_x-2\pi/d)D/2]}{(k_x-2\pi/d)D/2}+\frac{\alpha}{2}\frac{\sin[(k_x+2\pi/d)D/2]}{(k_x+2\pi/d)D/2}.$$

При больших размерах решетки напряженность поля дифрагировавшей волны $E(k_x)$ заметно отлична от нуля только вблизи тех направлений θ, которым соответствует обращение в нуль одного из знаменателей. Отсюда получаем три главных максимума: $k_x=0$ (т. е. $\theta=0$) и $k_x=\pm2\pi/d$ (т. е. $d\sin\theta=\pm\lambda$). В предельном случае бесконечной ($D\to\infty$) решетки с синусоидальным пропусканием и периодом d падающая плоская волна порождает помимо прошедшей прямо волны две плоские дифрагировавшие волны, соответствующие спектрам порядков $m=\pm1$. Этому физическому процессу можно сопоставить математическое разложение функции пропускания $1+\alpha\cos(2\pi x/d)$ в ряд Фурье, содержащий лишь три члена с пространственными частотами $k_x=0$, $k_x=\pm2\pi/d$. В случае более сложной функции пропускания (например, ступенчатой периодической функции щелевой решетки) ее разложение в ряд Фурье содержит и гармоники, соответствующие пространственным частотам $k_x=m(2\pi/d)$ с $m=\pm2$, ±3, ... Пространственная гармоника с некоторым номером m приведет к возникновению максимума порядка m в дифракционной картине, если длина волны λ падающей волны меньше, чем период этой гармоники d/m, т. е. $m_{max}=d/\lambda$. Интенсивность максимума определяется квадратом соответствующего коэффициента Фурье.

Техника изготовления дифракционных решеток доведена до высокой степени совершенства. Высококачественные отражательные решетки впервые были созданы в конце прошлого столетия Роулендом в США. В СССР усилиями Ф. М. Герасимова с сотрудниками разработана оригинальная аппаратура и налажен массовый выпуск превосходных дифракционных решеток. Современные автоматизированные делительные машины с интерференционными следящими устройствами позволяют изготовлять решетки с почти строго эквидистантным расположением штрихов. Размеры гравированных решеток достигают 40×40 см. В зависимости от области спектра они имеют от 4 (для инфракрасной) до 3600 штрихов на 1 мм (для

ультрафиолетовой области). Это, как правило, отражательные решетки с почти треугольным профилем штриха (рис. 6.28, *б*), концентрирующие до 80% падающего на решетку света в спектр какого-либо одного ненулевого порядка. Вследствие высокой стоимости оригинальных гравированных решеток получили распространение *реплики*, т. е. отпечатки с гравированной матрицы на специальных пластмассах, покрытые тонким металлическим отражающим слоем. По качеству реплики почти не уступают оригиналам.

От некоторых недостатков нарезных решеток свободны *голографические* (см. § 7.7) дифракционные решетки, которые получают путем создания в специальных светочувствительных материалах пространственно периодического распределения интенсивности, возникающего при интерференции когерентного лазерного излучения.

Контрольные вопросы

От каких параметров, характеризующих решетку, зависит положение главных максимумов дифракционной картины?
Какова наибольшая интенсивность побочных максимумов?
Каким условием определяется наибольший порядок спектра m_{max}?
Какую максимальную длину волны можно наблюдать в спектре решетки с периодом d?
В чем преимущество спектров низких порядков при использовании решетки в качестве диспергирующего элемента?
При каком отношении ширины a щели к периоду решетки d в дифракционной картине будет отсутствовать спектр 3-го порядка?
Какими преимуществами обладают дифракционные решетки с профилированным штрихом?
Как связано распределение интенсивности дифрагировавшего света по спектрам разных порядков m с разложением функции пропускания амплитудной решетки в ряд Фурье?

Задачи

1. Найти распределение интенсивности $I(\theta)$ при дифракции монохроматической плоской волны, падающей по нормали на фазовую решетку, профиль штрихов которой показан на рис. 6.28, *а*.

Для нахождения распределения интенсивности в фраунгоферовой дифракционной картине от одного штриха воспользуемся принципом Гюйгенса — Френеля. Поместим начало координат в середине штриха. Оптическая разность хода распространяющихся под углом θ вторичных волн от элементов волновой поверхности в начале координат и в точке x равна $\Delta = x\,[\sin\theta + (n-1)\alpha]$, где n — показатель преломления материала решетки, α — угол наклона плоскости штриха ($\alpha \ll 1$). Складывая вторичные волны от всех элементов в пределах штриха, т. е. интегрируя по x от $-a/2$ до $a/2$, получим

$$I_1(\theta) = I_0(\sin u/u)^2,\ u = (\pi a/\lambda)\,[\sin\theta + (n-1)\alpha].$$

Учитывая интерференцию волн от N штрихов, находим полную интенсивность света, дифрагировавшего под углом θ:

$$I(\theta)=I_1(\theta)\frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)},\quad \delta=(2\pi d/\lambda)\sin\theta.$$

Главный максимум $I_1(\theta)$ расположен при $u=0$, т. е. $\sin\theta=-(n-1)a$. Это соответствует углу отклонения луча при преломлении в призме с малым преломляющим углом α. Если в этом направлении расположен один из главных максимумов интерференционного сомножителя, то значительная часть дифрагировавшего света сконцентрируется в спектре одного порядка с $m\neq 0$. Для точного совпадения должно выполняться условие $d\sin\theta=m\lambda$, т. е. $d(n-1)\alpha=-m\lambda$. Например, концентрация света в спектр с $m=-1$ будет происходить при наклоне штрихов $\alpha=\lambda/[d(n-1)]$. Подставляя это значение α в выражение для u, получаем $u=(\pi a/\lambda)(\sin\theta+\lambda/d)$. Ширина штриха a практически совпадает с периодом решетки d, поэтому $u\approx\pi(d\sin\theta/\lambda+1)$. Для направлений θ_m, соответствующих главным максимумам интерференционного члена, $d\sin\theta_m=m\lambda$ и $u_m=\pi(m+1)$. При этих значениях u_m (за исключением $m=-1$) интенсивность $I_1(\theta_m)$ в дифракционной картине от одного штриха обращается в нуль, в том числе и для $m=0$. Таким образом, такая фазовая решетка дает единственный спектр порядка $m=-1$.

2. Найти распределение интенсивности дифрагировавшего монохроматического света для отражательной решетки с профилированным штрихом (рис. 6.28, *б*).

◇ Пусть падающая волна составляет угол θ' с нормалью к плоскости решетки, дифрагировавшая — угол θ. Плоскость грани штриха наклонена на угол ε, поэтому с нормалью к грани угол падения равен $\theta_1=\theta'+\varepsilon$, угол дифракции равен $\theta_2=\theta-\varepsilon$. Направим ось x вдоль грани (рис. 6.28, *в*) и выберем начало отсчета в середине штриха. Разность хода Δ вторичных волн от середины и от элемента с координатой x составляет $\Delta=x(\sin\theta_2-\sin\theta_1)$. Суммируя по принципу Гюйгенса — Френеля вторичные волны от всех элементов штриха (интегрируя от $-a/2$ до $a/2$), находим распределение интенсивности I_1 в дифракционной картине от одного штриха:

$$I_1=I_0(\sin^2 u)/u^2,\quad u=(\pi a/\lambda)(\sin\theta_2-\sin\theta_1).$$

Центральному максимуму соответствует $u=0$, т. е. $\theta_2=\theta_1$, — он расположен в направлении зеркального отражения падающей волны от плоскости грани (угол блеска). Переходя к углам θ и θ', это условие запишем в виде $\theta=\theta'+2\varepsilon$. Если в этом направлении находится какой-либо главный максимум интерференционного сомножителя $\sin^2(N\delta/2)/\sin^2(\delta/2)$, то практически весь дифрагировавший свет сконцентрируется в спектр одного порядка. Покажем это, ограничиваясь случаем малых углов θ, θ' и ε. Направление θ_m на максимум порядка m определяется из условия (6.40): $d(\theta_m-\theta')=m\lambda$. Направление зеркального отражения $\theta=\theta'+2\varepsilon$ совпадает с θ_m, если наклон граней ε таков, что $2\varepsilon=\theta_m-\theta'=m\lambda/d$. Тогда $u\approx(\pi a/\lambda)(\theta_2-\theta_1)=(\pi a/\lambda)(\theta-\theta'-2\varepsilon)=(\pi a/\lambda)(\theta-\theta'-m\lambda/d)$. Ширина a отдельного штриха для решетки рассматриваемого профиля (рис. 6.28, *б*) практически совпадает с ее периодом d. Поэтому $u\approx\pi[d(\theta-\theta')/\lambda-m]$. Отсюда следует, что для направлений на главные максимумы интерференционного сомножителя, т. е. когда $d(\theta-\theta')=m'\lambda$, получается $u\approx\pi(m'-m)$ и интенсивность I_1 обращается в нуль для всех m' кроме $m'=m$. Решетка, у которой $2\varepsilon=m\lambda/d$, дает только один главный максимум порядка m.

3. Найти распределение интенсивности $I(\theta)$ дифрагировавшего света при падении по нормали плоской монохроматической волны на амплитудную решетку, пропускание которой изменяется по синусоидальному закону в направлении оси x, перпендикулярной ее штрихам, так что амплитуда прошедшей волны сразу за решеткой зависит от x по закону $E(x)=E_0[1+\alpha\cos(2\pi x/d)]$; $\alpha<1$.

◇ Напряженность поля $E_1(\theta)$ в фраунгоферовой дифракционной картине от одного периода в соответствии с принципом Гюйгенса — Френеля определяется интегралом

$$E_1(\theta) \sim \int_{-d/2}^{d/2} [1 + \alpha\cos(2\pi x/d)]\, e^{-ik_x x}\, dx = d\frac{\sin(\delta/2)}{\delta/2}\left[1 - \frac{\alpha}{1-(2\pi/\delta)^2}\right],$$

где $\delta = k_x d = kd\sin\theta = (2\pi d/\lambda)\sin\theta$. В результате интерференции вторичных волн от всех N периодов решетки получается следующее распределение интенсивности:

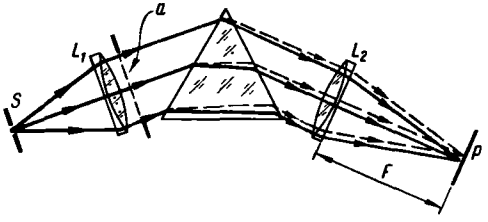
$$I(\theta) = I_1(\theta)\frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)} = I_0\frac{\sin^2(N\delta/2)}{(\delta/2)^2}\left[1 - \frac{\alpha}{1-(2\pi/\delta)^2}\right]^2.$$

Положение главных максимумов интерференционного сомножителя $\delta = 2\pi m$, за исключением $m=0$ и $m=\pm 1$, совпадает с минимумами $I_1(\theta)$. Этому соответствует сокращение $\sin^2(\delta/2)$ в числителе и знаменателе. Поэтому амплитудная дифракционная решетка с синусоидальным пропусканием дает центральный максимум при $\theta=0$ ($m=0$) и только два спектра порядка $m=\pm 1$ в направлениях, которые удовлетворяют условию $\delta = \pm 2\pi$, т. е. $d\sin\theta = \pm\lambda$. Их интенсивность найдем, раскрывая неопределенность при $\delta=0, \pm 2\pi$ в полученном выражении: $I(0) = I_0 N^2$, $I(\pm\theta_1) = I_0 N^2\alpha^2/4$. ◇

6.6. Спектральные приборы

Любой спектральный прибор выполняет *гармонический анализ* падающего на него излучения, т. е. физическое разложение излучения на монохроматические составляющие. При математическом разложении некоторой функции времени в интеграл Фурье находятся амплитуды и фазы составляющих ее гармонических колебаний. В спектральном приборе устанавливается распределение энергии исследуемого излучения по частотам, т. е. находятся интенсивности отдельных монохроматических составляющих.

Спектральные приборы можно классифицировать по характеру решаемых с их помощью задач. Прибор для фотографической регистрации спектра исследуемого излучения называется *спектрографом*. Измерения распределения энергии по спектру выполняются с помощью *спектрофотометров*. *Монохроматор* позволяет выделить в спектре падающего излучения для дальнейшего использования некоторый (обычно довольно узкий) интервал и непрерывно перемещать его по спектру. Спектральные приборы различают также по области спектра, для которой они предназначены.



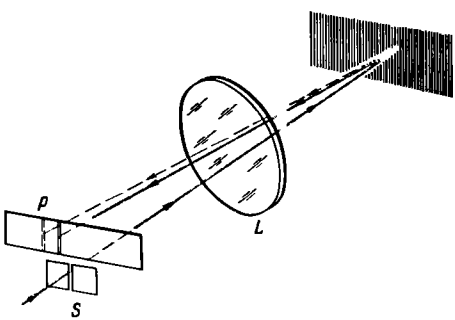
6.29
Схема призменного спектрального прибора

Действие спектральных приборов может быть основано на разных физических принципах. Основная часть прибора, называемая *диспергирующим элементом*, осуществляет обычно про-

странственное разделение излучения разных длин волн, отклоняя его на различные углы. В *призменных* спектральных приборах, берущих свое начало от известного опыта Ньютона по разложению солнечного света, используется зависимость показателя преломления от длины волны (дисперсия). Диспергирующим элементом *дифракционных* спектральных приборов служит дифракционная решетка (см. § 6.5). В приборах *высокой разрешающей силы* (интерферометр Фабри — Перо, см. § 5.7; интерферометр Майкельсона, см § 5.6, и др.) используется многолучевая или двухлучевая интерференция при очень больших разностях хода. Существуют и новые типы спектральных приборов, основанные на селективной частотной и амплитудной модуляции исследуемого излучения, возникающей при изменении разности хода интерферирующих пучков. Это фурье-спектрометры (см. § 5.6) и СИСАМы.

Принципиальная схема спектрального прибора приведена на рис. 6.29 на примере призменного спектрографа. Кроме диспергирующего элемента прибор должен иметь фокусирующую оптику. Освещаемая исследуемым излучением входная щель S находится в фокальной плоскости коллиматорного объектива L_1, назначение которого в том, чтобы направить на диспергирующий элемент параллельный пучок лучей. Выходящие из него параллельные пучки имеют для разных длин волн несколько различные направления. Второй (камерный) объектив L_2 фокусирует эти пучки. Создаваемые ими изображения входной щели — спектральные линии — получаются в разных местах фокальной плоскости объектива. Величина пространственного разделения изображений входной щели в разных длинах волн определяется *дисперсией* спектрального прибора. Угловая дисперсия D_θ характеризует изменение угла отклонения параллельного пучка света диспергирующим элементом при изменении длины волны: $D_\theta = d\theta/d\lambda$. Линейная дисперсия D_l зависит также от фокусного расстояния F_2 камерного объектива: $D_l = D_\theta F_2$.

Поместив в фокальной плоскости камерного объектива фотопластинку, можно сфотографировать спектр исследуемого излучения. Можно совместить с изображением щели S диафрагму (выходную щель) и вывести через нее излучение, соответствующее заданному участку спектра. Так работает монохроматор. Осуществляя непрерывную регистрацию энергии этого излучения с помощью фотоэлектрического приемника при перестройке (повороте) диспергирующего элемента или перемещении диафрагмы в фокальной



6.30
Автоколлимационная схема спектрального прибора

плоскости (сканирование), можно измерить распределение энергии по спектру.

Фокусирующая оптика спектрального прибора может быть как линзовой, так и зеркальной. В автоколлимационных приборах один и тот же объектив выполняет функции коллиматорного и камерного, а входная щель и ее цветные изображения лежат в одной его фокальной плоскости (рис. 6.30). В приборах с вогнутой отражательной дифракционной решеткой специальная фокусирующая оптика отсутствует. Изображение щели создает сама решетка.

Спектральный прибор можно считать идеальным, если распределение энергии на выходе не зависит от его особенностей, а определяется только свойствами источника излучения. В действительности любой спектральный прибор и любой приемник излучения вносят в спектр искажения. Рассмотрим сначала идеализированный случай, когда спектр падающего излучения состоит из нескольких монохроматических компонент. На выходе реального прибора вместо соответствующих этому излучению бесконечно узких спектральных линий будут наблюдаться линии, имеющие конечную ширину. Вносимое спектральным прибором уширение линий обусловлено следующими причинами: 1) наблюдаемая спектральная линия представляет собой изображение входной щели, имеющей конечную ширину; 2) дифракционные явления в диспергирующем элементе уширяют геометрическое изображение щели; 3) аберрации оптической системы и дефекты диспергирующего элемента также уширяют изображение; 4) система регистрации излучения дополнительно уширяет линию (из-за конечных размеров диафрагмы фотоприемника или зерна фотоэмульсии).

Относительный вклад каждой из этих причин зависит от конструктивных особенностей прибора. Но любой реальный прибор, регистрируя монохроматическое излучение, дает некоторый контур конечной ширины, описываемый функцией $f(\lambda)$. Эта функция определяется свойствами спектрального прибора и называется *аппаратной функцией* или *инструментальным контуром*. Каждой длине волны λ в приборе соответствуют некоторый угол отклонения φ и определенная точка x фокальной плоскости камерного объектива. Поэтому инструментальный контур можно записать также в виде $f(\varphi)$ или $f(x)$. Эта функция дает распределение интенсивности в фокальной плоскости прибора, создаваемое монохроматическим источником.

Чем уже инструментальный контур, тем меньше искажений вносит спектральный прибор в измеряемый спектр.

Аппаратную функцию принято нормировать на единицу, т. е. требовать, чтобы она удовлетворяла следующему условию нормировки:

$$\int f(\varphi)\mathrm{d}\varphi = 1 \quad \text{или} \quad \int f(x)\mathrm{d}x = 1. \tag{6.45}$$

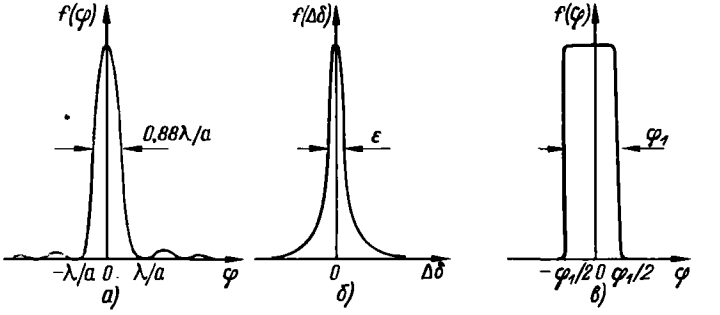
Тогда на интервал $(\varphi, \varphi+d\varphi)$ приходится световой поток $d\Phi = \Phi f(\varphi)d\varphi$, где Φ — полный световой поток.

Рассмотрим примеры инструментальных контуров разных спектрометров. Пусть входная щель бесконечно узка и аберрации оптической системы пренебрежимо малы. Тогда причина уширения изображения щели лежит в самом диспергирующем элементе. В призменном спектрографе изображение щели расширяется вследствие дифракции при ограничении ширины падающего на призму параллельного пучка диафрагмой или краями призмы. Угловое распределение интенсивности монохроматического дифрагировавшего света в этом случае будет таким же, как при дифракции Фраунгофера на щели, т. е. описывается формулой (6.20) с $u=(\pi a/\lambda)\sin\varphi$, где a — размер диафрагмы, ограничивающей ширину пучка. Угол φ отсчитывается здесь от направления на центр геометрического изображения щели. При достаточно большой ширине a пучка дифракционное уширение изображения невелико, т. е. существенны лишь малые углы $\varphi \ll 1$, для которых $\sin\varphi \approx \varphi$. Удовлетворяющая условию (6.45) аппаратная функция имеет вид

$$f(\varphi)=\frac{1}{\pi}\frac{\sin^2(\pi a\varphi/\lambda)}{(\pi a\varphi/\lambda)^2}. \qquad (6.46)$$

Такой обусловленный дифракцией аппаратный контур показан на рис. 6.31, *а*. Его ширина на половине высоты определяется корнем уравнения $x=\sqrt{2}\sin x$ ($x=1{,}39$) и равна $0{,}88\lambda/a$.

В приборе с дифракционной решеткой при работе в спектре m-го порядка вид аппаратного контура (при бесконечно узкой входной щели) определяется распределением интенсивности монохроматического дифрагировавшего света в окрестности главного максимума порядка m. При большом числе штрихов $N \gg 1$ максимумы очень узкие и поэтому в формуле (6.37) плавный сомножитель $I_1(\theta)$, описывающий распределение интенсивности в дифракционной картине от одного штриха, в пределах инструментального контура можно считать постоянным. Будем и здесь отсчитывать



6.31
Аппаратная функция спектрального прибора

угол φ от направления на центр геометрического изображения щели, т. е. от направления θ_m на m-й главный максимум. Так как существенны только значения $\varphi \ll 1$, то $\sin\theta = \sin(\theta_m + \varphi) \approx \sin\theta_m + \varphi\cos\theta_m$ и разность фаз δ волн от соседних штрихов в (6.37) можно представить в виде

$$\delta = (2\pi d/\lambda)(\sin\theta - \sin\theta') \approx 2\pi m + (2\pi d/\lambda)\cos\theta_m \cdot \varphi,$$

где учтено, что для направления на главный максимум $d(\sin\theta_m - \sin\theta') = m\lambda$ (θ' — угол падения). Подставляя это выражение для δ в (6.37) и учитывая, что в знаменателе можно заменить $\sin[(\pi d/\lambda)\cos\theta_m \cdot \varphi]$ его аргументом ($\varphi \ll 1$), получим для прибора с дифракционной решеткой такую же нормированную аппаратную функцию (6.46), что и для призмы, с $a = Nd\cos\theta_m$. Этого и следовало ожидать, так как Nd — это полный размер решетки, а $Nd \times \cos\theta_m$ — площадь сечения дифрагировавшего пучка, образующего изображение входной щели в спектре m-го порядка (см. рис. 6.25).

В интерферометре Фабри—Перо (см. § 5.7) распределение интенсивности при освещении монохроматическим светом описывается формулой Эйри (5.72). При высоком коэффициенте отражения R зеркал отдельные максимумы имеют лоренцевскую форму (5.75). Такую же форму будет иметь и аппаратная функция, т. е. регистрируемый в монохроматическом свете контур отдельной полосы равного наклона или сигнал фотоприемника при сканировании с использованием круглой диафрагмы очень малого диаметра (что соответствует бесконечно узкой щели в дифракционном или призменном приборе). Если ее рассматривать как функцию от $\Delta\delta = \delta - 2\pi m$, т. е. отклонения разности фаз δ от ее значения в соответствующем максимуме, то в соответствии с (5.75)

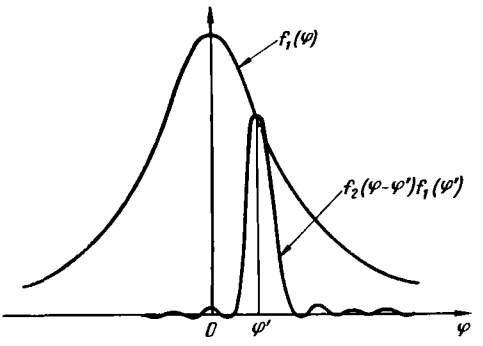
$$f(\Delta\delta) = \frac{2/(\pi\varepsilon)}{1 + (2\Delta\delta/\varepsilon)^2}. \qquad (6.47)$$

Здесь $\varepsilon = 2(1-R)/\sqrt{R}$ — ширина аппаратного контура на половине высоты. Вид этого контура показан на рис. 6.31, *б*.

Рассмотрим теперь случай широкой входной щели, когда ширина ее геометрического изображения в призменном или дифракционном спектрографе во много раз превышает ширину главного дифракционного максимума ($\sim\lambda/a$). В таких условиях дифракцией можно пренебречь и изображение щели в фокальной плоскости будет равномерно освещенным. Аналогично, в интерферометре Фабри—Перо с круглой (кольцевой) диафрагмой, размер которой много больше ширины отдельного максимума, сигнал при сканировании остается постоянным, пока светлое кольцо находится в пределах диафрагмы. Соответствующий инструментальный контур имеет прямоугольную форму (рис. 6.31, *в*).

В промежуточном случае, когда ширина геометрического изображения щели одного порядка є шириной дифракционного максимума, для нахождения аппаратного контура можно поступить

следующим образом. Рассмотрим входную щель как источник конечной ширины и разделим его мысленно на отдельные элементарные полоски. Считая их некогерентными бесконечно узкими источниками, найдем освещенность в фокальной плоскости как наложение описываемых формулой (6.46) распределений, центр каждого из которых расположен при $\varphi = \varphi'$, т. е. там, где находится геометрическое изображение соответствующего элемента щели:



6.32.
К нахождению результирующего аппаратного контура

$$F(\varphi)=\int_{-\varphi_1/2}^{\varphi_1/2} f(\varphi-\varphi')\,d\varphi'. \qquad (6.48)$$

Получаемое в монохроматическом свете изображение входной щели из-за дифракции оказывается размытым по краям. Обе причины, т. е. конечная ширина входной щели и дифракция в диспергирующем элементе, дают вклад в ширину результирующего аппаратного контура (6.48).

В более общем случае, когда одна из причин уширения дает контур, описываемый функцией $f_1(\varphi)$, а другая — функцией $f_2(\varphi)$, результирующий аппаратный контур можно получить тем же способом. Один из контуров, например $f_1(\varphi)$, нужно разделить на узкие элементы шириной $d\varphi'$. Отдельный элемент, лежащий при $\varphi=\varphi'$, под действием второй причины дает контур, описываемый функцией f_2, но смещенный на φ', т. е. $f_2(\varphi-\varphi')$. Его интенсивность пропорциональна $f_1(\varphi')d\varphi'$ (рис. 6.32). Наложение контуров от отдельных элементов

$$F(\varphi)=\int_{-\infty}^{\infty} f_1(\varphi')f_2(\varphi-\varphi')\,d\varphi' \qquad (6.49)$$

дает результирующую аппаратную функцию. Стоящий в правой части (6.49) интеграл называется *сверткой функций* f_1 и f_2. Когда есть несколько независимых причин уширения, аппаратный контур представляет собой свертку функций, каждая из которых описывает контур, обусловленный одной из причин уширения.

Аппаратный контур характеризует те искажения, которые вносит прибор в спектр монохроматического излучения, т. е. при регистрации бесконечно узкой спектральной линии.

Рассмотрим теперь немонохроматическое излучение, спектральное распределение интенсивности которого характеризуется некото-

рой функцией $I(\lambda)$. В случае квазимонохроматического света функция $I(\lambda)$ описывает контур его спектральной линии с некоторой шириной $\Delta\lambda \ll \lambda$. Примеры спектральных линий и приводящих к ним статистических моделей излучения были приведены в § 1.8, 5.4. Излучение сложного спектрального состава можно представить в виде совокупности отдельных взаимно некогерентных монохроматических компонент, отклик спектрального прибора на каждую из которых описывается аппаратным контуром пропорциональной интенсивности с центром при соответствующем ее длине волны значении φ. Поэтому реально наблюдаемый контур $I_{\text{набл}}(\varphi)$ представляет собой свертку спектрального контура $I(\varphi)$ исследуемого излучения и общего инструментального контура $F(\varphi)$ независимо от того, в результате каких причин уширения последний образован:

$$I_{\text{набл}}(\varphi)=\int_{-\infty}^{\infty} F(\varphi')I(\varphi-\varphi')\mathrm{d}\varphi'. \qquad (6.50)$$

Когда все спектральное распределение исследуемого излучения сосредоточено в интервале, узком по сравнению с шириной инструментального контура $F(\varphi)$, наблюдаемый контур $I_{\text{набл}}(\varphi)$ практически совпадает с инструментальным. Напротив, для того чтобы $I_{\text{набл}}(\varphi)$ совпадал с $I(\varphi)$, т. е. чтобы прибор непосредственно регистрировал спектральное распределение исследуемого излучения, аппаратный контур должен быть много у́же деталей этого распределения.

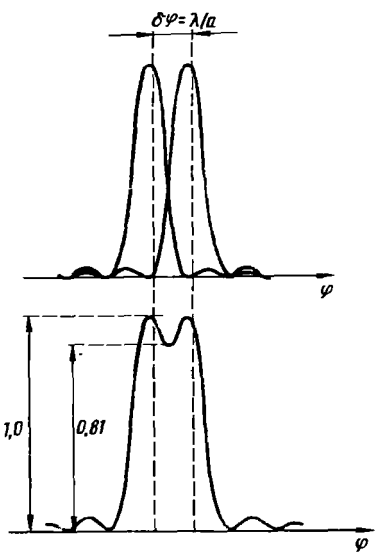
Цель спектральных измерений состоит в нахождении истинного, не искаженного прибором распределения энергии в спектре исследуемого излучения. Когда инструментальные искажения значительны, нахождение функции $I(\varphi)$ по известным функциям $I_{\text{набл}}(\varphi)$ и $F(\varphi)$ сводится к решению интегрального уравнения (6.50). Такая обратная оптическая задача, или задача редукции к идеальному прибору, в принципе разрешима даже при очень широком инструментальном контуре, если только функции $I_{\text{набл}}(\varphi)$ и $F(\varphi)$ известны совершенно точно. В действительности они могут быть получены лишь в результате измерений распределения интенсивности в фокальной плоскости прибора. Эти измерения неизбежно содержат ошибки (шумы), что накладывает ограничения на точность, с которой может быть восстановлена функция $I(\varphi)$. Восстановление $I(\varphi)$ относится к числу так называемых некорректных математических задач, когда малые ошибки в значениях $I_{\text{набл}}(\varphi)$ могут приводить к очень большим погрешностям при нахождении $I(\varphi)$.

Случайные ошибки измерений обусловлены не только погрешностями измерительных приборов, но и самой природой измеряемой величины — световой энергии, излучение и поглощение которой имеет квантовый характер. Для точного измерения спектрального распределения энергии нужно уменьшать спектральный интервал $\Delta\lambda$, но при этом возрастают статистические погрешности результатов измерений. Возрастают они и при уменьшении времени измерения, а также при уменьшении самой измеряемой интенсивности. В самом деле, пусть в выделенном интервале $\Delta\lambda$ за время

измерения регистрируется в среднем n квантов. Отклонение Δn числа зарегистрированных квантов в отдельном измерении от среднего значения n составляет в среднем $\sqrt{n}$. Поэтому относительная погрешность $\Delta n/n = 1/\sqrt{n}$ мала при $n \gg 1$, однако при $n \approx 1$ она приближается к 100%. Это налагает принципиальные ограничения на точность измерения распределения $I_{\text{набл}}(\varphi)$. В то же время для детального нахождения истинного спектрального распределения требуется тем точнее знать $I_{\text{набл}}(\varphi)$, чем шире инструментальный контур.

Таким образом, *реальные* возможности спектрального прибора более или менее детально исследовать спектр излучения в значительной степени определяются шириной и формой его инструментального контура. Для характеристики этих возможностей вводится понятие *разрешающей силы* или *разрешающей способности*. Так называется отношение длины волны λ к $\delta\lambda$ — наименьшей разности длин волн двух монохроматических спектральных линий, при которой спектральный прибор еще позволяет раздельно их наблюдать. Возможность разделить две монохроматические линии зависит от точности, с которой измерен инструментальный контур и может быть измерен наблюдаемый контур. Поэтому количественный критерий разрешающей способности можно сформулировать только условно в предположении определенной точности энергетических измерений распределения интенсивности.

Условный критерий, удобный для сравнения разрешающей силы различных спектральных приборов, был предложен Рэлеем для случая, когда инструментальный контур имеет дифракционную форму (6.46). Согласно *критерию Рэлея*, наименьший разрешимый интервал $\delta\lambda$ равен расстоянию между главным максимумом и ближайшим к нему минимумом аппаратной функции (6.46). Две монохроматические линии одинаковой интенсивности на таком расстоянии друг от друга (в угловой мере оно равно $\delta\varphi = \lambda/a$) дают суммарный контур с двумя максимумами (рис. 6.33), провал между которыми, как легко подсчитать, составляет около 20% от интенсивности в максимумах. Благодаря провалу такой контур воспринимается как двойная спектральная линия. Если считать критерием разрешения именно наличие провала, то можно обобщить критерий Рэлея на случаи, когда ап-



6.33
Критерий Рэлея

паратная функция отличается от дифракционной: при любой форме инструментального контура две монохроматические линии одинаковой интенсивности находятся на пределе разрешения $\delta\lambda$, если провал в суммарном контуре составляет 20%.

Подчеркнем еще раз условный характер критерия Рэлея. Если интенсивность одной из линий существенно больше другой, то провал в наблюдаемом контуре может отсутствовать даже тогда, когда расстояние между ними значительно больше, чем требует критерий Рэлея. С другой стороны, линии, расположенные ближе, могут быть разрешены, если погрешность измерения наблюдаемого распределения интенсивности меньше 20%. Фактически возможность разрешения близких спектральных линий, как уже отмечалось, ограничивается шумами при измерениях: линии можно разрешить, если наблюдаемое распределение отличается от распределения для одиночной линии больше чем на ошибку измерения.

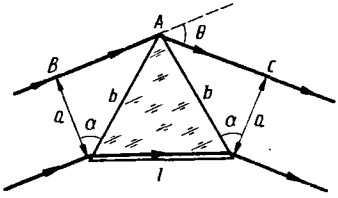
При дифракционной форме инструментального контура (6.46), когда в соответствии с критерием Рэлея $\delta\varphi=\lambda/a$, предел разрешения $\delta\lambda$ в длинах волн и разрешающую силу $\lambda/\delta\lambda$ можно выразить через угловую дисперсию прибора $D_\theta=\mathrm{d}\theta/\mathrm{d}\lambda$:

$$\delta\lambda=\delta\varphi/D_\theta=\lambda/(aD_\theta),\ \lambda/\delta\lambda=aD_\theta. \qquad (6.51)$$

Таким образом, теоретическая разрешающая сила равна произведению ширины a параллельного пучка, выходящего из диспергирующего элемента, и угловой дисперсии D_θ. Эта вспомогательная формула удобна для нахождения обусловленного дифракцией теоретического предела разрешающей способности разных спектральных приборов.

В призменных спектральных приборах наибольшая разрешающая сила достигается при симметричной установке призмы, когда угол θ отклонения параллельного пучка от первоначального направления минимален (рис. 6.34). В условиях минимального отклонения $\theta=2\alpha-A$, где α — угол падения, A — угол при вершине призмы. Поэтому для угловой дисперсии, обусловленной зависимостью показателя преломления материала призмы от длины волны, можно написать

$$D_\theta=\frac{\mathrm{d}\theta}{\mathrm{d}\lambda}=\frac{\mathrm{d}\theta}{\mathrm{d}n}\frac{\mathrm{d}n}{\mathrm{d}\lambda}=2\frac{\mathrm{d}\alpha}{\mathrm{d}n}\frac{\mathrm{d}n}{\mathrm{d}\lambda}. \qquad (6.52)$$



6.34
К определению дисперсии и разрешающей способности призмы

Учитывая, что оптическая длина пути луча, идущего вдоль основания призмы, равна сумме длин BA и AC, можем написать $nl=2b\sin\alpha$, где b — длина ребра призмы. Отсюда $l\mathrm{d}n=2b\cos\alpha\mathrm{d}\alpha$, т. е. $\mathrm{d}\alpha/\mathrm{d}n=l/(2b\cos\alpha)$. Но $b\cos\alpha=a$, где a — ширина пучка, ограниченная размерами призмы. Подставляя $\mathrm{d}\alpha/\mathrm{d}n=$

$=l/(2a)$ в (6.52), получаем следующее выражение для угловой дисперсии призмы:

$$D_\theta=(l/a)\,\mathrm{d}n/\mathrm{d}\lambda. \tag{6.53}$$

Разрешающая сила $\lambda/\delta\lambda$ в соответствии с (6.51)

$$\lambda/\delta\lambda=l\,\mathrm{d}n/\mathrm{d}\lambda \tag{6.54}$$

зависит от $\mathrm{d}n/\mathrm{d}\lambda$, т. е. дисперсии вещества, из которого она сделана, и пропорциональна длине l основания призмы, но не зависит от преломляющего угла A: призмы с разными углами A, но равными основаниями обладают одинаковой разрешающей силой. Однако с увеличением угла A уменьшается a (рис. 6.34) и, как ясно из (6.53), дисперсия призмы возрастает, что облегчает задачу практической реализации ее разрешающей способности (см. ниже). Поэтому предпочтение отдается призмам с большим преломляющим углом, близким к предельному значению $A_{\max}=2\arcsin(1/n)$ (при бо́льших углах любой луч, вошедший в призму, испытывает на второй грани полное отражение). Наиболее часто используют призмы с преломляющим углом $A\approx 60°$. Угол отклонения при этом близок к 45°.

Один из существенных недостатков призмы состоит в быстром уменьшении разрешающей способности и дисперсии при продвижении в длинноволновую часть спектра. Для тяжелого стекла сорта «флинт» в синей части спектра $\mathrm{d}n/\mathrm{d}\lambda\approx 3200\ \text{см}^{-1}$, в красной $\approx 1170\ \text{см}^{-1}$. При таких значениях разрешающая способность призмы с основанием $l=5$ см в синей части составляет ≈ 15000, в красной ≈ 5000. Для увеличения разрешающей способности можно применить систему из нескольких последовательно расположенных призм, что эквивалентно увеличению l в формуле (6.54).

Угловая дисперсия дифракционной решетки в спектре m-го порядка в соответствии с (6.41) равна $D_\theta=m/(d\cos\theta)$. Учитывая, что ширина a параллельного пучка, дифрагировавшего под углом θ, для решетки из N штрихов равна $Nd\cos\theta$ (см. рис. 6.25), из формулы (6.51) найдем теоретическую разрешающую способность:

$$\lambda/\delta\lambda=mN. \tag{6.55}$$

Эту формулу можно получить и непосредственно, применяя критерий Рэлея к многолучевой интерференции волн, дифрагировавших от отдельных штрихов. Направление на главный максимум порядка m для длины волны $\lambda+\delta\lambda$ определяется из условия, что разность хода Δ от соседних штрихов равна $m(\lambda+\delta\lambda)$. В этом же направлении для длины волны λ будет расположен минимум, ближайший к главному максимуму порядка m, если $\Delta=(m+1/N)\lambda$ (между двумя главными максимумами расположен $N-1$ минимум). Поэтому $m(\lambda+\delta\lambda)=(m+1/N)\lambda$, откуда $\lambda/\delta\lambda=mN$.

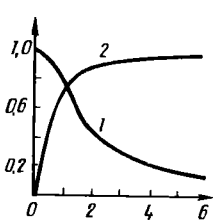
Отметим, что разрешающая способность решетки в спектре данного порядка m определяется только полным числом штрихов N в отличие от дисперсии, которая зависит от периода d, т. е. от числа штрихов на единицу длины решетки N/l. При заданном значении

постоянной решетки d разрешающая способность возрастает с увеличением длины l ее рабочей области. Для увеличения разрешающей силы стремятся к изготовлению дифракционных решеток больших размеров (до 15—20 см для видимой области и до 40 см для инфракрасной).

В соответствии с формулой (6.40) $m=d(\sin\theta_m-\sin\theta')/\lambda$, поэтому выражение (6.55) для разрешающей силы можно также записать в виде

$$\lambda/\delta\lambda=Nd(\sin\theta-\sin\theta')/\lambda, \qquad (6.56)$$

т. е. разрешающая сила равна числу длин волн, укладывающихся в разность хода лучей, дифрагировавших в направлении θ от крайних штрихов решетки (они находятся на расстоянии $l=Nd$). Ее максимальное значение будет в спектре наибольшего порядка m_{max}, который возможен при заданных d и λ. Так как $m_{max}\leqslant 2d/\lambda$, то максимальная разрешающая сила не превосходит $2Nd/\lambda$, т. е. числа длин волн, укладывающихся на удвоенной длине рабочей части решетки. Она не зависит от того, образована ли решетка большим числом штрихов малого периода или малым числом штрихов большого периода, если общая длина Nd одна и та же. Однако грубые решетки будут иметь ту же дисперсию и разрешающую силу только в спектрах высоких порядков m. Решетки с большим числом штрихов на единицу длины, предназначенные для работы в спектрах низких порядков m, обладают важным преимуществом — большой свободной областью дисперсии $\Delta\lambda=\lambda/m$ (6.42), что позволяет применять их для исследования излучения с широким спектральным составом.



6.35

Отношение разрешающей силы к ее значению при бесконечно узкой щели (кривая *1*) и интенсивность в центре линии (кривая *2*) в зависимости от ширины входной щели

По оси абсцисс отложено отношение ширины входной щели к ширине нормальной щели

Ширина и форма инструментального контура прибора определяются, как уже отмечалось, суммарным действием ряда факторов. К ним относятся конечная ширина входной щели, дефекты фокусирующей оптики, зернистое строение фотоэмульсии и рассеяние света в ней, конечная ширина диафрагмы при фотоэлектрической регистрации. Поэтому далеко не просто реализовать в эксперименте высокие значения теоретической разрешающей силы, обусловленной диспергирующим элементом. В частности, требуются высококачественные фокусирующие объективы большого диаметра и достаточно длиннофокусные, чтобы обеспечить необходимую линейную дисперсию $D_l=D_\theta=D_\theta F$. При малой линейной дисперсии будет сказываться зернистость фотоэмульсии.

При конечной ширине входной щели ширина и форма инструментального коитура в соответствии с (6.48) зависят от соотиошения

между шириной φ_1 геометрического изображения щели и расстоянием $\delta\varphi = \lambda/a$ между двумя линиями, разрешимыми согласно критерию Рэлея. Если $\varphi_1 = \delta\varphi$, щель называется *нормальной*. Зависимость практической разрешающей силы от ширины щели (для некогерентного освещения) показана на рис. 6.35. При расширении щели, пока она остается уже нормальной, ширина контура растет очень медленно. Так же медленно уменьшается разрешающая способность: для нормальной щели она составляет более 75% от значения при бесконечно узкой щели. Когда ширина щели превышает нормальную более чем вдвое, ширина инструментального контура становится примерно равной ширине геометрического изображения щели. На том же рисунке показана зависимость интенсивности в центре монохроматического изображения щели от ее ширины. Для узкой щели интенсивность растет пропорционально ширине, так как растущий световой поток приходится на ту же площадь дифракционного изображения щели. После достижения нормальной ширины рост интенсивности резко замедляется и она приближается к постоянному значению, соответствующему бесконечно широкой щели, ибо одинаково быстро растут как световой поток, так и площадь изображения щели, по которой он распределяется. Для многих целей оптимальная ширина щели близка к нормальной. В тех случаях, когда дифракционный предел разрешающей способности не достигается из-за аберраций оптической системы или зерна фотоэмульсии, оптимальная ширина щели может быть в несколько раз больше нормальной.

Как показывает формула (6.55), разрешающая способность спектрального прибора равна произведению порядка интерференции m на число N интерферирующих световых пучков. Высокая разрешающая сила хороших дифракционных решеток достигается за счет больших значений общего числа штрихов N при низких порядках интерференции ($m = 1, 2, 3$). В интерференционных спектральных приборах, наоборот, число пучков сравнительно невелико ($N \approx 30$ для интерферометра Фабри—Перо, $N = 2$ для интерферометра Майкельсона), а большое разрешение достигается за счет высоких порядков интерференции m.

Для интерферометра Фабри—Перо с аппаратной функцией (6.47) пределом разрешения $\delta\lambda$ можно считать ширину контура на половине высоты. Провал в наблюдаемом контуре от двух находящихся на таком расстоянии монохроматических линий составляет около 17%, т. е. это условие практически совпадает с обобщенным критерием Рэлея. Ширине контура соответствует изменение разности фаз на $\varepsilon = 2(1-R)/\sqrt{R}$ [см. (5.74)]. Разность фаз δ интерферирующих волн в максимуме m-го порядка равна $2\pi m$. Изменению ее на ε соответствует изменение длины волны на $\delta\lambda = [\varepsilon/(2\pi m)]\lambda$, откуда для разрешающей силы $\lambda/\delta\lambda$ находим

$$\lambda/\delta\lambda = m(2\pi/\varepsilon) = mF, \qquad (6.57)$$

где $F = 2\pi/\varepsilon = \pi\sqrt{R}/(1-R)$ — резкость интерференционных полос

(5.76). Сравнивая (6.57) с (6.55) видим, что резкость F играет роль эффективного числа пучков в интерферометре Фабри—Перо: такое число пучков равной интенсивности обеспечивает ту же разрешающую способность, что и бесконечная последовательность пучков убывающей интенсивности. При $R=0{,}9$ эффективное число пучков $F\approx 30$. Порядок интерференции m для центра системы колец равен $m=2h/\lambda$. При толщине $h\approx 1$ см для $\lambda=500$ нм $m\approx \approx 4\cdot 10^4$ и теоретическая разрешающая сила превышает 1 млн. Увеличивая толщину h, можно добиться еще больших значений $\lambda/\delta\lambda$, но это приведет к пропорциональному уменьшению свободной области дисперсии $\Delta\lambda=\lambda/m=\lambda^2/(2h)$, что целесообразно лишь при исследовании очень узких спектральных линий.

Другая возможность увеличения разрешающей способности интерферометра Фабри—Перо заключается в повышении коэффициента отражения R зеркал. Однако в реальном приборе такая возможность ограничена несовершенством его поверхностей. Инструментальный контур неидеального интерферометра уширяется из-за наложения смещенных относительно друг друга контуров Эйри, создаваемых разными участками его поверхностей (см. § 5.7). При очень высоком коэффициенте отражения R контуры от отдельных участков становятся столь узкими, что форма результирующего контура будет целиком определяться дефектами поверхностей. Дальнейшее увеличение R в таких условиях нецелесообразно, так как разрешающая способность не возрастает, а количество пропускаемой световой энергии убывает из-за сужения контуров от отдельных участков и получается лишь ухудшение отношения сигнала к шуму. Картина здесь аналогична той, что получается при сужении входной щели спектрографа, когда ее ширина меньше нормальной: разрешающая способность остается прежней, а интенсивность уменьшается.

При фотоэлектрической регистрации измеряется световой поток через круглую или кольцевую диафрагму, помещенную в плоскости, где фокусируются интерференционные кольца. Из-за конечных размеров диафрагмы получающийся при сканировании инструментальный контур отличается от кривой Эйри (6.47). Сужение диафрагмы до величины, значительно меньшей ширины отдельного максимума теоретического контура (6.47), не дает заметного увеличения разрешающей силы, но резко сокращает попадающий на фотоприемник световой поток, что снижает отношение сигнала к шуму. При компромиссном решении, когда спектральная ширина диафрагмы выбирается равной ширине $\delta\lambda$ контура Эйри, разрешающая способность в $\sqrt{2}$ раз меньше теоретического значения mF (6.57).

Интерферометр Фабри—Перо проще в обращении и обеспечивает более высокую разрешающую силу, чем приборы с большими дифракционными решетками. Осиовной его недостаток — малая величина свободной области дисперсии. Система из двух последовательиых иитерферометров, толщины которых находятся в простом кратиом отношении (мультиплекс), имеет область дисперсии, ха-

рактерную для тонкого интерферометра, а разрешающую способность — для толстого. Для исследования широких участков спектра интерферометр Фабри—Перо «скрещивают» с призменным или дифракционным спектральным прибором (см. § 5.7).

Контрольные вопросы

- Какие функции выполняют диспергирующий элемент, коллиматорный и камерный объективы спектрального прибора?
- Какие причины вызывают уширение наблюдаемой спектральной линии при регистрации прибором монохроматического излучения?
- Что называется аппаратной функцией спектрального прибора?
- Какой вид имеют аппаратные функции для призменного и дифракционного спектрографов с узкой входной щелью; для интерферометра Фабри—Перо?
- Каким будет инструментальный контур при одновременном действии двух независимых причин уширения, действия которых описываются функциями $f_1(\varphi)$ и $f_2(\varphi)$?
- Как связан регистрируемый прибором спектр с истинным спектральным распределением исследуемого излучения?
- Что значит, что критерий Рэлея имеет условный характер?
- От каких величин, характеризующих призму, зависит разрешающая способность призменного спектрографа? Почему разрешающая способность максимальна при симметричной установке призмы?
- Чему равна разрешающая способность дифракционной решетки с N штрихами в спектре порядка m? Каким параметром решетки определяется наибольшая достижимая разрешающая сила для длины волны λ?
- Как влияет ширина входной щели на разрешающую силу спектрографа? Что такое нормальная щель?
- Какие параметры интерферометра Фабри—Перо определяют его разрешающую способность?

6

ГЕОМЕТРИЧЕСКАЯ ОПТИКА И РОЛЬ ДИФРАКЦИИ В ОПТИЧЕСКИХ ПРИБОРАХ

● Многие простые оптические явления, такие, например, как возникновение теней и образование изображений в оптических приборах, можно объяснить на основе законов геометрической (или лучевой) оптики. Геометрическая оптика использует представление о световых лучах — математических линиях, вдоль которых происходит распространение энергии световых колебаний. Пучки света рассматриваются как совокупности бесконечного числа независимых лучей, удовлетворяющих хорошо известным законам прямолинейного распространения в однородной среде, отражения и преломления на границе раздела двух сред.

● Законы геометрической оптики можно получить из волновой оптики в предельном случае исчезающе малых длин волн $\lambda \to 0$, когда волновая природа света становится несущественной. Геометрическая оптика оказывается хорошим приближением при описании распространения света и позволяет решать простыми средствами широкий круг практически важных задач именно потому, что длина световой волны представляет собой малую величину по сравнению с характерными размерами элементов оптических систем (зеркал, линз, диафрагм и т. п.). Однако для решения более тонких вопросов, таких, например, как распределение света вблизи фокуса или разрешающая способность оптических инструментов, требуется выход за рамки этого приближения. Обусловленные волновой природой света отклонения от простых законов геометрической оптики обнаруживают себя в явлениях дифракции. Теория дифракции устанавливает границы применимости этого приближения.

7.1. Основные положения геометрической оптики

Рассмотрим переход от волновой оптики к геометрической в предельном случае исчезающе малой длины волны ($\lambda \to 0$).

Плоская волна характеризуется тем свойством, что ее поверхности постоянной фазы (волновые поверхности) представляют собой неограниченные плоскости, а направление ее распространения и амплитуда везде одинаковы. В общем случае световые волны таким свойством не обладают. Тем не менее часто световую волну можно приближенно рассматривать как плоскую в каждом небольшом участке пространства. Это возможно тогда, когда амплитуда световых колебаний и направление распространения волны почти не изменяются на расстоянии порядка длины волны. Волновые поверхности при этом имеют небольшую кривизну и на небольших участках пространства можно, как и у плоской волны, говорить об определенном направлении распространения, нормальном к волновой поверхности. Для характеристики этого направления вводят понятие *лучей*, т. е. линий, касательная к которым в каждой точке совпадает с направлением распространения волны.

Геометрическая оптика, отвлекаясь от волновой природы света, описывает его распространение с помощью лучей. При этом оказывается, что поведение лучей при $\lambda \to 0$ определяется теми же законами, что и для плоских волн: законы преломления и отражения, установленные для плоской волны, падающей на плоскую границу раздела, справедливы в приближении геометрической оптики при более общих условиях. Например, при падении луча на поверхность линзы направление, интенсивность и состояние поляризации отраженного и преломленного лучей можно найти из соответствующих формул для плоских волн.

В плоской монохроматической волне, распространяющейся в изотропной однородной среде с показателем преломления n, зависимость напряженности поля от координат и времени имеет вид

$$E(\mathbf{r}, t) = E_0 e^{i(\mathbf{kr} - \omega t)} = E_0 e^{i(k_0 n \mathbf{sr} - \omega t)}. \tag{7.1}$$

Здесь введен единичный вектор $\mathbf{s}$ (*лучевой вектор*), указывающий направление распространения плоской волны: $\mathbf{s} = \mathbf{k}/k$. Поэтому $\mathbf{k} = k\mathbf{s} = k_0 n \mathbf{s}$, где $k_0 = \omega/c = 2\pi/\lambda_0$ — волновое число (для вакуума). В неоднородной среде показатель преломления зависит от координат: $n = n(\mathbf{r})$ — и выражение (7.1) уже не будет решением уравнений Максвелла. Можно искать решение в виде монохроматической волны более общего типа:

$$E(\mathbf{r}, t) = E_0(\mathbf{r})\, e^{i[k_0 S(\mathbf{r}) - \omega t]}, \tag{7.2}$$

где величина $S(\mathbf{r})$, называемая *эйконалом*, представляет собой вещественную скалярную функцию координат, а амплитуда $E_0(\mathbf{r})$ зависит от положения $\mathbf{r}$ рассматриваемой точки. Выражение (7.2) будет давать приближенное решение уравнений Максвелла в предельном случае больших k_0 (т. е. малых длин волн λ_0) при условии,

что функция $S(\mathbf{r})$ удовлетворяет некоторому дифференциальному уравнению (*уравнение эйконала*). Найдем его.

Напряженность поля $E(\mathbf{r}, t)$ в (7.2) при $\lambda_0 \to 0$, т. е. $k_0 \to \infty$, испытывает очень быстрые изменения в пространстве, но амплитуда $E_0(\mathbf{r})$ и эйконал $S(\mathbf{r})$ изменяются постепенно и при $k_0 \to \infty$ остаются конечными. В малых участках пространства $S(\mathbf{r})$ можно разложить в ряд, ограничившись членами первого порядка:

$$S(\mathbf{r}) = S(\mathbf{r}_0) + (\mathbf{r} - \mathbf{r}_0)(\partial S/\partial \mathbf{r})_{\mathbf{r}=\mathbf{r}_0} = S_0 + (\mathbf{r} - \mathbf{r}_0)\nabla S. \tag{7.3}$$

Подставляя это разложение в (7.2), представим напряженность поля в окрестности точки $\mathbf{r}_0$ в виде

$$E(\mathbf{r}, t) = E_0 e^{i(k_0 \nabla S \cdot \mathbf{r} - \omega t)}, \quad E_0 = E_0(\mathbf{r}_0) e^{ik_0[S(\mathbf{r}_0) - \mathbf{r}_0 \cdot \nabla S]}. \tag{7.4}$$

Это и значит, что в каждом малом участке пространства, в пределах которого справедливо разложение (7.3), произвольную волну (7.2) можно рассматривать как плоскую, так как в (7.4) зависимость напряженности от координат $\mathbf{r}$ такая же, как и в плоской волне (7.1). Сравнивая (7.4) с (7.1), получаем дифференциальное уравнение, которому должна удовлетворять функция $S(\mathbf{r})$:

$$n\mathbf{s} = \nabla S(\mathbf{r}), \text{ или } (\nabla S)^2 = n^2. \tag{7.5}$$

Поверхности равных фаз (*волновые поверхности*) монохроматической волны (7.2) определяются уравнением

$$k_0 S(\mathbf{r}) - \omega t = \text{const}. \tag{7.6}$$

Уравнение (7.5) показывает, что лучи, т. е. линии, касательные к которым в каждой точке совпадают с направлением распространения волны, задаваемым единичным вектором $\mathbf{s}$, ортогональны к волновым поверхностям. В общем случае при показателе преломления $n(\mathbf{r})$, изменяющемся от точки к точке, лучи будут искривлены. Поверхности равных фаз перемещаются в направлении луча $\mathbf{s}$ со скоростью $v = c/n$.

Чтобы выяснить, как искривляются лучи в оптически неоднородной среде, получим из уравнения эйконала (7.5) дифференциальное уравнение для лучей. Радиус-вектор $\mathbf{r}$ точки P, лежащей на луче, будем рассматривать как функцию длины дуги l. Тогда $d\mathbf{r}/dl = \mathbf{S}$ и из (7.5) находим $n d\mathbf{r}/dl = \nabla S$. Продифференцируем это уравнение по l и преобразуем правую часть следующим образом: $d/dl(\nabla S) = \nabla(dS/dl) = \nabla n$ [здесь мы воспользовались тем, что $dS/dl = \nabla S\ \mathbf{s} = n$]. Таким образом,

$$\frac{d}{dl}\left(n\frac{d\mathbf{r}}{dl}\right) = \nabla n. \tag{7.7}$$

В частности, в однородной среде $n = \text{const}$, $\nabla n = 0$ и (7.7) принимает вид $d^2\mathbf{r}/dl^2 = 0$. Его решение $\mathbf{r} = \mathbf{a}l + \mathbf{b}$, где $\mathbf{a}$ и $\mathbf{b}$ — постоянные векторы, представляет собой прямую линию, направленную по вектору $\mathbf{a}$ и проходящую через точку $\mathbf{r} = \mathbf{b}$. В однородной среде световые лучи прямолинейны. Но волновые поверхности $S(\mathbf{r}) = \text{const}$ при

этом могут и не быть плоскими. Например, одно из возможных решений уравнения эйконала $(\nabla S)^2=n^2$ при $n=\text{const}$ имеет вид $S(\mathbf{r})=n|\mathbf{r}-\mathbf{r}_0|$ (особая точка при $\mathbf{r}=\mathbf{r}_0$). В этом случае $\nabla S=n(\mathbf{r}-\mathbf{r}_0)/|\mathbf{r}-\mathbf{r}_0|$ и лучи света образуют семейство прямых, расходящихся из точки $\mathbf{r}=\mathbf{r}_0$, а волновые поверхности — концентрические сферы.

Для неоднородной среды $n=n(\mathbf{r})$ и уравнение (7.7) преобразуется к виду $(dn/dl)\mathbf{s}+n(d\mathbf{s}/dl)=\nabla n$. Так как $dn/dl=\nabla n\mathbf{s}$, то

$$\frac{d\mathbf{s}}{dl}=\frac{1}{n}[\nabla n-\mathbf{s}(\nabla n\,\mathbf{s})]. \tag{7.8}$$

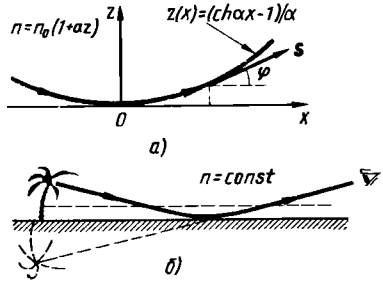
Производная единичного вектора $\mathbf{s}$ по длине луча l характеризует кривизну луча: $d\mathbf{s}/dl=\mathbf{N}/R$, где $\mathbf{N}$ — единичный вектор главной нормали к лучу, R — радиус его кривизны. Умножая обе части (7.8) скалярно на $\mathbf{N}$ и учитывая, что $\mathbf{N}\mathbf{s}=0$, получаем следующее выражение для радиуса кривизны луча:

$$1/R=\mathbf{N}(\nabla n)/n. \tag{7.9}$$

Отсюда, в частности, следует, что $\mathbf{N}\nabla n>0$, т. е. угол между $\mathbf{N}$ и ∇n острый — луч изгибается в область с бóльшим показателем преломления n.

Для демонстрации искривления световых лучей можно взять две смешивающиеся жидкости, например сероуглерод ($n=1{,}63$) и бензол ($n=1{,}50$), и расположить их слои один над другим. Граница между слоями вскоре пропадает вследствие диффузии, и получается среда с непрерывно изменяющимся показателем преломления. Исходя из уравнения (7.8), можно показать (см. задачу 1), что в среде, показатель преломления которой изменяется в каком-либо одном направлении по линейному закону $n(z)=n_0(1+\alpha z)$, лучи представляют собой цепные линии (рис. 7.1, *а*). Если начало координат выбрать там, где вектор $\mathbf{s}$ направлен горизонтально (вдоль оси x), уравнение луча имеет вид $z(x)=(\operatorname{ch}\alpha x-1)/\alpha$. В соответствии с (7.9) луч сильнее всего искривлен там, где его направление перпендикулярно градиенту показателя преломления.

Искривлением лучей света в неравномерно нагретом воздухе объясняется явление *миража*, когда в жаркой пустыне возникает иллюзия находящейся на горизонте водной глади или вдали на асфальте нагретого солнечными лучами шоссе видны «лужи», исчезающие при приближении к ним. Прилегающие к раскаленной земле слои воздуха нагреты сильнее, поэтому плотность воздуха и его показатель преломления возрастают



7.1
Искривление лучей света в неоднородной среде (*а*) и возникновение миража (*б*)

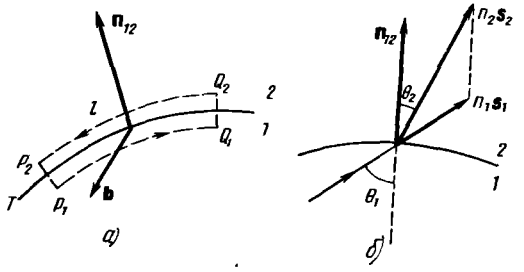
с высотой. Лучи света, входящие в такой слой под небольшими углами, искривляются и, не достигнув земли, выходят обратно под такими же углами (рис. 7.1, *б*). Мы настолько привыкли к прямолинейному распространению света, что подсознательно считаем источник расположенным на прямолинейном продолжении попадающих в глаз лучей даже тогда, когда они искривлены; мы видим расположенные над горизонтом удаленные предметы как бы отраженными горизонтальной зеркальной поверхностью («водной гладью»).

Другой пример искривления лучей дает явление *астрономической рефракции,* обусловленное тем, что плотность земной атмосферы и, следовательно, ее показатель преломления убывают с высотой. Наблюдаемая высота небесного светила над горизонтом оказывается больше истинной. Эффект особенно значителен, когда светило наблюдается у горизонта (рефракция при этом достигает 35′) и быстро убывает с увеличением высоты. Этим объясняется сплюснутая форма солнечного диска при восходе и закате. Благодаря рефракции мы видим Солнце в течение нескольких минут после того, как оно уже зашло.

Выше предполагалось, что показатель преломления представляет собой непрерывную функцию координат. Чтобы рассмотреть поведение лучей при переходе через резкую границу раздела сред с различными показателями преломления, можно мысленно заменить граничную поверхность тонким переходным слоем, в котором показатель преломления изменяется непрерывно. Тогда уравнение (7.5) останется в силе. Применяя к обеим его частям операцию rot и учитывая тождество rot grad $S=0$, находим соотношение, которому должен удовлетворять лучевой вектор s:

$$\mathrm{rot}(n\mathbf{s})=0. \tag{7.10}$$

Выберем небольшой прямоугольный контур l (рис. 7.2, *а*), стороны P_1Q_1 и P_2Q_2 которого параллельны поверхности раздела T двух сред, а P_1P_2 и Q_1Q_2 перпендикулярны T. Пусть **b** — единичный вектор нормали к плоскости контура. Умножим (7.10) скалярно на **b** и проинтегрируем по площади σ, ограниченной контуром l. Поток ротора по теореме Стокса преобразуется в интеграл по контуру l:



7.2

К выводу закона преломления лучей на искривленной границе раздела

$$\int_\sigma \mathbf{b}\,\mathrm{rot}(n\mathbf{s})\,d\sigma =$$

$$= \oint_l n\mathbf{s}\,d\mathbf{r} = 0.$$

Переходя к пределу, когда длины сторон P_1P_2 и Q_1Q_2 стремятся к нулю, получим

$$\mathbf{b}[\mathbf{n}_{12}\times(n_2\mathbf{s}_2-n_1\mathbf{s}_1)]=0,\ \text{или}\ n_2(\mathbf{n}_{12}\times\mathbf{s}_2)=n_1(\mathbf{n}_{12}\times\mathbf{s}_1),\qquad(7.11)$$

где $\mathbf{n}_{12}$ — единичный вектор нормали к поверхности раздела T. Из формулы (7.11) следует, что преломленный луч $\mathbf{s}_2$ лежит в плоскости падения (плоскости, образованной падающим лучом $\mathbf{s}_1$ и нормалью $\mathbf{n}_{12}$, рис. 7.2, *б*), а синусы углов падения и преломления связаны соотношением $n_2\sin\theta_2=n_1\sin\theta_1$. Как и следовало ожидать, закон преломления, установленный для плоских волн на плоской границе раздела, в приближении геометрической оптики ($\lambda_0\to 0$) справедлив для преломляющих поверхностей более общей формы. Фактически для этого достаточно, чтобы радиусы кривизны волновой поверхности падающей волны и поверхности раздела были велики по сравнению с λ_0.

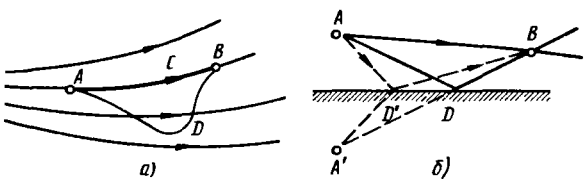
Нахождение траекторий лучей света в приближении геометрической оптики можно сформулировать как задачу вариационного исчисления, если воспользоваться *принципом Ферма,* согласно которому свет распространяется между двумя точками по такому пути, который требует для прохождения наименьшего времени. Принцип наикратчайшего оптического пути, сформулированный Пьером Ферма в середине XVII в., можно получить как следствие основного уравнения геометрической оптики (7.5). Рассмотрим некоторую область с показателем преломления $n(\mathbf{r})$, через каждую точку которой проходит только один луч (например, от точечного источника), т. е. эти лучи в рассматриваемой области не пересекаются. Пусть точки A и B (рис. 7.3, *а*) лежат на одном луче. Используя уравнение (7.5) $n\mathbf{s}=\nabla S(\mathbf{r})$, вычислим следующий интеграл вдоль произвольной кривой, соединяющей точки A и B:

$$\int\limits_{AB} n\mathbf{s}d\mathbf{r}=\int\limits_{AB}\nabla S d\mathbf{r}=S(B)-S(A).\qquad(7.12)$$

Этот интеграл равен разности значений эйконала в точках A и B и, следовательно, не зависит от пути интегрирования (интегральный инвариант Лагранжа.

Очевидно, что $\mathbf{s}d\mathbf{r}=dl\cos(\widehat{\mathbf{s},d\mathbf{r}})\leqslant dl$, поэтому

$$S(B)-S(A)=\int\limits_{AB} n\mathbf{s}d\mathbf{r}\leqslant\int\limits_{AB} n dl,\qquad(7.13)$$

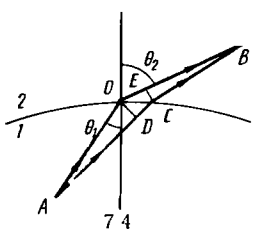


7.3
К обосованию принципа Ферма

причем знак равенства справедлив только в том случае, когда направления векторов s и dr совпадают в каждой точке рассматриваемой кривой, т. е. когда она представляет собой реальный луч (ACB на рис. 7.3, *а*). Для любой другой кривой, соединяющей точки A и B (например, ADB на рис. 7.3, *а*), правая часть (7.13), называемая *оптической длиной пути,* оказывается больше, чем для реального луча ACB. Поскольку $\mathrm{d}l = v\mathrm{d}t = (c/n)\mathrm{d}t$, оптическая длина равна произведению c на время, которое требуется свету для прохождения вдоль этой кривой. Таким образом, свет между A и B распространяется по тому пути, который требует наименьшего времени.

В приведенном выше доказательстве было использовано предположение о том, что через каждую точку рассматриваемой области проходит только один луч. Это условие во многих практически важных случаях не выполняется. Например, при отражении от зеркала света, испускаемого точечным источником A, через любую точку B проходят два луча (рис. 7.3, *б*). Чтобы охватить подобные случаи, принцип Ферма можно сформулировать в более слабой форме, но применимой в более широкой области: реальный луч отличается от остальных кривых, соединяющих две заданные точки, тем, что соответствующая ему оптическая длина имеет *стационарное* значение, т. е. малое изменение траектории (например, точки падения на зеркало; рис. 7.3, *б*) не приводит в первом порядке к изменеиию оптической длины. Эта формулировка вполне достаточна для практических приложений, ибо для нахождеиия луча можно ограничиться сравнением оптических длин для воображаемых путей, которые проходят бесконечно близко от действительного.

Проиллюстрируем применение принципа Ферма на примере преломления луча на границе однородных сред. Пусть AOB — истинный путь света из A в B (рис. 7.4). Тогда при малом изменении траектории — смещении точки преломления из O в C — оптическая длина в первом приближении должна остаться неизменной. Оптический путь в среде 1 увеличивается на $n_1|DC| = n_1|OC|\sin\theta_1$, в среде *2* — уменьшается на $n_2|OE| = n_2|OC|\sin\theta_2$. Приравнивая эти величины, опять получаем закон преломления: $n_1\sin\theta_1 = n_2\sin\theta_2$.



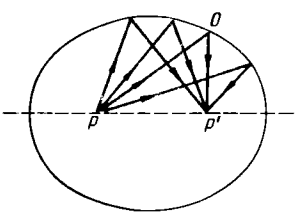
Применение принципа Ферма к преломлению лучей

Если пучок лучей, выходящих из какой-либо точки P, после отражений, преломлений на границах или искривлений в неоднородной среде сходится в точке P', то P' называется *действительным фокусом* геометрического схождения лучей. Его можно рассматривать как оптическое *изображение* точки P. Изображение называют *мнимым,* если в P' пересекаются не сами лучи, а их продолжения, проведенные в направлении, противоположном распространению света. Если источник P и его

изображение P' поменять местами, то форма всех лучей останется без изменения и лишь их направление изменится на противоположное (*принцип обратимости*). Поэтому точки P и P' называются *сопряженными*. Когда все лучи пересекаются строго в одной точке P', изображение называется *стигматическим* (точечным). Пучок лучей, выходящих из одной точки или сходящихся в одну точку, называется *гомоцентрическим*.

В соответствии с принципом Ферма оптическая длина всех лучей между сопряженными точками одинакова. В качестве примера рассмотрим зеркало в форме эллипсоида вращения (рис. 7.5). Сумма расстояний $|PO|+|OP'|$ от его фокусов до точки O имеет одно и то же значение при любом положении точки O на его поверхности. Если в один из фокусов поместить точечный источник, в другом фокусе пучок отраженных от зеркала лучей образует стигматическое изображение источника. Исходящие из фокуса эллипсоида гомоцентрические пучки лучей в результате отражения превращаются снова в гомоцентрические. Совершенно аналогично в фокусе параболического зеркала образуется стигматическое изображение находящегося на оси параболоида бесконечно удаленного точечного источника (параболоид можно рассматривать как предельный случай эллипсоида, когда второй его фокус удаляется в бесконечность). Такие параболические зеркала используются в астрономических телескопах-рефлекторах.

Эти примеры преобразования пучков света иллюстрируют скорее исключения, чем общее правило: обычно при отражении или преломлении пучок утрачивает свойство гомоцентричности и не образует стигматического изображения точечного источника. Например, отраженные параболическим зеркалом лучи от бесконечно удаленного источника, не лежащего на оси зеркала, пересекаются не в одной точке, а в некоторой ее окрестности, что ухудшает качество изображения. Используемые на практике оптические системы состоят из линз и зеркал, преломляющие и отражающие поверхности которых, как правило, сферические или плоские. Ход приосевых лучей и образование изображений в центрированных оптических системах рассматриваются в § 7.2. Искажения изображений, связанные с нарушением гомоцентричности пучков, называются *геометрическими* или *лучевыми аберрациями* оптических систем (см. § 7.4). Зависимость показателя преломления от длины волны приводит к появлению *хроматической аберрации* (см. § 7.4). Неизбежные в принципе погрешности отображения можно уменьшить до разумных пределов, используя многолинзовые конструкции. В этом отношении инструментальная оптика достигла замечательных результатов.



7.5
Эллиптическое зеркало

Но даже в тех случаях, когда по законам геометрической оптики пучок

лучей пересекается строго в одной точке, образование точечного изображения светящейся точки невозможно из-за дифракции света. Вблизи фокуса пучка лучей кривизна волновых поверхностей становится значительной. Изменение амплитуды волны здесь уже нельзя считать малым на протяжении длины волны, т. е. условия применимости геометрической оптики не выполняются, и распределение интенсивности вблизи фокуса обусловлено волновой природой света. *Волновые*, или *дифракционные*, искажения определяют теоретический предел разрешающей способности оптических инструментов (см. § 7.5).

Контрольные вопросы

- В каком предельном случае волновая оптика переходит в геометрическую? Приведите примеры, в которых условия применимости геометрической оптики не выполняются.
- Как из уравнения эйконала $n\mathbf{s}=\nabla S$ получить дифференциальное уравнение для лучей?
- В какую сторону искривляется луч в среде с зависящим от координат показателем преломления? Объясните явления миража и астрономической рефракции.
- Как из уравнения эйконала получить закон преломления лучей на границе раздела сред? Сформулируйте принцип Ферма. Как его доказать с помощью основного уравнения геометрической оптики? Приведите примеры применения принципа Ферма.
- Что такое сопряженные точки?
- Какую форму должно иметь зеркало для получения стигматического изображения светящейся точки? бесконечно удаленного точечного источника?

Задача

Показатель преломления среды изменяется в направлении оси z по линейному закону: $n(z)=n_0(1+\alpha z)$. Найти форму луча в такой среде $z=z(x)$, если в начале координат вектор $\mathbf{s}$ направлен вдоль оси x.

Обозначим $z'(x)=\operatorname{tg}\varphi\equiv p$ (см. рис. 7.1, *а*). Тогда $dl=dz\sqrt{1+p^2}/p$, $s_x=1/\sqrt{1+p^2}$, $s_z=p/\sqrt{1+p^2}$. Рассмотрим проекцию уравнения для лучей (7.8) на ось x:

$$\frac{p}{\sqrt{1+p^2}}\frac{d}{dz}\frac{1}{\sqrt{1+p^2}}=-\frac{p}{1+p^2}\frac{1}{n}\frac{dn}{dz}.$$

Отсюда $[p/(1+p^2)]\,dp=dn/n$. Интегрируя это уравнение с учетом того, что при $z=0$ луч направлен вдоль x, т. е. $p=0$, получаем $1+p^2=(n/n_0)^2$, или $1+z'^2=(1+\alpha z)^2$. Интегрирование этого уравнения с начальным условием $z(0)=0$ дает

$$z(x)=(\operatorname{ch}\alpha x-1)/\alpha.$$

При $\alpha x\ll 1$ эта цепная линия аппроксимируется параболой $z(x)\approx\approx\alpha x^2/2$.

7.2. Центрированные оптические системы

Большинство используемых на практике оптических инструментов относится к *центрированным системам,* у которых центры кривизны всех сферических преломляющих и отражающих поверхностей расположены на одной прямой, называемой *главной оптической осью.* Теория таких систем становится особенно простой, когда все распространяющиеся в них пучки лучей *параксиальны,* т. е. проходят на небольшом расстоянии от оптической оси системы и образуют с ней малые углы. Гомоцентрический пучок параксиальных лучей при прохождении через центрированную систему остается почти гомоцентрическим, поэтому для каждой точки протяженного предмета система формирует стигматическое (резкое) изображение. Образование изображений в параксиальных лучах было систематически исследовано Гауссом в 1841 г., поэтому теорию центрированных оптических систем в параксиальном приближении обычно называют *гауссовой оптикой.*

Преобразование луча в оптической системе удобно описывать с помощью специальных матриц. Достоинство матричного метода в том, что его можно использовать не только в геометрической оптике параксиальных лучей, но и при описании распространения гауссовых пучков с дифракционной расходимостью (лазерное излучение).

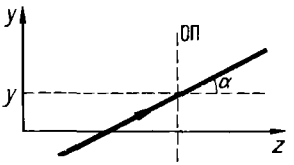
В оптической системе сферические (и плоские) поверхности служат границами раздела различных однородных сред (материал линз и промежутки между ними). Траектория луча состоит из отрезков прямых линий. Будем рассматривать только меридиональные лучи, т. е. лучи, лежащие в одной плоскости с главной оптической осью (ось z на рис. 7.6). Пусть это будет плоскость yz. Выберем некоторую плоскость $z=\text{const}$, перпендикулярную оптической оси, и назовем ее опорной плоскостью (ОП). Любой меридиональный луч можно определить заданием двух параметров: координаты y точки его пересечения с опорной плоскостью и угла α, который он составляет с осью z (рис. 7.6). Однако в дальнейшем для характеристики направления луча удобно вместо α использовать параметр $V=n\alpha$, т. е. произведение показателя преломления среды на угол α.

Преобразование параметров y и V луча при переходе от одной опорной плоскости $ОП_1$ к другой $ОП_2$ в параксиальном приближении будет линейным, т. е. для любой пары опорных плоскостей оно имеет вид

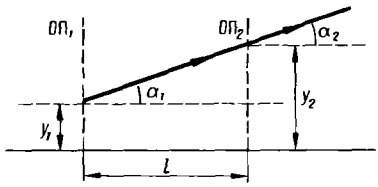
$$y_2=Ay_1+BV_1,\quad V_2=Cy_1+DV_1.$$

Это преобразование можно записать в матричной форме:

$$\begin{pmatrix} y_2 \\ V_2 \end{pmatrix}=\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} y_1 \\ V_1 \end{pmatrix}.$$



7.6
Параметры y, α меридионального луча



7.7

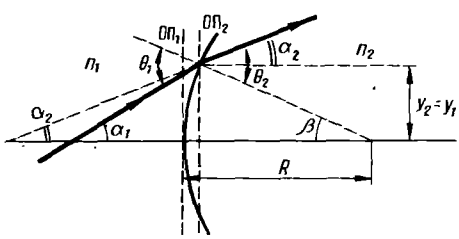
К выводу матрицы преобразования лучей для оптического промежутка

Опорные плоскости можно выбирать в разных местах оптической системы. Для данной пары плоскостей $ОП_1$ и $ОП_2$ преобразоваиие параметров любого параксиального луча описывается одной и той же матрицей, сопоставляемой промежутку между $ОП_1$ и $ОП_2$. Ее элементы A, B, C и D зависят от свойств этого промежутка, т. е. от того, какие преломляющие поверхности и какие среды находятся между $ОП_1$ и $ОП_2$. Матрица, описывающая преобразование лучей всей оптической системой, получается перемножением матриц, сопоставляемых отдельным промежуткам.

Для исчерпывающего исследования поведения параксиального луча в центрированной оптической системе достаточно получить матрицы преобразования для трех основных элементов: оптического промежутка (т. е. участка однородной среды), преломляющей и отражающей поверхностей. Оптический промежуток между $ОП_1$ и $ОП_2$ (рис. 7.7) характеризуется толщиной l и показателем преломления n. Преобразование параметра y находится из рис. 7.7: $y_2 = y_1 + l\operatorname{tg}\alpha_1$. В параксиальном приближении углы наклона лучей считаются малыми. Тогда $y_2 \approx y_1 + l\alpha_1$ (при $\alpha_1 < 0{,}1$, т. е. $\alpha_1 < 6°$, погрешность не превышает 1%). Переходя от α_1 к $V_1 = n\alpha_1$, можем написать $y_2 \approx y_1 + (l/n)n\alpha_1 = y_1 + LV_1$, где $L = l/n$ — приведенная толщина оптического промежутка. Наклон луча при переходе от $ОП_1$ к $ОП_2$ не изменяется, поэтому $V_2 = V_1$. Таким образом, преобразование параметров луча оптическим промежутком можно описать с помощью следующей матрицы $\mathcal{T}$:

$$\begin{pmatrix} y_2 \\ V_2 \end{pmatrix} = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ V_1 \end{pmatrix}, \quad \mathcal{T} = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}. \tag{7.15}$$

Сферическая преломляющая поверхность, разделяющая две среды с показателями преломления n_1 и n_2 (рис. 7.8), характеризуется радиусом кривизны R. Чтобы одни и те же формулы были справедливы для выпуклой и для вогнутой поверхностей, значение R считают положительным, если центр кривизны лежит на оси z справа от границы, и отрицательным в противном случае. Для нахождения матрицы преломления выберем опорные плоскости $ОП_1$ и $ОП_2$ по обе стороны в непосредственной близости от



7.8

К выводу матрицы преломления на сферической поверхности

преломляющей поверхности. Расстояние между ними $R(1-\cos\beta)\approx \approx R\beta^2/2$ в параксиальном приближении пренебрежимо мало. Поэтому можно считать, что луч пересекает их на одном и том же расстоянии от оси, т. е. $y_2\approx y_1$. Чтобы найти, как преобразуется параметр $V=n\alpha$, воспользуемся законом преломления $n_1\sin\theta_1 = = n_2\sin\theta_2$, что для параксиальных лучей сводится к $n_1\theta_1=n_2\theta_2$. По теореме о внешнем угле треугольника можно написать (рис. 7.8) $\theta_1=\alpha_1+\beta$, $\theta_2=\alpha_2+\beta$. Умножая первое из этих равенств на n_1, второе на n_2 и применяя закон преломления $n_1\theta_1=n_2\theta_2$, находим $n_2\alpha_2=n_1\alpha_1+\beta(n_1-n_2)$. Подставляя сюда $n_2\alpha_2=V_2$, $n_1\alpha_1=V_1$, $\beta=y_1/R$, получаем закон преобразования параметра V: $V_2= = V_1-y_1(n_2-n_1)/R$ или $V_2=V_1-Py_1$, где введена величина $P=(n_2-n_1)/R$, называемая *оптической силой* преломляющей поверхности. Вместе с $y_2=y_1$ это дает закон преобразования параметров луча при преломлении, описываемый с помощью матрицы преломления $\mathscr{R}$:

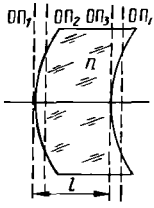
$$\begin{pmatrix} y_2 \\ V_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -P & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ V_1 \end{pmatrix}, \quad \mathscr{R} = \begin{pmatrix} 1 & 0 \\ -P & 1 \end{pmatrix}. \tag{7.16}$$

В частном случае плоской поверхности $R\to\infty$, $P=0$ и преломление описывается единичной матрицей, что заранее очевидно, так как при этом, по закону преломления, $n_2\alpha_2=n_1\alpha_1$, т. е. $V_2=V_1$.

Чтобы включить в рассмотрение отражающие поверхности (плоские и сферические), вводят следующее правило: когда луч распространяется в отрицательном направлении оси z, показатель преломления среды, через которую он проходит, считается отрицательным $(-n)$. Тогда закон отражения $\theta_2=-\theta_1$ формально можно рассматривать как частный случай закона преломления при $n_2=-n_1$. Матрица преобразования параметров луча при отражении от сферической поверхности имеет точно такой же вид (7.16), как и матрица преломления, если в выражении для оптической силы P заменить n_2 на $-n_1$: $P=-2n_1/R$. Для выпуклого зеркала $R>0$ и оптическая сила отрицательна $(P<0)$, для вогнутого — положительна $(P>0)$.

Толщина оптического промежутка $l=z_2-z_1$ между опорными плоскостями $z=z_1$ и $z=z_2$ положительна при $z_2>z_1$ и отрицательна при $z_2<z_1$, т. е. $l<0$ для лучей, распространяющихся влево. Так как для таких лучей и показатель преломления нужно считать отрицательным, приведенная толщина $L=l/n$ будет положительной. Поэтому матрица $\mathscr{T}$ (7.15) для оптического промежутка не зависит от направления лучей.

Параметры луча на выходе сложной оптической системы могут быть выражены через параметры входного луча путем последовательного применения преобразований этих параметров ее отдельными элементами. Рассмотрим в качестве примера линзу, т. е. однородную среду с показателем преломления n (стекло), ограниченную сферическими поверхностями с радиусами кривизны R_1 и R_2 (рис. 7.9). Расстояние между поверхностями вдоль оптической оси равно l



7.9
Толстая
линза

(толщина линзы). Обозначим: $K_1(y_1, V_1)$ — параметры некоторого луча на $ОП_1$ (перед входом в линзу), K_2 — на $ОП_2$ (после преломления на первой поверхности), K_3 — на $ОП_3$ (после прохождения оптического промежутка приведенной толщины $L=l/n$, $K_4(y_4, V_4)$ — на $ОП_4$ после преломления на задней поверхности. Пусть $\mathcal{T}$ — матрица оптического промежутка (7.15), а $\mathcal{R}_1$ и $\mathcal{R}_2$ — матрицы преломления (7.16) соответственно для первой и второй поверхностей. В них $P_1=(n-1)/R_1$, $P_2=(1-n)/R_2$ (считаем, что линзу окружает среда с показателем преломления $n_0=1$). Очевидно, что

$$K_4=\mathcal{R}_2K_3=\mathcal{R}_2(\mathcal{T}K_2)=\mathcal{R}_2\mathcal{T}(\mathcal{R}_1K_1)=\mathcal{R}_2\mathcal{T}\mathcal{R}_1K_1. \qquad (7.17)$$

Таким образом, матрица $\mathcal{M}$ преобразования луча в сложной системе равна произведению матриц для ее отдельных элементов, взятых в обратном порядке: $\mathcal{M}=\mathcal{R}_2\mathcal{T}\mathcal{R}_1$. Выполняя перемножение, находим

$$\mathcal{M}=\begin{pmatrix}1 & 0\\ -P_2 & 1\end{pmatrix}\begin{pmatrix}1 & L\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ -P_1 & 1\end{pmatrix}=\begin{pmatrix}1-P_1L & L\\ -(P_1+P_2-P_1P_2L) & 1-P_2L\end{pmatrix}. \qquad (7.18)$$

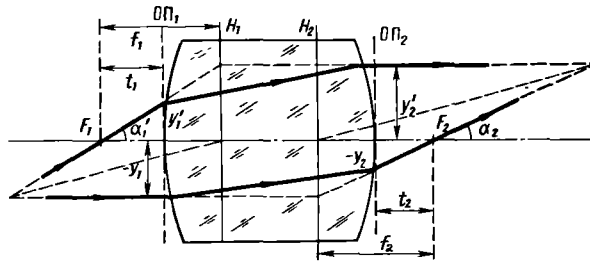
В частном случае *тонкой линзы*, когда толщину оптического промежутка l между преломляющими поверхностями можно считать пренебрежимо малой ($L\to 0$), матрица $\mathcal{T}$ вырождается в единичную и полная матрица $\mathcal{M}$ имеет такой же вид, как и матрица $\mathcal{R}$ преломления на одной поверхности (7.16), но с оптической силой $P=P_1+P_2$:

$$P=(n-1)(1/R_1-1/R_2). \qquad (7.19)$$

Этим способом можно найти полную матрицу $\mathcal{M}$ преобразования параметров параксиального луча для произвольной центрированной оптической системы, если известны кривизна и взаимное расположение ее преломляющих и отражающих поверхностей и значения показателей преломления. Введем обозначения A, B, C и D для ее элементов:

$$\mathcal{M}=\begin{pmatrix}A & B\\ C & D\end{pmatrix}. \qquad (7.20)$$

Рассмотрим луч, входящий в оптическую систему параллельно оптической оси на некоторой высоте y_1 (рис. 7.10). Для такого луча $\alpha_1=0$ и, следовательно, $V_1=0$. На выходе из системы луч имеет параметры $y_2=Ay_1$ и $V_2=Cy_1$. Угол его наклона к оптической оси $\alpha_2=V_2/n_2$, поэтому луч (или его продолжение) пересечет оптическую ось в точке F_2, отстоящей от последней преломляющей поверхности $ОП_2$ на расстоянии $t_2=-y_2/\alpha_2=-n_2y_2/V_2$. Подставляя



7.10
К нахождению кардинальных точек оптической системы

сюда y_2 и V_2, получаем $t_2 = -n_2A/C$. Расстояние t_2 не зависит от y_1, т. е. все лучи, входящие в систему параллельно оптической оси, проходят (в параксиальном приближении) через одну и ту же точку F_2, называемую *задней фокальной точкой* или *задним главным фокусом* оптической системы.

Пересечение продолжений входящего параллельно оптической оси луча и выходящего луча происходит в задней *главной плоскости* H_2. Определим фокусное расстояние f_2 как смещение вдоль оси от H_2 до F_2. Тогда $f_2 = -y_1/\alpha_2 = -n_2y_1/V_2$ (рис. 7.10). Подставляя $V_2 = Cy_1$, получаем $f_2 = -n_2/C$, т. е. фокусное расстояние определяется элементом C матрицы $\mathcal{M}$ оптической системы (7.20).

Чтобы найти положение передней фокальной точки F_1, рассмотрим луч, идущий через нее под некоторым углом α_1'. На выходе из системы он должен быть параллелен оптической оси, т. е. для него $V_2' = 0$. Поэтому $V_2' = Cy_1' + DV_1' = 0$. Подставив сюда $V_1' = n_1\alpha_1'$, найдем $y_1' = -Dn_1\alpha_1'/C$. Из рис. 7.10 видно, что $t_1 = -y_1'/\alpha_1' = n_1D/C$. Мы получили, что расстояние t_1 не зависит от α_1', т. е. все лучи из F_1 после прохождения через систему будут параллельны оптической оси. Рассматривая продолжения падающего и выходящего лучей, определяем положение передней главной плоскости H_1 и переднее фокусное расстояние, отсчитываемое от H_1 до F_1: $f_1 = n_1/C$ (при этом учтено, что для матрицы $\mathcal{M}$, образованной произведением любого числа матриц $\mathcal{R}$ и $\mathcal{T}$, $\det\mathcal{M} = AD - BC = 1$). Когда показатели преломления сред по обе стороны от системы одинаковы ($n_1 = n_2$), ее переднее и заднее фокусные расстояния равны по модулю, но противоположны по знаку: $f_2 = -f_1$.

Фокусы F_1 и F_2 и точки пересечения главных плоскостей H_1 и H_2 с оптической осью называются *кардинальными точками* оптической системы. Их положение полностью определяет преобразование любого параксиального луча оптической системой. Если оно известно, можно построить выходящий из системы луч, не рассматривая реального хода лучей в системе. Для удобства нахождения кардинальных точек по известным элементам матрицы $\mathcal{M}$ оптической системы полученные выше результаты сведены в таблицу.

Таблица 7.1

Расстояние		при $n_1=n_2=1$
$ОП_1-F_1$	n_1D/C	D/C
H_1-F_1	$f_1=n_1/C$	$f_1=1/C$
$ОП_1-H_1$	$n_1(D-1)/C$	$(D-1)/C$
$ОП_2-F_2$	$-n_2A/C$	$-A/C$
H_2-F_2	$f_2=-n_2/C$	$f_2=-1/C$
$ОП_2-H_2$	$n_2(1-A)/C$	$(1-A)/C$

Фокусное расстояние оптической системы полностью определяется элементом C матрицы преобразования лучей: $f_2=-1/C$ (при $n_1=n_2=1$). Как и у тонкой линзы, этот элемент, взятый с противоположным знаком, называется *оптической силой* системы: $P=-C$. Для толстой линзы, как видно из матрицы (7.18), $P=P_1+P_2-P_1P_2L$. Подставляя выражения для P_1, P_2 и L, находим

$$P=1/f_2=(n-1)[1/R_1-1/R_2+(n-1)l/(nR_1R_2)]. \qquad (7.21)$$

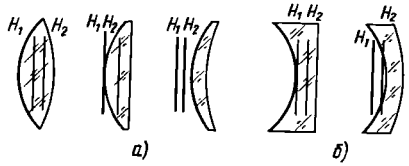
При $P>0$ линза называется собирающей, при $P<0$ — рассеивающей. Используя диагональные элементы A и D матрицы $\mathcal{M}$ (7.18), по формулам таблицы легко найти положение главных плоскостей толстой линзы. На рис. 7.11 *а, б* показаны главные плоскости некоторых типичных линз.

До сих пор мы рассматривали матрицу $\mathcal{M}$ (7.20) преобразования параметров лучей между передней ($ОП_1$) и задней ($ОП_2$) преломляющими поверхностями оптической системы. Теперь легко получить матрицу $\mathcal{M}_H$ преобразования луча между двумя главными плоскостями H_1 и H_2:

$$\mathcal{M}_H=\mathcal{T}_2\mathcal{M}\mathcal{T}_1,$$

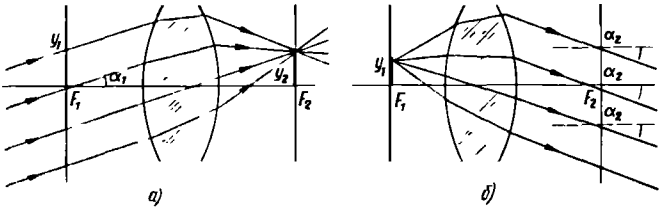
где $\mathcal{T}_1$ — матрица (7.15) для оптического промежутка толщиной $L_1=(1-D)/C$ от H_1 до $ОП_1$, $\mathcal{T}_2$ — от $ОП_2$ до H_2 с $L_2=(1-A)/C$. Перемножая матрицы и учитывая, что $\det\mathcal{M}=AD-BC=1$, получаем

$$\mathcal{M}_H=\begin{pmatrix}1 & 0\\ C & 1\end{pmatrix}=\begin{pmatrix}1 & 0\\ -1/f_2 & 1\end{pmatrix}. \qquad (7.22)$$



7.11
Расположение главных плоскостей для толстых линз

Таким образом, матрица $\mathcal{M}_H$ преобразования между двумя главными плоскостями произвольной центрированной системы совпадает с матрицей $\mathcal{R}$ для тонкой линзы, имеющей оптическую силу $P=-C$. Любой луч, падающий на H_1, выйдет из H_2 на той же высоте, так как $y_2=y_1$.



7.12

Преобразование лучей между фокальными плоскостями

Совершенно аналогично можно получить матрицу $\mathcal{M}_F$ преобразования луча между двумя фокальными плоскостями, проходящими через F_1 и F_2:

$$\mathcal{M}_F = \begin{pmatrix} 0 & -1/C \\ C & 0 \end{pmatrix} = \begin{pmatrix} 0 & f_2 \\ -1/f_2 & 0 \end{pmatrix}. \tag{7.23}$$

Эта матрица показывает, что все параллельные лучи, пересекающие переднюю фокальную плоскость на разной высоте, пересекут заднюю фокальную плоскость в одной точке, высота y_2 которой зависит только от наклона α_1 падающих лучей ($y_2 = f_2\alpha_1$) и не зависит от y_1 (рис. 7.12, *а*). Если же лучи идут из одной точки y_1 передней фокальной плоскости под разными углами α_1, то заднюю плоскость все они пересекут под одним и тем же углом $\alpha_2 = -y_1/f_2$ (рис. 7.12, *б*).

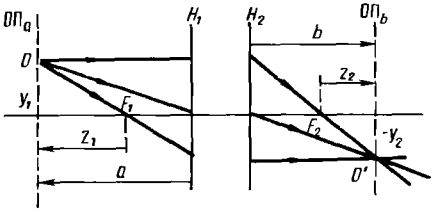
Используя матрицу $\mathcal{M}_H$ или $\mathcal{M}_F$, не представляет труда найти матрицу $\mathcal{M}_{ab}$ преобразования луча между произвольными плоскостями $ОП_a$ и $ОП_b$ (рис. 7.13). Пусть a — расстояние от H_1 до $ОП_a$ ($a<0$, если $ОП_a$ слева от H_1), b — от H_2 до $ОП_b$. Иногда бывает удобно определять положение $ОП_a$ и $ОП_b$ относительно фокальных точек F_1 и F_2, вводя расстояние z_1 от F_1 до $ОП_a$ ($z_1 = a - f_1$) и z_2 от F_2 до $ОП_b$ ($z_2 = b - f_2$). Считая, что оптическую систему окружает среда с показателем преломления $n_1 = n_2 = 1$, получаем

$$\mathcal{M}_{ab} = \begin{pmatrix} 1-b/f_2 & -a+ab/f_2+b \\ -1/f_2 & 1+a/f_2 \end{pmatrix} = \begin{pmatrix} -z_2/f_2 & z_1z_2/f_2+f_2 \\ -1/f_2 & z_1/f_2 \end{pmatrix}. \tag{7.24}$$

Пусть расстояния a и b (или z_1 и z_2) выбраны так, что правый верхний элемент матрицы $\mathcal{M}_{ab}$ равен нулю, т. е.

$$-1/a + 1/b = 1/f_2 \text{ или } z_1z_2 = -f_2^2. \tag{7.25}$$

Тогда $y_2 = (1 - b/f_2)y_1$. Это значит, что все лучи, выходящие под любым углом α_1 из точки O с координатой y_1 на плоскости $ОП_a$, пройдут через одну и ту же точку O' с координатой y_2 на плоскости $ОП_b$, т. е. гомоцентрический пучок параксиальных лучей, идущих из O, на выходе из системы будет гомоцентрическим и образует в точке O' *стигматическое изображение* точки O. Так как для любой точки на плоскости $ОП_a$ существует сопряженная точка на $ОП_b$,



7.13

Образование изображения в оптической системе

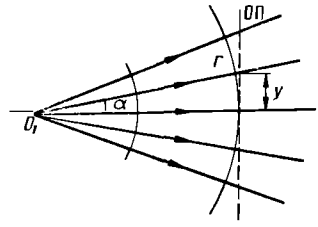
оптическая система *отображает* плоскость ОП$_a$ на ОП$_b$ с *увеличением* $\gamma = y_2/y_1 = 1 - b/f_2 = b/a$. Положение сопряженных плоскостей связано основным соотношением (7.25).

При перемещении плоскости ОП$_a$ на расстояние Δa сопряженная с ней плоскость ОП$_b$ перемещается на некоторое расстояние Δb, которое можно рассматривать как длину изображения отрезка Δa, параллельного оптической оси. При малом Δa из (7.25) получаем следующее соотношение: $\Delta a/a^2 = \Delta b/b^2$. Отсюда *продольное увеличение* $\beta = \Delta b/\Delta a = (b/a)^2$. Оно всегда положительно, т. е. Δb и Δa совпадают по направлению. Продольное увеличение равно квадрату поперечного ($\beta = \gamma^2$), поэтому изображение пространственного предмета сохраняет геометрическое подобие с предметом только при $|\gamma| = 1$. Такое изображение (в натуральную величину) получается, если $a = -b$, т. е., как видно из (7.25), при $a = -2f_2$, и расположено при $b = 2f_2$. Изображение при этом перевернуто, так как $\gamma = -1$.

Если нижний левый элемент C матрицы $\mathcal{M}$ (7.20) преобразования лучей оптической системой обращается в нуль, то (см. табл.) фокальные точки лежат в бесконечности. Такая система называется *телескопической* или *афокальной*. Примером может служить зрительная труба, установленная на бесконечность, когда задняя фокальная плоскость объектива совмещена с передней фокальной плоскостью окуляра. При $C = 0$ наклон выходящего луча $\alpha_2 = D\alpha_1$ не зависит от y_1, т. е. все лучи, падающие на систему параллельно друг другу, дадут на выходе также параллельный пучок лучей. Отношение углов наклона выходящих и входящих лучей $\alpha_2/\alpha_1 = D$ характеризует *угловое увеличение* телескопической системы. Оно определяется элементом D матрицы $\mathcal{M}$. Угловое увеличение зрительной трубы показывает, во сколько раз угол, под которым бесконечно удаленный предмет виден в трубу, больше угла, под которым он был бы виден невооруженным глазом.

Рассмотрим теперь преобразование гомоцентрического пучка параксиальных лучей, идущих из некоторой точки O_1 на оптической оси системы (рис. 7.14). Соответствующие такому пучку волновые поверхности представляют собой сферы с центрами в O_1. Для любой опорной плоскости ОП отношение высоты y луча к углу наклона α одинаково у всех лучей пучка и равно радиусу кривизны r волновой поверхности в том месте, где она пересекает ОП: $r = y/\alpha$. Поэтому весь пучок лучей можно характеризовать на ОП одним параметром, в качестве которого удобно использовать приведенное значение радиуса кривизны волновой поверхности $R = r/n = y/V$. Когда

пучок пересекает плоскую границу раздела двух сред, значение R остается неизменным, так как у каждого луча при пересечении границы параметры y и $V=n\alpha$ остаются прежними. Значение R для пучка изменяется при пересечении сферической границы и при прохождении оптического промежутка однородной среды. Изменение параметра R легко выразить через элементы матрицы $\mathscr{M}$ (7.20) преобразования луча в оптической системе. Так как $y_2=Ay_1+BV_1$, $V_2=Cy_1+DV_1$, то



7.14
Гомоцентрический пучок лучей

$$R_2=\frac{y_2}{V_2}=\frac{Ay_1+BV_1}{Cy_1+DV_1}=\frac{A(y_1/V_1)+B}{C(y_1/V_1)+D}=\frac{AR_1+B}{CR_1+D}. \qquad (7.26)$$

Соотношение (7.26), называемое *правилом ABCD*, позволяет вычислить приведенный радиус кривизны волновой поверхности с центром на оптической оси при переходе от одной ОП к другой. Например, при прохождении оптического промежутка приведенной толщины $L=l/n$, когда $\mathscr{M}$ совпадает с матрицей $\mathscr{T}$ (7.15), из (7.26) найдем $R_2=R_1+L$. При преломлении пучка в тонкой линзе с оптической силой P, когда преобразование лучей описывается матрицей $\mathscr{R}$ (7.16), правило $ABCD$ дает $R_2=R_1/(1-PR_1)$ или $1/R_2=1/R_1-P$ — кривизна волновой поверхности уменьшается на величину, равную оптической силе линзы.

Замечательно, что простое правило $ABCD$ применимо и для описания *гауссовых пучков* (см. § 6.4) при их распространении в центрированных оптических системах и тем самым дает удобный способ решения разнообразных задач, связанных с фокусировкой лазерного излучения. Гауссов пучок, пересекающий расположенную в некоторой точке z опорную плоскость, характеризуется двумя параметрами: радиусом кривизны волновой поверхности $R(z)$ и радиусом поперечного сечения $w(z)$. Изменение этих параметров при распространении пучка в свободном пространстве описывается формулами (6.33). При прохождении через тонкую линзу ширина $w(z)$ пучка остается прежней, а кривизна волновой поверхности в соответствии с формулой (6.35) уменьшается на величину, равную оптической силе линзы: $1/R_2=1/R_1-P$.

Чтобы следить сразу за изменением обоих параметров R и w при распространении гауссова пучка через оптическую систему, вводят следующую их комбинацию $q(z)$, называемую *комплексным радиусом кривизны:*

$$1/q=1/R+2i/(kw^2). \qquad (7.27)$$

Именно такая величина стоит множителем при $(x^2+y^2)/2$ в выражении (6.32) для напряженности поля в гауссовом пучке.

Вещественная часть $1/q$ равна кривизне волновой поверхности $1/R$, а мнимая пропорциональна $1/w^2$ (характеризует ширину пучка). Используя формулы (6.33), легко убедиться, что $q = q_0 + z$, где $q_0 = kw_0^2/(2i)$ — значение $q(z)$ при $z = 0$ (перетяжка пучка). Поэтому при распространении пучка в свободном пространстве или через оптический промежуток приведенной толщины $L = l/n$ преобразование параметра q происходит по формуле $q_2 = q_1 + L$ (точно так же, как для вещественных R по правилу $ABCD$). При преломлении на сферической границе раздела (или в тонкой линзе) ширина пучка не изменяется и поэтому $w_2 = w_1$, т. е. преобразование не затрагивает мнимой части параметра q. И здесь для комплексного q формула преобразования в точности такая же, как и для вещественного R: $1/q_2 = 1/q_1 - P$. Отсюда следует, что при прохождении гауссова пучка через произвольную центрированную оптическую систему, преобразование параметров параксиального луча в которой дается матрицей $\mathcal{M}$, изменение комплексного радиуса кривизны q можно находить с помощью того же правила $ABCD$, что и для вещественного R:

$$q_2 = (Aq_1 + B)/(Cq_1 + D); \qquad (7.28)$$

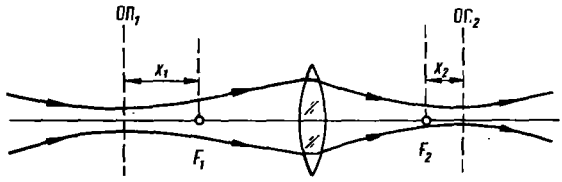
здесь параметр q_2 на выходе из системы выражается через q_1 и элементы матрицы $\mathcal{M}$.

Применим эти результаты для нахождения положения перетяжки и ее радиуса после прохождения гауссова пучка через тонкую линзу (рис. 7.15). Выберем $ОП_1$ в перетяжке исходного пучка, $ОП_2$ — в перетяжке преобразованного. Их положения будем определять расстояниями x_1 и x_2 соответственно от передней F_1 и задней F_2 фокальных точек линзы. Матрица $\mathcal{M}$ преобразования лучей от $ОП_1$ до $ОП_2$ получается перемножением двух $\mathcal{T}$-матриц (7.15) для оптических промежутков и $\mathcal{M}_F$-матрицы (7.23) преобразования между фокальными плоскостями линзы:

$$\mathcal{M} = \mathcal{T}_2 \mathcal{M}_H \mathcal{T}_1 = \begin{pmatrix} 1 & x_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & f \\ -1/f & 0 \end{pmatrix} \begin{pmatrix} 1 & -x_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -x_2/f & x_1x_2/f + f \\ -1/f & x_1/f \end{pmatrix} \qquad (7.29)$$

В перетяжках пучков волновые поверхности плоские (R_1, $R_2 \to \infty$), поэтому параметр q на $ОП_1$ и $ОП_2$ чисто мнимый. Его модуль пропорционален квадрату радиуса перетяжки w_0^2:

$$q_1 = -ikw_{01}^2/2 = -iz_{01}, \quad q_2 = -ikw_{02}^2/2 = -iz_{02}.$$



7.15

Преобразование гауссова пучка

Здесь z_{01} и z_{02} — радиусы дифракционной расходимости исходного и преобразованного пучков. Применяя правило $ABCD$ (7.28) с элементами матрицы $\mathscr{M}$ (7.29), получаем

$$-iz_{02} = \frac{f^2 + x_1x_2 + ix_2z_{01}}{x_1 + iz_{01}} = \frac{f^2x_1 + x_1^2x_2 + z_{01}^2x_2 - iz_{01}f^2}{x_1^2 + z_{01}^2}.$$

Приравнивая отдельно вещественные и мнимые части, получаем две формулы, по которым параметры x_2 и $z_{02} = kw_{02}^2/2$, определяющие положение и радиус перетяжки преобразованного пучка, выражаются через соответствующие параметры исходного пучка и фокусное расстояние линзы:

$$x_2 = -x_1f^2/(x_1^2 + z_{01}^2),\quad z_{02} = z_{01}f^2/(x_1^2 + z_{01}^2). \tag{7.30}$$

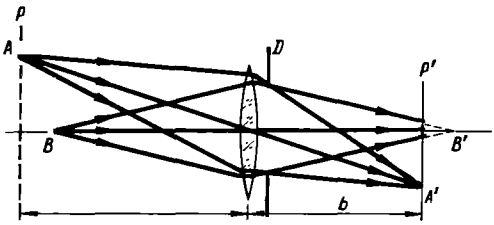
В частности, при $f^2 = x_1^2 + z_{01}^2$ формулы (7.30) дают $z_{02} = z_{01}$ и $x_2 = -x_1$, т. е. после прохождения через линзу с таким фокусным расстоянием воспроизводится геометрия исходного гауссова пучка.

Контрольные вопросы

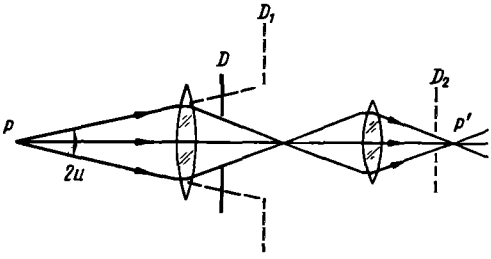
- Какие лучи называют параксиальными? Какими параметрами задают луч на опорной плоскости? Приведите вид матриц преобразования параметров луча, сопоставляемых: 1) оптическому промежутку; 2) сферической преломляющей поверхности; 3) отражающей поверхности; 4) тонкой линзе.
- Какими свойствами обладают фокальные точки оптической системы? Что такое главные плоскости?
- Как можно найти преобразование параксиального луча в оптической системе, если известно положение ее кардинальных точек?
- Что такое сопряженные плоскости? Как найти в пространстве изображений плоскость, сопряженную с определенной плоскостью в пространстве предметов?
- При каком условии создаваемое оптической системой изображение трехмерного предмета геометрически подобно самому предмету?
- Как в оптической системе изменение радиуса кривизны волновой поверхности выражается через элементы матрицы $\mathscr{M}$ преобразования луча?
- Какими параметрами характеризуется гауссов пучок? Что такое комплексный радиус кривизны? Как он преобразуется при прохождении пучка через оптическую систему?

7.3. Ограничение световых пучков в оптических системах

В оптических инструментах обычно требуется получить изображение некоторого предмета на какой-либо поверхности, чаще всего на плоскости, перпендикулярной оптической оси (например, на фотопластинке в фотоаппарате). Предмет же в большинстве случаев бывает пространственным, а не плоским. Однако даже идеальная оптическая система (центрированная система в параксиальном приближении) дает на некоторой плоскости P' резкое изображение лишь тех точек предмета, которые лежат в сопряженной с ней плоскости P (точка A на рис. 7.16). Положе-



7.16
К зависимости глубины резкости от ограничения пучков



7.17
Апертурная диафрагма (D), входной (D_1) и выходной (D_2) зрачки

ние этой *плоскости установки* (или плоскости наводки) определяется формулой (7.25). Когда точка предмета не лежит в плоскости установки, ее изображение получается в виде *кружка рассеяния*. Размер кружка, соответствующего изображению точки B в плоскости P', как видно из рис. 7.16, зависит от смещения B из плоскости установки P (или от смещения B' из сопряженной с P плоскости P'), а также от угловой ширины реального пучка лучей, формирующих изображение. Эта ширина определяется ограничивающей пучок диафрагмой D или краями линзы. Ниже мы увидим, что ограничение реальных световых пучков в оптических системах играет доминирующую роль во всех вопросах, касающихся отображения.

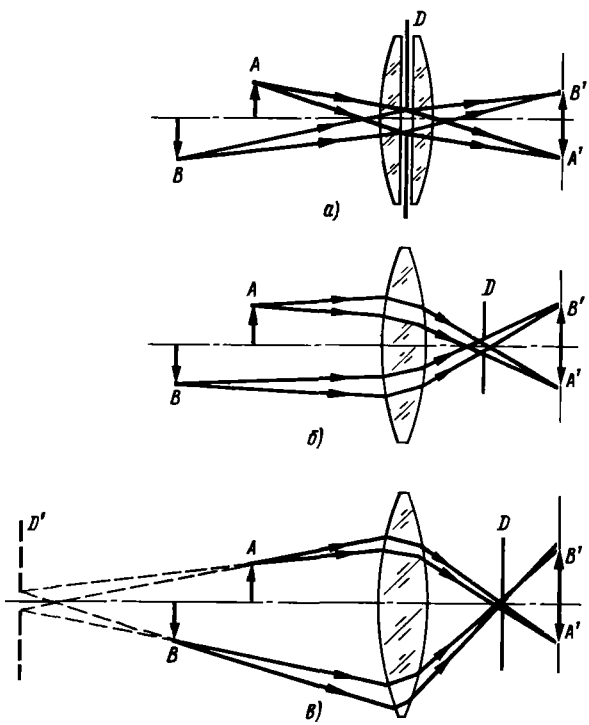
Для получения удовлетворительного изображения размер кружка рассеяния не должен превышать определенного предела. Этим объясняется существование некоторой глубины пространства, резко отображаемого на плоскости. Очевидно, что *глубина резкости* возрастает при уменьшении диаметра диафрагмы.

Ограничение пучков в оптических системах в общем случае осуществляется по-разному для лучей, исходящих из разных точек предмета. Диафрагма, которая ограничивает пучок лучей, формирующий изображение расположенной на оси системы точки предмета, называется *апертурной* или *действующей*. Ее роль может выполнять оправа какой-либо линзы или специальная диафрагма, расположенная либо между линзами, либо перед системой или после нее. Изображение апертурной диафрагмы (действительное или мнимое), создаваемое находящейся перед ней частью оптической системы, называется *входным зрачком*, а изображение, создаваемое расположенными за диафрагмой оптическими элементами, — *выходным зрачком* (рис. 7.17). Входным зрачком определяется *апертура* — угол раскрытия пучка лучей, идущего из точки предмета P через систему.

В телескопах и зрительных трубах (см. рис. 7.19) апертурной

диафрагмой обычно служит край объектива. Так как никаких линз перед ним нет, то он будет одновременно и входным зрачком, а выходным будет его изображение, создаваемое окуляром.

Плоские изображения пространственно протяженных предметов всегда передают геометрическую *перспективу* (определенное соотношение между размерами изображений предметов, лежащих на различном удалении). Например, на фотоснимке получается центральная проекция фотографируемых предметов с центром проекции в середине объектива, так как идущие через центр линзы лучи не отклоняются. Для получения правильного пространственного впечатления при рассматривании фотоснимка нужно, чтобы видимые глазом угловые размеры изображений предметов были такими же, как и при непосредственном наблюдении. Это условие выполняется, если рассматривать снимок одним глазом с такого расстояния, на каком (от пластинки) находился объектив при фотографировании. Для n-кратно увеличенных по сравнению с негативом фотоснимков это расстояние также следует увеличить в n раз. В большой аудитории (кинозал) такое условие выполняется для немногих мест. При рассматривании с неправильного расстояния фотография создает пространственное впечатление с искаженной перспективой: при слишком большом расстоянии глубина снимка кажется увеличенной, а при слишком малом — уменьшенной. Искажение перспективы заметно и при непосредственном наблюдении в зрительную трубу или бинокль: при сильном увеличении все предметы и расстояния



7.18
Влияние положения апертурной диафрагмы D на характер перспективы:
а — нормальная перспектива; *б* — телецентрическая перспектива; *в* — гиперцентрическая

кажутся укороченными в глубину. Еще большее впечатление производит обратный опыт: если смотреть в бинокль, направив на предмет не объектив, а окуляр, то протяженность предмета в глубину кажется сильно преувеличенной.

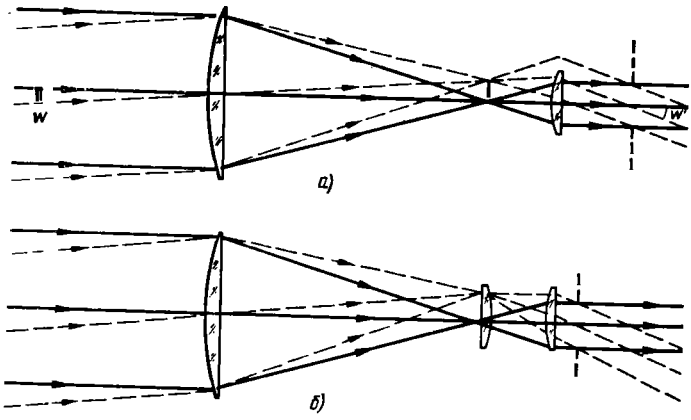
Ограничение световых пучков оказывает решающее влияние не только на глубину пространства, резко изображаемого на плоскости, но и на характер передаваемой этим изображением перспективы. В общем случае перспектива определяется положением входного зрачка системы, ибо именно он служит центром проекции предмета на плоскость наводки. Проиллюстрируем это следующим примером. Пусть с помощью тонкой линзы на плоском экране создается изображение одинаковых предметов A и B (рис. 7.18), находящихся на разных расстояниях. Если апертурная диафрагма D расположена рядом с линзой (рис. 7.18, *а*), ее изображение будет практически в том же месте, т. е. диафрагма служит одновременно и входным зрачком. Осевые лучи всех пучков, формирующих изображения предметов A и B, пересекаются в центре входного зрачка. На плоскости экрана изображение близкого предмета A крупнее, чем изображение такого же, но более удаленного предмета B (нормальная перспектива).

Когда апертурная диафрагма D расположена вблизи фокальной точки линзы (рис. 7.18, *б*), изображение диафрагмы в пространстве предметов удалено в бесконечность: через центр диафрагмы пройдут те лучи, которые до линзы шли параллельно оптической оси. Главные лучи (через центр диафрагмы) действующих световых пучков, формирующих изображения предметов A и B, перед линзой направлены параллельно оптической оси. Поэтому на экране изображения предметов A и B, лежащих на разных расстояниях, имеют одинаковые размеры (телецентрическая перспектива). Изменение расстояния до предмета влияет в этом случае только на резкость изображения, но не на его размеры. Такая перспектива применяется в измерительных микроскопах.

Еще необычнее случай, когда апертурная диафрагма D расположена за фокусом (рис. 7.18, *в*). Ее изображение D' (входной зрачок, служащий центром проекции) лежит перед предметом. Изображение на экране близкого предмета A имеет меньший размер, чем изображение такого же удаленного предмета B. Перспектива получается обращенной (гиперцентрической).

От апертурной диафрагмы зависит не только глубина резкости и характер передаваемой перспективы, но и энергия света, проходящего через систему, т. е. фотометрические характеристики изображения (см. § 7.5).

Кроме апертурной диафрагмы, ограничивающей действующие пучки света от лежащих на оптической оси точек предмета, в оптической системе существует *диафрагма поля зрения*. Поясним ее роль следующей аналогией: при наблюдении невооруженным глазом через окно поле зрения ограничено рамой окна. Диафрагма поля зрения определяет ту часть поверхности протяженного предмета, которая отображается прибором. Рассмотрим этот вопрос на примере астрономической зрительной трубы (телескопа-рефрактора). Обычно задняя фокальная плоскость объектива, где образуется действительное изображение удаленных предметов, совпадает с передней фокальной плоскостью окуляра, так что падающий на объектив параллельный пучок лучей идет из окуляра также параллельным (рис. 7.19, *а*). Апертурной диафрагмой и входным зрачком служит оправа объектива, выходным зрачком — ее изображение, создаваемое окуляром. Наклонные пучки лучей, идущие от не лежащих на оптической оси точек предмета, частично или полностью срезаются краями окуляра. Край окуляра в этом случае служит диафрагмой поля зрения. Его изображение, создаваемое объективом, расположено в пространстве предметов и называется *входным окном (входным люком)* системы (на рис. 7.19 он слева за пределами



7.19
Ограничение поля зрения и затенение изображения оправой окуляра (*а*) и применение полевой линзы (*б*)

чертежа). Угол $2w$, под которым входной люк виден из центра входного зрачка, называется *углом поля зрения.*

Из-за частичного срезания диафрагмой поля зрения (краями окуляра) наклонных пучков лучей от внеосевых точек предмета освещенность видимого глазом изображения удаленного протяженного предмета постепенно уменьшается к краям поля зрения. Такой эффект называется *затенением* или *виньетированием.* Виньетирования не будет, когда входной люк лежит в плоскости предмета. В рассматриваемом примере зрительной трубы устранить виньетирование и сделать границы поля зрения резкими можно, поместив диафрагму в фокальной плоскости объектива вблизи промежуточного изображения. Но лучше в этой плоскости поместить дополнительную линзу (рис. 7.19, *б*), называемую *коллективом* или *полевой линзой.* При правильном выборе фокусного расстояния полевой линзы ее оправа служит диафрагмой поля зрения. Этим достигается одновременно и устранение виньетирования, и увеличение поля зрения. Апертура, определяемая диаметром объектива, и угловое увеличение трубы остаются прежними, изменяется лишь положение выходного зрачка. Практически полевую линзу располагают позади плоскости первичного изображения, чтобы сделать незаметными загрязнения и дефекты ее поверхности и чтобы в плоскость изображения можно было внести измерительную шкалу или крест нитей.

Понимание роли ограничения световых пучков совершенно необходимо при использовании любых оптических приборов. Ограничение пучков определяет не только пропускаемый системой световой поток, но и такие характеристики, как поле зрения, глубина резкости, передаваемая изображением перспектива. Ограничение пучков для всех оптических приборов определяет и предельную (ди-

фракционную) разрешающую способность (см. § 7.6). Для телескопа это наименьшее угловое расстояние между воспринимаемыми раздельно удаленными точечными источниками, для микроскопа — наименьший различимый размер предмета, для спектрального аппарата — наименьшее регистрируемое различие длин волн и т. п.

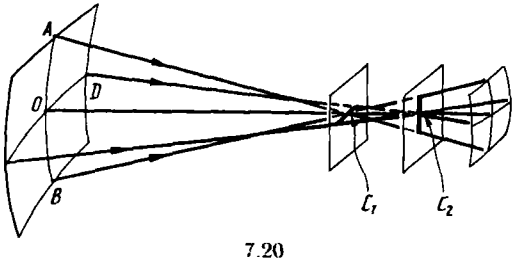
Контрольные вопросы

- ☐ Что называется апертурной диафрагмой оптической системы? входным и выходным зрачками? В каком случае входной зрачок совпадает с апертурной диафрагмой?
- ☐ С какого расстояния нужно рассматривать фотоснимок, чтобы получить правильное пространственное впечатление?
- ☐ Как положение апертурной диафрагмы влияет на характер перспективы, получающейся на плоском изображении?
- ☐ Что называется диафрагмой поля зрения, входным и выходным окнами оптической системы?
- ☐ Какую роль в оптической системе выполняет полевая линза (коллектив)?

7.4. Аберрации оптических систем

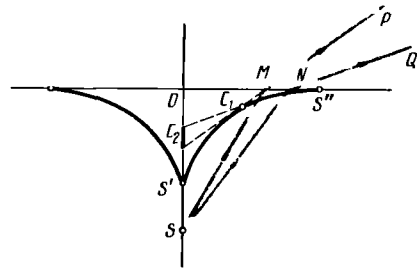
До сих пор предполагалось, что все действующие лучи параксиальны и, следовательно, гомоцентрический пучок лучей, идущих из некоторой точки предмета, при прохождении через оптическую систему остается почти гомоцентрическим. В практической оптике условия применимости этой идеализации для реальных пучков лучей, участвующих в формировании изображения предмета, выполняются крайне редко. Нарушение гомоцентричности пучков приводит к появлению погрешностей (*аберраций*) изображения, создаваемого оптической системой.

Если при преломлении или отражении пучок лучей перестает быть гомоцентрическим, нормальная к лучам волновая поверхность уже не будет сферической. Как известно из дифференциальной геометрии, для любой точки O произвольной гладкой поверхности (рис. 7.20) существует два взаимно перпендикулярных направления AOB и COD, которым соответствуют наименьшее R_1 и наибольшее R_2 значения радиуса кривизны. Лучи, проходящие через точки A, O и B, пересекаются в центре кривизны C_1, лежащем на расстоянии R_1 от поверхности; лучи через C, O и D — в центре C_2 на расстоянии R_2. При $R_2 \neq R_1$ пучок лучей называется *астигматическим*.



7.20
Астигматический пучок лучей

В отличие от гомоцентрического бесконечно узкий астигматический пучок дает два точечных изображения C_1 и C_2, удаленных вдоль оси пучка на расстояние R_2-R_1, называемое *астигматической разностью*. В пучке конечной ширины все лучи пересекаются на двух взаимно перпендикулярных *фокальных отрезках*, проходящих через C_1 и C_2 (рис. 7.20). Произвольный широкий пучок лучей можно разложить на элементарные астигматические пучки, каждому из которых соответствует пара фокальных точек C_1 и C_2. Множество фокальных точек образует двухлистную поверхность, называемую *каустикой* пучка лучей.

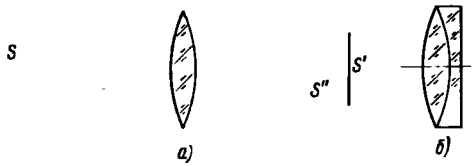


21
Преломление на плоской границе

Рассмотрим, например, преломление гомоцентрических пучков лучей от точечного источника S (рис. 7.21) на плоской границе раздела прозрачных сред. Получающиеся в результате преломления пучки во второй среде будут астигматическими. На рис. 7.21 показаны два близких меридиональных луча SMP и SNQ. Их продолжения пересекаются в точке C_1, координаты которой зависят от угла падения и могут быть найдены с помощью закона преломления. Чтобы получить узкий пространственный пучок лучей, мысленно повернем рисунок на небольшой угол вокруг оси симметрии SO. Точка C_1 прочертит при этом небольшую дугу, перпендикулярную плоскости рисунка. Это будет меридиональная фокальная линия астигматического пучка. Вторая (сагиттальная) фокальная линия представляет собой отрезок идущей через источник S нормали к границе раздела. С увеличением угла падения возрастает астигматическая разность преломленного пучка, так как фокальная точка C_1 перемещается по некоторой кривой $S'C_1S''$. Поэтому при рассматривании предметов, находящихся, например, под водой, четкость изображения ухудшается из-за астигматизма пучков при отклонении направления наблюдения от нормали к поверхности. Каустика меридиональных лучей широкого преломленного пучка представляет собой воронкообразную поверхность, получающуюся при вращении кривой $S'C_1S''$ вокруг нормали SO. Каустика сагиттальных лучей вырождается в отрезок перпендикуляра $S'O$.

Дадим краткую характеристику основных видов геометрических аберраций.

Сферическая аберрация. Пусть точечный источник S расположен на оптической оси собирающей линзы (рис. 7.22, *а*). Выходящие из S параксиальные лучи после преломления в линзе образуют изображение S' на расстоянии b от линзы, которое определяется формулой (7.25). Но лучи, пересекающие линзу вблизи ее краев, преломляются сильнее и сойдутся на оптической оси в точке S'',



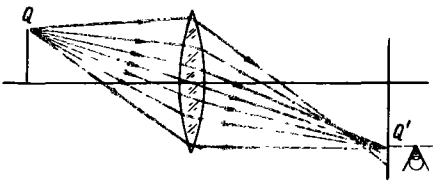
Сферическая аберрация простой линзы (*а*) и сложная линза, исправленная на сферическую аберрацию (*б*)

которая лежит ближе, чем S'. Расстояние $S'S''$ называется продольной сферической аберрацией. На совмещенном с S' экране будет наблюдаться светлый кружок. Его радиус называется поперечной сферической аберрацией. Наиболее отчетливым изображение источника S будет тогда, когда экран помещен посередине между S' и S''. У несимметричной линзы аберрация зависит от того, какой стороной линза обращена к источнику.

Для заданной пары сопряженных точек сферическая аберрация может быть исправлена выбором более сложной формы преломляющих поверхностей. Но на практике для уменьшения сферической аберрации используют комбинацию собирающей и рассеивающей линз со сферическими преломляющими поверхностями (рис. 7.22,*б*). Метод основан на том, что рассеивающая линза обладает сферической аберрацией противоположного знака. Сферическую аберрацию удается устранить лишь для определенного расстояния до предмета. Для зрительных труб и обычных фотообъективов выбирают удаленный предмет, объективы микроскопов коррегируют для положения предмета непосредственно перед фокусом. Сферическую аберрацию создают не только сферические, но и плоские поверхности. Поэтому объективы микроскопов коррегируют для вполне определенной толщины плоскопараллельных покровных стекол. Поверхностям зеркал телескопов-рефлекторов для устранения сферической аберрации придают форму параболоида вращения.

Все остальные виды геометрических аберраций (в монохроматическом свете) появляются лишь при отображении точек предмета, не лежащих на оптической оси системы.

Кома. Эта аберрация возникает при отображении широкими пучками лучей внеосевых точек предмета. Лучи из точки Q, проходящие ближе к краям линзы, преломляются сильнее, чем требуется для их схождения в точку Q' пересечения параксиальных лучей (рис. 7.23), и образуют в плоскости параксиальных изображений кружки рассеяния, которые заполняют область между прямыми, выходящими из Q'. Это схематически показано в правой части рис. 7.23. Такое «изображение» точки напоминает комету, чем и объясняется название «кома».



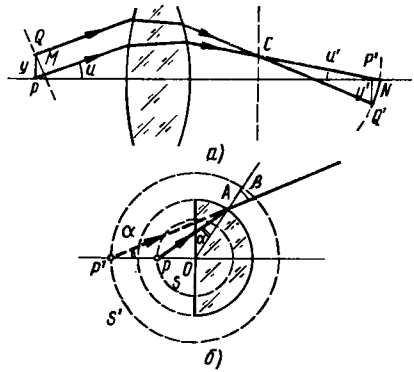
7.23
Кома

Аберрация комы отсутствует у систем, удовлетворяющих *условию синусов Аббе*. Изобра-

жение протяженного предмета будет резким, если оптические длины лучей между всеми парами сопряженных точек предмета и его изображения одинаковы (см. задачу 2). Пусть небольшой предмет PQ длиной y расположен перпендикулярно оптической оси (рис. 7.24, *а*). Рассмотрим два параллельных луча, выходящих из P и Q под произвольным углом u к оптической оси. После преломления эти лучи пересекутся в точке C фокальной плоскости и пройдут через сопряженные с P и Q точки P' и Q'. Если предмет PQ мал, то лучи PP' и QQ' и после преломления почти параллельны, т. е. можно считать, что они наклонены к оптической оси под одним и тем же углом u'. Проведем ортогональные к лучам волновые поверхности QM и $Q'N$. Оптические длины всех лучей между ними одинаковы. Поэтому оптические длины лучей PP' и QQ' будут равны при условии равенства оптических длин отрезков PM и $P'N$:

$$ny\sin u = n'y'\sin u', \qquad (7.31)$$

где n и n' — показатели преломления сред со стороны предмета и изображения. Таким образом, для получения стигматических изображений малых участков поверхности, перпендикулярной оси, с помощью широких пучков синусы углов u и u' раскрытия пучков должны удовлетворять условию (7.31). Отображение, при котором выполнено условие синусов, принято называть *апланатическим*. Оптическая система может дать апланатическое отображение только для определенных расстояний до предмета и изображения, на которые она рассчитана при изготовлении. Например, сопряженные апланатические точки P и P' (рис. 7.24, *б*) для преломления на поверхности стеклянного шара радиусом R лежат на расстояниях nR и R/n от его центра (см. задачу 3). Это свойство шара используется в *иммерсионных объективах* микроскопов, когда между покровным стеклом и передней линзой объектива вводится слой жидкости (иммерсия) с таким же, как у стекла, показателем преломления (кедровое масло). Передняя линза представляет собой стеклянное полушарие, плоской стороной обращенное к предмету, который помещается в точке P на расстоянии R/n от центра O (рис. 7.24, *б*). Благодаря иммерсии лучи не испытывают преломления на плоской границе и P будет апланатической точкой. Мнимое изображение предмета находится в сопряженной апланатической точке P' на расстоянии nR от центра O.

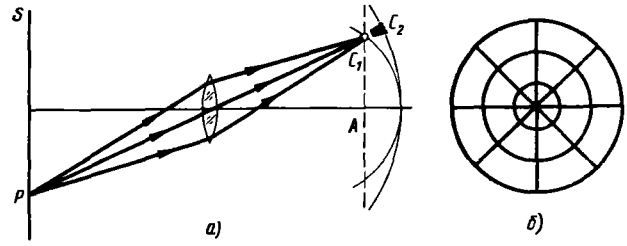


7.24
К выводу условия синусов (*а*) и апланатнческне точкн при преломленнн на сферичсской поверхности (*б*)

Таким способом уменьшаются углы наклона лучей к оптической оси. Получение апланатического изображения, т. е. устранение сферической аберрации и комы, особенно существенно для микроскопов с большим увеличением, где используются максимально широкие апертуры (вплоть до теоретического предела 180°), что необходимо для достижения высокой разрешающей способности (см. § 7.6).

Астигматизм наклонных пучков и кривизна поля изображения. Даже узкие пучки лучей при прохождении через оптическую систему, как правило, утрачивают гомоцентричность и становятся астигматическими, если они составляют значительный угол с оптической осью. Такие пучки вместо одного фокуса дают две фокальные линии (см. рис. 7.20). Пусть центральный луч преломленного пучка лежит в меридиональной плоскости (плоскости чертежа на рис. 7.25, *а*). Меридиональные лучи этого пучка пересекаются на фокальной линии C_1, перпендикулярной плоскости чертежа. Сагиттальные (т. е. лежащие в перпендикулярной чертежу плоскости) лучи пересекаются на фокальной линии C_2, лежащей в плоскости чертежа. Расстояние между фокальными линиями (астигматическая разность) возрастает с увеличением угла наклона пучка. При отображении плоскости множество отрезков C_1 и C_2, которые можно рассматривать как «изображения» точек плоскости S меридиональными и сагиттальными лучами, образует две искривленные поверхности с симметрией вращения относительно оси системы, касающиеся друг друга в точке пересечения с оптической осью.

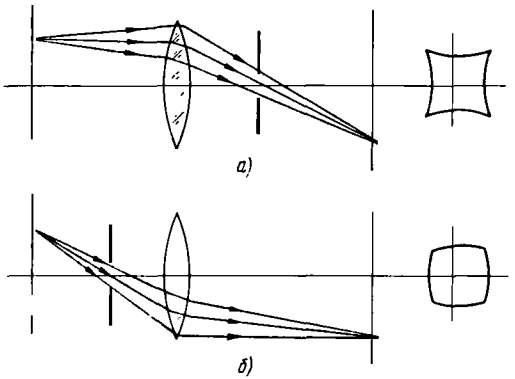
Аберрация астигматизма ярко проявляется при получении изображения плоского объекта, имеющего форму «колеса со спицами» (рис. 7.25, *б*), центр которого лежит на оптической оси. При перемещении плоского экрана A (рис. 7.25, *а*) вдоль оси можно получить резкое изображение определенной окружности, когда экран совмещается с положением соответствующих меридиональных фокальных линий C_1. Эти фокальные линии дают черточки изображения, ориентированные по дугам окружности, и при наложении друг на друга образуют ее резкое изображение, в то время как изобра-



7.25

Астигматизм наклонных пучков и кривизна поля изображения (*а*) и объект для испытания линз на астигматизм (*б*)

жения радиусов («спиц») окажутся размытыми. Если же плоскость экрана совпадает с положением вытянутых вдоль радиусов фокальных линий C_2 для сагиттальных лучей, на изображении будут резкими соответствующие участки «спиц», а окружности будут размыты.



7.26
Дисторсия изображения при разных положениях диафрагмы.
Справа показано изображенне квадрата

Подбирая комбинацию линз из разных сортов стекла и с разной кривизной преломляющих поверхностей, можно приблизительно совместить поверхности меридиональных и сагиттальных фокальных линий (т. е. уменьшить астигматизм) и одновременно до некоторой степени выпрямить их, т. е. сделать поле изображения достаточно плоским. Это особенно важно для фотообъективов, которые должны давать резкое изображение на плоской поверхности. Исправленные на астигматизм системы называют *анастигматами*.

Астигматизм пучков, параллельных оптической оси, возникает при нарушении осевой симметрии системы, например когда кривизна преломляющей поверхности неодинакова в различных сечениях. Таким астигматизмом нередко обладает человеческий глаз, что проявляется в неспособности видеть одинаково резко взаимно перпендикулярные полосы на испытательных таблицах. Для коррекции этого недостатка применяют очки с цилиндрическими линзами.

Дисторсия. В отличие от рассмотренных выше аберраций, ухудшающих резкость изображения, в оптических системах возможно искажение (дисторсия) геометрической формы изображения протяженного предмета. Если линейное увеличение растет по мере удаления от оптической оси к краям поля зрения, изображение квадрата приобретает вид «подушки» (рис. 7.26, *а*). Так бывает при расположении ограничивающей пучки диафрагмы позади линзы. Если диафрагма находится перед линзой, увеличение по краям поля зрения меньше, чем в центре, и изображение квадрата приобретает вид «бочки» (рис. 7.26, *б*). В системе двух линз при расположении диафрагмы между линзами можно добиться почти полного уничтожения дисторсии, так как подушкообразная дисторсия, создаваемая первой линзой, компенсируется бочкообразной дисторсией второй линзы.

В телескопах и приборах для визуальных наблюдений дисторсия не имеет большого значения, так как она не влияет на резкость изображения. Но в фотообъективах (особенно для аэрофотосъемки)

и в приборах для геодезических измерений дисторсия вредна и должна быть по возможности исправлена.

Хроматическая аберрация. Все оптические стекла обладают дисперсией, и потому угол отклонения луча при преломлении в линзе зависит от длины волны. При использовании белого света оптическая система создает множество монохроматических изображений, которые не совпадают ни по положению, ни по размерам. В результате их наложения изображение предмета получается нерезким и с окрашенными краями. Это явление называется хроматической аберрацией.

Для уменьшения хроматической аберрации используют комбинации собирающих и рассеивающих линз из сортов стекла с различающейся дисперсией. Устранить ее для всего спектра невозможно*. Обычно стремятся совместить изображения для каких-либо двух длин волн, выбор которых определяется назначением оптической системы. В приборах для визуальных наблюдений ахроматизацию производят для фраунгоферовых спектральных линий $F(\lambda_F = 480$ нм) и C ($\lambda_C = 656$ нм), при этом во всей видимой части спектра аберрация будет значительно ослаблена.

Рассмотрим способы ахроматизации тонких линз, ограничиваясь параксиальным приближением. Согласно формуле (7.19), для линзы с определенными значениями радиусов кривизны R_1 и R_2 произведение $(n-1)f$ не зависит от длины волны. Отсюда легко выразить изменение δf ее фокусного расстояния при изменении δn показателя преломления n:

$$\delta f/f = -\delta n/(n-1). \qquad (7.32)$$

Для характеристики разных сортов стекла вводят величину

$$\Delta = (n_F - n_C)/(n_D - 1), \qquad (7.33)$$

где n_F, n_C и n_D — показатели преломления для F, C и D — линий Фраунгофера ($\lambda_D = 589$ нм). У флинта $\Delta \approx 1/30$, у крона $\approx 1/60$. Полагая $\delta n = n_C - n_F$, $n \approx n_D$, формулу (7.32) можем записать в виде

$$\delta f/f = \Delta. \qquad (7.34)$$

Фокусное расстояние системы из двух тонких линз, находящихся на расстоянии l друг от друга, в соответствии с (7.18) определяется выражением

$$1/f = 1/f_1 + 1/f_2 - l/(f_1 f_2)$$

Отсюда следует, что для ахроматизации фокусного расстояния системы ($\delta f = 0$) должно выполняться условие

$$\delta f_1/f_1^2 + \delta f_2/f_2^2 - [l/(f_1 f_2)](\delta f_1/f_1 + \delta f_2/f_2) = 0,$$

* Полностью свободны от хроматической аберрации зеркальные системы (телескоп-рефлектор), так как лучи всех длин волн отражаются одинаково.

которому с помощью (7.34) можно придать вид

$$f_2\Delta_1 + f_1\Delta_2 = l(\Delta_1 + \Delta_2). \qquad (7.35)$$

Один из способов ахроматизации состоит в использовании двух соприкасающихся тонких линз, одна из которых сделана из крона, другая — из флинта. В этом случае $l = 0$ и из (7.35) $f_2\Delta_1 + f_1\Delta_2 = 0$. Так как для всех сортов стекла $\Delta > 0$, то фокусные расстояния f_1 и f_2 должны иметь противоположные знаки. Если задано общее фокусное расстояние f, то для данных сортов стекла f_1 и f_2 определяются однозначно:

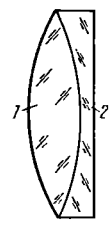
$$1/f_1 = (1/f)\Delta_2/(\Delta_2 - \Delta_1),\ 1/f_2 = -(1/f)\Delta_1/(\Delta_2 - \Delta_1).$$

Чтобы общее фокусное расстояние f было положительно, собирающая линза должна изготовляться из крона, рассеивающая — из флинта (рис. 7.27). Так как f_1 и f_2 зависят от трех радиусов кривизны, то один из них можно выбрать произвольно. Наличие свободного параметра позволяет одновременно с ахроматизацией значительно уменьшить сферическую аберрацию такой составной линзы.

Другая возможность получения ахроматической системы состоит в использовании двух линз из одинакового стекла ($\Delta_1 = \Delta_2$), находящихся одна от другой на расстоянии, равном полусумме их фокусных расстояний: $l = (f_1 + f_2)/2$. Ахроматизация фокусного расстояния такой системы сразу для всего спектра непосредственно следует из формулы (7.35). Однако это лишь частичная ахроматизация, так как она обеспечивает совпадение углового увеличения изображений в разных цветах, но не их местоположения (из-за различия в положении главных плоскостей). Такой способ применяется в окулярах зрительных труб.

Путем построения более сложных систем можно добиться совпадения фокусов для трех различных цветов. Таковы наиболее совершенные объективы микроскопов (апохроматы), разработанные Аббе. Геометрооптические качества 10-линзового апохромата Аббе настолько высоки, что он позволяет достичь теоретического дифракционного предела разрешающей способности, обусловленного волновой природой света (см. § 7.6).

Уменьшение многочисленных аберраций возможно лишь путем устройства сложных специально рассчитанных оптических систем. Однако одновременное устранение всех аберраций может оказаться неразрешимой задачей. Поэтому приходится идти на компромисс: рассчитывая оптику для определенной цели, стремятся избавиться от особенно вредных недостатков и мирятся с неполным устранением других. Каждый оптический прибор должен отвечать своему специальному назначению. К объективам телескопов, микроскопов и фотообъективам предъявляются совсем разные требования. Если



Ахроматическая линза: *1* — крон; *2* — флинт

для телескопа с малым угловым размером поля зрения достаточно устранить хроматическую и сферическую аберрации, то для объективов микроскопов и фотообъективов с широким полем зрения необходимо еще устранение дисторсии и кривизны поля изображения. Для наблюдения объектов малой яркости необходимы светосильные объективы с большим относительным отверстием, и это вынуждает мириться с некоторыми аберрациями, неизбежными при использовании широких пучков. Объектив коллиматора спектрографа должен быть исправлен на сферическую аберрацию и хорошо ахроматизирован, так как для получения параллельного пучка входная щель должна находиться в общем для всех длин волн фокусе. Для камерного объектива ахроматизация вообще не обязательна, так как различие фокусных расстояний для красных и фиолетовых лучей можно скомпенсировать наклонным расположением кассеты с пластинкой, но зато он должен быть хорошо исправлен на астигматизм и кому, так как лучи разных длин волн выходят из диспергирующего элемента под разными углами к его оптической оси, и т. д.

В построении систем, свободных от тех или иных аберраций, инструментальная оптика достигла поразительных результатов.

Контрольные вопросы

1. Какими способами устраняют сферическую аберрацию в зеркальных и линзовых объективах телескопов?
2. Какое условие должно выполняться для того, чтобы оптическая система создавала резкое изображение точек предмета, не лежащих на оптической оси?
3. Где расположены сопряженные апланатические точки для сферической преломляющей поверхности?
4. В чем заключается аберрация астигматизма?
5. Какое физическое явление вызывает хроматическую аберрацию? Какими способами удается ее уменьшить?
6. Какие аберрации должны быть устранены в первую очередь у объективов: а) телескопа; б) фотоаппарата; в) микроскопа; г) коллиматора спектрографа; д) камеры спектрографа?

Задачи

1. Точечный источник S находится под водой на глубине h от поверхности (см. рис. 7.21). Найти каустическую поверхность пучка выходящих из воды преломленных лучей.

◇ Направим ось x вдоль поверхности воды, ось y — вдоль нормали к поверхности. В узком астигматическом пучке преломленных лучей, выходящем под углом α к нормали, меридиональные лучи пересекаются в точке C_1, координаты x и y которой даются формулами

$$x(\alpha)=h(n^2-1)\sin^3\alpha/(n^2-\sin^2\alpha)^{3/2},\quad y(\alpha)=-hn^2\cos^3\alpha/(n^2-\sin^2\alpha)^{3/2}.$$

Это параметрическое уравнение каустики меридиональных лучей. При $\alpha=0$ точка C_1 лежит на одной нормали с источником S на глубине h/n:

$x=0$, $y=-h/n$ (точка S'). При $\alpha=\pi/2$ (скользящий вдоль поверхности пучок преломленных лучей) $x=h/\sqrt{n^2-1}$, $y=0$ (точка S''). Сагиттальные лучи всех пучков пересекаются в точках нормали SO. Каустика таких лучей вырождается в отрезок $S'O$ длиной h/n.

2. Доказать, что в оптической системе, создающей резкое отображение плоскости предмета на плоскость в пространстве изображений, оптические длины лучей от точки предмета до ее изображения одинаковы для всех пар сопряженных точек.

3. Доказать, что для преломления света на поверхности шара точки P и P', лежащие на расстояниях R/n и nR от центра O, образуют сопряжениую пару апланатических точек (рис. 7.24, *б*).

По условию, $|AO|/|OP|=n$ и $|OP'|/|AO|=n$. Это зиачит, что треугольники APO и $AP'O$ подобны. Из подобия треугольников APO и $AP'O$ следует, что угол $AP'O$ равен углу падения α на поверхность шара луча, выходящего из точки P. Из треугольника $AP'O$ получаем, что $\sin\beta/\sin\alpha=|OP'|/|AO|=n$, т. е. прямая $P'A$ дает направление преломлеиного луча. Таким образом, если в P поместить точечный источник, то после преломления получится широкий гомоцентрический пучок лучей, расходящихся из точки P'. Ввиду сферической симметрии рассматриваемой системы не только точкн P и P', но и целиком сферы S и S' отображаются стигматически друг в друга широкими пучками лучей. Так как радиусы сфер S' и S относятся как n^2, то поперечное линейное увелнчение при таком аплаиатическом отображении равно n^2. ◇

7.5. Яркость и освещенность оптических изображений

Оптический прибор формирует изображение предмета, которое рассматривается глазом или воздействует на какой-либо иной приемник излучения. Точки предмета до сих пор считались математическими точками, но в действительности излучение с конечной энергией испускается элементом поверхности, который имеет конечные размеры. Будем считать, что поверхность предмета излучает по закону Ламберта (см. § 1.10), т. е. характеризуется яркостью B, не зависящей от направления. Допустим, что небольшой элемент поверхности, имеющий площадь σ, расположен перпендикулярно оптической оси системы (рис. 7.28) и отображается апланатически с выполнением условия синусов (7.31). Найдем световой поток Φ от этого элемента, проходящий через систему и достигающий изображения. Поток в элементарный телесный угол $d\Omega=\sin\theta d\theta d\varphi$ равен

$$d\Phi=\sigma B\cos\theta d\Omega=\sigma B\cos\theta\sin\theta d\theta d\varphi.$$

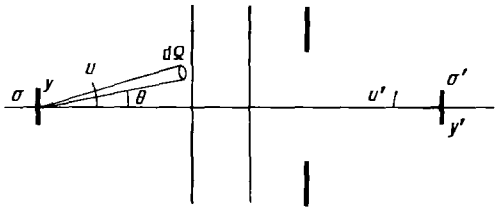
Интегрируя выражение по всей входной апертуре, т. е. по φ в пределах от 0 до 2π и по θ от 0 до u, находим

$$\Phi=\pi\sigma B\sin^2 u. \qquad (7.36)$$

Аналогично, для светового потока Φ', исходящего от площадки — изображения σ', можно написать

$$\Phi'=\pi\sigma' B'\sin^2 u', \qquad (7.37)$$

где B' — яркость изображения; $2u'$ — апертура со стороны изображения. Если потерями излучения из-за отражений, рассеяния и поглощения в стекле пренебречь, то потоки через входной и выходной зрачки одинаковы: $\Phi=\Phi'$.



К расчету освещенности изображения

Приравнивая правые части (7.36) и (7.37) и учитывая, что из условия синусов (7.31) при $n'=n$ (одинаковая среда по обе стороны системы) следует $\sigma\sin^2 u = \sigma'\sin^2 u'$, получаем $B'=B$ — яркость изображения равна яркости предмета. Совпадение яркостей (при отсутствии потерь) обусловлено тем, что при апланатическом отображении увеличение площади изображения сопровождается уменьшением в то же число раз телесного угла, в котором распространяется световой поток. С помощью оптических приборов невозможно увеличить яркость. А практически неизбежные потери света из-за отражений, рассеяния и поглощения приводят к тому, что яркость создаваемого оптической системой изображения может быть только меньше яркости самого предмета.

Формулу (7.37) можно трактовать и иначе, а именно как поток излучения, приходящего на площадку σ' от воображаемого протяженного источника с поверхностной яркостью B', как бы заполняющего выходной зрачок системы. Но, как было показано выше, $B'=B$, т. е. поверхность этого воображаемого источника излучает с той же яркостью, что и сам предмет. В то же время видимый из σ' угловой размер этого источника, равный выходной апертуре $2u'$, может быть много больше углового размера самого предмета. Это легко продемонстрировать следующим простым опытом. Источник с маленькой светящейся поверхностью помещают в фокус большой линзы или в фокус автомобильной фары. Тогда большая поверхность линзы или отверстие зеркала фары излучает с той же яркостью, что и маленькая поверхность источника. Хотя этот опыт и тривиален, он производит сильное впечатление.

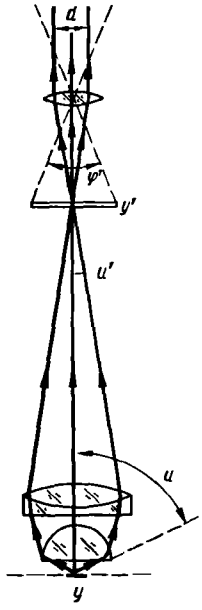
Если в плоскость изображения поместить экран, освещенность площадки σ' может значительно превышать ту освещенность, которую создавал бы источник σ при отсутствии оптической системы. Это легко понять на примере всем знакомого детского «зажигательного» стекла. Чтобы получить непосредственно от Солнца такую же освещенность, как в фокусе линзы, пришлось бы приблизиться к Солнцу настолько, чтобы солнечный диск был виден невооруженным глазом под тем же углом, что и поверхность линзы из ее фокуса.

Освещенность E площадки σ' равна Φ'/σ'. Поэтому из (7.37) с учетом того, что $B'=B$, получаем $E=\pi B\sin^2 u'$. Для систем, формирующих изображения удаленных предметов (зрительные трубы и телескопы, фотообъективы), $\sin u$ в параксиальном прибли-

жении пропорционален отношению диаметра D входного зрачка к фокусному расстоянию f. Отношение D/f называется *относительным отверстием*. Освещенность изображения практически не зависит от расстояния до предмета, а определяется его поверхностной яркостью B и квадратом относительного отверстия $(D/f)^2$, принимаемым за *светосилу* такого объектива. Для характеристики объективов микроскопов часто используют *числовую апертуру* (Ч. А.), которая определяется как произведение синуса половины входной апертуры на показатель преломления среды в пространстве предметов: Ч. А.$=n\sin u$. От нее зависит не только светосила, но и предельная разрешающая способность микроскопа (см. § 7.6).

При визуальных наблюдениях оптический прибор и глаз наблюдателя образуют единую систему, все элементы которой должны быть согласованы друг с другом. Это требование налагает определенные условия на выбор разумного увеличения. Напомним, что увеличением прибора называется отношение углов, под которыми протяженный предмет виден через прибор и при наблюдении невооруженным глазом. Например, для зрительной трубы (см. рис. 7.19) увеличение $\Gamma=w'/w$ равно отношению фокусных расстояний объектива и окуляра: $\Gamma=f_1/f_2$. Уменьшая фокусное расстояние окуляра, можно получить с данным объективом большее увеличение. Однако не всегда следует стремиться только к получению большого увеличения. При наблюдении протяженных предметов малой яркости нужно, чтобы освещенность их изображения, получающегося на сетчатке глаза, была как можно больше. Для этого диаметр выходного зрачка трубы не должен быть меньше входного зрачка глаза, чтобы именно зрачок глаза служил апертурной диафрагмой всей системы. Тогда освещенность изображения на сетчатке будет максимальной — такой же (в пренебрежении потерями света на отражение и рассеяние), как и при наблюдении невооруженным глазом. Диаметр d выходного зрачка трубы с данным объективом диаметра D зависит от увеличения: как видно из рис. 7.19, *а* $D/d=f_1/f_2=\Gamma$. Увеличение называется *нормальным* или *равнозрачковым,* когда диаметр d выходного зрачка прибора равен диаметру d_0 зрачка глаза. При бо́льших увеличениях $d<d_0$ и освещенность изображения уменьшится. При увеличениях, не превосходящих нормального, освещенность изображения максимальна и не зависит от увеличения. Однако применять увеличение меньше нормального нецелесообразно, так как при этом используется только центральная часть объектива: при $d>d_0$ периферийные лучи не попадают в зрачок глаза.

Диаметр d_0 зрачка глаза зависит от условий освещения и изменяется от 6—8 мм в полной темноте до 2 мм при ярком дневном свете. Поэтому при работе с телескопом, имеющим определенный диаметр D объектива, следует учитывать условия наблюдения, влияющие на зрачок глаза, и выбирать окуляр, обеспечивающий оптимальное увеличение $\Gamma=D/d_0$.



7.29

К определению нормального увеличения микроскопа

Увеличением микроскопа называют отношение угла φ', под которым малый предмет виден в микроскоп, к углу φ, под которым он был бы виден невооруженным глазом с расстояния ясного зрения L. Предмет помещают в передней апланатической точке объектива (рис. 7.29), а его действительное изображение получается в сопряженной апланатической точке. Так как она находится на большом расстоянии, апертура выходящих из объектива лучей мала ($u' \ll 1$) и условие синусов можно записать в виде $ny\sin u = y'u'$, где y — размер предмета, y' — его изображения. Окуляр с фокусным расстоянием f располагается как и в зрительной трубе таким образом, чтобы изображение предмета находилось в его фокальной плоскости. Поэтому ширина d выходящего из окуляра пучка лучей, т. е. диаметр выходного зрачка, составляет $2fu' = 2fny\sin u/y'$. Выразим y/y' через увеличение микроскопа $\Gamma = \varphi'/\varphi$, учитывая, что $\varphi = y/L$, $\varphi' = y'/f$. В результате получим $d = 2Ln\sin u/\Gamma$. Для нормального увеличения фокусное расстояние окуляра f следует выбрать так, чтобы диаметр d выходного зрачка был равен диаметру d_0 зрачка глаза. Отсюда получаем следующее выражение для нормального увеличения:

$$\Gamma = 2Ln\sin u/d_0. \qquad (7.38)$$

Для нормального глаза обычно полагают $d_0 = 2$ мм, $L = 250$ мм. Наибольшее достижимое значение числовой апертуры $n\sin u$ у сухих объективов равно 1, у иммерсионных — 1,5. Соответствующие этим случаям нормальные увеличения по (7.38) должны составлять $\Gamma = 250$ и $\Gamma = 375$. В условиях хорошего освещения предмета допустимо применять увеличение, превосходящее нормальное в 2—4 раза. Таким путем нельзя выявить новые детали рассматриваемого предмета, но можно обеспечить менее напряженные условия для глаза, чтобы не заставлять его работать на пределе разрешающей способности. Однако увеличения, значительно превосходящие нормальное, не только бесполезны, но и вредны, так как сужение выходящего пучка может внести в наблюдаемое изображение сильные искажения из-за дифракции. Поэтому предельное увеличение в лучших сухих системах можно принять равным 1000, в иммерсионных — 1500.

Вопрос об оптимальном выборе увеличения оптических приборов рассмотрен в § 7.6 в связи с их предельной разрешающей способностью, обусловленной волновой природой света.

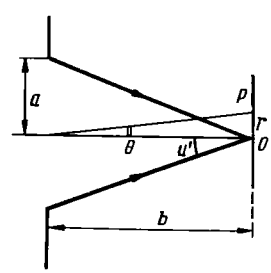
Контрольные вопросы

- ☐ Объектив фотоаппарата создает в фокальной плоскостн действительное изображение удаленного протяженного источнннка света, поверхность которого излучает по закону Ламберта. Как будут изменяться при изменении светосилы объектнва (диаметра диафрагмы) яркость изображения источника и освещенность фотопластинки в том месте, где получается изображение?
- ☐ Объясните, почему (при отсутствии потерь света) яркость, создаваемого оптической системой изображенння нсточника равна яркости самого источника.
- ☐ В каком случае увелнчение прнбора, предназначенного для визуальных наблюдений, называется нормальным?
- ☐ Какнм должен быть диаметр D линз объективов бинокля с десятикратным увеличением, если диаметр зрачка глаза равен $d_0=5$ мм?

7.6. Разрешающая способность оптических инструментов

В оптической системе с исправленными геомоцентрический пучок лучей, расходящихся от точечного источника, превращается в сходящийся гомоцентрический пучок, фокус которого и служит совершенным с точки зрения геометрической оптики изображением источника. Однако на сравнимых с длиной волны расстояниях от фокуса пучка кривизна волновых поверхностей становится значительной и условия применимости геометрической оптики перестают выполняться. Создаваемое системой изображение точечного источника имеет вид некоторой дифракционной картины, т. е. светлого пятнышка конечного размера. Это обстоятельство ограничивает возможность различать на изображении близкие точки предмета. Таким образом, обусловленное волновой природой света и принципиально неустранимое явление дифракции определяет теоретический предел достижимой разрешающей способности оптических приборов, формирующих изображение.

Распределение интенсивности света в плоскости, проходящей через фокус сходящегося пучка, определяется дифракцией Фраунгофера (см. § 6.3). Вид дифракционной картины зависит от формы отверстия диафрагмы, ограничивающей поперечное сечение пучка. Для оптических инструментов наиболее важен случай круглого отверстия (такую форму обычно имеют диафрагмы, оправы линз и объективов), когда дифракционная картина состоит из концентрических светлых и темных колец (см. рис. 6.17, *б*). Радиальное распределение интенсивности в ней выражается по формуле (6.27) через функцию Бесселя первого порядка $J_1(z)$ (см. рис. 6.18) от аргумента $z=2\pi a\theta/\lambda$, где a — радиус выход-



7.30
К нахождению распределения интенсивности вблизи фокуса сходящегося пучка лучей

ного зрачка, а угол θ характеризует направление из его центра на точку наблюдения P (рис. 7.30). Пусть r — расстояние от фокуса O пучка лучей до точки P, а b — до плоскости выходного зрачка. Тогда $\theta \approx r/b$ и $z = 2\pi r a/(\lambda b) = 2\pi r u'/\lambda$, где $u' = a/b$ — половина выходной апертуры. Функция $J_1(z)$ обращается в нуль при значениях аргумента $z_1 = 3{,}83$; $z_2 = 7{,}02$; $z_3 = 10{,}17$; .... Первый минимум интенсивности соответствует z_1, откуда для радиуса r_1 первого темного кольца, окружающего центральный максимум, получаем

$$r_1 = 0{,}61\lambda/u'. \qquad (7.39)$$

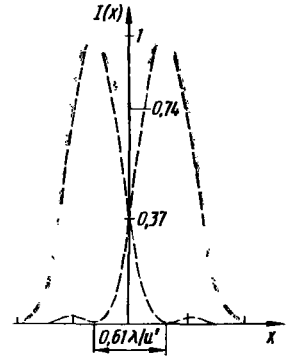
На центральный максимум, называемый *диском Эйри*, приходится 84% светового потока, и его можно считать изображением точечного источника, создаваемым оптической системой. Размер этого изображения [см. (7.39)] определяется выходной апертурой u'.

Протяженный предмет можно рассматривать как совокупность точечных источников, каждый из которых отображается системой в виде диска Эйри с окружающими его дифракционными кольцами. Если соседние точки предмета можно считать некогерентными источниками, то испускаемые ими волны не интерферируют и происходит сложение интенсивностей, т. е. результирующее изображение находится как простое наложение дифракционных картин от отдельных точек. Этот случай реализуется для самосветящихся (или некогерентно освещенных) объектов и важен в теории телескопа. Другой предельный случай когерентно освещенных объектов может быть реализован при наблюдении в микроскоп. Здесь для нахождения изображения требуется сложить напряженности полей в дифракционных картинах от отдельных точек предмета.

Пусть предмет состоит из двух одинаковых точечных источников S_1 и S_2. Если расстояние между центрами их изображений в оптическом приборе мало́ по сравнению с размерами дисков Эйри, то результирующая картина практически не отличается от изображения одного точечного источника. В таком случае говорят, что прибор *не разрешает* рассматриваемые точки. Если увеличивать расстояние между S_1 и S_2, то расстояние между центрами их изображений S_1' и S_2' также будет увеличиваться при неизменном размере соответствующих им дисков Эйри. Начиная с некоторого расстояния $|S_1'S_2'| = l_{\min}$ на графике суммарного распределения интенсивности вдоль линии $S_1'S_2'$ в середине появится провал. Такая картина будет восприниматься как раздельное изображение двух точек. В этом случае говорят, что прибор *разрешает* точки S_1 и S_2, а величину $l_{\min}$ называют разрешаемым расстоянием.

Количественный критерий разрешающей способности, как и в случае спектрального прибора (см. § 6.6), может быть сформулирован только условно, так как возможность отличить дифракционное изображение двух близких точек от изображения одиночного точеч-

ного источника зависит от той точности, с которой производится измерение наблюдаемого распределения интенсивности. Такой условный критерий был предложен Рэлеем: как и в случае узких спектральных линий (см. § 6.6), два точечных некогерентных источника считаются разрешенными, если центр дифракционной картины от одного из них совпадает с ближайшим к центру минимумом картины от другого. Это соответствует расстоянию l_{min} между центрами изображений, равному радиусу диска Эйри (7.39): $l_{min} = 0{,}61\lambda/u'$. Результирующий контур показан на рис. 7.31.



Критерий Рэлея разрешения изображений точечных некогерентных источников

Значение функции $[2J_1(z)/z]^2$ при $z = z_1/2 = 1{,}91$ равно 0,37, поэтому интенсивность в провале составляет 74% от максимальной.

Для телескопа предположение о некогерентности точек предмета всегда оправдано. Такими «точками» могут быть, например, две близкие звезды. Из-за очень большого удаления от Земли звезды можно считать точечными источниками, несмотря на их гигантские линейные размеры. Изображение звезды в фокальной плоскости объектива телескопа никак не отражает реальной формы и размеров звезды, а представляет собой лишь дифракционную картину, создаваемую круглой оправой объектива. Радиус r_1 первого темного кольца в ней равен $0{,}61\lambda/u' = 1{,}22\lambda f/D$, где f — фокусное расстояние объектива, D — его диаметр. Отсюда для минимального углового расстояния θ_{min} между двумя звездами, разрешаемыми телескопом, получаем

$$\theta_{min} = r_1/f = 1{,}22\lambda/D. \qquad (7.40)$$

Величину $1/\theta_{min}$, обратную предельному разрешаемому угловому расстоянию, называют *разрешающей способностью* телескопа. Она пропорциональна действующему диаметру объектива. Этим отчасти объясняется стремление строить большие телескопы. Телескоп с диаметром главного зеркала $D = 5$ м при условии устранения геометрических аберраций может обеспечить угловое разрешение (для света с длиной волны $\lambda = 560$ нм) $\theta_{min} = 1{,}4 \cdot 10^{-7} = 0{,}028''$.

Попытки повышения разрешающей способности телескопов путем сооружения гигантских механических конструкций имеют естественный предел, вытекающий из прочности конструкционных материалов. Этот предел практически уже достигнут. Принципиально новые возможности повышения разрешающей способности связаны с направлением, получившим название *апертурного синтеза* или *адаптивной оптики*, суть которого состоит в построении большой оптической системы из элементов сравнительно небольшого размера.

В случае оптико-механического апертурного синтеза формируется параболическое составное зеркало, большая апертура которого тесно заполняется прилегающими друг к другу малыми зеркалами. Специальная автоматическая система юстировочных механизмов поддерживает взаимное положение сегментов так, чтобы они образовали единую поверхность. Использование таких «активных» систем для телескопов позволяет существенно уменьшить массу зеркала заданного диаметра.

Систему с *разреженной апертурой* образует совокупность малых зеркал, не прилегающих друг к другу. Простейший пример такой системы — звездный интерферометр Майкельсона (см. § 5.5). Наименьшее угловое расстояние, доступное измерению, определяется не диаметром D объектива (или зеркала) телескопа, на котором он смонтирован, а максимальным расстоянием между внешними подвижными зеркалами M_3 и M_4 (см. рис. 5.22), которое может значительно превосходить D. Предельное разрешение разреженной апертуры близко к разрешению такой же по размерам сплошной апертуры. К недостаткам систем с разреженной апертурой следует отнести потери энергии и значительное усложнение формы изображения точечного источника (аппаратной функции), связанное с тем, что по мере «разбавления» апертуры возрастает относительная интенсивность боковых максимумов дифракционной картины. В частности, в предельном случае разрежения апертуры, т. е. в звездном интерферометре, боковые максимумы сравниваются по интенсивности с центральным, образуя систему одинаковых интерференционных полос. Поэтому он пригоден лишь для измерения комплексной степени когерентности излучения и угловых размеров источника, а не для регистрации оптического изображения.

Измеряемое распределение интенсивности в создаваемом прибором изображении некоторого объекта можно представить как свертку *аппаратной функции* (изображения точечного источника) и *функции объекта* (распределения интенсивности, которое создавалось бы идеальным прибором). Чем больше ширина аппаратной функции и чем сложнее ее форма, тем большие искажения вносит прибор в функцию объекта. Однако даже при широкой, но точно известной аппаратной функции путем математической обработки измеряемого распределения можно восстановить вид функции объекта, иначе говоря, произвести редукцию к идеальному прибору. Успех решения этой *обратной задачи* определяется погрешностями при измерениях, т. е. *уровнем шумов*. Анализ показывает, что при наличии шумов прибор с узкой аппаратной функцией обеспечивает лучшее восстановление функции объекта и, следовательно, характеризуется более высокой разрешающей способностью.

Таким образом, реальные возможности получения более или менее детального изображения объекта в значительной степени определяются шириной и формой аппаратной функции прибора. Ее вид в общем случае обусловлен не только дифракцией, но и геомет-

рическими аберрациями, полное устранение которых невозможно. У зеркал больших телескопов аберрации возникают из-за деформаций отражающей поверхности, вызываемых трудноустранимыми механическими напряжениями. Полной реализации разрешающей способности препятствует также наличие земной атмосферы. Возникающие в ней неоднородности сопровождаются локальными изменениями показателя преломления воздуха. Связанные с этим искажения изображения и ухудшение разрешения особенно существенны у больших телескопов, поэтому обсерватории для них стремятся строить в высокогорных районах. По указанным причинам главное преимущество больших телескопов заключается не в высокой разрешающей способности, а в увеличении формирующего изображение светового потока, что позволяет обнаруживать и фотографировать слабые небесные объекты.

При визуальных наблюдениях телескоп и глаз образуют единую систему. Для реализации разрешающей силы объектива требуется согласование всех элементов системы, что достигается выбором окуляра, обеспечивающего оптимальное увеличение телескопа. Остановимся на этом вопросе подробнее.

При рассматривании удаленных предметов глаз действует так же, как объектив телескопа: свет от точечного источника образует на сетчатке дифракционную картину, угловой радиус центрального максимума которой определяется той же формулой (7.40), если заменить в ней D на диаметр d_0 зрачка глаза. Поэтому разрешаемое невооруженным глазом угловое расстояние θ_0 при $\lambda=550$ нм, $d_0=4$ мм составляет $1{,}22\lambda/d_0=1{,}7\cdot10^{-4}$ рад$=0{,}5'$*. Это значение очень близко к остроте зрения нормального глаза, которая определяется расстоянием между соседними чувствительными элементами (колбочками) в центральной части сетчатки, где плотность их размещения наибольшая. Это значит, что, совершенствуясь в процессе эволюции, наш орган зрения фактически достиг максимума, принципиально допустимого законами физики.

Предположим, что угловое расстояние между двумя удаленными точками как раз равно предельному значению $\theta_{min}=1{,}22\lambda/D$ (7.40), которое еще может разрешить объектив телескопа. В телескоп с увеличением Γ эти точки видны под углом $\theta=\Gamma\theta_{min}$. Чтобы точки воспринимались глазом как раздельные, угол θ не должен быть меньше угла $\theta_0=1{,}22\lambda/d_0$, разрешаемого глазом. Отсюда находим $\Gamma\geqslant D/d_0$. Знак равенства здесь соответствует введенному в § 7.5 *нормальному увеличению,* при котором наиболее эффективно используется световой поток, попадающий в объектив телескопа. При увеличениях, меньших нормального, апертурной диафрагмой служит зрачок глаза и используется только центральная часть объек-

* В условиях плохого освещения диаметр d_0 зрачка глаза может увеличиться до 8 мм, но это не приводит к увеличению разрешающей способности из-за геометрических аберраций.

тива, диаметр которой меньше D. Это приводит к уменьшению разрешающей способности всей системы.

Применение увеличений больше нормального лишь снижает освещенность изображения на сетчатке глаза и не улучшает разрешения, так как не может выявить новых деталей рассматриваемого объекта, отсутствующих в первичном изображении, которое представляет собой совокупность дифракционных картин от оправы объектива. Однако в некоторых случаях (см. § 7.5) допустимо применять увеличения, в 2—4 раза превосходящие нормальное, чтобы создать менее напряженные условия для работы глаза.

Угловые размеры почти всех звезд много меньше углового разрешения даже самых больших телескопов. Поэтому, как уже отмечалось выше, изображение звезды в фокальной плоскости объектива неотличимо от изображения точечного источника и представляет собой дифракционный кружок. Диаметр этого кружка настолько мал, что при использовании нормального увеличения он, как и сама звезда, для глаза неотличим от точечного источника. Это значит, что размер дифракционного пятна на сетчатке глаза не зависит от того, наблюдается ли звезда в телескоп или непосредственно. Но световой поток, приходящийся на это дифракционное пятно, и, следовательно, освещенность изображения при наблюдении в телескоп во столько раз больше, чем при наблюдении невооруженным глазом, во сколько раз площадь отверстия объектива больше площади зрачка глаза. В то же время освещенность изображения протяженных предметов (фона), как было показано в § 7.5, не изменяется. Этим объясняется, почему в телескоп звезды на фоне неба видны и днем.

Разрешающую способность микроскопа в случае самосветящихся объектов можно найти тем же способом, что и для телескопа. Предмет расположен в передней апланатической точке объектива (см. рис. 7.30). Получающееся в сопряженной плоскости изображение отдельной точки предмета представляет собой фраунгоферову дифракционную картину, возникающую при дифракции на круглой апертурной диафрагме. По критерию Рэлея, две точки предмета считаются разрешенными, если расстояние между центрами их изображений не меньше радиуса диска Эйри r_1 (7.39): $l' \geqslant 0{,}61\lambda/u'$. Величина l' связана с расстоянием l между точками самого предмета условием синусов (7.31), которое при $u' \ll 1$ (что практически всегда выполняется) и $n'=1$ принимает вид $l n \sin u = l' u'$. Поэтому для минимального расстояния $l_{\min}$ между двумя точками предмета, разрешаемыми микроскопом, получаем

$$l_{\min} = 0{,}61\lambda/(n \sin u). \qquad (7.41)$$

В отличие от телескопа разрешающая способность микроскопа определяется не диаметром объектива, а углом раствора $2u$ конуса исходящих от предмета лучей, улавливаемых объективом. Стоящая в знаменателе формулы (7.41) *числовая апертура* $n \sin u$ должна

б. ть как можно больше для достижения высокого разрешения. В лучших объективах апертура приближается к своему теоретическому пределу $2u=\pi$, а для увеличения числовой апертуры применяют иммерсию (жидкость с высоким показателем преломления n), вводимую между покровным стеклом и передней линзой объектива. Наличие n в формуле (7.41) связано с тем, что разрешаемое расстояние должно определяться длиной волны не в вакууме, а в той среде, где находится предмет, т. е. λ/n. При $n=1{,}5$ максимальное значение числовой апертуры составляет 1,5, что в соответствии с (7.41) дает $l_{\min}=0{,}4\lambda$. Этот предел разрешения обусловлен волновой природой света. Никакие технические усовершенствования микроскопа не позволят его превзойти. Для видимого света ($\lambda\approx$ $\approx 5\cdot10^{-5}$ см) $l_{\min}=2\cdot10^{-5}$ см. Улучшить разрешение можно в ультрафиолетовых лучах, но это требует использования кварцевых объективов и перехода к фотографической регистрации.

Радикальное увеличение разрешающей способности достигается в электронном микроскопе, где вместо световых лучей используются электроны. Соответствующая электронам длина волны де Бройля $\lambda=h/(mv)$ при ускоряющем напряжении 10 кВ равна 10^{-11} м, что уже меньше размеров атома ($\sim10^{-10}$ м). В формирующих изображение системах электронных линз (магнитных и электростатических) из-за больших аберраций используются только узкие параксиальные пучки с малыми апертурами ($\sim0{,}01\div0{,}1$), и все же разрешающая способность электронного микроскопа в сотни раз больше, чем оптического. Это позволяет разрешать детали, всего в несколько раз превосходящие размеры отдельных атомов.

При визуальных наблюдениях для полной реализации разрешающей способности объектива требуется правильный выбор увеличения микроскопа. Минимальное расстояние $l_{\min}$, разрешаемое объективом, с расстояния ясного зрения L видно невооруженным глазом под углом $\varphi=l_{\min}/L$. В микроскоп с увеличением Γ то же расстояние видно под углом $\varphi'=\Gamma\varphi=\Gamma l_{\min}/L$. Угол φ' не должен быть меньше предельного угла $\theta_0=1{,}22\lambda/d_0$, разрешаемого глазом. Используя для $l_{\min}$ выражение (7.41), условие $\varphi'\geqslant\theta_0$ запишем в виде

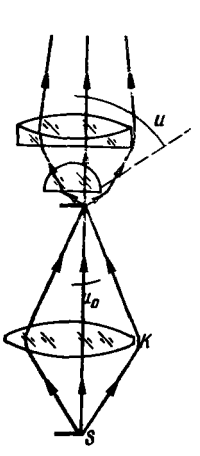
$$\Gamma\geqslant 2Ln\sin u/d_0. \qquad (7.42)$$

Знаку равенства здесь соответствует введенное в § 7.5 нормальное (равнозрачковое) увеличение (7.38), при котором весь проходящий через микроскоп световой поток попадает в глаз. Таким образом, нормальное увеличение совпадает с наименьшим увеличением, при котором полностью используется разрешающая способность объектива. Как и в случае телескопа, применение увеличений, превышающих нормальное, не может выявить новых деталей изображения и лишь снижает освещенность, но, как уже отмечалось, в некоторых случаях оно оказывается целесообразным по причинам физиологического характера.

Если не требовать геометрического подобия между предметом и его изображением, то с помощью микроскопа можно обнаружить

частицы, размер которых меньше предельного l_{min}, устанавливаемого формулой (7.41). Для этого сильный боковой пучок света фокусируют на кювету с жидкостью, в которой во взвешенном состоянии находятся «ультрамикроскопические» частицы (например, коллоидные частицы металлов размером около 5 нм), и наблюдают в микроскоп рассеянный частицами свет на черном фоне. Изображение частицы, как и изображение звезды в телескопе, имеет вид дифракционного кружка от оправы объектива. Этот кружок дает представление о положении и перемещении частицы, но не о ее форме и размерах.

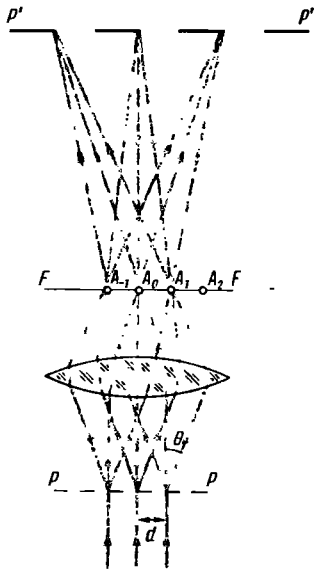
До сих пор при обсуждении условий разрешения предполагалось, что две точки предмета S_1 и S_2 представляют собой некогерентные точечные источники, и в плоскости создаваемого оптической системой изображения происходит простое наложение дифракционных картин от каждого из них. Несамосветящийся объект должен быть освещен каким-либо источником света. Если этот источник точечный, то световые колебания в точках S_1 и S_2 освещаемого им предмета когерентны. Любой реальный источник имеет конечные размеры, поэтому в общем случае световые колебания в близких точках S_1 и S_2 освещаемого предмета будут частично когерентны. Степень пространственной когерентности γ_{12} световых колебаний в точках S_1 и S_2 зависит от расстояния l между ними и от угловых размеров источника света (см. § 5.5). Когда применяется оптическая осветительная система (конденсор), отображающая светящуюся поверхность источника на плоскость объекта (рис. 7.32), роль углового размера источника играет выходная апертура $2u_0$ осветителя: в пределах центрального максимума дифракционной картины от его оправы световые колебания частично когерентны, ибо каждая точка источника отображается конденсором в виде кружка конечных размеров. Радиус этого кружка, т. е. размер области когерентности, порядка λ/u_0. Если апертура осветителя мала по сравнению с числовой апертурой объектива микроскопа, то расстояние l_{min} между точками S_1 и S_2, лежащими на пределе разрешения, много меньше ширины дифракционного кружка от оправы конденсора и световые колебания в S_1 и S_2 можно считать полностью когерентными.



7.32
К вопросу о когерентности освещения в микроскопе

Теория разрешающей способности микроскопа для предельного случая полностью когерентного освещения была развита Эрнстом Аббе. Теория Аббе дает наглядное представление о дифракционном характере формирования изображения в оптической системе. Чтобы облегчить учет интерференции лучей от разных мест протяженного объекта, освещаемого когерентным све-

том, будем в качестве объекта рассматривать простую щелевую решетку (рис. 7.33). Свет от точечного источника, помещенного в фокус конденсора, образует плоскую волну, которая испытывает дифракцию на решетке-объекте PP. В фокальной плоскости FF объектива возникает фраунгоферова дифракционная картина, положение главных максимумов которой определяется периодом d решетки: в случае нормального падения $d \sin \theta_m = m\lambda$. На оптической оси находится максимум нулевого порядка ($m = 0$), по обе стороны от него на расстоянии $f \operatorname{tg} \theta_1$ расположены максимумы первого порядка A_1 и A_{-1} и т. д. Минуя эти точки, когерентные лучи встречаются в сопряженной с PP плоскости $P'P'$, где интерферируют между собой и образуют более или менее верное изображение решетки-объекта.



К дифракционной теории образования изображения в микроскопе

Если с помощью диафрагмы, помещенной в фокальную плоскость FF объектива, перекрыть часть главных максимумов A_m, то изображение решетки в плоскости $P'P'$ ухудшается, так как оно соответствует решетке, у которой при дифракции эти максимумы не возникают. Когда диафрагма пропускает только один пучок A_0, достигающий плоскости $P'P'$ свет будет состоять только из одной волны, как было бы при замене решетки PP прозрачной плоскопараллельной пластинкой. Никакой информации о структуре решетки эта волна не несет и в плоскости изображения $P'P'$ будет наблюдаться равномерно освещенное поле. Когда диафрагма в плоскости FF открывает кроме главного максимума A_0 еще и максимумы первого порядка A_1 и A_{-1}, интерферирующие волны создадут в плоскости $P'P'$ изображение решетки с плавным переходом от прозрачных к непрозрачным участкам. Легко понять, что точно такая картина наблюдалась бы в случае, когда на месте щелевой решетки PP находилась бы решетка того же периода d, но с синусоидальным пропусканием. Как было показано в § 6.5 (задача 3), при дифракции на такой решетке возникают главные максимумы только порядков $m = 0, \pm 1$.

Пропускание щелевой решетки характеризуется ступенчатым профилем. Функцию пропускания можно разложить в ряд Фурье, т. е. представить решетку как наложение синусоидальных решеток с периодами d, $d/2$, $d/3$ и т. д., каждая из которых дает только по одному главному максимуму справа и слева от A_0. Главные максимумы порядка $\pm m$ при дифракции на щелевой решетке обусловлены соответствующими фурье-компонентами ее функции про-

волна оказывается промодулированной в плоскости xy объекта по амплитуде и фазе, так что ее комплексная амплитуда дается некоторой функцией $E(x, y)$. Поле световой волны в плоскости xy можно разложить в двухмерный интеграл Фурье, т. е. представить в виде суперпозиции бесконечного числа плоских дифрагировавших волн. Комплексная амплитуда дифрагировавшей волны с волновым вектором $\mathbf{k}$ пропорциональна соответствующей фурье-компоненте (пространственной гармонике) $E(k_x, k_y)$ функции $E(x, y)$:

$$E(k_x, k_y)=\frac{1}{(2\pi)^2}\iint E(x, y)\,e^{-i(k_x x+k_y y)}\,dx\,dy. \qquad (7.44)$$

Каждой точке фокальной плоскости FF объектива (рис. 7.33) соответствует вполне определенное направление (θ, φ) дифрагировавшей волны: $k_x=k\sin\theta\cos\varphi$, $k_y=k\sin\theta\sin\varphi$. Поэтому распределение амплитуды света в фокальной плоскости представляет собой фурье-образ (7.44) функции $E(x, y)$. В оптической системе с исправленными аберрациями оптические пути между всеми парами сопряженных точек одинаковы. Поэтому при интерференции волн в некоторой точке плоскости $P'P'$ изображения их сложение происходит с теми же относительными фазами, какие они имели в соответствующей точке (x, y) плоскости PP объекта. Распределение амплитуды света $E(x', y')$ в плоскости изображения $P'P'$ описывается (с точностью до масштаба) таким же интегралом Фурье, что и в плоскости x, y:

$$E(x', y')=\iint E(k_x, k_y)\,e^{i(k_x x'+k_y y')}\,dk_x\,dk_y, \qquad (7.45)$$

где интегрирование по k_x и k_y, т. е. по направлениям θ и φ дифрагировавших волн, производится в пределах, определяемых апертурой оптической системы. Так как в (7.45) входят только те фурье-компоненты функции $E(x, y)$ объекта, которым соответствуют прошедшие через оптическую систему дифрагировавшие волны, то распределение амплитуды света $E(x', y')$ в плоскости изображения не дает точного воспроизведения функции объекта. Пространственные гармоники, соответствующие мелким деталям объекта, отсутствуют в формируемом системой изображении. Поэтому наименьший размер различимых деталей объекта зависит от апертуры оптической системы.

Если в фокальной плоскости FF объектива поместить экран с небольшим отверстием в центре, то он пропустит только дифрагировавший свет от крупных деталей объекта. Если же, напротив, экранировать центральный участок фокальной плоскости, то будет проходить свет, соответствующий пространственным гармоникам сравнительно малого периода, что увеличит контраст мелких деталей на изображении. Помещая в фокальную плоскость специальные фильтры, можно изменять относительные амплитуды и фазы разных фурье-компонент и тем самым влиять на характер изображения (подобно тому, как это делается в методе фазового контрас-

та). Такой метод *фильтрации пространственных гармоник* применяется в оптических системах обработки информации.

Обычно для передачи сообщений (например, по радио или телевидению) используют несущую волну высокой частоты, которую *модулируют во времени* по амплитуде или фазе существенно более низкими частотами (например, звуковыми частотами при передаче речи или музыки). В спектре модулированной волны по обе стороны от несущей частоты возникают «боковые полосы», которые и содержат весь объем передаваемой информации. Для получения информации приемник должен по возможности полно воспринять их.

Оптическое изображение также можно интерпретировать как передачу информации. В этом случае отображаемый объект осуществляет *пространственную модуляцию* световой волны, вызывая появление дифрагировавших волн. Эти отклоненные на разные углы волны, подобно боковым полосам при временно́й модуляции, несут информацию о структуре объекта, о его пространственных гармониках. Информация передается тем точнее, чем полноценнее используется частотный спектр при временно́й модуляции и угловой спектр — при пространственной.

Контрольные вопросы

- Какие величины, характеризующие оптическую систему, определяют размер создаваемого ею дифракционного изображения точечного источника (диска Эйри)?
- Когда два одинаковых точечных источника считаются, по Рэлею, разрешенными на изображении? В чем заключается условность этого критерия?
- Чему равно минимальное угловое расстояние между одинаковыми звездами, разрешаемое объективом телескопа с диаметром D? Почему оно не зависит от фокусного расстояния объектива?
- Какими преимуществами и недостатками обладают системы с разреженной апертурой, подобные звездному интерферометру?
- Как нужно выбирать окуляр оптической системы для визуальных наблюдений, чтобы реализовать разрешающую способность ее объектива?
- Почему в телескоп звезды видны днем, когда их наблюдение невооруженным глазом невозможно?
- С какой целью в сильных микроскопах применяют иммерсию?
- Как теория Аббе объясняет зависимость разрешающей способности микроскопа от числовой апертуры объектива при когерентном освещении?
- Можно ли с помощью оптического микроскопа обнаружить частицы, размер которых меньше длины световой волны?
- Объясните идею метода фазового контраста и метода фильтрации пространственных гармоник.

7.7. Физические принципы голографии

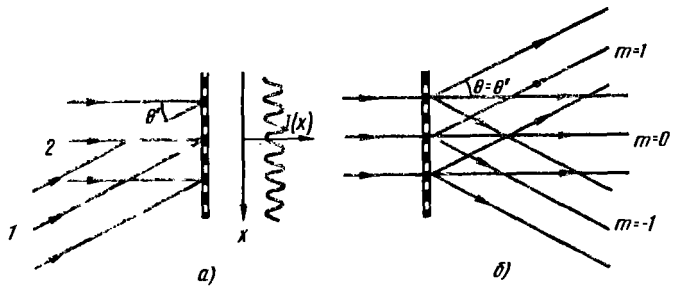
Голографией называют метод записи и последующего восстановления структуры световых волн, основанный на явлениях интерференции и дифракции когерентных пучков света. Как и фотография, она обеспечивает возможность записи, хранения и воспроизведения зрительных образов предметов. Однако обычная фотография дает

лишь плоское изображение объемной картины, видимое из определенной точки. Рассматривая фотоснимок, невозможно заглянуть за предметы, находящиеся на переднем плане. В отличие от фотографии голография позволяет записать и восстановить не двухмерное распределение освещенности в плоскости снимка, а рассеянные предметом световые волны со всеми их характеристиками — направлением распространения, амплитудой, фазой, длиной волны. Восстановленные голограммой световые волны создают полную иллюзию реальности наблюдаемых предметов. Голограмма представляется наблюдателю окном, сквозь которое видно снятую объемную сцену во всей ее глубине. Близкие и далекие предметы видны одинаково четко. Изменяя точку зрения, можно видеть предметы в разных ракурсах.

На голограмме регистрируется не оптическое изображение объекта, а *интерференционная картина,* возникающая при наложении световой волны, рассеянной объектом, и когерентной с ней опорной волны. Эта интерференционная картина фиксирует информацию о распределении амплитуд и фаз в предметной волне. Освещение голограммы восстанавливающей волной, идентичной с той, что служила опорной при регистрации, вызывает появление дифрагировавших волн, одна из которых представляет собой более или менее точную копию волны, рассеянной предметом. Попадая в глаз наблюдателя, она создает такие же ощущения, как и при непосредственном рассматривании предмета.

Идеи, лежащие в основе голографической записи и восстановления зрительной информации, были высказаны и продемонстрированы на опыте Габором (1947—1948). Для практической реализации голографии необходимы источники света с высокой пространственной и временной когерентностью. Поэтому широкое распространение она получила после создания лазеров, начиная с работ Лейта и Упатниекса (1963) и Ю. Н. Денисюка (1962—1963), предложившего записывать голограммы на толстослойных фотоэмульсиях, что позволяет восстанавливать изображение в белом свете.

Уяснить принцип голографии легче всего, рассматривая простейшие объекты. Наиболее прост для понимания случай голографической записи и восстановления плоской волны. Пусть такая волна *1,* исходящая от предмета, падает на фотопластинку под углом θ' к нормали (рис. 7.34, *а*). Мгновенное распределение *фаз* световых колебаний на поверхности пластинки зависит от направления волны, но светочувствительный слой способен зарегистрировать лишь среднее за время экспозиции распределение *освещенности.* В результате пластинка окажется равномерно почерневшей. По степени почернения можно судить об амплитудах световых колебаний, но информация об их фазах полностью теряется. Определить направление воздействовавшей на фотопластинку волны *1* таким способом невозможно.



7.34

Получение (*а*) и восстановление (*б*) голограммы плоской волны

Теперь представим себе, что на ту же фотопластинку одновременно с *предметной волной 1* падает по нормали когерентная с ней плоская волна *2* (*опорная,* или *референтная, волна*). В результате их интерференции на поверхности фотопластинки устанавливается стационарное распределение освещенности в виде системы параллельных эквидистантных полос, ориентированных перпендикулярно плоскости чертежа на рис. 7.34, *а*. Зависимость интенсивности от координаты x (см. § 5.1) выражается формулой

$$I(x)=I_1+I_2+2\sqrt{I_1 I_2}\cos k\Delta(x), \qquad (7.46)$$

где I_1 и I_2 — интенсивности предметной и опорной волн, $k=2\pi/\lambda$ — волновое число, а $\Delta(x)=x\sin\theta'$ — разность их хода (точка $x=0$ выбрана там, где $\Delta=0$). Расстояние между соседними полосами, как следует из (7.46), равно $d=2\pi/(k\sin\theta')=\lambda/\sin\theta'$. При правильном выборе экспозиции и режима обработки пластинки ее пропускание будет зависеть от x по тому же закону (7.46), что и $I(x)$, т. е. мы получим на голограмме дифракционную решетку с синусоидальным пропусканием. Структура зарегистрированных на голограмме интерференционных полос содержит информацию о распределении фаз световых колебаний в предметной волне.

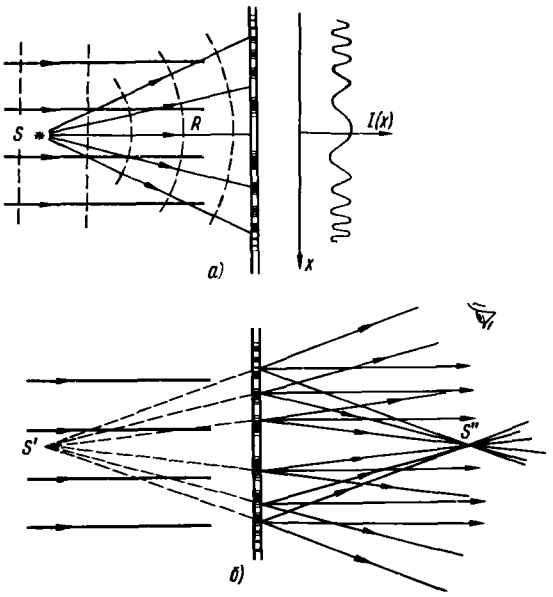
Легко понять, как с помощью такой голограммы можно воспроизвести записанную на ней предметную волну. Направим на голограмму восстанавливающую волну, идентичную с опорной волной *2*, использовавшейся при записи (рис. 7.34, *б*). В результате ее дифракции на решетке с синусоидальным пропусканием возникают (см. § 6.5) три плоские волны: одна из них соответствует главному максимуму порядка $m=0$ и распространяется в направлении падающей волны, две другие — главным максимумам порядка $m=\pm 1$. Наибольший интерес для голографии представляет волна с $m=1$. Ее направление θ определяется условием $d\sin\theta=\lambda$. Так как $d=\lambda/\sin\theta'$, то $\theta=\theta'$, т. е. направление этой волны (как и все остальные характеристики) точно такое же, как у предметной волны *1* при записи голограммы. Попадая в глаз наблюдателя при рассматривании голограммы, восстановленная волна вызывает при

отсутствии предмета такие же ощущения, какие вызвала бы предметная волна при непосредственном его наблюдении.

При голографировании сложного объекта его освещают когерентным лазерным пучком. Рассеянное объектом волновое поле можно в соответствии с теоремой Фурье представить в виде совокупности плоских волн. Каждая из них при интерференции с опорной волной, получаемой из того же лазерного пучка, создает на фотопластинке свою систему интерференционных полос с характерными для нее ориентацией и периодом. После проявления на голограмме образуется совокупность дифракционных решеток с синусоидальным пропусканием. Каждая из этих решеток на этапе восстановления при дифракции пучка, идентичного с опорным, формирует соответствующую ей исходную элементарную плоскую волну. Это главный дифракционный максимум с $m=1$. Все восстановленные элементарные волны находятся в таких же амплитудных и фазовых соотношениях, как и при записи голограммы. Их совокупность воссоздает полное рассеянное объектом световое поле и вызывает те же зрительные образы, что и при непосредственном наблюдении объекта. Другими словами, в том месте, где находился объект при записи голограммы, возникает его мнимое изображение. Кроме того, каждая элементарная система дифракционных полос (решетка) формирует еще две волны, соответствующие главным максимумам с $m=0$ и $m=-1$. Волны с $m=0$ распространяются в направлении опорной волны и не попадают в глаз наблюдателя при надлежащем его расположении. Волны с $m=-1$ формируют, как показано ниже, еще одно (действительное) изображение объекта.

Кроме совокупности дифракционных решеток, возникших при интерференции элементарных предметных волн с опорной, голограмма содержит дополнительную структуру, возникающую из-за интерференции элементарных предметных волн между собой. Обычно опорная волна значительно интенсивнее предметных, так что эта структура выражена слабо. При восстановлении она приводит к образованию дополнительных дифрагировавших волн, концентрирующихся вблизи направления опорной волны. Они не мешают наблюдению восстановленного голограммой мнимого изображения объекта, если угол падения опорной волны в достаточной степени отличается от углов падения предметных волн.

К объяснению принципа голографии можно подойти и иначе, рассматривая процесс записи и восстановления сферических волн, рассеиваемых отдельными точками объекта. Интерференция сферической волны, исходящей из точки S (рис. 7.35, a), и когерентной с ней плоской опорной волны, падающей по нормали на фотопластинку, приводит к образованию стационарной картины в виде концентрических колец. Радиальное распределение интенсивности в интерференционной картине опять дается формулой (7.46), но разность хода $\Delta(x)$ между плоской опорной волной и сферической

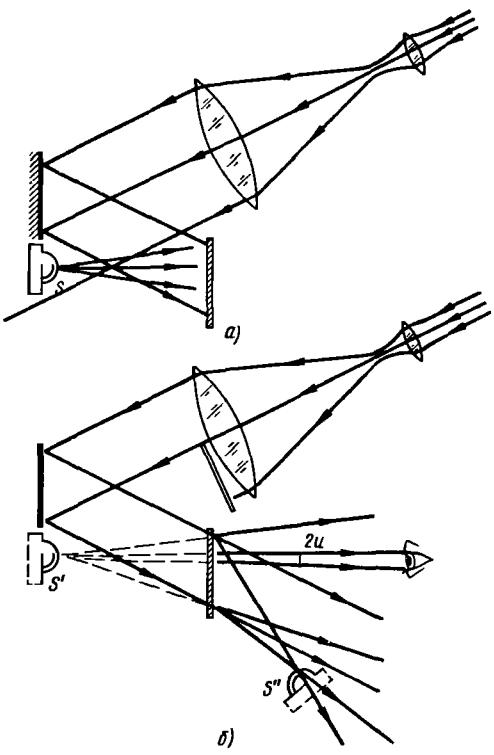


7.35

Запись (*а*) и восстановленне (*б*) голограммы точечного нсточннка

предметной зависит от x по закону $\Delta(x)=(R^2+x^2)^{1/2}-R\approx x^2/(2R)$ (если в точке $x=0$ разность хода $\Delta=0$). Радиус x_n n-го светлого кольца находится из условия $\Delta(x_n)=n\lambda$, откуда $x_n=\sqrt{2Rn\lambda}$. Как раз такими должны быть радиусы колец в зонной пластинке Френеля (см. § 6.1) с фокусным расстоянием $f=R$. Отличие голограммы точечного источника от зонной пластинки, показанной на рис. 6.5, состоит только в том, что переход от светлых колец к темным происходит в соответствии с (7.46) плавно, по синусоидальному закону.

На стадии восстановления полученную голограмму освещают плоской волной, идентичной с опорной (рис. 7.35, *б*). Как и в случае зонной пластинки, в результате дифракции возникают кроме проходящей прямо волны сходящаяся и расходящаяся сферические волны. Из-за плавного перехода от светлых колец к темным здесь образуются, как и у дифракционной решетки с синусоидальным пропусканием, только три главных максимума с $m=0, \pm 1$. Центр расходящейся дифрагировавшей волны S' расположен как раз в том месте, где находился точечный источник S при записи голограммы. В самом деле, когда продолжения дифрагировавших лучей пересекаются в S', разность хода между лучами от соседних светлых колец голограммы равна длине волны λ. Это эквивалентно отсутствию разности хода вообще, и такие лучи будут восприниматься глазом как выходящие из точки S'. Они образуют мнимое изображение источника. Наблюдатель видит сквозь голограмму находящийся в S' точечный источник, хотя никакого источника там нет. Возникающая при дифракции восстанавливающего пучка



7.36
Схема получения (*а*) и восстановления (*б*) голограммы

на голограмме расходящаяся сферическая волна создает такой же зрительный образ, как и исходная предметная волна. Сходящаяся дифрагировавшая волна образует в точке S'' действительное изображение точечного источника. Так как здесь пересекаются сами дифрагировавшие лучи, а не их продолжения, действительное изображение можно получить на помещенном в это место экране в виде светлого дифракционного кружка. Оно формируется голограммой без помощи линз или зеркал.

В схеме на рис. 7.35, *б* для наблюдения мнимого изображения S'' глаз должен располагаться выше или ниже голограммы, так как в противном случае прошедшая прямо или сходящаяся волна создают сильные помехи. Чтобы избежать этого, используют угловое разделение прошедшей и дифрагировавшей волн при наклонном падении опорной волны. Возможная схема записи и восстановления приведена на рис. 7.36, *а, б*. Параллельный пучок лазерного излучения, предварительно расширенный простой оптической системой, направляется на объект и расположенное рядом с ним зеркало. Отраженная зеркалом волна когерентна с волнами, рассеянными предметом, и используется в качестве опорной. При восстановлении прошедшая прямо волна и дифрагировавшие волны, образующие мнимое S' и действительное S'' изображения, оказываются хорошо разделенными в пространстве, что позволяет без помех наблюдать мнимое изображение.

Сложный объект, рассеивающий когерентный лазерный свет, можно рассматривать как совокупность точечных источников, испускающих сферические волны. В результате их интерференции с опорной волной на голограмме возникает сложный узор в виде наложения элементарных зонных решеток. На этапе восстановления при дифракции опорной волны каждая такая решетка формирует волну, расходящуюся от центра, где находилась соответствующая рассеивающая свет точка предмета при записи голограммы. Совокуп-

ность этих волн, создающих мнимое изображение всего объекта, неотличима от волн, рассеянных объектом при записи, так как характеризуется тем же распределением амплитуд и фаз. Поэтому голограмма полностью восстанавливает его объемную структуру и передает не только видимое пространственное расположение предметов, но и эффект параллакса, заключающийся в изменении этого расположения при перемещении точки наблюдения.

Совокупность сходящихся волн, возникающих при дифракции восстанавливающей волны на голограмме, без всяких линз формирует действительное изображение предмета. Увидеть его можно лишь при определенном положении глаза, когда изображение находится между голограммой и глазом. При этом изображение предмета рассматривается с задней стороны и выпуклые места предмета выглядят на нем как вогнутые, и наоборот, т. е. изображение будет как бы вывернутым (псевдоскопическим).

Важное свойство голограмм заключается в том, что восстановить предметную волну можно с помощью небольшого участка голограммы. В случае голограммы плоской волны это свойство очевидно: если закрыть часть дифракционной решетки, то направление дифрагировавших волн останется прежним. При этом лишь уменьшится их интенсивность и несколько увеличится ширина главных максимумов. Этот вывод справедлив и для зонной пластинки, небольшую часть которой можно рассматривать как дифракционную решетку с искривленными штрихами и постепенно изменяющимся периодом. Изменение шага решетки приводит к локальным изменениям направления дифрагировавших пучков, так что любой участок зонной пластинки восстанавливает часть одной и той же сферической волны. В отличие от обычной фотографии, где информация о какой-либо точке предмета фиксируется в одной определенной точке, каждый участок голограммы содержит в закодированной форме информацию о всех точках предмета, так как при ее записи свет, рассеянный каждой точкой предмета, обычно падает на всю поверхность фотопластинки. По мере уменьшения размеров голограммы лишь ухудшается разрешающая способность и сужается поле зрения.

Наблюдение мнимого изображения фактически эквивалентно рассматриванию самого предмета через отверстие, совпадающее с рабочей частью голограммы. При фиксированном положении глаза используется только часть дифрагировавшего излучения, ограниченная действующим конусом лучей, попадающих в глаз. Ясно, что при рассматривании определенной точки предмета в этом конусе распространяется свет, претерпевший дифракцию на небольшом участке голограммы (рис. 7.36). Если изменить положение глаза, изображение той же точки будет восстанавливаться другим участком голограммы.

Создаваемое голограммой изображение (мнимое или действительное) точки предмета, как и в любой другой оптической системе, имеет вид дифракционного пятна, размеры и форма которого опре-

деляются раствором (угловой апертурой) пучка дифрагировавшего на голограмме света, формирующего изображение точки. Минимальное расстояние l_{min} между близкими точками предмета, разрешимыми на изображении, как и для случая когерентного освещения в микроскопе, дается выражением (7.43): $l_{min}=\lambda/\sin u$ (или $l_{min}=\lambda/(2\sin u)$ при наклонном падении опорной волны). Здесь $2u$ — угол, под которым из места расположения точек предмета видна действующая часть голограммы. При визуальном наблюдении мнимого изображения глаз пропускает лишь часть восстановленной голограммой волны (рис. 7.36), действующая часть голограммы (и апертура $2u$) ограничена не размером голограммы, а конусом попадающих в глаз дифрагировавших лучей. В этом случае предельное разрешение определяется глазом, т. е. разрешающая способность голограммы полностью не используется.

До сих пор предполагалось, что светочувствительный слой, регистрирующий интерференционную картину, полностью передает все ее детали. Для этого требуется фотографическая эмульсия высокого качества. Если опорная волна падает на пластинку под углом θ, а предметная — приблизительно по нормали, то соседние интерференционные полосы проходят на расстоянии $d=\lambda/\sin\theta$. Чтобы при восстановлении прямая волна ($m=0$) не мешала наблюдению мнимого изображения, угол θ должен быть достаточно большим. Уже при $\theta\approx 20°$ расстояние между полосами $d\approx 2$ мкм, т. е. 500 полос на 1 мм, а при встречном направлении опорной и предметной волн (метод Ю. Н. Денисюка, см. ниже) $d=\lambda/2$, что соответствует примерно 3000 полос на 1 мм. Применяемые в обычной фотографии фотоматериалы позволяют зарегистрировать около 100 штрихов на 1 мм. Чтобы размер зерна не ограничивал разрешающей способности голограмм, специально для голографии разработаны мелкозернистые фотоэмульсии, регистрирующие несколько тысяч линий на 1 мм.

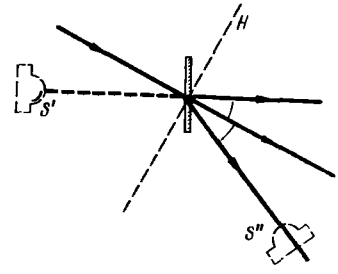
Для восстановления изображения в равной мере пригодны как позитив, так и негатив голограммы. В случае зонной пластинки это очевидно: действие ее одинаково как при четных, так и при нечетных открытых зонах Френеля. В общем случае это следует из теоремы Бабине, согласно которой дополнительные экраны создают одинаковые дифракционные картины в тех местах, куда не попадает прямая волна (восстанавливающий пучок). Важно только, чтобы амплитудное пропускание голограммы линейно зависело от освещенности зарегистрированной на ней интерференционной картины. Тогда при записи плоской волны получается дифракционная решетка с синусоидальным пропусканием, которая даст при восстановлении главные максимумы только порядков $m=0, \pm 1$. В противном случае функция пропускания будет иметь пространственные гармоники более высоких порядков $m=\pm 2, \pm 3, \ldots$, которые приведут на стадии восстановления к возникновению соответствующих главных дифракционных максимумов, т. е.

дополнительных мнимых и действительных изображений предмета. В голографии они играют роль помех.

Наряду с рассмотренными выше *амплитудными* голограммами применяют и прозрачные *фазовые* голограммы, в которых интерференционный узор зафиксирован в изменениях не пропускания, а показателя преломления (оптической толщины) регистрирующего слоя. Это достигается соответствующей химической обработкой (отбеливанием) светочувствительного материала. Такая голограмма приводит к пространственной модуляции фазы восстанавливающей волны, из-за чего возникают дифрагировавшие волны. Они создают, как и в случае амплитудной голограммы, мнимое и действительное изображения предмета.

Опорная волна при записи голограммы должна быть когерентна со светом, рассеянным всеми точками объекта. Для получения голограммы большого объекта необходимо излучение с высокой степенью временно́й и пространственной когерентности. Длина когерентности должна превосходить максимальную разность хода между опорной и предметными волнами, которая для трехмериого объекта практически совпадает с его размерами. Размеры области пространственной когерентности должны быть больше размеров голограммы. Одновременное выполнение этих условий возможно только при использовании лазерного излучения. Для получения четкой интерференционной картины при записи голограммы необходимо также обеспечить во время экспозиции неподвижность всех элементов с точностью до долей длины волны.

Когда пучок света при восстановлении идентичен опорному пучку, использованному при записи, мнимое изображение расположено там же, где находился предмет, и полностью ему подобно. Но в стадии восстановления можно использовать свет с другим направлением распространения. Если восстанавливающий пучок повернуть на небольшой угол, направления пучков, образующих действительное S' и мнимое S'' изображения, повернутся на такой же угол. Изображения S' и S'' расположены симметрично относительно плоскости H, перпендикулярной направлению восстанавливающего пучка (рис. 7.37).

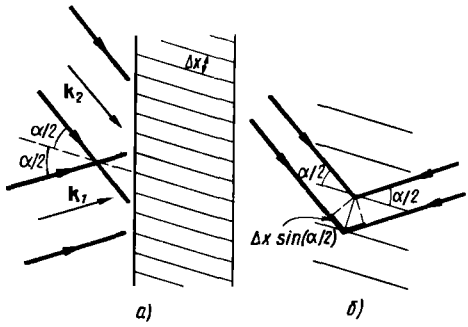


7.37
Расположение изображений объекта при восстаиовлении голограммы

На стадии восстановления можно использовать монохроматический свет с другой длиной волны, нежели при записи. Тогда линейный размер восстановленного изображения отличается от размера предмета и оно будет находиться от голограммы на ином расстоянии. В качестве опорного и восстанавливающего излучений можно использовать не только плоские, но и сферические волны. Однако во всех

случаях, когда восстанавливающая волна не идентична опорной волне, использованной при записи, пучки дифрагировавших на голограмме лучей, формирующие изображения отдельных точек предмета, перестают быть гомоцентрическими. Восстановленному изображению в большей или меньшей степени присущи сферическая аберрация, кома, астигматизм, дисторсия и хроматизм, аналогичные соответствующим аберрациям оптических систем (см. § 7.5).

При изложении предполагалось, что толщина светочувствительного слоя много меньше пространственного периода регистрируемой на голограмме интерференционной картины. Такие голограммы называют *плоскими* (двухмерными). Если толщина слоя регистрирующей среды существенно превосходит этот период, говорят об *объемной* (трехмерной) записи. Для уяснения особенностей толстослойных голограмм рассмотрим простейший случай интерференции плоских опорной и предметной волн с волновыми векторами $\mathbf{k}_2$ и $\mathbf{k}_1$. Как было показано в § 5.1, поверхности равной интенсивности представляют собой эквидистантные плоскости $\mathbf{Kr}=\text{const}$, перпендикулярные вектору $\mathbf{K}=\mathbf{k}_2-\mathbf{k}_1$ (см. рис. 5.2) и отстоящие на расстоянии $\Delta x=\lambda/[2\sin(\alpha/2)]$, где α — угол между $\mathbf{k}_1$ и $\mathbf{k}_2$ [см. (5.5)]. После проявления в толще фотоэмульсии на месте максимумов интенсивности образуется система полупрозрачных зеркальных плоскостей, расположенных параллельно биссектрисе угла между направлениями опорной и предметной волн (рис. 7.38, *а*). Будем для простоты пренебрегать преломлением света в материале эмульсии, считая расстояние между плоскостями равным Δx. Пусть на голограмму падает плоская волна с длиной волны λ' (рис. 7.38, *б*), направление которой, как и у опорной, составляет угол $\alpha/2$ с зеркальными плоскостями. Легко видеть, что разность хода волн, отраженных соседними плоскостями под таким же углом $\alpha/2$, составляет $2\Delta x\sin(\alpha/2)$. Если подставить сюда $\Delta x=\lambda/[2\sin(\alpha/2)]$, то получим, что разность хода равна λ, т. е. длине волны излучения, использованного при записи голограммы. Чтобы эти волны при многолучевой интерференции усиливали друг друга, должно выполняться условие $\lambda'=\lambda$. Если $\lambda'\neq\lambda$, то отраженные разными плоскостями волны имеют всевозможные разности фаз и не дают конструктивной интерференции. При освещении пучком белого света в восстановлении изображения участвуют только волны с $\lambda'=\lambda$. Таким образом, на стадии восстановления толстослойная голограм-

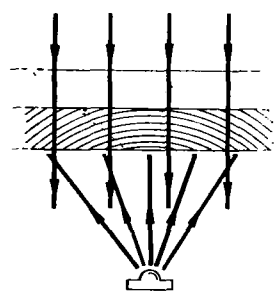


7.38
К объяснению свойств толстослойных голограмм

ма действует как *интерференционный фильтр*, выделяя из падающего света излучение с той длиной волны, которая использовалась при записи. Конечно, такой фильтр характеризуется ограниченным спектральным разрешением, и выделяемое им излучение не столь монохроматично, как при записи. Чем больше отражающих слоев содержит объемная голограмма, тем выше степень монохроматичности восстановленного пучка.

Если восстановление производится монохроматическим светом с той же длиной волны, что и при записи, то отраженные зеркальными слоями волны лишь тогда будут находиться в фазе и при интерференции усилят друг друга, когда направление восстанавливающего пучка совпадает с опорным. Голограмма действует как *оптический коллиматор*. Отраженные волны, как видно из рис. 7.38, *б*, имеют при этом то же направление, что и предметная волна. Поэтому толстослойная голограмма восстанавливает лишь одно (мнимое) изображение предмета. Для получения действительного изображения восстанавливающий пучок должен иметь направление, противоположное опорному, т. е. должен освещать голограмму с обратной стороны. При этом восстанавливается волна с такой же формой волновых поверхностей, что и предметная, но с противоположным направлением распространения. Иначе можно сказать, что в процессе восстановления реализуется *обращение волнового фронта* предметной волны. Изображение получается в том же месте, где находился предмет.

Особенности объемной голограммы как интерференционного фильтра проявляются наиболее сильно, когда отражающие поверхности располагаются почти параллельно границам эмульсионного слоя, т. е. когда опорная и предметные волны распространяются почти навстречу. Такое расположение (рис. 7.39) используется в методе, предложенном Ю. Н. Денисюком. Плоская опорная волна лазерного излучения падает на фотопластинку со стороны стекла и, проходя через фотослой толщиной 15—20 мкм, освещает предмет. Рассеянные предметом волны распространяются почти навстречу опорной волне и при интерференции с ней образуют в толще фотоэмульсии системы из нескольких десятков параллельных полупрозрачных отражающих слоев. При освещении белым светом такая голограмма восстанавливает только одно изображение. Из-за усадки фотоэмульсии в процессе химической обработки пространственный период регистрируемой трехмерной интерференционной структуры уменьшается и цвет восстановленного изображения отличается от опорного лазерного излучения сдвигом в сторону синего конца спектра. При изменении направления восстанавливающего пучка цвет изображения изменяется.



7.39
Схема получения толстослойной голограммы по методу Ю. Н. Денисюка

Изображение в натуральных цветах можно получить, если на одной объемной голограмме зарегистрировать интерференционные картины при освещении предмета излучением, имеющим в своем спектре три монохроматические линии (красную, зеленую и синюю), которые вместе вызывают ощущение белого света. При восстановлении с помощью источника белого света возникают три совмещенных изображения предмета в трех спектральных цветах, воспринимаемые глазом как одно объемное изображение, передающее натуральную окраску предмета.

Из многочисленных практических применений голографии отметим прежде всего *голографическую интерферометрию,* позволяющую наблюдать интерференцию волн, зарегистрированных в разные моменты времени. Используя один и тот же опорный пучок, на одной фотопластинке можно дважды последовательно запечатлеть рассеянные предметом волны. Если между экспозициями какие-то части предмета несколько сместились или деформировались, то при восстановлении две одновременно возникающие когерентные предметные волны будут иметь определенную разность хода и изображение поверхности предмета будет покрыто системой интерференционных полос (рис. 7.40), аналогичных обычным полосам равной толщины. По расположению этих полос можно судить об изменениях предмета между экспозициями. Замечательно, что изучаемый предмет может при этом отражать свет диффузно, иметь сложный рельеф и шероховатую поверхность, так как все эти факторы одинаково влияют на обе восстанавливаемые предметные волны. Несмотря на очень сложную форму волновых поверхностей, эти волны вполне подобны и создают простые и легко наблюдаемые интерференционные полосы.

7.40
Интерференционные полосы, наблюдаемые при восстановлении голограммы, полученной методом двойной экспозиции

Методом двойной экспозиции можно исследовать и изменения, происходящие в прозрачных (фазовых) объектах (рис. 7.40). На пути просвечивающего объект пучка ставят прозрачный рассеивающий экран (матовое стекло), чтобы свет от каждой точки объекта попадал на всю поверхность голограммы. Так как этот экран и объект неподвижны и экспонируются дважды, их оптические неоднородности, сколь сложными бы они ни были, никак не влияют на расположение интерференционных полос, наблюдаемых на восстановленном изображении. Эти полосы характеризуют локальные изменения оптической толщины объекта между двумя экспозициями. В противоположность этому в традиционной интерферометрии для выполнения аналогичных измерений

поверхности объекта должны обладать высокими оптическими качествами.

В другом варианте метода на голограмме регистрируют волну, рассеянную объектом только в некотором начальном состоянии. Затем при восстановлении полученной голограммы объект не удаляют, а освещают так же, как и при регистрации голограммы. В результате возникают две волны: распространяющаяся от самого объекта в данный момент и восстановленная голограммой предметная волна, соответствующая начальному состоянию объекта. Непрерывно наблюдая создаваемую этими когерентными волнами интерференционную картину, можно судить о происходящих с течением времени изменениях состояния объекта. Такой метод называют *голографической интерферометрией реального времени.*

Применение голографии в микроскопии позволяет преодолеть серьезный недостаток микроскопа при сильном увеличении — очень малую глубину резкости изображения. Вместо того чтобы регистрировать изображение, можно записать на голограмме проходящую через микроскоп предметную волну. При восстановлении такой голограммы можно наблюдать находящиеся в разных плоскостях детали предмета, перемещая только оптическую систему наблюдения.

Используя вместо когерентного света ультразвуковые волны, можно получить акустическую голограмму. Звук проникает в оптически непрозрачные предметы. Поэтому акустическая голограмма позволяет восстановить трехмерное изображение внутренних частей предмета, например органов человеческого тела или глубин океана, что открывает широкие перспективы для применений в медицине, в подводных исследованиях, геофизике, археологии.

Специально изготовленные голограммы могут использоваться в качестве определенных оптических элементов. Голограмма-зонная решетка может выполнять некоторые функции линзы, голограмма-дифракционная решетка — служить диспергирующим элементом спектрального прибора, толстослойная голограмма с параллельными отражающими слоями — служить интерференционным фильтром и т. п.

С помощью одной голограммы можно записать и восстановить огромное количество информации. Большое число независимых сведений, регистрируемых голограммой, внешне проявляется в чрезвычайной сложности ее структуры. Под микроскопом такая голограмма производит впечатление хаотического набора пятен всевозможной формы и ориентации в отличие от регулярной структуры голограммы простейшего объекта. Очень важно, что декодирование этой огромной информации на этапе восстановления происходит просто и чрезвычайно быстро. Способность голографии к регистрации, хранению и быстрому преобразованию информации открывает перспективы создания новых систем памяти ЭВМ, оптических систем обработки данных, систем распознавания образов и символов с помощью оптической фильтрации сигналов.

Контрольные вопросы

- ☐ В чем заключается основная идея голографической записи зрительных образов?
- ☐ Как выполняется декодирование информации, зарегистрированной на голограмме?
  Какими преимуществами обладает голография по сравнению с обычной фотографией?
- ☐ Объясните физическую сущность записи и восстановления голограммы на примере плоской предметной волны.
- ☐ Какой вид имеют интерференционные полосы на голограмме точечного источника? В чем сходство и различие такой голограммы и зонной пластинки Френеля?
  Почему целесообразно использовать наклонное падение опорной волны на голограмму?
- ☐ Какие требования предъявляются в голографии к источнику света и к регистрирующей среде (фотоэмульсии)?
  В каких случаях возникают аберрации восстанавливаемого голограммой изображения?
- ☐ Какими преимуществами обладают толстослойные голограммы?
- ☐ Что дает применение голографии в интерферометрии и микроскопии? в системах оптической обработки информации?

ОПТИКА ДВИЖУЩИХСЯ ТЕЛ

● В этой главе основное внимание сосредоточено на оптических явлениях, в которых или источник света, или наблюдатель, или среда движутся относительно друг друга. Закономерности таких явлений могут быть существенно иными, чем в случае, когда эти тела покоятся. Длительный путь развития оптики движущихся тел привел в начале нашего столетия к созданию частной теории относительности — фундаментальной физической теории, лежащей в основе современной научной картины физического мира. Исчерпывающее решение основных проблем оптики движущихся тел фактически содержится уже в работе Эйнштейна «К электродинамике движущихся тел» (1905).

● Теория относительности базируется не на гипотезах, а на твердо установленных на опыте принципах (постулатах). Большую роль в их становлении сыграли оптические опыты с движущимися телами и попытки теоретической интерпретации их результатов. Но наше убеждение в полноте экспериментального обоснования теории относительности основано не на отдельном опыте или группе опытов, а на всей совокупности опытов, результаты которых подтверждают ее постулаты и следствия из них. Наиболее точные опыты относятся к физике высоких энергий (например, столкновения релятивистских частиц, ядерные превращения), а не к оптике движущихся тел. В наши дни сомневаться в справедливости теории относительности — это то же самое, что сомневаться в реальности ускорителей заряженных частиц или атомных электростанций. Но фундаментальность представлений теории относительности и парадоксальность ее следствий до сих пор приводят к попыткам ее пересмотра. Поэтому не прекращается и постановка новых опытов (в том числе и оптических) с использованием усовершенствованных приборов и новых эффектов для проверки основных принципов теории относительности.

8.1. Безуспешные поиски «светоносной среды»

Вопрос о законах оптических явлений в случае движущихся тел возник задолго до создания электромагнитной теории света. Высказанная Гюйгенсом мысль о волновой природе света требовала введения особой среды, в которой могли бы распространяться световые колебания. Эту среду Гюйгенс назвал *эфиром*. Так в физике возникло понятие, несостоятельность которого была окончательно установлена лишь теорией относительности.

Убедительное подтверждение волновая теория света получила в начале XIX в., когда на ее основе было дано исчерпывающее объяснение явлениям интерференции и дифракции. Открытие поляризации света свидетельствовало о поперечности световых волн. В рамках механической волновой теории, рассматривавшей свет по аналогии со звуковыми волнами, эфир пришлось наделить механическими свойствами твердого тела, так как поперечные упругие волны могут распространяться только в твердых телах. Конечно, это была странная среда: заполняя все пространство и пронизывая все тела, она при этом никак не влияла на их движение.

Но даже если оставить в стороне трудный вопрос непротиворечивого объяснения столь экзотических свойств эфира, со всей остротой вставала задача обнаружения связанной с ним системы отсчета. Движение тел относительно эфира можно было бы рассматривать как некоторое абсолютное движение.

Пришедшая на смену старой волновой теории электромагнитная теория света практически не внесла ничего нового в постановку этого вопроса. Рассматривая свет как частный вид электромагнитных волн, она позволила обойтись без противоречивых механических представлений об эфире, но не затронула предположения о возможности определять движение тел относительно эфира. Считалось, что уравнения Максвелла справедливы в определенной системе отсчета, за которой и сохранилось название эфира. Задача экспериментального обнаружения этой привилегированной системы отсчета по-прежнему оставалась актуальной. Предполагалось, что при переходе к другой инерциальной системе отсчета уравнения Максвелла в отличие от уравнений механики Ньютона должны изменить свой вид. Другими словами, считалось, что принцип относительности, т. е. утверждение об эквивалентности всех инерциальных систем отсчета, выполняется только для механических явлений и не справедлив для электромагнитных и оптических явлений.

Гипотеза о существовании эфира как выделенной системы отсчета выдвигала постановку ряда опытов с целью выяснения законов распространения света в телах, движущихся относительно эфира, и опытов, связанных с движением наблюдателя относительно эфира. Результаты этих опытов вскрыли *противоречия в самом понятии эфира* и привели в конечном счете к отказу от представлений о возможности определить «абсолютное» движение тел с помощью оптических явлений. Принцип относительности был распространен не только на механические явления, но и на все явления

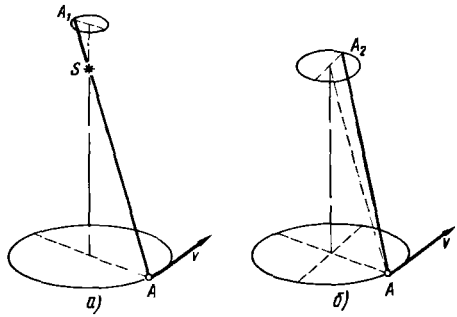
вообще. Было признано также, что скорость света в пустоте имеет универсальное значение, одинаковое во всех инерциальных системах отсчета.

Эти утверждения — принцип относительности, распространенный на все явления, и существование предельной скорости распространения сигналов, совпадающей со скоростью света в пустоте и одинаковой во всех системах отсчета, — и составляют содержание двух принципов, или постулатов, лежащих в основе частной теории относительности.

Влияние движения тел на оптические явления впервые было обнаружено в астрономических наблюдениях. В 1676 г. Ремер при наблюдении периодических затмений галилеевых спутников Юпитера заметил, что времена их обращения возрастают, когда Земля в своем орбитальном движении удаляется от Юпитера, и уменьшаются, когда она приближается. Ремер объяснил изменения наблюдаемых периодов обращения конечной скоростью света (см. § 2.11).

Орбитальное движение Земли приводит также к явлению *звездной аберрации*, которое было открыто Брэдли (1725—1728). В своих наблюдениях он пытался обнаружить годичный параллакс, т. е. кажущуюся траекторию, которую описывает проекция звезды на небесный свод из-за изменения положения наблюдателя при движении Земли по орбите (рис. 8.1, *а*). В общем случае такая траектория должна быть эллипсом, вырождающимся в окружность для звезды, расположенной вблизи полюса эклиптики (как на рис. 8.1, *а*), или в отрезок прямой для звезды, лежащей в плоскости эклиптики. Брэдли нашел, что звезда действительно описывает эллипс, большая ось которого равна 41″, однако направление углового отклонения звезды совершенно иное, чем должно быть при параллаксе (рис. 8.1, *б*): когда Земля находится в точке A, ее наблюдаемое положение смещено не в точку A_1, а в точку A_2, т. е. отклонение происходит в направлении движения Земли. Кроме того, отклонение не зависит от расстояния до звезды и значительно больше, чем параллактическое смещение даже ближайших звезд. Существование параллакса неподвижных звезд было твердо установлено Бесселем лишь сто лет спустя.

Найденное Брэдли объяснение аберрации опирается на то, что свет распространяется с конечной скоростью. При движении наблюдателя кажущееся направление на звезду отличается от истинного подобно тому, как от-



8.1
Отклонение наблюдаемого положения звезды из-за параллакса (*а*) и аберрации (*б*)

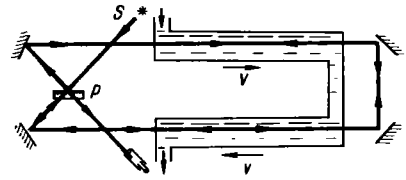
весно падающий дождь кажется косым наблюдателю в движущемся вагоне. Угол отклонения равен v/c (при $v \ll c$). Через полгода направление движения Земли изменится на противоположное и кажущееся положение звезды будет смещено на такой же угол в противоположную сторону. Разность видимых направлений составит $2v/c$. Учитывая, что $v = 30$ км/с, $c = 3 \cdot 10^5$ км/с, получаем $2v/c = 2 \cdot 10^{-4} = 41''$, что совпадает с измеренным значением полного смещения. Из своих наблюдений Брэдли сделал важный вывод о равенстве скоростей света от разных звезд. Отметим, что наблюдение аберрации возможно лишь потому, что Земля в своем орбитальном движении изменяет направление скорости. Наблюдаема не сама аберрация, а ее изменение. Если бы Земля двигалась прямолинейно и равномерно, ни о каком обнаружении аберрации земным наблюдателем не могло быть и речи.

Брэдли в своем объяснении аберрации исходил из корпускулярной картины распространения света. Такой же результат получается и из волновой теории (впервые это было показано Юнгом), если считать, что «светоносная среда» покоится в гелиоцентрической системе отсчета. К представлению о «неподвижном эфире» приводят и наблюдения за изменением периодов затмений спутников Юпитера, т. е. измерения скорости света по методу Ремера, выполненные при разных положениях Юпитера на эклиптике.

Если эфир неподвижен в гелиоцентрической системе, то орбитальное движение Земли должно вызывать «эфирный ветер». Первые попытки его обнаружения на опыте были сделаны Араго в 1810 г. Араго наблюдал преломление света от неподвижной звезды в призме, когда Земля, а вместе с ней и призма, двигалась к звезде и от звезды. При разных направлениях движения призмы относительно неподвижного эфира должен быть различен ее показатель преломления. Араго не обнаружил ожидаемого эффекта, и хотя точность его опытов была недостаточна, он интуитивно сделал правильный вывод о том, что никакой зависимости от движения Земли нет.

Этот отрицательный результат легко объяснить, предположив, что эфир полностью увлекается движущейся Землей, но тогда возникают непреодолимые трудности в объяснении аберрации. Выход из противоречий был найден Френелем на основе гипотезы о том, что плотность эфира в прозрачных телах больше, чем в пустоте (при неизменной упругости). Такое предположение вместе с требованием неразрывности эфира приводило к заключению о том, что внутри движущегося тела, где эфир уплотнен, его скорость относительно тела в n^2 раз меньше, чем снаружи. Иначе об этом можно говорить как о частичном увлечении эфира движущимся телом с коэффициентом увлечения $1 - 1/n^2$. Объяснение аберрации при этом не претерпевает никаких изменений (при $n = 1$ эфир остается неподвижным) и в то же время движение тел (вместе с наблюдателем) относительно эфира не оказывает влияния на любые оптические явления в первом порядке по v/c. Во втором порядке это

влияние должно проявиться, но при максимальной доступной скорости тел относительно эфира — скорости орбитального движения Земли — $(v/c)^2 = 10^{-8}$, что лежит за пределами достижимой точности таких экспериментов. Таким образом, френелевский коэффициент увлечения объясняет отрицательные результаты опытов, в которых среда движется вместе с наблюдателем относительно эфира. При этом безразлично, делается ли опыт со светом от звезды или от земного источника, так как по волновой теории (в отличие от корпускулярной) скорость света не зависит от движения источника*.



8.2
Схема опыта Физо

Прямое измерение коэффициента увлечения возможно в опытах, где изучается распространение света в телах, движущихся относительно наблюдателя Такие опыты с движущейся водой были выполнены Физо в 1851 г. и повторены с большей точностью Майкельсоном и Морли в 1886 г. Луч света от источника S расщепляется полупрозрачной пластинкой P (рис. 8.2) на два луча, один из которых распространяется в трубах по направлению течения воды, другой — против. Затем эти лучи соединяются и дают систему интерференционных полос, наблюдаемую в зрительную трубу. Смещение полос по сравнению с их положением при неподвижной воде позволяет измерить вносимую движением воды разность хода и тем самым найти скорость света в движущейся воде. В покоящейся воде скорость света c/n. Результаты опытов совершенно определенно показали, что в воде, движущейся со скоростью v, скорость света определяется выражением

$$u = c/n + v(1 - 1/n^2), \qquad (8.1)$$

полностью соответствующем выводу Френеля о частичном увлечении эфира.

Электромагнитная теория света, отказавшись от механического эфира, сохранила представление о существовании выделенной системы отсчета, в которой справедливы уравнения Максвелла и скорость света в пустоте по всем направлениям равна c. Изменение скорости света в неподвижном веществе $u = c/n$ в электронной теории Лоренца объяснялось как макроскопический эффект, обусловленный вынужденными колебаниями входящих в его состав зарядов. Введенное Френелем чисто феноменологически частичное

* Вопрос о зависимости скорости света от движения источника, на который волновая теория давала определенно отрицательный ответ, вновь возник значительно позже, в полемике Ритца с Эйнштейном по основам теории относительности. В 1913 г. де-Ситтер указал, что независимость скорости света от движения источника с высокой точностью вытекает из наблюдений за двойными звездами (безотносительно к принятой теоретической концепции — волновой или корпускулярной)

увлечение эфира движущимися телами получает при этом простое физическое истолкование. По Лоренцу, эфир неувлекаем и внутри тел; увлекаются, т. е. движутся вместе с телом, только заряды, из которых оно построено. Макроскопически это движение зарядов приводит к появлению френелевского коэффициента увлечения $1 - 1/n^2$.

Еще одно важное открытие, хорошо укладывавшееся в схему неподвижного, неувлекаемого эфира, — теоретически предсказанный в 1842 г. Х. Доплером эффект влияния движения источника волн или наблюдателя на воспринимаемую частоту. Элементарное объяснение эффекта предполагает определенную скорость c распространения волн относительно эфира, подобно скорости звука в неподвижной среде (например, в воздухе). Тогда частота ω воспринимаемых наблюдателем колебаний связана с частотой ω_0 колебаний в источнике соотношением

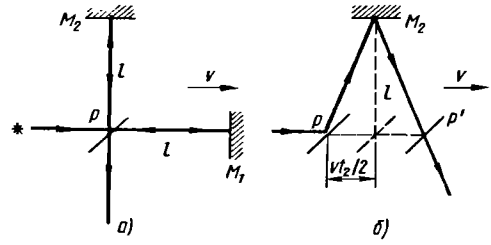
$$\omega = \omega_0 \frac{1 - v_1/c}{1 - v_2/c}, \tag{8.2}$$

где v_1 и v_2 — проекции скорости наблюдателя и источника относительно среды на направление от источника к наблюдателю. Формула (8.2) дает разные результаты в зависимости от того, движется источник или наблюдатель, но это различие сказывается лишь начиная с членов второго порядка по v_1/c и v_2/c. В первом порядке сдвиг частоты $\Delta\omega = \omega - \omega_0$ определяется только относительной скоростью $v = v_1 - v_2$: $\Delta\omega/\omega \approx -(v_1 - v_2)/c = -v/c$. Эффект Доплера получил опытное подтверждение как в астрономических, так и в лабораторных наблюдениях. О некоторых его применениях сказано в § 8.3.

Глубокие противоречия концепции неподвижного эфира со всей отчетливостью проявились лишь в оптических опытах второго порядка по v/c. Идея эксперимента, который давал бы принципиальную возможность непосредственно обнаружить движение Земли относительно эфира, была высказана Максвеллом в 1878 г. Но Максвелл полагал, что из-за малости ожидаемого эффекта (при орбитальном движении Земли $(v/c)^2 = 10^{-8}$) необходимая точность измерений недостижима. Однако в 1881 г. такой опыт с использованием интерференции света был осуществлен Майкельсоном и повторен им совместно с Морли в 1887 г. с большей точностью. Опыту Майкельсона суждено было стать самым знаменитым в оптике движущихся тел, и впоследствии он многократно воспроизводился разными исследователями с улучшенной экспериментальной техникой и в новых вариантах.

Чтобы понять идею опыта Майкельсона по обнаружению абсолютного движения Земли, нужно встать на точку зрения физики конца XIX в., согласно которой скорость света только в одной системе отсчета одинакова по всем направлениям и равна c. Пусть интерферометр Майкельсона (см. § 5.3) ориентирован так, что одно

из его одинаковых плеч, например PM_1, параллельно скорости орбитального движения Земли (рис. 8.3, *а*). Тогда промежутки времени t_1 и t_2, которые затрачивает свет для прохождения туда и обратно одинаковых расстояний вдоль плеч PM_1 и PM_2, оказываются различными. В самом деле, используя гелиоцентрическую систему отсчета, в которой гипотетический эфир неподвижен, для t_1 находим



8.3
Схема опыта Майкельсона

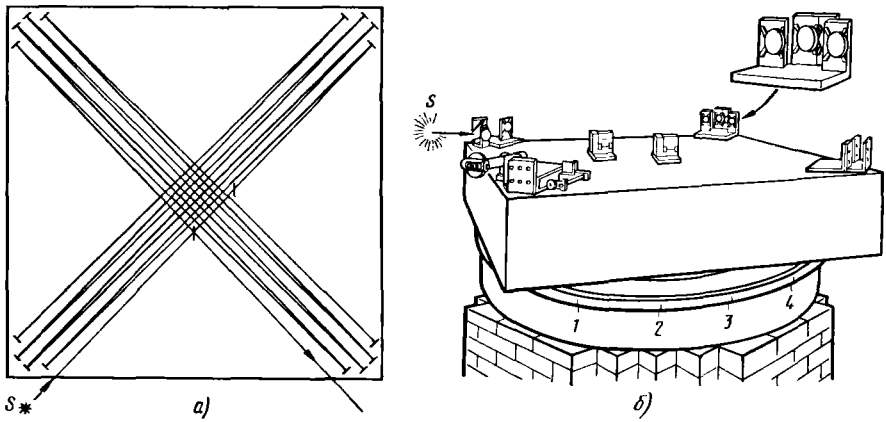
$$t_1 = \frac{l}{c-v} + \frac{l}{c+v} = \frac{2l}{c}\frac{1}{1-v^2/c^2}. \qquad (8.3)$$

Свет, распространяющийся в перпендикулярном направлении, за время t_2 проходит путь PM_2P' (рис. 8.3, *б*), длина которого $2\sqrt{l^2+(vt_2/2)^2}$. С другой стороны, эта же длина равна ct_2, так как свет проходит ее за время t_2 со скоростью c. Решая полученное уравнение для t_2, находим

$$t_2 = \frac{2l}{c}\frac{1}{\sqrt{1-v^2/c^2}}. \qquad (8.4)$$

К этому же результату можно прийти, используя связанную с Землей систему отсчета, если считать, что скорость света в ней складывается из скорости c относительно эфира и скорости орбитального движения Земли.

Времена t_1 и t_2 различаются на величину, зависящую от квадрата малого отношения $\beta = v/c$: $\Delta t = t_1 - t_2 = (l/c)(v/c)^2 = (l/c)\beta^2$. Если повернуть прибор на 90°, то эта разность изменит знак и наблюдаемые интерференционные полосы должны сместиться. Ожидаемое смещение полос ΔN определяется удвоенным значением Δt: $\Delta N = 2\Delta t/T = 2(l/\lambda)\beta^2$. При $v = 30$ км/с $\beta^2 = 10^{-8}$ и смещения интерференционной картины на одну полосу можно ожидать при длине плеча интерферометра $l = 5\cdot 10^7\lambda = 30$ м. В опытах 1887 г. эффективная длина плеч с помощью системы зеркал была доведена до 11 м (рис. 8.4, *а*). Для уменьшения вибраций и изгибов при повороте интерферометр был смонтирован на квадратной каменной массивной плите (рис. 8.4, *б*) со стороной 1,5 м, кольцеобразная подставка которой плавала в сосуде с ртутью. Измерения положения полос производились при непрерывном медленном равномерном вращении установки. В пределах точности измерений (0,01 полосы) никакого смещения полос не наблюдалось, хотя ожидаемое смещение составляло более трети полосы. Майкельсон пришел к совершенно определенному заключению об отрицательном результате опыта.



8.4

Оптическая схема (*а*) и общий вид интерферометра Майкельсона (*б*)

Отсутствие эфирного ветра было подтверждено во всех повторениях опыта Майкельсона. Например, Иос в 1930 г. получил отрицательный результат, используя интерферометр, рассчитанный на обнаружение эфирного ветра со скоростью 1,5 км/с. В опытах Кеннеди и Торндайка (1932) наблюдения велись непрерывно в течение длительного времени, так что Земля при движении вокруг Солнца успевала значительно изменить направление скорости. Это можно рассматривать как переход из одной инерциальной системы отсчета в другую. Если бы скорость света в вакууме при этом изменялась, то возникло бы смещение полос. Прибор был настолько чувствителен, что позволял обнаружить изменение на 2 км/с. Но никакого смещения полос не наблюдалось, что свидетельствовало о неизменности скорости света при переходе в другую систему отсчета.

Изобретение лазеров (см. § 9.4) позволило упростить постановку подобных опытов и значительно повысить их точность. В 1964 г. Таунс, Джаван и другие использовали установку с двумя одинаковыми газовыми лазерами, расположенными перпендикулярно друг другу на поворотной платформе. Частота генерации в соответствии с формулой (9.39) зависит от скорости света c. С позиций гипотезы неподвижного эфира поворот установки на 90° должен вызвать из-за орбитального движения Земли изменение разности частот лазеров приблизительно на 3 МГц, что легко наблюдать при смещении их излучения на фотокатоде приемника. На основании таких опытов было установлено, что скорость эфирного ветра не превышает 30 м/с (при скорости орбитального движения Земли 30 км/с). В опытах с использованием гамма-излучения и эффекта Мёссбауэра этот предел был уменьшен до 5 м/с.

Отрицательный результат опыта Майкельсона проще всего было объяснить, предположив полное увлечение эфира движущейся Землей. Но такое предположение противоречило наблюдаемому эффекту аберрации и результатам опыта Физо с движущейся водой. Чтобы объяснить опыт Майкельсона и сохранить эфир, Лоренц и независимо от него Фитцджеральд предположили, что все тела при движении относительно неподвижного эфира испытывают сокращение в $\sqrt{1-v^2/c^2}$ раз в направлении движения. Причину такого сокращения Лоренц находил в изменении электромагнитных сил и, следовательно, положений равновесия образующих тело заряженных частиц из-за движения в эфире*. Но гипотеза лоренцевского сокращения обнаруживает свою несостоятельность при интерпретации отрицательного результата выполненного Кеннеди и Торндайком несколько видоизменного опыта Майкельсона с разной длиной плеч интерферометра. К такому же заключению приводит анализ более поздних (и более точных) опытов на основе эффекта Мёссбауэра.

Хотя подобные эксперименты не могли быть выполнены во времена Лоренца, он предвидел, что, подобно опыту Майкельсона, они могут дать отрицательный результат. Поэтому он ввел еще одну гипотезу об изменении хода часов при их движении относительно эфира. Уравнения для изменения длины и времени таковы, что при измерении скорости света в любой системе отсчета будет получаться одна и та же величина. Новая теория не только объяснила отрицательный результат опыта Майкельсона, но и приводила к принципиальной невозможности обнаружения движения относительно эфира. Этим Лоренц лишил эфир последнего и единственного свойства, которое в электромагнитной теории еще позволяло называть его некоторой средой, — свойства локализации по отношению к определенной системе отсчета. Спасая привычные классические представления о пространстве и времени, Лоренц пришел к теории, не удовлетворяющей критерию наблюдаемости: нельзя представить себе хотя бы принципиальных возможностей ее опытного подтверждения или опровержения. Трудно поверить в то, что эфир существует, если природа устроена так, что его в принципе невозможно обнаружить.

Непротиворечивое объяснение всех оптических опытов с движущимися телами дала частная теория относительности. Релятивистский подход к их истолкованию обсуждается в § 8.3. Здесь же отметим, что преодоление трудностей оказалось возможным благодаря отказу от представлений об эфире как выделенной системе отсчета и распространению принципа относительности на все явления. Полная эквивалентность инерциальных систем означает,

* Теория относительности также предсказывает сокращение тел в направлении движения, но там это чисто кинематический эффект зависимости результатов измерений от относительного движения (см. § 8.2). Здесь же сокращение Лоренца—Фитцджеральда обусловлено движением относительно эфира, т. е. привилегированной системы отсчета.

что равномерное прямолинейное движение замкнутой системы тел по отношению к инерциальной системе отсчета не влияет ни на какие физические явления в системе тел. С таких позиций отрицательный результат опыта Майкельсона очевиден и не требуется придумывать гипотезы ad hoc (т. е. только для данного случая). Но преимущество теории относительности не только в гораздо более ясной и удовлетворительной постановке всего вопроса, а прежде всего в плодотворности, эвристичности: она предсказывает новые явления, и все выполненные по сей день эксперименты оправдывают эти предсказания.

Контрольные вопросы

- ☐ Поясните исторнческие предпосылки возникновения в физике концепции эфира.
- ☐ Какне изменения понятие эфира претерпело после создания электромагнитной теории света?
- ☐ Какую роль играет нзменение направления орбнтального движения Земли в наблюдениях аберрации звезд?
- ☐ Как результаты опыта Физо объясняются в теории эфира и в электроиной теорин Лоренца?
- ☐ Изложите идею опыта Майкельсона.
- ☐ Почему предложенное Лоренцем объясненне отрицательного результата опыта не может счнтаться удовлетворительным?

8.2. Основные положения частной теории относительности

Как уже отмечалось, частная теория относительности основана на двух принципах, или постулатах, в которых из всего огромного экспериментального материала выделены бесспорные, не вызывающие сомнений положения. Первый постулат — это *принцип относительности*, распространенный на все физические явления. Он утверждает *полное равноправие всех инерциальных систем отсчета*. Другими словами, неускоренное движение замкнутой физической системы как целого не влияет на законы любых явлений. Второй постулат утверждает *существование предельной скорости распространения сигналов*, совпадающей со скоростью света в пустоте.

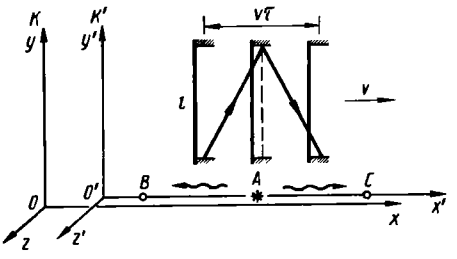
Эти принципы содержат очень сильные и общие утверждения. Например, второй постулат утверждает, что любые взаимодействия между телами распространяются в пустоте с одной и той же конечной скоростью, не зависящей от движения тел. В соответствии с первым постулатом эта скорость одинакова во всех инерциальных системах. В итоге скорости света в вакууме придается универсальное значение, не связанное с физической природой взаимодействия, а отражающее некоторое общее свойство пространства и времени. Поэтому едва ли можно говорить о каких-либо «решающих» опытах, доказывающих справедливость этих принципов. Нельзя считать, что они вытекают из рассмотренных выше оптических опытов. Но, как уже отмечалось, оптические опыты с движущимися телами

сыграли большую роль в их становлении. Неудивительно, что именно с позиций теории относительности вся совокупность этих опытов получает исчерпывающее и в то же время наиболее простое и естественное объяснение.

Признание существования абсолютной, одинаковой во всех системах отсчета скорости потребовало отказа от классических ньютоновских представлений о пространстве и времени и замены их новыми релятивистскими представлениями. Это, в свою очередь, привело к созданию релятивистской механики, которая сводится к ньютоновской только в предельном случае движения тел со скоростями, малыми по сравнению со скоростью света.

Напомним кратко основные положения частной теории относительности. Анализ операций измерения промежутков времени и расстояний, проведенный в рамках постулатов теории, привел к необходимости отказаться от представлений классической физики об абсолютном характере таких понятий, как *одновременность событий, промежуток времени между событиями, расстояние между точками в пространстве*. Проиллюстрируем зависимость этих понятий от системы отсчета на простых примерах.

Рассмотрим две инерциальные системы отсчета K и K', причем K' движется относительно K с постоянной скоростью v. Условимся, как это обычно принято, выбирать направления одноименных осей координат в K и K' одинаковыми, причем оси x и x' направим вдоль вектора $\mathbf{v}$ (рис. 8.5). Пусть в некоторой точке A на оси x' происходит короткая вспышка света, т. е. из A в противоположных направлениях отправляются сигналы. Рассмотрим приход сигналов в точки B и C системы K', равноудаленные от A. Так как скорость света не зависит от направления, эти события в K' происходят одновременно. Легко видеть, однако, что эти же два события — приход сигналов в B и C — уже не будут одновременными в системе отсчета K. В самом деле, в соответствии с постулатами скорость сигналов в системе K также не зависит от направления, но точка B движется относительно K навстречу сигналу, а точка C — от посланного в нее сигнала. Поэтому с точки зрения наблюдателя в K сигналу, распространяющемуся с конечной скоростью, приходится на пути в C преодолевать большее расстояние, чем на пути в B, и приход сигнала в C произойдет позже, чем в B. Значит, *понятие одновременности событий зависит от системы отсчета*. Классическое представление об абсолютном характере одновременности оказывается приближенным. Оно



8.5
К обоснованию относительности одновременности событий

возникло потому, что в повседневном опыте приходится иметь дело с движениями, скорости которых малы по сравнению со скоростью света.

Исходя из постулатов теории относительности легко показать *относительный характер промежутков времени между событиями* и найти закон их преобразования при переходе из одной системы отсчета в другую. Рассмотрим для этого следующий мысленный опыт. На концах стержня длиной l, ориентированного перпендикулярно оси x и неподвижного в системе отсчета K', закреплены параллельные зеркала (рис. 8.5). Между зеркалами движется короткий световой импульс. Рассмотрим один цикл таких «световых часов», т. е. выход светового импульса от нижнего зеркала и его возвращение после отражения от верхнего зеркала, с точки зрения наблюдателей в каждой из систем отсчета. В K' оба события происходят в одной и той же точке, и измеренный в K' промежуток времени между ними называется собственным временем τ_0. Очевидно, что $\tau_0 = 2l/c$. В системе K стержень находится в движении и световой импульс движется зигзагообразно (рис. 8.5). Свет проходит за один цикл больший путь, и, следовательно, промежуток времени τ между теми же событиями в K оказывается больше, чем в K', ибо скорость света, согласно второму постулату, одинакова в K и K'. Как видно из рис. 8.5, для определения τ можно написать уравнение $c\tau = 2\sqrt{l^2 + (v\tau/2)^2}$. Отсюда, учитывая, что $2l/c = \tau_0$, находим

$$\tau = \tau_0 / \sqrt{1 - v^2/c^2}. \qquad (8.5)$$

При любой скорости $v \neq 0$, как видно из (8.5), $\tau > \tau_0$. Поэтому с точки зрения наблюдателя в K движущиеся часы (т. е. часы в K') идут медленнее, чем его собственные, идентичные по устройству. Подчеркнем, что речь здесь идет о сравнении показаний одних и тех же движущихся часов с показаниями разных (находящихся в разных точках) неподвижных часов. Синхронизированные для неподвижного наблюдателя, эти часы не будут синхронизированы для движущегося (из-за относительного характера одновременности событий).

Рассмотренный релятивистский эффект замедления времени, как того требует равноправие инерциальных систем отсчета K и K', является взаимным: с точки зрения наблюдателя K', медленнее идут часы, связанные с системой K.

Аналогичным образом непосредственно на основе постулатов можно найти *релятивистский закон преобразования пространственных расстояний*. Для этого можно, например, рассмотреть мысленный эксперимент с теми же «световыми часами», но ориентированными иначе — параллельно направлению относительной скорости систем отсчета. В результате получим, что длина их стержня (и, следовательно, любого тела) относительна, т. е. зависит от системы отсчета. Наибольшее значение l_0 длина имеет в системе отсчета, где тело покоится (*собственная длина*). В другой системе, относительно

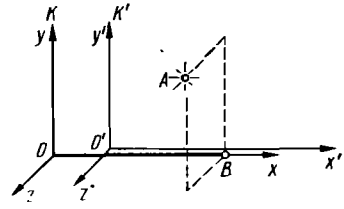
которой тело движется со скоростью v, его длина l в направлении движения сокращается:

$$l = l_0 \sqrt{1 - v_2/c^2}. \qquad (8.6)$$

Этот релятивистский эффект, как и эффект замедления времени, становится заметным при движениях со скоростями, сравнимыми со скоростью света. При $v \ll c$ расстояние между точками и промежуток времени между событиями приобретают практически абсолютный смысл в полном соответствии с классическими представлениями о пространстве и времени, сформировавшимися на основе опыта наблюдений над сравнительно медленными движениями.

Отметим, что релятивистские эффекты сокращения длины и замедления времени означают лишь, что *измеряемый* промежуток времени и *измеряемая* длина зависят от относительного движения. Теория относительности предсказывает *влияние движения наблюдателя на результаты измерений.* Собственное время и собственная длина (т. е. длина тела в системе отсчета, где оно покоится), по определению, абсолютны. Для объяснения замедления времени и сокращения длины нет необходимости искать какие-то процессы или изменения в движущихся часах и движущемся стержне: эти эффекты непосредственно следуют из анализа самого процесса измерения.

Используя формулу (8.6), связывающую расстояние между двумя точками в разных системах отсчета, можно написать релятивистский закон преобразования координат и времени произвольного события при переходе от одной системы отсчета к другой Рассмотрим описание некоторого события A в двух инерциальных системах отсчета: K и K'. Пусть координаты и время этого события в K равны x, y, z и t, а в K' — x', y', z' и t' (рис. 8.6). Будем считать, что при $t=0$ точки O и O' совпадают. Расстояния в направлении, перпендикулярном относительной скорости $\mathbf{v}$ систем отсчета, одинаковы в K и K', поэтому $y=y'$ и $z=z'$. Координату x можно рассматривать как собственную длину L_0 отрезка OB, неподвижного в системе K. Длина L этого же отрезка в системе K', где измерение производится в момент времени t', равна $x'+vt'$. В соответствии с (8.6) $L=L_0\sqrt{1-v^2/c^2}$, поэтому $x'+vt'=x\sqrt{1-v^2/c^2}$, откуда $x=(x'+vt')/\sqrt{1-v^2/c^2}$. Эта формула выражает координату x события A в системе K через координату x' и время t' этого же события к K'.



8.6
К выводу преобразований Лоренца

Чтобы получить формулу для момента времени t через x' и t', рассмотрим отрезок $O'B$, неподвижный в системе K' (рис. 8.6). Собственная длина l_0 этого отрезка — это координата x' события A. Длина l этого же отрез-

ка в системе отсчета K (относительно которой он движется со скоростью v), измеряемая в момент времени t, равна $x - vt$. Учитывая снова релятивистское сокращение длины (8.6), можем написать $x - vt = x'\sqrt{1 - v^2/c^2}$. Подставляя сюда полученное выше выражение для x, после простых преобразований находим выражение для t через x' и t', которое вместе с формулой для x дает искомый закон преобразования координат и времени произвольного события:

$$x = \frac{x' + vt'}{\sqrt{1 - v^2/c^2}}, \quad y = y', \quad z = z', \quad t = \frac{t' + (v/c^2)x'}{\sqrt{1 - v^2/c^2}}. \tag{8.7}$$

Это и есть релятивистские *преобразования Лоренца*. Обратные преобразования от K и K' можно получить из (8.7), разрешив их относительно x' и t'. Разумеется, они отличаются от (8.7) лишь заменой $v \to -v$ и перестановкой штрихов в полном соответствии с принципом относительности, утверждающим равноправие систем отсчета K и K'.

Преобразования Лоренца (8.7) в предельном случае $v \ll c$ переходят в *преобразования Галилея*, выражающие классические представления о пространстве и времени:

$$x = x' + vt, \quad y = y', \quad z = z', \quad t = t'. \tag{8.8}$$

Это значит, что теория относительности не отвергает классическую нерелятивистскую физику, а ограничивает ее применимость областью сравнительно медленных движений. Промежуток времени между событиями и расстояние между точками в классической физике абсолютны, т. е. одинаковы во всех системах отсчета. Другими словами, это инварианты преобразований Галилея. Теория относительности перевела эти понятия из абсолютных в относительные.

Преобразования Лоренца (8.7) для координат и времени события можно рассматривать как геометрические преобразования в четырехмерном пространстве-времени. При переходе от одной системы отсчета к другой пространственные координаты x, y, z и время t события изменяются, но определенная их комбинация

$$s = \sqrt{c^2t^2 - (x^2 + y^2 + z^2)}, \tag{8.9}$$

называемая *интервалом*, остается неизменной. Интервал обобщает понятия расстояния в трехмерном пространстве и промежутка времени. Его можно считать расстоянием в четырехмерном пространстве-времени от начала координат до мировой точки (ct, x, y, z), изображающей данное событие. По аналогии с трехмерным евклидовым пространством, координаты мировой точки (ct, x, y, z) можно рассматривать как проекции четырехмерного радиуса-вектора события, а интервал s (8.9) — как его длину. Преобразования Лоренца (8.7) дают закон изменения этих проекций при переходе от одной системы отсчета к другой, подобно тому, как

проекции радиуса-вектора точки в трехмерном пространстве преобразуются при повороте системы координат. Но в отличие от евклидовой геометрии пространства, где инвариант выражается через сумму квадратов проекций, в псевдоевклидовой геометрии пространства — времени инвариантная «длина» вектора выражается через разность квадратов временно́й и пространственных проекций.

Для анализа некоторых оптических экспериментов с движущимися телами полезен релятивистский закон преобразования скорости. Пусть $\mathbf{u}=\mathrm{d}\mathbf{r}/\mathrm{d}t$ — скорость некоторой частицы относительно системы отсчета K, а $\mathbf{u}'=\mathrm{d}\mathbf{r}'/\mathrm{d}t'$ — скорость той же частицы относительно K'. Рассматривая движение частицы как непрерывную последовательность событий $(ct, \mathbf{r})$, можно найти связь между $\mathbf{u}$ и $\mathbf{u}'$:

$$\mathbf{u}=\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}=\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t'}\,\frac{1}{\mathrm{d}t/\mathrm{d}t'}\,. \tag{8.10}$$

Входящие сюда производные находятся дифференцированием правых частей формул преобразований Лоренца (8.7). При этом нужно учесть, что x и t зависят от t' не только явно, но и через x'. Подставляя производные в (8.10), получаем

$$u_x=\frac{u'_x+v}{1+vu'_x/c^2},\quad u_y=\frac{u'_y\sqrt{1-v^2/c^2}}{1+vu'_x/c^2},\quad u_z=\frac{u'_z\sqrt{1-v^2/c^2}}{1+vu'_x/c^2}\,. \tag{8.11}$$

При $v\ll c$ формулы (8.11) переходят в классический закон сложения скоростей, но в релятивистском случае преобразование скорости (8.11) при переходе в другую систему отсчета отнюдь не сводится к векторному сложению относительной и переносной скоростей. В то же время при разложении скорости частицы на составляющие в какой-либо одной системе отсчета она ведет себя как обычный трехмерный вектор (т. е. равна векторной сумме своих составляющих по разным направлениям). Формулы для обратного преобразования скорости от K к K' получаются из (8.11) изменением знака скорости v.

Контрольные вопросы

- ☐ Сформулируйте постулаты частной теории относительности. В какой мере оптические опыты с движущимися телами подтверждают их справедливость?
- ☐ С помощью простых мысленных экспериментов покажите, что из постулатов следует относительность таких понятий, как одновременность событий и промежуток времени между событиями.
- ☐ Докажите с помощью преобразований Лоренца инвариантность пространственно-временно́го интервала между событиями.
- ☐ Получите формулы для преобразования энергии и импульса частицы при переходе к другой системе отсчета.
- ☐ Как преобразуется скорость частицы при переходе к другой системе отсчета?

8.3. Теория относительности и оптика движущихся тел

Отрицательные результаты всех опытов, в которых предпринимались попытки обнаружить влияние равномерного движения Земли на оптические явления, с релятивистских позиций представляются самоочевидными. Об этом прямо говорит первый постулат теории — принцип относительности, утверждающий эквивалентность всех инерциальных систем отсчета. Равномерное движение лаборатории относительно гелиоцентрической системы отсчета не может быть обнаружено как в оптических опытах первого порядка по v/c, берущих свое начало с опыта Араго по преломлению световых лучей в призме, так и в опытах второго порядка, к которым относятся опыт Майкельсона и его многочисленные модификации. В лабораторной системе отсчета, где прибор покоится, скорость света, как и в гелиоцентрической системе, одинакова по всем направлениям и равна c. Вид наблюдаемой интерференционной картины обусловлен только геометрической разностью хода лучей, которая не изменяется при повороте установки.

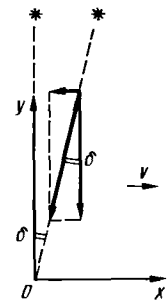
Что касается электромагнитных и оптических явлений в равномерно движущихся телах, то их законы могут быть найдены из соответствующих законов для неподвижных тел путем чисто кинематического пересчета к другой системе отсчета на основе преобразований Лоренца (8.7) или вытекающих из них формул (8.11) преобразования скорости. Рассмотрим с этой точки зрения интерпретацию результатов *опыта Физо* по измерению скорости света в движущейся воде. Пусть вода покоится в системе отсчета K'. Тогда скорость света в K' равна $u'=c/n$. Подставляя это значение в (8.11), найдем скорость u распространяющегося вдоль направления течения воды света в лабораторной системе отсчета K:

$$u=\frac{c/n+v}{1+v/(cn)}\approx\left(\frac{c}{n}+v\right)\left(1-\frac{v}{cn}\right)\approx\frac{c}{n}+v\left(1-\frac{1}{n^2}\right) \qquad (8.12)$$

Здесь отброшены малые члены порядка v^2/c. Френелевское «частичное увлечение эфира», описываемое формулой (8.1), можно рассматривать как простое следствие релятивистской кинематики. Тот факт, что электронная теория дает такое же выражение для скорости света в движущейся среде, совсем не удивителен, так как уравнения электродинамики удовлетворяют принципу относительности (сохраняют свой вид при преобразованиях Лоренца).

Закон преобразования скорости (8.11) дает релятивистское объяснение и явлению *звездной аберрации*. Пусть в гелиоцентрической системе K направление на некоторую удаленную звезду составляет прямой угол с направлением скорости орбитального движения Земли. В каком направлении видит эту звезду находящийся на Земле наблюдатель? Выберем ось x в направлении скорости Земли, ось y — в направлении на звезду (рис. 8.7). Тогда для скорости света от звезды в системе K можно написать: $u_x=0$, $u_y=-c$, $u_z=0$. Переходя в систему отсчета K', связанную с Землей,

с помощью формул (8.11) находим: $u'_x = -v$, $u'_y = -c\sqrt{1-v^2/c^2}$, $u'_z = 0$. Поэтому для земного наблюдателя кажущееся направление на звезду составляет с направлением в гелиоцентрической системе угол δ (угол аберрации) такой, что $\sin\delta = |u'_x|/\sqrt{u'^2_x + u'^2_y} = v/c$. (Классическая теория для угла аберрации дает выражение $\operatorname{tg}\delta = v/c$, которое при $v/c \ll 1$ совпадает с релятивистским.) Заметим, что в явлении аберрации наблюдается не сам угол аберрации, а его изменение при изменении направления орбитальной скорости Земли. Поэтому наблюдаемая аберрация одинакова для всех звезд и не зависит от их движения относительно Солнца.



8.7
К выводу формулы для аберрации света ($\sin\delta = v/c$)

Столь же простое кинематическое объяснение получает в теории относительности и *эффект Доплера*. Для волн в среде, например для звука в воздухе, скорость волн зависит от системы отсчета: если в системе отсчета, где среда покоится, скорость волн равна u, то в любой другой системе отсчета, движущейся со скоростью v, скорость волн зависит от их направления и изменяется от $u+v$ до $u-v$ (при $v \ll c$). Сдвиг частоты при эффекте Доплера зависит от скоростей источника и наблюдателя относительно среды. Для распространения электромагнитных волн (света) в пустоте никакой среды не требуется и их скорость во всех системах отсчета одинакова. Поэтому сдвиг частоты определяется только относительной скоростью источника и наблюдателя.

Приведем сначала элементарный вывод релятивистского выражения для продольного эффекта Доплера, когда относительная скорость источника и приемника направлена вдоль соединяющей их линии. Пусть, например, источник находится в начале координат системы K', его координата $x'=0$, а приемник — в начале координат $x=0$ системы K. Источник посылает сигналы через одинаковые промежутки времени, которые равны τ_0 по часам системы K', где он покоится. Найдем промежутки времени T между моментами приема последовательных сигналов по часам системы K, где покоится приемник.

Пусть для определенности первый сигнал, например кратковременная вспышка света, посылается в момент времени $t=0$, когда начала координат K и K' совпадают, т. е. источник и приемник находятся рядом. Сигнал будет принят в тот же момент времени $t=0$. Следующий сигнал посылается из точки $x'=0$ спустя промежуток времени τ_0 по часам в K', т. е. в момент времени $\tau = \tau_0/\sqrt{1-v^2/c^2}$ по часам системы K [см. (8.5)]. Источник в этот момент находится в точке $x=v\tau$. Поэтому сигнал, распространяющийся со скоростью c, достигнет приемника в момент $T=\tau+x/c = \tau+v\tau/c$. Подставляя сюда выражение для τ через τ_0, находим промежуток времени между принимаемыми сигналами: $T=\tau_0\times$

$\times \sqrt{(1+v/c)/(1-v/c)}$. Частота $\omega=2\pi/T$ принимаемых периодических сигналов связана с собственной частотой источника $\omega_0=2\pi/\tau_0$ соотношением

$$\omega=\omega_0\sqrt{\frac{1-v/c}{1+v/c}}=\omega_0\sqrt{\frac{c-v}{c+v}}. \tag{8.13}$$

Здесь относительная скорость v источника и приемника положительна, если они удаляются друг от друга, и отрицательна, если они сближаются.

Из приведенного вывода видно, что возникновение эффекта Доплера связано, во-первых, с изменением расстояния между источником и приемником и, во-вторых, с преобразованием промежутков времени при переходе от одной системы отсчета к другой. Первое обстоятельство не имеет отношения к теории относительности, и именно оно объясняет существование эффекта Доплера в нерелятивистской теории. Зависимость же скорости волн от системы отсчета приводит к тому, что в случаях движения источника относительно среды или движения наблюдателя классическая формула (8.2), справедливая для упругих волн, дает для сдвига частоты разные результаты при одной и той же относительной скорости v источника и наблюдателя. Различие исчезает только в первом порядке по v/c, т. е. в предельном случае медленного (по сравнению со скоростью волн) движения. Тогда $\omega=\omega_0(1-v/c)$ или $\Delta\omega/\omega=-v/c$. Релятивистская формула (8.13) в первом порядке по v/c дает то же самое.

При сближении источника и наблюдателя регистрируемая частота больше собственной, при удалении — меньше («красное смещение»). Предсказав этот эффект теоретически, Доплер пытался объяснить им различие в цвете компонент двойных звезд. Однако в случае сплошного спектра малый сдвиг частоты не приводит к заметным изменениям. В 1848 г. Физо высказал мысль, что по смещению спектральных линий можно обнаружить движение небесных тел относительно Земли. Эта идея сыграла важнейшую роль в астрономии. После открытия спектрального анализа и установления тождественности химических элементов на Земле и небесных телах эффект Доплера стали применять для определения радиальных (лучевых) скоростей небесных тел. Впервые таким способом в 1868 г. была измерена радиальная скорость Сириуса по смещению водородной линии в его спектре относительно той же линии в спектре разрядной трубки.

Для экспериментальной проверки принципа Доплера в оптике необходимо независимое измерение скорости источника. В первых попытках использовался свет от краев солнечного диска, скорость которых может быть найдена из движения солнечных пятен. Первое убедительное подтверждение принципа Доплера в условиях земной лаборатории было получено только в 1900 г. русским астрофизиком А. А. Белопольским. В опытах Штарка, начатых в 1905 г., наблю-

далось смещение спектральных линий, испускаемых быстродвижущимися ионами и атомами в каналовых лучах. Определенная по принципу Доплера скорость излучающих частиц оказалась в хорошем согласии с оценками этой скорости другими методами.

Расщепление спектральных линий некоторых звезд на две компоненты позволяет сделать вывод о том, что наблюдается система двух звезд, обращающихся вокруг центра масс. Если обычными или интерференционными методами эти звезды разрешить не удается, то систему называют спектрально-двойной звездой. С помощью эффекта Доплера можно определить скорости компонент и период обращения.

Измеряя доплеровское смещение линий в спектрах галактик, американский астроном Э. Хаббл сделал в 1929 г. на самом большом в то время телескопе с диаметром зеркала 2,5 м важнейшее открытие в астрофизике. Он установил, что удаленные галактики разбегаются, причем их скорость v растет пропорционально расстоянию R до них в соответствии с соотношением $v=HR$, получившим название *закона Хаббла*. В модели однородной и изотропной Вселенной закон Хаббла соответствует равномерному расширению, когда скорость удаления любых двух точек пропорциональна расстоянию между ними. Картина разбегания галактик выглядит одинаково из любой точки. Расширение Вселенной подтверждает нестационарную космологическую модель, построенную в 1922 г. советским ученым А. А. Фридманом на основе общей теории относительности.

В спектроскопии проявление эффекта Доплера состоит в том, что хаотическое тепловое движение испускающих свет атомов или ионов приводит к уширению наблюдаемых спектральных линий. В случае максвелловского распределения атомов по скоростям обусловленная эффектом Доплера форма спектральной линии описывается колоколообразной функцией Гаусса (см. § 1.8). Доплеровская ширина линии зависит от температуры ($\sim\sqrt{T}$), что используется в спектроскопических методах измерения температуры светящегося газа.

Первое экспериментальное подтверждение правильности квадратичных по v/c членов в релятивистской формуле (8.13) было получено Айвсом и Стилуэллом в 1938—1941 гг. Измерялся сдвиг частоты определенной спектральной линии света, испускаемого быстродвижущимися ($v/c\approx 5\cdot 10^{-3}$) атомами водорода в пучке каналовых лучей. Для выделения эффекта второго порядка по v/c в спектрограф одновременно направлялся свет, испускаемый атомами по направлению их движения, и отраженный от неподвижного зеркала свет, испускаемый против движения. Из-за квадратичных членов смещение соответствующих частот ω_1 и ω_2 относительно частоты ω_0 света, испускаемого неподвижными атомами, будет несимметричным:

$$\omega_{1,2}=\omega_0\sqrt{(1\pm v/c)/(1\mp v/c)}\approx\omega_0\left[1\pm v/c+{}^1/_2(v/c)^2\right]. \qquad (8.14)$$

Согласие экспериментальных результатов с вычислениями по релятивистской формуле (8.14) было очень хорошим. Фактически в этих опытах был впервые подтвержден релятивистский закон преобразования промежутков времени (8.5), так как с ним связан вклад в квадратичные по v/c члены доплеровского сдвига частоты.

Релятивистское преобразование промежутков времени проявляется не только в квадратичных членах продольного эффекта Доплера, но и в появлении *поперечного эффекта*. Когда направление на источник составляет прямой угол с его скоростью (угол должен быть прямым в системе отсчета, связанной с приемником), расстояние остается неизменным и весь эффект уменьшения наблюдаемой частоты связан с тем, что регистрируемый наблюдателем промежуток времени τ между двумя событиями больше собственного времени τ_0 между ними (т. е. времени в той системе, где они происходят в одном месте) в соответствии с формулой (8.5). Поэтому

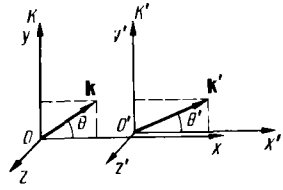
$$\omega = \omega_0 \sqrt{1 - v^2/c^2} \approx \omega_0 [1 - {}^1/_2 (v^2/c^2)]. \tag{8.15}$$

Сдвиг частоты в поперечном эффекте Доплера проявляется лишь с квадратичных по v/c членов. Напомним, что по классическим представлениям время абсолютно и потому поперечного эффекта Доплера не должно быть.

Приведем релятивистское объяснение эффекта Доплера и аберрации света с несколько иной точки зрения, рассматривая одну и ту же плоскую монохроматическую электромагнитную волну в двух инерциальных системах отсчета. При преобразовании от одной инерциальной системы отсчета к другой векторы $\mathbf{E}$ и $\mathbf{B}$ электромагнитного поля волны в каждой пространственно-временной точке $(\mathbf{r}, t)$ изменяются, но их скалярное произведение $\mathbf{EB}$ и разность квадратов $E^2 - c^2B^2$ остаются неизменными. Поэтому свойство ортогональности векторов $\mathbf{E}$ и $\mathbf{B}$ в бегущей электромагнитной волне в вакууме, а также соотношение (1.31) между их модулями выполняются сразу во всех инерциальных системах отсчета. Одно и то же значение во всех системах отсчета имеет, разумеется, и скорость c волны, но направление распространения и частота, вообще говоря, различны. Эффект Доплера можно рассматривать как преобразование частоты волны, а аберрацию — как преобразование ее направления при переходе от одной инерциальной системы отсчета к другой.

Состояние электромагнитного поля рассматриваемой волны в некоторой мировой точке $(ct, \mathbf{r})$, например максимум или нуль напряженности, не может зависеть от выбора системы отсчета. Так как это состояние определяется фазой волны $(\omega t - \mathbf{kr})$, то фаза должна быть инвариантом преобразований Лоренца. Инвариантность фазы можно пояснить еще и следующим образом. Представим себе цуг электромагнитных волн с одинаковой длиной волны, имеющий конечную протяженность. Число отдельных волн, т. е. периодов в этом цуге, определяется разностью значений фазы, соответствующих началу и концу цуга. Но число периодов, укладываю-

щихся на протяжении данного цуга, одинаково для всех наблюдателей. Поэтому фаза должна оставаться неизменной при переходе от одной системы отсчета к другой



8.8
Преобразование волнового вектора при переходе из одной системы отсчета в другую

Инвариантность фазы $(\omega t - \mathbf{kr})$ относительно преобразований Лоренца позволяет рассматривать это выражение как скалярное произведение 4-векторов: четырехмерного радиуса-вектора $(ct\ \mathbf{r})$ и четырехмерного волнового вектора $(\omega/c, \mathbf{k})$, пространственной компонентой которого служит трехмерный волновой вектор $\mathbf{k}$, а временно́й — частота волны ω, деленная на c. Для электромагнитной волны в вакууме $k=\omega/c$, поэтому четырехмерный волновой вектор имеет нулевую инвариантную длину.

Введение четырехмерного волнового вектора удобно потому, что закон преобразования его проекций при переходе из одной инерциальной системы отсчета в другую позволяет сразу найти преобразование частоты волны и ее направления, т. е. получить релятивистские выражения для эффекта Доплера и аберрации. Проекции четырехмерного волнового вектора волны в системе K $(\omega/c, k_x, k_y, k_z)$ выражаются через проекции в системе K' $(\omega'/c, k'_x, k'_y, k'_z)$ по формулам преобразований Лоренца (8.7), если в них сделать замену $ct \to \omega/c$, $x \to k_x$, $y \to k_y$, $z \to k_z$. Пусть в системе отсчета K направление волны образует угол θ с осью x (рис. 8.8), частота волны равна ω. Тогда $k=\omega/c$ и $k_x=(\omega/c)\cos\theta$, $k_y=(\omega/c)\sin\theta$, $k_z=0$. Подставляя эти величины в формулы (8.7), получаем:

$$\frac{\omega}{c}\cos\theta=\frac{\omega'}{c}\,\frac{\cos\theta'+v/c}{\sqrt{1-v^2/c^2}},\quad \frac{\omega}{c}\sin\theta=\frac{\omega}{c}\sin\theta', \tag{8.16}$$

$$\omega=\frac{\omega'+vk'_x}{\sqrt{1-v^2/c^2}}=\omega'\,\frac{1+(v/c)\cos\theta'}{\sqrt{1-v^2/c^2}}. \tag{8.17}$$

Будем для определенности систему отсчета K' связывать с источником волны, а систему K — с наблюдателем (т. е. с приемником). Формулу для аберрации можно получить, разделив почленно второе из равенств (8.16) на первое:

$$\operatorname{tg}\theta=\frac{\sin\theta'\sqrt{1-v^2/c^2}}{\cos\theta'+v/c}. \tag{8.18}$$

Если в системе K' направление волны перпендикулярно скорости наблюдателя $(\theta'=90°)$, то для системы K, связанной с наблюдателем, формула (8.18) дает $\operatorname{tg}\theta=\sqrt{c^2/v^2-1}$. Отсюда для угла аберрации $\delta=\pi/2-\theta$ получаем $\sin\delta=v/c$, что совпадает со значением, найденным выше из релятивистского закона преобразования скорости.

Формула (8.17) выражает воспринимаемую наблюдателем частоту ω волны через собственную частоту $\omega_0=\omega'$. Если источник

удаляется от наблюдателя вдоль соединяющей их линии, то $\theta'=\pi$ и (8.17) дает $\omega=\omega_0\sqrt{(1-v/c)/(1+v/c)}$, что совпадает с относящейся к этому случаю формулой (8.13). В общем случае вместо (8.17) удобнее иметь формулу, в которой частота ω выражалась бы не через угол θ', а через угол θ, характеризующий направление волны в системе отсчета K, связанной с наблюдателем. Для этого выразим $\cos\theta'$ с помощью формулы (8.16), подставив в нее предварительно ω/ω' из (8.17):

$$\cos\theta'=\frac{\cos\theta-v/c}{1-(v/c)\cos\theta}. \tag{8.19}$$

Подставляя это выражение* в (8.17), окончательно находим

$$\omega=\omega_0\frac{\sqrt{1-v^2/c^2}}{1-(v/c)\cos\theta}. \tag{8.20}$$

Эта формула описывает релятивистский эффект Доплера при произвольном направлении волны. Случай $\theta=0$ (или $\theta=\pi$) соответствует сближению (или удалению) источника и приемника вдоль одной прямой, т. е. продольному эффекту Доплера. Если волна приходит к наблюдателю по направлению, перпендикулярному скорости источника, то в формуле (8.20) следует положить $\theta=\pi/2$, что дает $\omega=\omega_0\sqrt{1-v^2/c^2}$. Это совпадает с приведенным выше выражением (8.15) для поперечного эффекта Доплера.

Контрольные вопросы

- ☐ Как в теории относительности объясняются отрицательные результаты всех опытов, в которых делались попытки обнаружить равномерное движение Земли с помощью оптических явлений?
- ☐ Как получить законы распространения света в равномерно движущихся относительно наблюдателя телах? Приведите релятивистское объяснение результатов опыта Физо.
- ☐ Какие причины вызывают изменеиие периода принимаемых сигналов при движении источиика или приемиика? В чем состоит главное отличие эффекта Доплера для электромагиитных волн в вакууме и для волн в среде? С чем связано возникновение поперечиого эффекта Доплера?
- ☐ Перечислите основные проявлеиия и применения эффекта Доплера в астрофизике и спектроскопии.
- ☐ Какие свойства и характеристики плоской электромагнитной волны в вакууме изменяются при переходе в другую систему отсчета и какие остаются без изменеиия? Дайте обосиование инвариантности фазы.
- ☐ Получнте релятивистские формулы для эффекта Доплера и аберрации, рассматривая преобразование четырехмерного волнового вектора.

* Заметим, что эта формула эквивалентна (8.18), т. е. также описывает аберрацию.

Задача

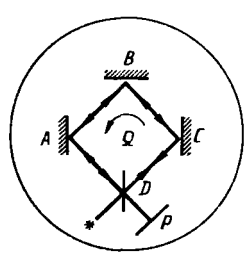
Пусть в системе отсчета K направление светового луча составляет угол θ с осью x. Направление этого луча в системе K' определяется углом $\theta' = \theta + \delta\theta$. Используя релятивистскую формулу (8.19), покажите, что в первом порядке по v/c угол аберрации $\delta\theta$ равен $(v/c)\sin\theta$.

8.4. Эффект Саньяка. Лазерный гироскоп

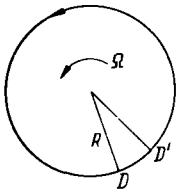
До сих пор в этой главе рассматривались оптические опыты, в которых движение тел происходило с постоянной по модулю и направлению скоростью относительно некоторой инерциальной системы отсчета (например, гелиоцентрической). В полном соответствии с принципом относительности все опыты, в которых делаются попытки обнаружить влияние *равномерного* движения лаборатории на оптические явления, дают отрицательный результат. Но *ускоренное* движение, в частности *вращение* относительно инерциальной системы отсчета, сказывается на характере протекающих явлений и может быть обнаружено как механическими, так и оптическими опытами.

В опыте Саньяка (1913) три зеркала A, B и C и полупрозрачная пластинка D вместе с источником света и фотопластинкой P смонтированы на платформе, которую можно приводить во вращение (рис. 8.9). Пластинка D делит свет от источника на два пучка, которые обегают контур $ABCD$ во встречных направлениях и затем образуют на фотопластинке интерференционную картину. Луч света, направление которого совпадает с направлением вращения, затратит на обход контура большее время, чем луч противоположного направления. Поэтому при вращении установки возникает разность фаз между интерферирующими лучами, пропорциональная угловой скорости платформы. В результате интерференционные полосы оказываются смещенными относительно их положения при неподвижной платформе.

Количественный расчет сдвига полос при вращении проведем, предположив для простоты, что свет распространяется по круговому контуру радиуса R (рис. 8.10). Результат будет справедлив и для контура произвольной формы. Точка D, где пучок от источника разделяется на два, обегающих контур в противоположных направлениях, движется в лабораторной системе отсчета по окружности со скоростью $v = \Omega R$, где Ω — угловая скорость платформы. Так как рассматриваемый эффект проявляет себя уже в первом порядке по v/c, то можно не принимать во внимание релятивистские эффекты сокращения длины и замедления времени, ибо они квадратичны по v/c. За время τ_1 обхода контура светом, распространяющимся по направлению вращения, точка D перемещается в положение D' и поэтому пройденное светом расстояние $c\tau_1$ равно



8.9
Схема опыта Саньяка



8.10
К расчету запаздывания встречных волн

$2\pi R + v\tau_1$. Отсюда

$$\tau_1 = \frac{2\pi R}{c-v} = \frac{2\pi R}{c}\,\frac{1}{1-v/c} \approx \frac{2\pi R}{c}\left(1+\frac{v}{c}\right).$$

Аналогично, время τ_2 обхода контура светом, распространяющимся навстречу, равно $(2\pi R/c)\times$ $\times(1-v/c)$. Разность $\tau_1-\tau_2$ дает относительное запаздывание встречных волн при вращении платформы:

$$\tau = (4\pi R/c)(v/c) = 4\pi R^2\Omega/c^2 = 4S\Omega/c^2, \qquad (8.21)$$

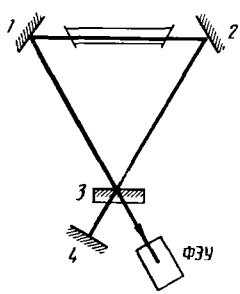
где $S=\pi R^2$ — площадь, охватываемая контуром. Разделив это запаздывание на период световых колебаний $T=\lambda/c$, найдем число полос ΔN, на которое происходит сдвиг интерференционной картины из-за вращения платформы:

$$\Delta N = 4S\Omega/(c\lambda). \qquad (8.22)$$

В условиях опыта Саньяка расчет по формуле (8.22) дает $\Delta N=$ $=0{,}079$; эксперимент дал 0,077.

В 1925 г. Майкельсон и Гэйл поставили опыт, в котором роль вращающейся платформы играла Земля. Из-за малой угловой скорости Ω в этом случае требуется большая площадь контура S. Майкельсон проложил около двух километров труб, охватывавших площадь 0,208 км^2, из которых был выкачан воздух для получения устойчивой интерференционной картины. Ожидаемое смещение составляло в таких условиях 0,236 полосы, а среднее значение по большому числу наблюдений оказалось равным $0{,}235\pm0{,}005$. Опыт Майкельсона—Гэйла можно рассматривать как оптический аналог механического опыта Фуко с маятником, цель которого состояла в обнаружении суточного вращения Земли.

В наши дни эффект Саньяка находит практическое применение для обнаружения и измерения вращательного движения без всяких внешних ориентиров. Основным элементом *оптического гироскопа* является *кольцевой лазер*, резонатор которого образован тремя (реже четырьмя) зеркалами (рис. 8.11). В отличие от линейного лазера с двухзеркальным резонатором, в котором формируется стоячая волна, кольцевой лазер может генерировать распространяющиеся по замкнутому контуру бегущие волны. При вращении платформы с лазером резонансные частоты для волн, обегающих контур во встречных направлениях, становятся различными. В самом деле, в условиях генерации за время обхода контура в каждой из волн



8.11
Кольцевой лазер с трехзеркальным резонатором

должно совершиться (одно и то же) целое число q полных колебаний: $\tau_1 = qT_1 = 2\pi q/\omega_1$, $\tau_2 = qT_2 = 2\pi q/\omega_2$. Подставляя $\tau_1 - \tau_2 = \tau$ в формулу (8.21), получаем

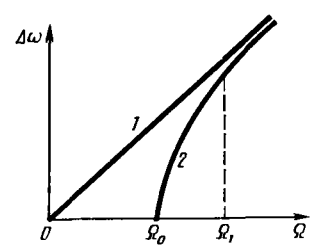
$$4S\Omega/c^2 = 2\pi q\,(1/\omega_1 - 1/\omega_2) \approx 2\pi q\Delta\omega/\omega^2,$$

где $\Delta\omega = \omega_2 - \omega_1$ — разность частот генерации для встречных волн. Учитывая, что $2\pi c/\omega = \lambda$ и $q = L/\lambda$ (L — периметр контура), находим выражение для $\Delta\omega$:

$$\Delta\omega = 8\pi S\Omega/(\lambda L). \qquad (8.23)$$

Таким образом, разность частот $\Delta\omega$ встречных волн в кольцевом лазере пропорциональна угловой скорости его вращения. Для измерения $\Delta\omega$ выходящие через частично прозрачное зеркало волны с помощью дополнительного зеркала *4* (рис. 8.11) совмещаются по направлению распространения и направляются на катод фотоумножителя. Ток ФЭУ содержит составляющую на частоте биений интенсивности, равной разности частот встречных волн. Чтобы составить представление о порядке величины расщепления частот, рассмотрим равносторонний трехзеркальный резонатор с периметром $L = 1$ м, работающий на длине волны $\lambda = 632$ нм и вращающийся вместе с Землей с угловой скоростью $\Omega = 15$ град/ч. По формуле (8.23) для этого случая имеем $\Delta\nu = \Delta\omega/(2\pi) = 21{,}7$ Гц.

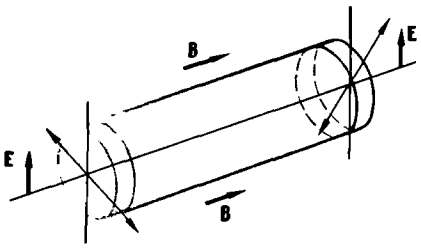
При практическом измерении угловой скорости вращения с помощью кольцевого лазера необходимо учитывать ряд явлений, усложняющих зависимость расщепления частот встречных волн от угловой скорости. Выражаемая формулой (8.23) линейная зависимость $\Delta\omega$ от Ω получена в предположении отсутствия связи встречных волн. Неизбежные отражения от неоднородностей зеркал, активной среды и других элементов, размещенных в резонаторе, приводят к тому, что часть энергии одной волны передается другой волне. В результате при относительно медленном вращении кольцевого лазера возникает *затягивание частоты,* проявляющееся в нелинейном характере зависимости $\Delta\omega$ от Ω (рис. 8.12), и взаимная *синхронизация* (захват) встречных волн, когда частота их биений обращается в нуль. С увеличением связи волн полоса синхронизации расширяется. Для расширения диапазона измерения малых значений угловых скоростей стремятся выбором конструкции кольцевого лазера обеспечить по возможности малую связь встречных волн либо используют метод «частотной подставки», позволяющий вывести тем или иным способом рабочую точку Ω_1 за пределы полосы синхронизации. Такую «подставку» можно получить дополнительным принудитель-



8.12

Зависимость расщепления частот встречных волн от угловой скорости Ω:

1 — теоретическая линейная зависимость (8.23); *2* — экспериментальная кривая



8.13
Схема ячейки Фарадея

ным вращением кольцевого лазера с заданной угловой скоростью либо включением в резонатор *невзаимных элементов,* обеспечивающих различие частот генерации для встречных волн уже при неподвижном резонаторе.

Наибольшее распространение в современных конструкциях лазерных гироскопов получили невзаимные устройства, основанные на различии фазовых скоростей волн правой и левой круговой поляризаций при их распространении в прозрачной среде, помещенной в продольное магнитное поле (ячейки Фарадея). Поскольку волны в резонаторе кольцевого лазера поляризованы линейно, на торцы стеклянного стержня (рис. 8.13) наклеены четвертьволновые пластинки, превращающие линейную поляризацию вне стержня в круговую внутри него. Магнитное поле в стержне создается с помощью соленоида или постоянного магнита. Оптическая длина такой ячейки различна для волн, распространяющихся навстречу. При помещении в резонатор с периметром 1 м ячейки длиной 1 см, находящейся в поле с индукцией 10^{-2} Тл, частота расщепления встречных волн составляет 65 кГц. Этого вполне достаточно для работы вдали от зоны синхронизации.

Лазерные гироскопы обладают рядом преимуществ по сравнению с гироскопами других типов и используются в системах навигации, стабилизации и управления кораблями, самолетами и космическими аппаратами, а также в геодезии и в измерительной технике.

Контрольные вопросы

Объясните, почему в опыте Саньяка возникает сдвиг интерференционных полос при вращении установки. Можно ли обнаружить вращение с помощью интерферометра Майкельсона?
Как в кольцевом лазере связана разность частот генерации для встречных волн с угловой скоростью вращения и с параметрами резонатора?
Каково назначение невзаимных элементов в конструкциях лазерных гироскопов?
Объясните принцип действия и устройство ячейки Фарадея.

ТЕРМОДИНАМИКА ИЗЛУЧЕНИЯ СВЕТОВЫЕ КВАНТЫ

● Электромагнитная теория света, дополненная представлениями электронной теории о взаимодействии света с веществом, позволила к концу прошлого столетия с исключительной простотой и ясностью разрешить многие проблемы, оказавшиеся непреодолимыми для старой волновой теории. Но в то же время возникли новые задачи, решение которых в рамках классической электромагнитной теории было невозможным. Так, например, безуспешны были все попытки получить количественное описание спектрального распределения света, излучаемого нагретыми телами. Безупречный с позиций классической физики вывод формулы Рэлея—Джинса приводил к противоречащему данным опыта и физически абсурдному результату. Решение проблемы было найдено Планком в конце 1900 г. введением гипотезы о квантовании энергии осциллятора, несовместимой с принципами классической физики. Эта гипотеза с очевидностью указывала на ограниченность классических представлений и дала толчок революционному процессу ломки старых понятий, завершившемуся созданием квантовой физики.

● Эйнштейн, развивая гипотезу Планка, в 1905 г. пришел к выводу о наличии у излучения корпускулярных свойств. Введенное им представление о квантах излучения — световых частицах, получивших впоследствии название фотонов, — позволило объяснить непонятные в рамках электромагнитной теории экспериментальные закономерности фотоэффекта. Проявление у излучения волновых и корпускулярных свойств, несовместимых с точки зрения классической физики, привело к формированию новых представлений о природе света, получивших название корпускулярно-волнового дуализма.

9.1. Тепловое излучение в замкнутой полости. Черное тело

Испускаемый источником свет уносит с собой энергию. В зависимости от того, где черпается эта энергия, различаются и виды свечения. Излучение, сопровождающееся химическими превращениями, называют *хемилюминесценцией*. Примером может служит свечение фосфора, медленно окисляющегося на воздухе. Испускание лучистой энергии в этом случае происходит за счет уменьшения внутренней энергии тела при изменении его химического состава. *Электролюминесценцией* называют свечение, возникающее при электрическом воздействии. В полупроводниках излучение может происходить при рекомбинации инжектированных через *p-n*-переход носителей тока (инжекционная электролюминесценция). Другой пример — свечение газов или паров в электрическом разряде. В этом случае необходимая для излучения энергия сообщается атомам или молекулам газа электронами, ускоряемыми электрическим полем разряда. Бомбардировка электронным пучком может вызвать также свечение твердых тел, например сернистого цинка, нанесенного на экран электронно-лучевой трубки (*катодолюминесценция*). Процессы излучения, вызываемые предварительным или одновременным освещением тела, объединяются под названием *фотолюминесценции*. При этом необходимая для излучения энергия доставляется светом от внешнего источника.

В тех случаях, когда необходимая энергия сообщается нагреванием, т. е. подводом теплоты, излучение называется *тепловым* или *температурным*. Среди различных видов свечения оно занимает особое место. В противоположность всем видам люминесценции это единственный вид излучения, который может находиться в состоянии термодинамического равновесия с телами. Поэтому физика теплового излучения представляет собой связующее звено между термодинамикой и оптикой. Объединение статистической механики и электромагнитной теории в проблеме теплового излучения разорвало рамки классической физики и дало начало одной из величайших революций в физике.

Чтобы составить представление о характере теплового излучения, рассмотрим несколько тел, нагретых до различной температуры и помещенных в замкнутую полость, стенки которой полностью отражают падающее на них излучение. Опыт показывает, что такая система рано или поздно приходит в состояние теплового равновесия, при котором все тела приобретают одинаковую температуру. Так происходит и в том случае, когда внутри полости абсолютный вакуум и тела могут обмениваться энергией только путем испускания и поглощения электромагнитных волн. Испускаемая телами лучистая энергия за каждый промежуток времени становится в среднем равной поглощаемой энергии, и плотность энергии излучения в пространстве между телами достигает определенной величины, соответствующей установившейся температуре тел. Макроскопически такое состояние излучения в полости остается неизменным во времени.

Это излучение, находящееся в термодинамическом равновесии с телами, имеющими определенную температуру, называется *равновесным* или *черным излучением*.

Плотность энергии равновесного излучения и его спектральный состав совершенно не зависят от размеров и формы полости и от свойств находящихся в ней тел. Свойства равновесного излучения зависят только от температуры. Поэтому можно говорить о температуре самого излучения, считая ее равной температуре тел, с которыми оно находится в тепловом равновесии. Равновесное излучение однородно, изотропно и неполяризовано, т. е. в каждой точке имеет одинаковую плотность и спектральный состав, а все направления распространения и все направления колебаний напряженности поля представлены с одинаковой вероятностью.

В пустой полости с идеально отражающими стенками поглощения и испускания света не происходит. Если каким-либо образом ввести в нее излучение, то направление его распространения будет изменяться при отражении от стенок, но спектральный состав сохранится прежним. Такое излучение неравновесно и неустойчиво. Сколь угодно малого отклонения свойств стенок от идеальности уже достаточно для того, чтобы излучение пришло к равновесию. Это произойдет и при идеальных стенках, если ввести в полость сколь угодно малое тело («пылинку»), способное поглощать и излучать. Такая «пылинка», не сказываясь на энергетическом балансе, переведет излучение в полости из любого неравновесного состояния в равновесное. От ее свойств зависит только время установления.

Для экспериментального изучения спектрального состава равновесного излучения можно проделать небольшое отверстие в стенках полости, поддерживаемых при определенной температуре. Излучение, выходящее наружу через отверстие, обладает точно таким же спектральным составом. От равновесного оно отличается только тем, что распространяется в пределах некоторого телесного угла в одном направлении, т е. оно не изотропно.

С увеличением температуры внутри полости будет возрастать энергия выходящего из отверстия излучения. При этом изменяется и его спектральный состав так, что максимум смещается в область более коротких волн: выходящий из горячей печи свет имеет красноватый оттенок при сравнительно невысокой температуре и становится желтым и даже белым по мере ее роста. Это отражает изменение объемной плотности U энергии равновесного излучения и ее спектрального распределения при повышении температуры. Для характеристики распределения энергии по длинам волн или по частотам вводят *спектральную плотность* излучения U_λ или U_ω, так что величина $U_\lambda d\lambda$ дает энергию единицы объема излучения с длинами волн в интервале от λ до $\lambda+d\lambda$, а $U_\omega d\omega$ — с частотами в интервале от ω до $\omega+d\omega$. Если это один и тот же спектральный интервал, то $U_\lambda d\lambda = U_\omega d\omega$. Учитывая, что $\lambda = 2\pi c/\omega$, отсюда легко выразить U_λ

через U_ω или наоборот (см. задачу 1). Очевидно, что

$$U=\int_0^\infty U_\lambda d\lambda=\int_0^\infty U_\omega d\omega. \tag{9.1}$$

В случае равновесного излучения спектральная плотность U_ω (или U_λ) представляет собой универсальную функцию только частоты (или длины волны) и температуры. Основная задача теории теплового излучения состоит в определении этой функции $U_\omega(T)$ (или $U_\lambda(T)$).

Универсальный характер спектральной плотности равновесного излучения, как впервые показал Кирхгоф в 1860 г., непосредственно вытекает из второго начала термодинамики. В самом деле, предположим противное, т. е. что $U_\omega(T)$ каким-то образом зависит от природы тела, с которым излучение находится в равновесии, например зависит от особенностей его спектра поглощения и испускания. Возьмем две полости, в которых излучение находится в равновесии с разными телами, имеющими одинаковую температуру. Соединим полости небольшим отверстием так, чтобы они могли обмениваться излучением. Если плотности энергии излучения в них различны, то возникнет направленный перенос лучистой энергии, который приведет к самопроизвольному нарушению теплового равновесия между телами (т. е. к установлению некоторой разности температур). Это противоречит второму началу термодинамики.

В то время как спектральное распределение энергии излучения, выходящего из отверстия в полости, имеет универсальный характер, для теплового излучения с открытой поверхности тела это не так: его спектральное распределение зависит не только от температуры, но и от материала поверхности. Для количественной характеристики этого спектрального распределения вводят понятие *испускательной способности* тела r_ω (или r_λ), т. е. спектральной плотности потока энергии излучения, испускаемого единичной площадкой поверхности по всем направлениям, так что $r_\omega d\omega$ (или $r_\lambda d\lambda$) представляет собой поток излучения в соответствующем спектральном интервале. Полный поток излучения всех длин волн представляет собой энергетическую светимость R поверхности (см. § 1.10). Очевидно, что

$$R=\int_0^\infty r_\omega d\omega=\int_0^\infty r_\lambda d\lambda. \tag{9.2}$$

Подчеркнем, что испускательная способность r_ω характеризует только тепловое излучение тела (все виды люминесценции исключаются). Для данного тела вид функции r_ω зависит только от его температуры. Испускательная способность тела не зависит от окружающей среды, в частности от того, находится тело в равновесии с излучением или нет.

Падающее на поверхность тела излучение частично отражается или рассеивается поверхностью, частично проходит через поверх-

ность. Ограничимся случаем, когда тело непрозрачно, т. е. толщина его достаточна для того, чтобы прошедшее в него излучение успевало полностью поглотиться, не достигнув противоположной поверхности. Тогда можно условно говорить о поглощении самой поверхностью. Поглощение, как правило, имеет селективный характер и зависит от температуры поверхности. Так, например, тонкий слой сажи практически полностью поглощает видимый свет, но в значительно меньшей степени инфракрасное излучение. Плавленый кварц прозрачен в широком интервале длин волн, но начинает заметно поглощать свет при температуре около 1500°С. Для характеристики таких свойств вводят *поглощательную способность* a_ω (или a_λ) тела, под которой понимают безразмерную величину, показывающую, какая доля энергии падающего излучения данной частоты (или длины волны) поглощается поверхностью. По определению, $a_\omega \leqslant 1$.

Опыт показывает, что чем больше поглощательная способность тела в некотором спектральном интервале, тем интенсивнее его тепловое излучение при той же температуре и для тех же длин волн. Например, кварц, окрашенный небольшими примесями редких земель, имеет полосы поглощения в видимой области. При нагревании такого кварца в спектре его теплового излучения можно наблюдать светлые полосы, их максимумы соответствуют областям поглощения при температуре, близкой к температуре свечения.

Существование упомянутой связи между тепловым излучением и поглощением следует из общих принципов термодинамики. Так как испускательная и поглощательная способности характеризуют само тело и не зависят от окружения, то для нахождения связи между ними можно рассмотреть частный случай, когда тело окружено равновесным излучением с той же температурой, например заключено в полость, стенки которой либо идеально отражают излучение, либо поддерживаются при той же температуре, что и тело. Выразим поток Φ_ω этого излучения в единичном спектральном интервале вблизи частоты ω, падающий на единичную площадку поверхности тела, через спектральную плотность U_ω равновесного излучения. Так как оно изотропно, то в пределах телесного угла $d\Omega = \sin\theta d\theta d\varphi$ распространяется энергия, составляющая $d\Omega/(4\pi)$ часть от всей энергии. Если выбранное направление образует угол θ с нормалью к поверхности, то поток $d\Phi_\omega$, падающий в пределах $d\Omega$ на единицу площади, равен $cU_\omega \cos\theta d\Omega/(4\pi)$. Для нахождения потока со всех направлений это выражение нужно проинтегрировать по φ в пределах от 0 до 2π и по θ в пределах от 0 до $\pi/2$:

$$\Phi_\omega = [cU_\omega/(4\pi)] \int_0^{2\pi} d\varphi \int_0^{\pi/2} \sin\theta \cos\theta d\theta = cU_\omega/4. \qquad (9.3)$$

Часть этого потока, равная $a_\omega \Phi_\omega$, поглощается, остальная часть $(1-a_\omega)\Phi_\omega$ отражается и рассеивается. С отраженным и рассеянным

потоком складывается поток собственного теплового излучения поверхности, равный испускательной способности r_ω поверхности. В соответствии с принципом детального равновесия полный исходящий от поверхности поток излучения $(1-a_\omega)\Phi_\omega+r_\omega$ должен быть равен падающему потоку Φ_ω из этого же спектрального интервала. Приравнивая эти выражения, находим $r_\omega/a_\omega=\Phi_\omega$, или [см. (9.3)]

$$r_\omega/a_\omega=cU_\omega/4 \qquad (9.4)$$

— отношение испускательной способности тела к его поглощательной способности с точностью до множителя $c/4$ совпадает со спектральной плотностью энергии равновесного излучения $U_\omega(T)$, которая, как уже отмечалось, представляет собой универсальную функцию частоты и температуры. Каждая из величин r_ω и a_ω может изменяться чрезвычайно сильно при переходе от одного тела к другому, но отношение r_ω/a_ω одинаково для всех тел. В этом заключается содержание закона Кирхгофа (1860).

Тело, которое поглощает целиком всю энергию падающего на него излучения, называется *черным*. Его поглощательная способность a_ω равна единице для всех частот и температур. Из закона Кирхгофа (9.4) следует, что испускательная способность черного тела только постоянным множителем $c/4$ отличается от спектральной плотности равновесного излучения: $r_\omega=cU_\omega/4$. Спектральное распределение теплового излучения черного тела будет таким же, как у равновесного излучения при той же температуре. Поэтому равновесное излучение иначе называют *черным излучением**.

Сажа и платиновая чернь в видимой области имеют поглощательную способность a_ω, близкую к единице. Но в далекой инфракрасной области a_ω и для них заметно меньше единицы. Абсолютно черных тел, как и других идеализированных объектов, в природе не существует. Но можно создать устройство, по своим свойствам сколь угодно близкое к абсолютно черному телу. Это уже рассмотренная выше замкнутая полость, в стенке которой имеется малое отверстие. Падающее извне излучение через отверстие проникает внутрь, попадает на стенки полости, частично поглощается ими, частично отражается или рассеивается и вновь попадает на стенки. Из-за малых размеров отверстия это произойдет многократно, прежде чем какая-то часть излучения снова попадет на отверстие. Поэтому практически весь падающий свет любой частоты полностью поглотится. Материал стенок полости значения не имеет. Отверстие полости в отношении падающего на него излучения и, следовательно, в отношении выходящего из него теплового излучения ведет себя как поверхность абсолютно черного тела с температурой, равной температуре стенок полости.

Описываемую модель черного тела можно проиллюстрировать

* Впрочем, как заметил Планк, его с ие меньшим основанием можио назвать и белым, имея в виду надлежащее обобщеиие того, что понимают под «совершенно белым светом».

простыми демонстрациями. Если взять коробку, покрашенную внутри белой краской, то малое отверстие в ней будет казаться совершенно черным даже на фоне выкрашенной черной краской наружной стенки. Черными выглядят снаружи открытые окна зданий, несмотря на то что в помещении стены и потолок светлые. Если же раскалить стенки полости из материала с малой поглощательной способностью, например фарфора, то ее отверстие будет ярко светиться на тусклом фоне прямого излучения самих стенок.

Источник в виде полости с малым отверстием, стенки которой поддерживаются при постоянной температуре, используется при количественных измерениях излучения черного тела. Аналогичным образом устроен стандартный излучатель, служащий эталоном единицы силы света (канделы).

На основе анализа экспериментальных данных Стефан в 1879 г. пришел к заключению, что полная испускаемая телом энергия при тепловом излучении пропорциональна четвертой степени термодинамической температуры. В 1884 г. Больцман, используя вытекающее из электродинамики Максвелла выражение для давления изотропного излучения $p=U/3$, с помощью термодинамических соображений теоретически получил эту закономерность для излучения черного тела:

$$R=\int_0^\infty r_\omega \, d\omega=\sigma T^4. \tag{9.5}$$

Соотношение (9.5) между энергетической светимостью черного тела и термодинамической температурой, называемое *законом Стефана—Больцмана,* было впоследствии подтверждено экспериментально. Числовое значение постоянной σ по современным измерениям составляет $5{,}67\cdot10^{-8}$ Вт/(м$^2\cdot$К^4). Следует отметить, что этот закон справедлив только для черного тела. Для нечерных тел не существует простой аналитической зависимости светимости от температуры.

Следующий важный шаг в теоретическом изучении равновесного излучения был сделан Вином в 1893 г. Рассматривая адиабатное сжатие равновесного излучения в полости с зеркальными стенками и учитывая изменение частоты света при отражении от движущегося зеркала (эффект Доплера), Вин показал, что спектральная плотность $U_\omega(T)$ должна иметь вид

$$U_\omega(T)=\omega^3 f(\omega/T), \tag{9.6}$$

где f — некоторая функция отношения частоты к температуре, конкретный вид которой в рамках электромагнитной теории и феноменологической термодинамики установить невозможно. Соотношение (9.6) сводит задачу нахождения универсальной функции $U_\omega(T)$, зависящей от двух аргументов, к задаче нахождения функции $f(\omega/T)$, зависящей только от одного аргумента ω/T. Это значит, что если известно спектральное распределение энергии равновесного

излучения при какой-либо одной произвольной температуре, то с помощью формулы (9.6) можно найти его при любой другой температуре.

Закон Вина (9.6) включает в себя как следствие и закон Стефана—Больцмана. Выражая испускательную способность черного тела r_ω через U_ω с помощью соотношения (9.4) и переходя при интегрировании по частотам к новой переменной $x=\omega/T$, получим

$$R=\frac{1}{4}c\int_0^\infty U_\omega(T)\mathrm{d}\omega=\frac{1}{4}c\int_0^\infty \omega^3 f\left(\frac{\omega}{T}\right)\mathrm{d}\omega=\frac{1}{4}cT^4\int_0^\infty x^3 f(x)\mathrm{d}x=\sigma T^4$$

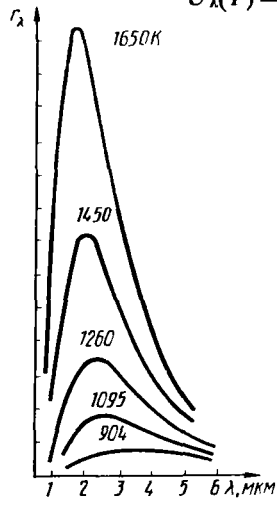
— интегральная светимость пропорциональна T^4, поскольку интеграл по x не зависит от T.

Измерения распределения по длинам волн интенсивности излучения из отверстия в полости, стенки которой поддерживаются при постоянной температуре (Люммер и Прингсгейм, 1899 г.), показали, что при некотором значении длины волны λ_m кривая $r_\lambda(T)$ имеет ярко выраженный максимум (рис. 9.1). С увеличением температуры максимум смещается в область более коротких длин волн, причем так, что произведение температуры на длину волны, отвечающую максимуму, остается прежним:

$$\lambda_m T=b. \qquad (9.7)$$

Значение константы $b=0{,}29$ см·К определено из опытных данных. Соотношение (9.7) получило название *закона смещения*. Оно непосредственно следует из термодинамической формулы Вина (9.6). Переходя в ней от частот к длинам волн, получим (см. задачу 1)

$$U_\lambda(T)=[(2\pi c)^4/\lambda^5]f(2\pi c/(\lambda T))=T^5 f_1(\lambda T), \qquad (9.8)$$



9.1
Испускательная способность черного тела

где $f_1(\lambda T)=(2\pi c)^4/(\lambda T)^5 f(2\pi c/(\lambda T))$ — новая неопределенная функция от аргумента λT. Обозначив $x=\lambda T$, при фиксированной температуре для определения положения максимума зависимости $U_\lambda(T)$ от λ получим уравнение

$$\mathrm{d}f_1(\lambda T)/\mathrm{d}\lambda=T\mathrm{d}f_1(x)/\mathrm{d}x, \text{ т. е. } \mathrm{d}f_1(x)/\mathrm{d}x=0,$$

которое имеет некоторый корень $x=x_m$. Это значит, что при любой температуре максимум зависимости $U_\lambda(T)$ от λ получается при одном и том же значении аргумента $x=x_m$, т. е. $\lambda_m T=x_m=\mathrm{const}$, что и доказывает закон смещения. Из (9.8) следует, что значение спектральной плотности $U_\lambda(T)$ в максимуме, т. е. при $\lambda=\lambda_m$, пропорционально пятой степени термодинамической температуры.

При практически достижимых температурах максимум испускательной способности черного тела лежит в инфракрасной области. Так, при $T = 3000$ К $\lambda_m = 0{,}96$ мкм. Только при $T > 5000$ К максимум попадает в зеленую область спектра, к которой глаз наиболее чувствителен. Светимость черного тела быстро растет с температурой ($\sim T^4$) и при высоких температурах достигает очень больших значений. При $T = 6000$ К (температура поверхности Солнца) с каждого квадратного сантиметра черного тела излучается поток мощностью 7,4 кВт, но лишь небольшая часть этой мощности приходится на видимую область спектра.

Рассмотренные выше закономерности теплового излучения используются в оптических методах измерения температуры (оптическая пирометрия).

Контрольные вопросы

- □ При каких условиях тепловое излучение будет термодинамически равновесным? Перечислите основные свойства равновесного излучения. Будет ли равновесным излучение, выходящее из маленького отверстия в полости?
- □ Как испускательная способность тела связана с его энергетической светимостью? Докажите с помощью второго закона термодинамики, что отношение испускательной и поглощательной способностей одинаково для всех тел.
- □ Покажите, что испускательная способность черного тела с точностью до множителя $c/4$ совпадает со спектральной плотностью равновесного излучения.
- □ Как на опыте можно реализовать черное тело?
- □ Как энергетическая светимость абсолютно черного тела зависит от температуры?
- □ Каким образом по измеренной спектральной зависимости испускательной способности черного тела r_ω при некоторой температуре T построить график r_{ω_1} для другой температуры T_1?
- □ Как изменяется положение максимума на кривых r_{ω_1} и r_λ при изменении температуры черного тела?

Задача

Выразить спектральную плотность U_λ энергии равновесного излучения в шкале длин волн через спектральную плотность U_ω в шкале частот.

◇ Объемная плотность энергии излучения в интервале частот от ω до $\omega + d\omega$ равна $U_\omega d\omega$. Эта же энергия, выраженная через длины волн, равна $U_\lambda d\lambda$. Подставим в $U_\omega d\omega$ $\omega = 2\pi c/\lambda$ и $d\omega = -(2\pi c/\lambda^2)\, d\lambda$. Знак минус здесь означает только то, что с возрастанием длины волны частота убывает, поэтому его можно опустить. Таким образом, $U_\lambda = (2\pi c/\lambda^2) U_\omega$. Если для U_ω воспользоваться законом Вина (9.6), выражающим ее через универсальную функцию f, то $U_\lambda(T) = (2\pi c)^4/\lambda^5 f(2\pi c/(\lambda T))$. ◇

9.2. Спектральная плотность равновесного излучения. Формула Планка

Изложенными в § 9.1 результатами исчерпывается все то, что может дать электромагнитная теория и феноменологическая термодинамика в проблеме теплового излучения. Термодинамическая формула Вина (9.6) теоретически обос-

новывает законы смещения (9.7) и Стефана — Больцмана (9.5), но числовые значения входящих в них постоянных b и σ оставляет неопределенными, так как выражает спектральную плотность равновесного излучения $U_\omega(T)$ через некоторую функцию $f(\omega/T)$, вид которой в этих рамках установить невозможно. Для решения основной задачи теории теплового излучения — нахождения испускательной способности черного тела $r_\omega(T)$ или связанной с ней соотношением $r_\omega = cU_\omega/4$ спектральной плотности энергии равновесного излучения U_ω — необходимо применение статистических методов с привлечением какой-либо конкретной физической модели.

Из термодинамических соображений ясно, что вид искомой функции $U_\omega(T)$ не должен зависеть от того, с какими телами излучение находится в тепловом равновесии. Поэтому, следуя Планку, рассмотрим простейший пример излучающего тела — линейный гармонический осциллятор с собственной частотой ω_0, зарядом e и массой m (электрон, связанный квазиупругой силой).

Пусть осциллятор находится в замкнутой полости, заполненной равновесным излучением с температурой T. Под действием поля излучения со сплошным спектром $U_\omega(T)$ он совершает вынужденные колебания. Благодаря резонансным свойствам осциллятора эти колебания будут иметь заметно отличную от нуля амплитуду лишь в узкой области частот вблизи собственной частоты осциллятора ω_0. При этом поглощаемая осциллятором мощность $P_{\text{погл}}$ может быть выражена через значение спектральной плотности излучения U_{ω_0} на частоте ω_0. В динамическом равновесии с излучением поглощаемая мощность $P_{\text{погл}}$ в среднем равна испускаемой осциллятором мощности $P_{\text{исп}}$, которая, в свою очередь, может быть выражена через среднюю энергию $\langle\varepsilon\rangle$ осциллятора при температуре T. Таким путем можно связать $U_{\omega_0}(T)$ со средней энергией $\langle\varepsilon\rangle$ теплового возбуждения осциллятора. Последняя вычисляется методами статистической механики. Так как все это справедливо для осциллятора с произвольным значением ω_0, то такой путь позволяет рассчитать спектральную плотность равновесного излучения на всех частотах.

Вынужденное движение осциллятора под действием гармонической внешней силы $F_x(t) = eE_x(t) = \operatorname{Re}[eE_{0x}\exp(-i\omega t)]$ рассматривалось в § 2.3 при изучении теории дисперсии. Установившиеся колебания происходят на той же частоте ω: $x(t) = \operatorname{Re}[x_0\exp(-i\omega t)]$, где комплексная амплитуда x_0 в соответствии с (2.32) равна

$$x_0 = \frac{e/m}{\omega_0^2 - \omega^2 - 2i\gamma\omega} E_{0x}. \tag{9.9}$$

Здесь постоянная затухания γ дается формулой (1.73).

Среднюю поглощаемую осциллятором мощность (т. е. работу, совершаемую над ним в 1 с внешним полем) $P_{\text{погл}} = \langle eE_x(t)\dot{x}(t)\rangle$ проще всего найти, если учесть, что при установившихся колебаниях она равна средней мощности силы радиационного трения

$F_{тр}(t)=2m\gamma\dot{x}(t)$: $P_{погл}=\langle F_{тр}(t)\dot{x}(t)\rangle=2m\gamma\langle\dot{x}^2(t)\rangle$. Так как $\dot{x}(t)=$ $=\mathrm{Re}\,[-i\omega x_0\exp(-i\omega t)]$, то $\langle x^2(t)\rangle=\omega^2|x_0|^2/2$. Поэтому

$$P_{погл}=m\gamma\omega^2|x_0|^2=\frac{\gamma\omega^2e^2/m}{(\omega_0^2-\omega^2)^2+4\gamma^2\omega^2}E_{0x}^2. \tag{9.10}$$

Нас интересует случай, когда действующее на осциллятор поле излучения изотропно и не поляризовано. Выразим квадрат его проекции на ось осциллятора E_{0x}^2 через объемную плотность энергии $U=\varepsilon_0\langle E^2(t)\rangle=\varepsilon_0(\langle E_x^2(t)\rangle+\langle E_y^2(t)\rangle+\langle E_z^2(t)\rangle)$. Так как в изотропном поле все направления колебаний вектора $\mathbf{E}(t)$ представлены одинаково, то $U=3\varepsilon_0\langle E_x^2(t)\rangle={}^3/_2\varepsilon_0E_{0x}^2$ и в правую часть (9.10) можно подставить $E_{0x}^2={}^2/_3U/\varepsilon_0$.

В немонохроматическом поле такое выражение для поглощаемой мощности справедливо для каждой из его монохроматических составляющих, не скоррелированных между собой. Мощность, поглощаемая осциллятором в спектральном интервале от ω до $\omega+d\omega$, получится из (9.10), если заменить U на $U_\omega d\omega$. Тогда полная поглощаемая мощность в интервале частот от 0 до ∞

$$P_{погл}=\frac{2\gamma e^2}{3m\varepsilon_0}\int_0^\infty\frac{\omega^2}{(\omega_0^2-\omega^2)^2+4\gamma^2\omega^2}U_\omega d\omega. \tag{9.11}$$

Для всех частот ω_0, лежащих в оптическом диапазоне, выполняется сильное неравенство $\gamma\ll\omega_0$. При этом условии первый сомножитель в подынтегральном выражении в (9.11) имеет резко выраженный резонансный характер: он имеет острый максимум при $\omega=\omega_0$ и быстро приближается к нулю, как только ω отклоняется от ω_0 более чем на γ. Поэтому заметный вклад в интеграл дает лишь узкая область частот вблизи $\omega=\omega_0$. Второй сомножитель U_ω — искомое спектральное распределение излучения — представляет собой плавную функцию частоты, так что в пределах этой области его можно считать постоянным и вынести за интеграл, заменив значением при $\omega=\omega_0$. Кроме того, можно заменить в первом сомножителе ω на ω_0, а $\omega_0^2-\omega^2$ на $2\omega_0(\omega_0-\omega)$. При $\gamma\ll\omega_0$ нижний предел интегрирования можно распространить до $-\infty$. Тогда

$$P_{погл}=\frac{\gamma e^2}{6m\varepsilon_0}U_{\omega_0}\int_{-\infty}^{\infty}\frac{d\omega}{(\omega_0-\omega)^2+\gamma^2}=\frac{\pi e^2}{6m\varepsilon_0}U_{\omega_0}. \tag{9.12}$$

Теперь нужно рассчитать испускаемую осциллятором мощность $P_{исп}$. Как следует из (9.9), вынужденные колебания осциллятора под действием излучения с непрерывным спектром U_ω имеют заметную амплитуду только в узкой полосе частот вблизи собственной частоты ω_0 осциллятора. Поэтому при расчете $P_{исп}$ можно считать, что осциллятор совершает гармонические колебания с частотой $\omega=\omega_0$, и воспользоваться результатами § 1.5. Согласно форму-

лам (1.72) — (1.73), $P_{\text{исп}}=2\gamma\langle\varepsilon\rangle$, где $\langle\varepsilon\rangle$ — средняя энергия осциллятора, а $\gamma=[1/(4\pi\varepsilon_0)]\,e^2\omega_0^2/(3mc^3)$. Таким образом,

$$P_{\text{исп}}=\frac{1}{4\pi\varepsilon_0}\frac{2e^2\omega_0^2}{3mc^3}\langle\varepsilon\rangle. \tag{9.13}$$

Так как в состоянии динамического равновесия с излучением энергия осциллятора в среднем остается неизменной, поглощаемая мощность должна быть равна испускаемой. Приравнивая правые части (9.12) и (9.13), получаем

$$U_\omega=\omega^2\langle\varepsilon\rangle/(\pi^2c^3). \tag{9.14}$$

Здесь опущен индекс у ω, так как это условие должно выполняться для осциллятора с любой собственной частотой.

Соотношение (9.14) связывает спектральную плотность $U_\omega(T)$ равновесного излучения со средней энергией $\langle\varepsilon\rangle$ осциллятора при температуре T. Заряд e и масса m осциллятора, т. е. его частные характеристики, выпали из этой формулы, ибо в тепловом равновесии как $U_\omega(T)$, так и $\langle\varepsilon\rangle$ определяются только температурой.

Классическая статистическая механика дает для средней энергии линейного осциллятора при температуре T значение $\langle\varepsilon\rangle=kT$, где $k=1{,}38\cdot10^{-23}$ Дж/К — постоянная Больцмана. Это частный случай закона классической статистики о *равнораспределении,* согласно которому в тепловом равновесии на каждую степень свободы в среднем приходится $^1/_2kT$ кинетической энергии. Для осциллятора, совершающего колебания на собственной частоте, средние значения кинетической и потенциальной энергии одинаковы, так что средняя энергия теплового возбуждения каждой колебательной степени свободы составляет kT:

$$\langle\varepsilon\rangle=\langle\varepsilon_{\text{к}}\rangle+\langle\varepsilon_{\text{п}}\rangle=2\langle\varepsilon_{\text{к}}\rangle=kT. \tag{9.15}$$

Если подставить классическое значение средней энергии осциллятора (9.15) в соотношение (9.14), то для спектральной плотности равновесного излучения $U_\omega(T)$ получается выражение, известное под названием *формулы Рэлея — Джинса:*

$$U_\omega(T)=\omega^2kT/(\pi^2c^3). \tag{9.16}$$

Это соотношение согласуется, как и следовало ожидать, с формулой Вина (9.6), которая как непосредственное следствие термодинамики должна выполняться независимо от конкретной модели. В самом деле, (9.16) совпадет с (9.6), если там неопределенную функцию f положить равной $f(\omega/T)=kT/(\pi^2c^3\omega)$. В длинноволновой области, т. е. при малых ω, формула Рэлея — Джинса (9.16) хорошо согласуется и с экспериментальными результатами: в этой области испускательная способность черного тела растет пропорционально квадрату частоты. Но при дальнейшем увеличении

частоты обнаруживается резкое расхождение с экспериментом, так как (9.16) предсказывает такой рост спектральной плотности для всех частот, вплоть до бесконечно больших значений при $\omega \to \infty$. В соответствии с (9.16) в полости при любой конечной температуре должны преобладать ультрафиолетовые и рентгеновские лучи. Из-за неограниченного роста U_ω с частотой получается, что объемная плотность U энергии излучения бесконечно велика: интеграл в (9.1) с U_ω из (9.16) расходится. Если же считать полную энергию конечной, то по этой теории получается, что при установлении равновесия между телом и излучением вся энергия тела перейдет в энергию высоких частот излучения, т. е. термодинамическое равновесие при конечной плотности энергии излучения вообще невозможно. В то же время опыт показывает, что равновесие устанавливается, плотность энергии излучения U остается при этом конечной, а ее спектральное распределение U_ω растет с частотой, достигает максимума при некоторой частоте и затем снова падает.

Таким образом, безупречный с точки зрения классической физики вывод дает очевидно абсурдную формулу (9.16), находящуюся в разительном противоречии с опытом. Такое положение П. Эренфест назвал «ультрафиолетовой катастрофой». По выражению Лоренца, уравнения классической физики оказались неспособными объяснить, почему угасшая печь не испускает синих лучей наряду с излучением больших длин волн.

В 1900 г. Планк получил формулу для спектральной плотности $U_\omega(T)$ равновесного излучения, хорошо согласующуюся с опытом при всех частотах. Оказалось, что для теоретического вывода этой формулы необходима гипотеза, коренным образом противоречащая представлениям классической физики. Планк предположил, что энергия осциллятора может принимать не любые, а только вполне определенные дискретные значения ε_n, отделенные друг от друга конечными интервалами. Переход осциллятора из одного состояния в другое сопровождается поглощением или испусканием конечной порции (кванта) энергии излучения. В такой системе с дискретным энергетическим спектром среднюю энергию $\langle\varepsilon\rangle$ в тепловом равновесии при температуре T уже нельзя находить по формуле (9.15). Вероятность p_n того, что осциллятор находится в состоянии с энергией ε_n, в соответствии с распределением Больцмана пропорциональна $\exp[-\varepsilon_n/(kT)]$, но при вычислении средних значений интегралы заменяются суммами:

$$\langle\varepsilon\rangle = \sum_n \varepsilon_n p_n = \sum_n \varepsilon_n \exp[-\varepsilon_n/(kT)] / \sum_n \exp[-\varepsilon_n/(kT)]. \qquad (9.17)$$

Еще одна гипотеза необходима для установления значений энергии ε_n разрешенных состояний осциллятора. По предположению Планка, гармонический осциллятор имеет эквидистантный энергетический спектр, так что энергия ε_n в состоянии с номером n составляет целое кратное наименьшей порции энергии ε_0: $\varepsilon_n = n\varepsilon_0$,

где $n=0, 1, 2, ...$*. Тогда стоящая в знаменателе формулы (9.17) сумма представляет собой просто бесконечную геометрическую прогрессию со знаменателем $\exp(-\beta\varepsilon_0)$, где $\beta=1/(kT)$:

$$S(\beta)=\sum_{n=0}^{\infty}\exp(-\beta n\varepsilon_0)=[1-\exp(-\beta\varepsilon_0)]^{-1}. \tag{9.18}$$

Легко видеть, что сумма, стоящая в числителе (9.17), равна производной $dS(\beta)/d\beta$, взятой с противоположным знаком:

$$\sum_{n=0}^{\infty} n\varepsilon_0\exp(-\beta n\varepsilon_0)=-\frac{d}{d\beta}\sum_{n=0}^{\infty}\exp(-\beta n\varepsilon_0)=$$
$$=[1-\exp(-\beta\varepsilon_0)]^{-2}\exp(-\beta\varepsilon_0)\varepsilon_0. \tag{9.19}$$

Подставляя (9.19) и (9.18) в (9.17), находим

$$\langle\varepsilon\rangle=\varepsilon_0[\exp(\beta\varepsilon_0)-1]^{-1}=\varepsilon_0/(e^{\varepsilon_0/(kT)}-1). \tag{9.20}$$

Этот результат совпадает с классическим выражением (9.15) только в предельном случае при $\varepsilon_0/(kT)\to 0$.

Подставляя полученное значение средней энергии квантового осциллятора** (9.20) в (9.14), получаем для спектральной плотности равновесного излучения вместо формулы Рэлея — Джинса следующее выражение:

$$U_\omega=\frac{\omega^2\varepsilon_0}{\pi^2c^3}\frac{1}{\exp[\varepsilon_0/(kT)]-1}. \tag{9.21}$$

Чтобы оно не противоречило термодинамической формуле Вина (9.6), согласно которой температура может входить в U_ω только в комбинации ω/T, необходимо принять, что

$$\varepsilon_0=\hbar\omega, \tag{9.22}$$

где $\hbar$ — универсальная постоянная (постоянная Планка), имеющая размерность произведения энергии на время. В результате для спектральной плотности равновесного излучения получаем *формулу Планка:*

$$U_\omega(T)=\frac{\hbar\omega^3}{\pi^2c^3}\frac{1}{\exp[\hbar\omega/(kT)]-1}. \tag{9.23}$$

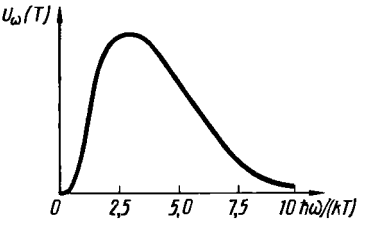
График функции $U_\omega(T)$ приведен на рис. 9.2. Формула Планка прекрасно согласуется с опытом при всех частотах и температурах. Для малых частот (длинных волн) и высоких температур, когда $\hbar\omega/(kT)\ll 1$, экспоненту можно разложить в ряд по степеням $\hbar\omega/(kT)$. Если ограничиться линейным членом этого разложения,

---

* Современная квантовая механика дает для уровней энергии гармонического осциллятора значения $\varepsilon_n=(n+1/2)\hbar\omega$.

** В оригинальной работе Планка 1900 г. средняя энергия осциллятора находилась путем вычисления средней энтропии на основе соотношения Больцмана $S=k\ln W$ между энтропией и термодинамической вероятностью W.

то формула Планка (9.23) перейдет в формулу Рэлея — Джинса (9.16). Но в области высоких частот и низких температур формула Планка предсказывает совершенно иное поведение $U\omega(T)$, нежели классическая теория. При $\hbar\omega/(kT)\gg 1$ экспонента в знаменателе гораздо больше единицы. Поэтому поведение $U_\omega(T)$ в области коротких волн приближенно описывается формулой



9.2
График функции $U_\omega(T)$, построенный по формуле Планка (9.23)

$$U_\omega(T)\approx\frac{\hbar\omega^3}{\pi^2c^3}\exp\left(-\frac{\hbar\omega}{kT}\right), \tag{9.24}$$

которая была предложена Вином в 1896 г. для описания экспериментальных результатов в этой области спектра. Экспоненциальное убывание спектральной плотности с ростом частоты устраняет «ультрафиолетовую катастрофу» классической теории.

Между предельными случаями, соответствующими применимости формул Рэлея — Джинса (9.16) и Вина (9.24), лежит обширная область, в которой и находится максимум кривой спектрального распределения. При повышении температуры этот максимум в согласии с законом смещения (9.7) сдвигается в сторону коротких волн, причем значение постоянной b в (9.7) может быть теперь найдено из решения трансцендентного уравнения (см. задачу 1):

$$b=1{,}265\hbar c/k. \tag{9.25}$$

Объемная плотность U энергии равновесного излучения (9.1) при интегрировании по всем частотам с U_ω из формулы Планка (9.23) получается конечной (см. задачу 2). Константа σ в законе Стефана — Больцмана (9.5) приобретает теперь вполне определенное теоретическое выражение (см. задачу 2):

$$\sigma=\pi^2k^4/(60c^2\hbar^3). \tag{9.26}$$

Соотношения (9.25) и (9.26) выражают константы b и σ, непосредственно измеряемые на опыте, через фундаментальные физические постоянные $\hbar$ и k. Исторически первое определение числового значения $\hbar$, как и наиболее точное для того времени (1900 г.) определение значения постоянной Больцмана k, было сделано Планком на основе закона Стефана — Больцмана (8.5) и закона смещения (9.7). Полученные Планком значения очень близки к принятым в настоящее время:

$\hbar=1{,}05\cdot10^{-34}$ Дж·с; $h=2\pi\hbar=6{,}626\cdot10^{-34}$ Дж·с;

$k=1{,}38\cdot10^{-23}$ Дж/К.

Поучительна история открытия закона излучения (9.23). Подлинная история создания теории далеко не всегда отражает логику ее построения. Формула Ви-

ла (9.24), квантовая по своей сути, была установлена раньше, чем классическая формула Рэлея — Джинса (9.16). (Работа Джинса, вызвавшая «ультрафиолетовую катастрофу», датируется 1905 г.) Вин исходил из аналогии между наблюдаемой на опыте экспоненциальной зависимостью U_ω от $\hbar\omega/(kT)$ в области высоких частот и максвелловским распределением молекул газа по скоростям. В теоретическом выводе формулы (9.24) Вин, развивая идеи русского физика В. А. Михельсона, делает ряд произвольных допущений, содержащих в скрытом виде квантовую гипотезу о связи частоты излучения с энергией элементарного процесса испускания света. Но в области низких частот полученная Вином формула не давала хорошего описания экспериментальных данных. К формуле (9.23) Планк пришел сначала полуэмпирическим путем, стараясь улучшить согласие с экспериментом в области низких частот, т. е., как это ни парадоксально, именно в той области, где квантовые свойства несущественны. Первое сообщение Планка 19 октября 1900 г. так и называлось: «Об одном улучшении закона излучения Вина». Полученная им формула была проверена на опыте Рубенсом, а также Луммером и Прингсгеймом. Хорошее согласие с опытом побудило Планка к настойчивым поискам истинного содержания своей формулы. И через два месяца, 14 декабря 1900 г., Планк делает доклад, в котором приводится полное решение поставленной задачи. Эту дату можно считать днем рождения квантовой теории. Планк доказал, что формулу (9.23) можно вывести только в том случае, если допустить квантование энергии, противоречащее классическим представлениям.

Нелегко было примириться с таким отказом от классических представлений, и Планк, совершив великое открытие, еще в течение нескольких лет пытался понять квантование энергии с позиций классической теории. Безуспешность этих попыток привела его к окончательному выводу, что в рамках классической физики излучение черного тела понять невозможно.

Постоянная Планка $\hbar$ играет в квантовой физике такую же роль, как скорость света c в релятивистской физике. Эти фундаментальные мировые константы определяют *границы применимости классического описания.*

В масштабах макромира числовое значение постоянной Планка чрезвычайно мало. Этим объясняется широкая применимость классической физики с лежащей в ее основе концепцией непрерывности к описанию макроскопических явлений. Решение проблемы теплового излучения исторически было первым шагом на пути к разгадке «тайны потерянной константы». Впоследствии ограниченность представлений классической физики обнаружила себя при исследовании фотоэффекта (см. § 9.5) и при попытках объяснения устойчивости атомов и закономерностей в спектрах из излучения. В начале века была создана так называемая «старая квантовая теория», в основе которой лежат гипотеза Планка о дискретном характере испускания и поглощения света осциллятором, введенное Эйнштейном представление о квантах света (фотонах) и уравнение фотоэффекта, построенная Бором теория простейших атомов. Но старая квантовая теория не представляла собой стройной, логически замкнутой науки. Удачно описав некоторые экспериментальные факты, она не могла дать правильного объяснения и количественного описания всего многообразия явлений микромира. С наступлением второй четверти нашего столетия начинается период создания современной квантовой теории с ее надежными логически непротиворечивыми основными положениями и адекватным математическим аппаратом.

Фундаментальная константа $\hbar$ — постоянная Планка, играющая выдающуюся роль в современной физике, — может быть определена экспериментально не только с помощью законов излучения черного тела, но и другими, более прямыми и точными методами. Некоторые из них рассмотрены ниже. Значения $\hbar$, полученные на основе разных физических явлений (тепловое излучение, фотоэффект, коротковолновая граница сплошного рентгеновского спектра, эффект Джозефсона в сверхпроводимости и др.), хорошо согласуются друг с другом.

Контрольные вопросы

- ☐ Как связана мощность, поглощаемая одномерным гармоническим осциллятором в поле равновесного излучения, со спектральной плотностью энергии поля?
- ☐ Как испускаемая осциллятором мощность выражается через его среднюю энергию?
- ☐ Получите выражение, связывающее спектральную плотность энергии равновесного излучения со средней энергией осциллятора, находящегося в термодинамическом равновесии с излучением.
- ☐ К каким теоретическим результатам для зависимости U_ω от частоты и для объемной плотности энергии U приводит применение классической статистической механики?
- ☐ Как найти среднюю энергию теплового возбуждения осциллятора при температуре T, если предположить, что его энергия может принимать только дискретные значения $\varepsilon_n = n\varepsilon_0$? В каком случае результат совпадает с тем, что дает теорема о равнораспределении?
- ☐ Какую зависимость U_ω от частоты дает формула Планка? Рассмотрите также предельные случаи низких и высоких частот.
- ☐ Какую роль в современной физике играет постоянная Планка?

Задачи

1. Вычислить константу b в законе смещения, основываясь на формуле Планка (9.23).

◇ Переходя в формуле Планка от частот к длинам волн (см. задачу 1 в § 9.1), получим

$$U_\lambda = (2\pi c/\lambda^2)U_\omega = (8\pi hc/\lambda^5)/\{\exp[hc/(\lambda kT)] - 1\}.$$

Для нахождения максимума функции U_λ введем вместо λ безразмерную переменную $x = hc/(\lambda kT)$. Тогда задача сводится к нахождению положения минимума функции $f_\lambda(x) = x^{-5}(e^x - 1)$. Приравнивая к нулю ее первую производную, получаем трансцендентное уравнение $x = 5(1 - e^{-x})$, корень которого легко найти методом последовательных приближений: $x_0 = 5$; $x_1 = 5(1 - e^{-5}) = 4{,}966$; $x_2 = 5(1 - e^{-x_1}) = 4{,}965$. Отсюда $b = \lambda_m T = hc/(kx_m) = 0{,}201hc/k = 2{,}90 \cdot 10^{-3}$ м/К.

Если же рассматривать спектральную плотность U_ω в шкале частот, то ее максимум соответствует минимуму функции $f_\omega(x) = x^{-3}(e^x - 1)$, где $x = \hbar\omega/(kT) = hc/(\lambda kT)$ имеет прежний смысл. Положение минимума находится из трансцендентного уравнения $x = 3(1 - e^{-x})$, которое имеет корень $x'_m = 2{,}821$. Отсюда $\omega_m/T = kx'_m/\hbar = 2{,}821\, k/\hbar$. Максимуму функции U_ω соответствует длина волны $\lambda'_m = 2\pi c/\omega_m = hc/(kTx'_m) > \lambda_m$. Таким образом, максимум спектральной плотности в шкале частот сдвинут в длинноволновую сторону по отношению к максимуму спектральной плотности в шкале длин волн:

$$\lambda'_m/\lambda_m = x'_m/x_m = 1{,}760. \qquad \diamond$$

2. Пользуясь формулой Планка, найти значение постоянной σ в законе Стефана — Больцмана.

◇ Для нахождения светимости R абсолютно черного тела нужно проинтегрировать испускательную способность $r_\omega = cU_\omega/4$ по всем частотам:

$$R = \frac{c}{4}\int_0^\infty U_\omega d\omega = \frac{\hbar}{4\pi^2c^2}\int_0^\infty \frac{\omega^3 d\omega}{\exp[\hbar\omega/(kT)] - 1} = \frac{k^4T^4}{4\pi^2c^2\hbar^3}\int_0^\infty \frac{x^3 dx}{e^x - 1}.$$

Здесь введена безразмерная переменная интегрирования $x=\hbar\omega/(kT)$. Рассматривая $1/(e^x-1)=e^{-x}/(1-e^{-x})$ как сумму геометрической прогрессии $e^{-x}(1+e^{-x}+e^{-2x}+\ldots)$, почленным интегрированием получаем

$$\int_0^\infty \frac{x^3\,dx}{e^x-1}=6\left(1+\frac{1}{2^4}+\frac{1}{3^4}+\ldots\right)=\frac{\pi^4}{15}.$$

Таким образом, $\sigma=\pi^2k^4/(60c^2\hbar^3)=5{,}67\cdot10^{-8}$ Вт/(м$^2\cdot$К^4).

9.3. Световые кванты. Спонтанное и вынужденное излучения

Гипотеза Планка о квантовании энергии осциллятора приводит, как было показано в § 9.2, к формуле (9.23), которая прекрасно подтверждается экспериментом. Однако эта гипотеза не сразу встретила признание. Считалось, что новые идеи представляют собой лишь искусственный математический прием, которому можно найти интерпретацию в рамках классической физики. Но все попытки такого рода оказались неудачными.

Революционные идеи Планка были оценены по достоинству и получили дальнейшее развитие прежде всего в работах Эйнштейна. Он первый указал на то, что кроме теплового излучения существуют и другие явления, которые можно объяснить на основе квантовой гипотезы. В частности, поведение теплоемкости твердых тел при низких температурах (отклонения от закона Дюлонга и Пти) получает объяснение, если для средней энергии осциллятора использовать квантовое выражение (9.20). Основанная на этих идеях количественная теория теплоемкости твердых тел была развита Дебаем.

В 1905 г. Эйнштейн выдвинул гипотезу световых квантов. Он предположил, что дискретный характер присущ не только процессам испускания и поглощения света, но и самому свету. Гипотеза о корпускулярных свойствах света позволила объяснить результаты экспериментов по фотоэффекту, совершенно непонятные с позиций классической электромагнитной теории (см. § 9.5). Однако представление о свете как потоке классических корпускул несовместимо с эмпирически совершенно явными волновыми свойствами света. Эйнштейн пришел к заключению, что «природа излучения должна быть не такой, какой мы ее считаем в настоящее время». За этими словами скрывается то, что теперь принято называть двойственной природой света или корпускулярно-волновым дуализмом (см. § 9.6). Корпускулярный аспект излучения проявляется наиболее отчетливо в коротковолновой части спектра, где для спектральной плотности $U_\omega(T)$ справедлива формула Вина (9.24), волновой аспект — в длинноволновой, где применима формула Рэлея—Джинса (9.16). Ни один из этих аспектов не дает полного представления об излучении, ибо для полного объяснения наблюдаемых явлений необходимо их сочетание. Закон излучения Планка (9.23), который содержит в себе предельные случаи, соответствующие корпускулярному и волновому аспектам, представляет собой

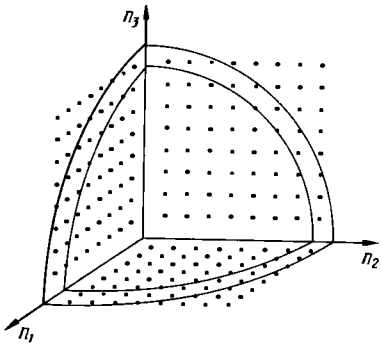
диалектический синтез двух, казалось бы, несовместимых теорий и снимает противоречие между ними.

Но вывод закона излучения по методу Планка, приведенный в § 9.2, в какой-то мере неудовлетворителен, поскольку он во многом основан на законах классической физики и лишь частично использует квантовые представления. В самом деле, формула (9.14), связывающая спектральную плотность энергии равновесного излучения $U_\omega(T)$ со средней энергией $\langle\varepsilon\rangle$ осциллятора, получена чисто классическим путем, так как поглощение и испускание света осциллятором рассчитывалось с помощью классической электродинамики, в то время как при нахождении $\langle\varepsilon\rangle$ использована квантовая гипотеза о дискретных энергетических уровнях осциллятора. Успех такой эклектической теории связан со спецификой выбранной модели: для осциллятора, как это уже отмечалось при обсуждении классической теории дисперсии (см. § 2.3), классическое и квантовомеханическое рассмотрение процессов поглощения и испускания приводит к одинаковым результатам.

Другая трактовка равновесного излучения, восходящая к Рэлею, состоит в том, чтобы само электромагнитное поле в полости рассматривать как набор осцилляторов. Можно говорить о собственных колебаниях этого поля и применить к ним методы статистической механики, а не вводить вспомогательный планковский осциллятор, взаимодействующий с излучением. Пусть для определенности полость имеет форму куба с ребром L, а ее стенки — зеркальные. Собственные нормальные колебания поля в таком объемном резонаторе представляют собой стоячие волны различных частот. Полное поле можно представить как суперпозицию таких стоячих волн, и в энергетическом отношении оно ведет себя как система невзаимодействующих гармонических осцилляторов. Для нахождения спектральной плотности энергии поля нужно подсчитать число независимых стоячих волн в полости с частотами в интервале от ω до $\omega+d\omega$. Как и в одномерном случае струны, закрепленной на концах, здесь для любого нормального колебания необходимо, чтобы вдоль каждого ребра укладывалось целое число полуволн. Пусть направление волны (нормаль к плоскостям равных фаз) образует углы α, β и γ с ребрами куба. Проекция любого ребра на это направление должна быть равна целому числу полуволн:

$$L\cos\alpha=n_1\lambda/2, \qquad L\cos\beta=n_2\lambda/2, \qquad L\cos\gamma=n_3\lambda/2.$$

Возводя в квадрат и складывая эти равенства, с учетом того, что $\cos^2\alpha+\cos^2\beta+\cos^2\gamma=1$, получаем $L^2=(n_1^2+n_2^2+n_3^2)\,(\lambda/2)^2$. Отсюда следует, что частота $\omega=2\pi c/\lambda=(n_1^2+n_2^2+n_3^2)^{1/2}\pi c/L$ каждого нормального колебания определяется суммой квадратов трех целых чисел n_1, n_2, n_3. Нормальное колебание удобно изобразить точкой в трехмерном пространстве. Совокупность таких точек образует кубическую решетку. Поэтому число dN различных колебаний с частотами в интервале от ω до $\omega+d\omega$ равно числу точек с цело-



9.3
К подсчету числа нормальных колебаний поля

численными координатами в пределах лежащего в первом октанте шарового слоя (рис. 9.3) радиусом $L\omega/(\pi c)$ и толщиной $L\mathrm{d}\omega/(\pi c)$:

$$\mathrm{d}N = L^3\omega^2\mathrm{d}\omega/(2\pi^2c^3).$$

Кроме того, следует учесть, что каждому нормальному колебанию, задаваемому целыми числами n_1, n_2, n_3, соответствуют две независимые стоячие волны с ортогональными состояниями поляризации.

Применяя гипотезу Планка к отдельному нормальному колебанию с частотой ω, будем считать, что его энергия может быть равна целому числу элементарных квантов $\hbar\omega$: $\varepsilon = n\hbar\omega$. Тогда в состоянии теплового равновесия средняя энергия, приходящаяся на одно нормальное колебание, выражается формулой (9.20) с $\varepsilon_0 = \hbar\omega$: $\langle\varepsilon\rangle = \hbar\omega/\{\exp[\hbar\omega/(kT)] - 1\}$. Умножая $\langle\varepsilon\rangle$ на число независимых колебаний поля $\omega^2\mathrm{d}\omega/(\pi^2c^3)$, приходящихся на 1 м^3 объема полости, получаем для спектральной плотности равновесного излучения $U_\omega(T)$ формулу Планка (9.23). Этот результат выведен здесь для кубической полости, но он должен быть верен для полости любой формы (если интересоваться волнами, длина волны которых мала по сравнению с ее размерами). В противном случае, вопреки тому, что нам известно о равновесном излучении, его плотность $U_\omega(T)$ зависела бы от формы полости.

Если же к осцилляторам поля применить теорему о равнораспределении энергии в состоянии теплового равновесия, т. е. положить среднюю энергию, приходящуюся на одно нормальное колебание, равной kT, то мы сразу придем к формуле Рэлея—Джинса (9.16). Причина «ультрафиолетовой катастрофы» проявляется здесь со всей отчетливостью: чем выше частота ω, тем большее число осцилляторов поля приходится на единичный интервал частот. Это число $\omega^2/(\pi^2c^3)$ растет неограниченно при $\omega \to \infty$. Ясно, что классический закон равнораспределения становится неприменимым для осцилляторов поля с высокими частотами.

В приведенном здесь выводе формулы Планка предполагалось, что каждое нормальное колебание (мода) электромагнитного поля в полости может обладать энергией $\varepsilon = n\hbar\omega$. Возникает вопрос, как нужно интерпретировать целое число n в рамках гипотезы световых квантов. Предположение, что энергия одного кванта света может быть равна $n\hbar\omega$, противоречит опыту (фотоэффект, см. § 9.5). Поэтому нужно считать n числом тождественных световых квантов (фотонов) в одной моде. Когда n уменьшается или увеличивается на единицу, говорят о поглощении одного фотона или испускании одного фотона

в данную моду. В состоянии теплового равновесия среднее число фотонов в моде

$$\langle n\rangle = \langle \varepsilon\rangle/(\hbar\omega) = \{\exp[\hbar\omega/(kT)]-1\}^{-1}. \qquad (9.27)$$

Такое распределение характерно для газа невзаимодействующих тождественных частиц, подчиняющихся статистике Бозе—Эйнштейна. Неклассический характер этой корпускулярной картины проявляется здесь в принципиальной неразличимости световых частиц — фотонов.

В отличие от рассмотренной Планком простейшей *модели* системы, взаимодействующей с излучением, — гармонического осциллятора — к процессам испускания и поглощения света *реальными атомами* классическое описание неприменимо. Исчерпывающее описание этих процессов дается квантовой электродинамикой, но их можно исследовать и на основе простой феноменологической теории, развитой Эйнштейном в 1916 г. В этой теории впервые было введено представление о вынужденном излучении, которое впоследствии нашло практическое применение в квантовых генераторах когерентного излучения — мазерах и лазерах.

Будем считать, что свободный атом может находиться только в стационарных состояниях с определенной энергией ε_1, ε_2, ...Переход атома из одного стационарного состояния в другое может происходить скачком в результате поглощения или испускания электромагнитного излучения, причем для такого элементарного процесса выполняется закон сохранения энергии: $\varepsilon_m - \varepsilon_n = \hbar\omega$ — энергия поглощаемого или испускаемого фотона равна разности энергий соответствующих стационарных состояний атома. Эти квантовые представления о строении атома и характере его взаимодействия с излучением, обобщающие гипотезу Планка о гармоническом осцилляторе, были введены Бором в 1913 г. и полностью подтверждаются современной квантовой теорией.

Пусть свободный, не подверженный внешним влияниям атом находится в возбужденном состоянии с энергией ε_2. С точки зрения классической физики атом начинает излучать сразу же после того, как он попадает в возбужденное состояние, и этот процесс длится в течение некоторого времени τ (см. § 1.5). По квантовым представлениям, самопроизвольный переход атома при отсутствии внешних воздействий из возбужденного состояния ε_2 в основное ε_1 с испусканием фотона *(спонтанное излучение)* происходит мгновенно, скачком. В какой именно момент произойдет этот переход, предсказать невозможно. Момент испускания фотона есть случайная величина, суждения о которой могут носить лишь статистический характер. Обозначим через A_{21} *вероятность спонтанного перехода* атома в единицу времени из возбужденного состояния в основное. Рассмотрим совокупность очень большого числа одинаковых атомов, которые образуют настолько разреженный газ, что взаимодействием между атомами можно пренебречь. Пусть в момент времени t в первом возбужденном состоянии ε_2 находится N_2 атомов. В тече-

ние промежутка времени от t до $t+\mathrm{d}t$ часть из них спонтанно перейдет в основное состояние. Невозможно указать, какие именно атомы совершат переход, но, зная вероятность A_{21}, можно указать среднее число $\mathrm{d}N_{21}$ таких переходов:

$$\mathrm{d}N_{21}=A_{21}N_2\mathrm{d}t. \tag{9.28}$$

Если при этом никаких процессов возбуждения атомов не происходит, то изменение числа N_2 возбужденных атомов за промежуток времени от t до $t+\mathrm{d}t$, согласно (9.28), равно

$$\mathrm{d}N_2=-A_{21}N_2\mathrm{d}t. \tag{9.29}$$

Коэффициент Эйнштейна A_{21} характеризует рассматриваемый атом и не зависит от времени. Поэтому решение уравнения (9.29) имеет вид

$$N_2(t)=N_{20}^1\exp(-A_{21}t)=N_{20}\exp(-t/\tau), \tag{9.30}$$

где N_{20} — число возбужденных атомов при $t=0$.

Таким образом, при отсутствии внешних воздействий в результате спонтанного излучения число возбужденных атомов убывает со временем по экспоненциальному закону. Промежуток времени $\tau=1/A_{21}$, в течение которого N_2 уменьшается в e раз, равен среднему времени жизни атома в возбужденном состоянии. По такому же экспоненциальному закону (9.30) должно убывать со временем свечение газа возбужденных атомов. Напомним, что радиационное затухание колебаний классического осциллятора формально описывается точно таким же законом (см. § 1.5). Однако физический смысл времени жизни τ в этих случаях совершенно различен. Согласно классической электродинамике, все излучающие осцилляторы одновременно совершают затухающие колебания и время τ одинаково для всех. По квантовым представлениям, спонтанное излучение — это совокупность независимых переходов: один из возбужденных атомов может вернуться в основное состояние через короткий промежуток времени, другой может прожить в возбужденном состоянии значительно дольше, но среднее для большой совокупности атомов время жизни τ имеет вполне определенную величину*.

Радиационное затухание собственных колебаний классического возбужденного осциллятора приводит к тому, что излучаемый при этом свет характеризуется не одной частотой, а узким спектральным распределением, заполняющим интервал частот $\Delta\nu\sim1/\tau$. Контур такой спектральной линии имеет лоренцевскую форму (см. § 1.7). На квантовом языке это означает, что спонтанному излучению атома при переходе из возбужденного состояния в основное соответствует узкий, но конечный интервал частот. Так как частота излучения определяется условием Бора $h\nu=\varepsilon_2-\varepsilon_1$, то

* Для высоколежащего возбужденного состояния атома с энергией ε_k обратное время жизни равно сумме вероятностей перехода в 1 с в основное состояние и все нижележащие возбужденные состояния с энергией $\varepsilon_i<\varepsilon_k$: $^1/_{\tau_k}=\Sigma A_{ki}$.

разброс частот свидетельствует о конечной ширине $\Delta\varepsilon$ энергетического уровня возбужденного состояния ε_2. Это заключение находится в полном соответствии с соотношением неопределенностей Гейзенберга между энергией и временем $\Delta\varepsilon\tau \gtrsim \hbar$, связывающим неопределенность в изменении энергии системы с неопределенностью момента времени, когда это изменение происходит.

Экспериментальная проверка закона распада (9.30) и определение среднего времени жизни τ возбужденного состояния наиболее непосредственно были произведены в опытах Вина (см. § 1.5) со свечением каналовых лучей. Этим методом для красной линии водорода (λ=656,2 нм) было получено $\tau = 1{,}5 \cdot 10^{-8}$ с, для резонансной линии ртути (λ=253,7 нм) $\tau = 9{,}8 \cdot 10^{-8}$ с.

Статистический, случайный характер процессов спонтанного излучения приводит к тому, что фазы, направления распространения и состояния поляризации световых волн, испускаемых отдельными атомами, не согласованы друг с другом. Это значит, что спонтанное излучение *некогерентно.*

В электромагнитном поле кроме спонтанного испускания будут происходить и процессы возбуждения атомов, т. е. переходы из основного состояния в возбужденное с поглощением фотонов с энергией $\hbar\omega = \varepsilon_2 - \varepsilon_1$. Вероятность такого перехода в единицу времени пропорциональна плотности энергии электромагнитного поля U_ω на частоте перехода ω и некоторому коэффициенту B_{12}, характеризующему вероятность возбуждения атома. Среднее число $\mathrm{d}N_{12}$ переходов из основного состояния в возбужденное за промежуток времени от t до $t+\mathrm{d}t$ пропорционально также числу N_1 атомов в основном состоянии:

$$\mathrm{d}N_{12} = N_1 B_{12} U_\omega \,\mathrm{d}t. \tag{9.31}$$

Коэффициент Эйнштейна B_{12} зависит от выбранных состояний ε_1 и ε_2 атома и может быть рассчитан методами квантовой механики. Здесь его можно рассматривать как феноменологическую постоянную. Заметим, что B_{12} определен так, что он не зависит от спектральной плотности энергии U_ω поля. Пропорциональность числа $\mathrm{d}N_{12}$ актов поглощения (9.31) величине U_ω соблюдается тогда, когда спектральная плотность U_ω плавно зависит от частоты ω вблизи частоты перехода.

Допустим теперь, что атомы находятся в термодинамическом равновесии с полем излучения. Тогда на основании принципа детального равновесия число переходов с испусканием фотонов и с поглощением должно быть одинаково. Основываясь на этом, можно, приравнивая правые части (9.28) и (9.31), найти вид функции U_ω, т. е. спектральную плотность энергии равновесного излучения. Однако при этом для U_ω получается не формула Планка (9.23), а ее предельный случай при $\hbar\omega/(kT) \gg 1$, т. е. формула Вина (9.24). Чтобы таким способом получить согласующуюся с опытом формулу Планка, необходимо предположить, как впервые показал

Эйнштейн, что электромагнитное поле вызывает не только переходы из основного состояния в возбужденное, но и обратные переходы из возбужденного состояния в основное, сопровождающиеся испусканием фотонов. Такие переходы под действием внешнего поля в отличие от спонтанных получили название *индуцированного* или *вынужденного* (стимулированного) излучения. Индуцированное излучение, как и спонтанное, имеет классический аналог. Осциллятор в поле световой волны будет совершать вынужденные колебания. В неустановившемся режиме вблизи резонанса в зависимости от соотношения фаз между колебаниями осциллятора и внешнего поля энергия может переходить как от поля к осциллятору (поглощение), так и от осциллятора к полю (вынужденное испускание).

Число вынужденных переходов $\mathrm{d}N_{21}$ за промежуток времени от t до $t+\mathrm{d}t$ пропорционально спектральной плотности энергии U_ω поля на частоте перехода, числу N_2 атомов в возбужденном состоянии ε_2 и некоторому коэффициенту B_{21}, характеризующему вероятность такого перехода в атоме. С учетом спонтанного излучения полное число переходов за $\mathrm{d}t$ из возбужденного состояния в основное равно

$$\mathrm{d}N_{21}=N_2(A_{21}+B_{21}U_\omega)\,\mathrm{d}t. \tag{9.32}$$

В состоянии термодинамического равновесия следует приравнять правые части выражений (9.31) и (9.32) и учесть, что населенности уровней ε_1 и ε_2 связаны соотношением Больцмана:

$$N_2/N_1=\exp[-(\varepsilon_2-\varepsilon_1)/(kT)]=\exp[-\hbar\omega/(kT)]. \tag{9.33}$$

Легко убедиться в том, что для данной пары уровней коэффициенты Эйнштейна B_{12} и B_{21} равны друг другу. В самом деле, при очень высокой температуре плотность энергии U_ω становится настолько большой, что в формуле (9.32) можно пренебречь первым слагаемым по сравнению со вторым. Это значит, что в равновесии при высокой температуре вынужденное испускание преобладает над спонтанным. Приравнивая для этих условий правые части (9.31) и (9.32), имеем $N_1B_{12}=N_2B_{21}$. Но в равновесии при $kT/(\hbar\omega)\to\infty$ населенности уровней, как видно из (9.33), выравниваются: $N_1=N_2$. Поэтому $B_{12}=B_{21}$. Коэффициенты B_{12} и B_{21} зависят только от свойств атома и не зависят от внешних условий, в которых происходят переходы. Поэтому равенство $B_{12}=B_{21}$, полученное для предельного случая $T\to\infty$, справедливо всегда, в том числе и в отсутствие теплового равновесия.

При произвольной температуре из равенства правых частей (9.31) и (9.32) с учетом (9.33) и $B_{12}=B_{21}$ для спектральной плотности равновесного излучения получаем

$$U_\omega=\frac{A_{21}/B_{21}}{\exp[\hbar\omega/(kT)]-1}. \tag{9.34}$$

Это выражение совпадает с формулой Планка (9.23) при $A_{21}/B_{21} = \hbar\omega^3/(\pi^2 c^3)$. Таким образом, все три коэффициента Эйнштейна связаны между собой.

Наиболее важный результат, который следует из вывода формулы Планка по Эйнштейну, — это заключение о существовании вынужденного излучения. Только в этом случае теория Эйнштейна не противоречит законам теплового излучения. При $\hbar\omega/(kT) \gg 1$ в (9.34) можно пренебречь единицей по сравнению с $\exp[\hbar\omega/(kT)]$ и (9.34) перейдет в формулу Вина. К такой формуле для U_ω мы бы пришли сразу, если бы не приняли во внимание вынужденное испускание. Физически это означает, что при $\hbar\omega \gg kT$ для поддержания термодинамического равновесия практически достаточно спонтанного испускания. Вынужденное испускание в таких условиях заметно меньше поглощения, так как при $kT \ll \hbar\omega$ термически возбужденных атомов мало. Напротив, при низких частотах и высоких температурах, когда $\hbar\omega/(kT) \ll 1$, в тепловом равновесии переходы с вынужденным испусканием происходят почти так же часто, как и переходы с поглощением.

Вынужденное излучение обладает замечательными свойствами, на что впервые обратил внимание П. Дирак в 1927 г., применив квантовую механику к полю излучения. В каждом акте вынужденного испускания происходит увеличение на единицу числа фотонов в той моде излучения, под действием которой произошел переход. Все фотоны одной моды тождественны. Это значит, что новый фотон неотличим от фотонов, вызывающих его испускание. Частота, фаза, направление распространения и поляризация волн, испущенных при вынужденных переходах, точно такие же, как у излучения, вызвавшего переходы.

С классической точки зрения пояснить когерентный характер вынужденного излучения можно следующим образом. Процесс вынужденного испускания обратен по отношению к поглощению (иногда его называют отрицательным поглощением). При распространении светового пучка в поглощающей среде происходит уменьшение интенсивности, но полностью сохраняются свойства когерентности. Это следует хотя бы из того, что в любых интерференционных опытах прохождение исходного пучка света через серый фильтр, уменьшающий интенсивность без изменения спектрального состава, не разрушает интерференционной картины. Поэтому можно ожидать, что при прохождении света через среду, содержащую возбужденные атомы, в результате вынужденного испускания будет происходить усиление распространяющейся волны при сохранении ее когерентности.

Коэффициенты Эйнштейна B_{12} и B_{21} характеризуют переходы между уровнями энергии ε_1 и ε_2 рассматриваемого атома под действием падающего излучения, спектральная плотность которого U_ω практически постоянна в пределах контура спектральной линии, соответствующей данному переходу. Обладающее сплошным спек-

тром равновесное излучение удовлетворяет этому требованию. Чтобы получить выражения для числа переходов dN_{12} и dN_{21} под действием монохроматического излучения с объемной плотностью U, нужно в формулах (9.31) и (9.32) заменить U_ω на $UF(\omega)$, где $F(\omega)$ есть нормированная на единицу функция формы спектральной линии рассматриваемого перехода (т. е. $F(\omega)d\omega$ дает ту долю полного числа переходов между данными уровнями, для которой частота лежит в интервале между ω и $\omega+d\omega$).

Рассмотрим связь коэффициентов Эйнштейна с макроскопическими оптическими параметрами среды. Пусть в разреженной среде распространяется параллельный пучок монохроматического излучения с частотой ω, лежащей в пределах контура спектральной линии для каких-либо двух уровней энергии атомов среды. Изменение числа dN фотонов в потоке при прохождении слоя толщиной $dx=c\,dt$ за счет процессов поглощения и вынужденного испускания в соответствии с (9.31) и (9.32) равно

$$dN=(N_2B_{21}-N_1B_{12})UF(\omega)\,dx/c. \qquad (9.35)$$

Вкладом спонтанного испускания в направленный зондирующий поток можно пренебречь, так как спонтанное излучение распространяется по всем направлениям и в фиксированное направление рассматриваемого пучка попадает ничтожная его часть. Формула (9.35) позволяет выяснить, какие условия необходимы для непосредственного экспериментального обнаружения вынужденного испускания. Так как $B_{21}=B_{12}$, то $dN=(N_2-N_1)B_{12}UF(\omega)\,dx/c$ и видно, что при распространении зондирующего пучка поглощение преобладает над вынужденным испусканием и интенсивность пучка убывает во всех случаях, когда $N_2<N_1$, т. е. число возбужденных атомов меньше числа атомов в основном состоянии. Так обычно и обстоит дело, если пучок распространяется в среде, находящейся в состоянии термодинамического равновесия или близком к нему. Чтобы наблюдать нарастание интенсивности зондирующего пучка (отрицательное поглощение) и тем самым экспериментально выявить вынужденное испускание, необходимо создать в среде неравновесное состояние, при котором число атомов на более высоком энергетическом уровне было бы больше, чем на низком ($N_2>N_1$). Первая попытка обнаружить вынужденное испускание в видимой области спектра на опыте в парах ртути, возбужденных электрическим разрядом в неравновесное состояние, была предпринята В. А. Фабрикантом (1939 г.). Им же впервые была высказана идея использования вынужденного излучения для усиления света.

Найдем закон изменения интенсивности пучка по мере его распространения в среде. Средняя плотность потока энергии S волны равна произведению объемной плотности U на скорость света: $S=cU$. Так как изменение потока фотонов на интервале dx равно $c\,dN$, то изменение потока энергии dS на том же отрезке dx рав-

но $\hbar\omega c\mathrm{d}N$. Умножая обе части равенства (9.35) на $\hbar\omega c$, найдем

$$\mathrm{d}S=(\hbar\omega/c)(N_2-N_1)B_{12}F(\omega)S\,\mathrm{d}x, \tag{9.36}$$

откуда

$$\mathrm{d}S/\mathrm{d}x=-\alpha S(x), \tag{9.37}$$

$$\alpha\equiv-(\hbar\omega/c)(N_2-N_1)B_{12}F(\omega).$$

Если населенности уровней N_1 и N_2 не зависят от x (однородная среда), то решение уравнения (9.37) имеет вид

$$S(x)=S_0\exp(-\alpha x). \tag{9.38}$$

Здесь S_0 — плотность потока энергии зондирующего пучка при $x=0$. Если $\alpha>0$, что бывает при $N_2<N_1$, то поток энергии экспоненциально убывает по мере распространения пучка. Мы приходим к *закону Бугера* (2.29), а формула (9.37) выражает коэффициент поглощения α через коэффициент Эйнштейна B_{12} и населенности уровней N_1 и N_2.

В тех случаях, когда в среде выполнено условие $N_2>N_1$, известное как *инверсия населенностей,* в (9.37) $\alpha<0$ (отрицательный коэффициент поглощения) и интенсивность волны в соответствии с (9.38) нарастает по мере ее распространения. Усиление падающего пучка света осуществляется за счет того, что при $N_2>N_1$ переходы с вынужденным испусканием фотонов происходят чаще, чем переходы с поглощением. Так как возникающие при вынужденном испускании фотоны тождественны с фотонами, вызвавшими испускание, когерентные свойства исходного пучка полностью сохраняются. Таков принцип действия *квантового усилителя* излучения. Различные способы создания необходимой для его работы среды с инверсией населенностей *(активной среды)* рассмотрены в § 9.4. Важно отметить, что для создания активной среды всегда требуется подведение извне дополнительной энергии, которая затем при вынужденном испускании частично преобразуется в энергию усиливаемого электромагнитного излучения.

Квантовый усилитель можно превратить в генератор излучения, если ввести в нем положительную обратную связь. Для этого часть выходящего из усилителя излучения должна возвращаться в активную среду. Тогда отпадает необходимость во входном сигнале, так как происходит самовозбуждение системы. Обратную связь можно осуществить, помещая активную среду в резонатор, образованный двумя параллельными зеркалами.

Бурно развивающаяся область науки и техники, занимающаяся исследованием и применением квантовых явлений для усиления, генерации и преобразования когерентного излучения, называется *квантовой электроникой.* Действие приборов квантовой электроники основано на явлении вынужденного излучения при квантовых переходах электронов, входящих в состав атомов, ионов, молекул и кристаллов.

Контрольные вопросы

- ☐ Как подсчитать число различных нормальных колебаний электромагнитного поля в кубической полости, приходящихся на интервал частот от ω до $\omega + d\omega$?
- ☐ Получите формулу Планка для спектральной плотности равновесного излучения, рассматривая излучение в полости как совокупность стоячих волн (нормальных колебаний).
- ☐ Покажите, что при экспоненциальном распаде возбужденного состояния по закону (9.30) вследствие спонтанного излучения среднее время жизни возбужденного атома равно $\tau = 1/A_{21}$.
- ☐ Зависят ли коэффициенты Эйнштейна A_{21}, B_{12}, B_{21} от спектральной плотности действующего на атом поля излучения?
- ☐ Укажите основные моменты вывода формулы Планка по Эйнштейну. Почему при этом необходимо предположить существование вынужденного излучения?
- ☐ Как связаны между собой коэффициенты Эйнштейна A_{21}, B_{12} и B_{21}? Почему эта связь, полученная для частного случая теплового равновесия, справедлива всегда?
- ☐ Какими свойствами характеризуется вынужденное излучение? Сравните его со спонтанным излучением.
- ☐ При каких условиях возможно нарастание интенсивности пучка монохроматического света по мере его распространения в среде? Поясните принцип действия квантового усилителя.

9.4. Лазеры

Принцип работы *лазера**, или *оптического квантового генератора* (ОКГ), основан на трех фундаментальных идеях, родившихся в разное время в различных областях физики. Первая идея связана с использованием *вынужденного испускания света* атомными системами, открытого Эйнштейном в 1917 г. при теоретическом изучении некогерентного теплового излучения. Как показал Дирак, испускаемые при этом фотоны неотличимы от тех, что вызвали испускание. Вторая идея заключается в применении *термодинамически неравновесных систем,* в которых возможно усиление, а не поглощение света. Она была высказана В. А. Фабрикантом в 1940 г. Третья идея, берущая начало в радиофизике, состоит в использовании *положительной обратной связи* для превращения усиливающей системы в генератор когерентного излучения.

Первые квантовые генераторы, работающие в микроволновом диапазоне (мазеры), были созданы в 1954 г. Н. Г. Басовым и А. М. Прохоровым в СССР и Гордоном, Цайгером и Таунсом в США. Активной средой в них служил пучок молекул аммиака; инверсия населенностей на рабочем переходе достигалась пространственным разделением молекул в различных квантовых состояниях при прохождении пучка через неоднородное электрическое поле. За разработку нового принципа усиления и генерации электромагнит-

* Light Amplification by Stimulated Emission of Radiation — усиление света индуцированным испусканием излучения.

ных волн и создание первых молекулярных генераторов когерентного микроволнового излучения Н. Г. Басову, А. М. Прохорову и Ч. Таунсу была присуждена Нобелевская премия 1964 г. по физике. Первый оптический квантовый генератор (лазер), работающий в импульсном режиме, был создан в 1960 г. Мейманом. Активной средой в нем служил стержень из кристалла рубина, возбуждаемый светом от лампы-вспышки. Годом позже Джаваном, Беннетом и Эрриотом был построен первый лазер непрерывного действия с газообразной активной средой (смесь неона и гелия).

Исключительные свойства когерентного лазерного излучения, коренным образом отличающие его от некогерентного излучения традиционных для оптической области тепловых, газоразрядных и люминесцентных источников света, обусловили бурное развитие лазерной техники и широкое применение лазеров в научных исследованиях и в практике. В настоящее время существует много типов лазеров, отличающихся способами возбуждения активной среды, спектральной областью, мощностью, временны́ми и спектральными характеристиками излучения и т. п.

Существуют разные способы получения необходимой для работы лазера усиливающей излучение *активной среды*. Преобладание процессов вынужденного излучения над поглощением осуществляется при инверсии населенностей ($N_2 > N_1$) рабочих уровней энергии ε_1 и ε_2 (см. § 9.3). В импульсных твердотельных лазерах используется оптическая накачка светом мощной газоразрядной лампы-вспышки. В полупроводниковых лазерах непрерывного действия неравновесное состояние достигается при пропускании электрического тока через p-n-переход. В газовых лазерах атомы или ионы рабочего вещества возбуждаются в условиях электрического разряда. Во всех случаях затраченная на это энергия внешнего источника в конечном свете частично преобразуется в энергию когерентного излучения.

Для осуществления положительной обратной связи часть генерируемого излучения должна оставаться внутри рабочего вещества и вызывать вынужденное испускание все новыми и новыми возбужденными атомами. С этой целью активную среду помещают в *оптический резонатор* (см. § 6.4), образованный двумя параллельными плоскими или сферическими зеркалами, одно из которых полупрозрачно. Возникшая в каком-либо месте в результате спонтанного излучения возбужденного атома световая волна усиливается за счет вынужденного испускания при распространении через активную среду. Эффективно усиливаются только те волны, направление распространения которых совпадает с осью резонатора, так как при всех других направлениях волна быстро покидает пределы активной среды. Дойдя до полупрозрачного зеркала, волна частично выходит наружу, а частично отражается назад. Отраженная волна, проходя вдоль резонатора от одного зеркала до другого путь L через активную среду, в соответствии с форму-

лой (9.38) усиливается в $\exp(-\alpha L)$ раз, где $(-\alpha)$ — коэффициент усиления (9.37). Отразившаяся от второго зеркала волна снова на длине резонатора усиливается в $\exp(-\alpha L)$ раз и опять падает на полупрозрачное зеркало. Пусть на двойной длине резонатора укладывается целое число длин волн: $2L=m\lambda$. Этим обеспечивается положительная обратная связь: при втором и всех следующих прохождениях через резонатор волна согласована по фазе с исходной, т. е. фактически неотличима от нее. Все элементарные волны, возникающие в разные моменты времени при вынужденных переходах атомов, расположенных в разных точках активной среды, создают одну когерентную волну. Таким образом, оптический резонатор увеличивает эффективное расстояние, которое распространяющаяся вдоль его оси волна проходит в активной среде, и способствует формированию когерентного монохроматического излучения, выходящего через полупрозрачное зеркало.

Оптический резонатор можно рассматривать как колебательную систему, в которой собственные нормальные колебания (моды) имеют вид стоячих электромагнитных волн с узлами на зеркалах. Частоты ω_q таких мод определяются условием $2L=q\lambda$, где q — целое число, т. е. они образуют эквидистантный спектр:

$$\omega_q=2\pi c/(n\lambda_q)=q[\pi c/(nL)], \tag{9.39}$$

где n — показатель преломления активной среды. Поэтому лазер представляет собой автоколебательную систему, в которой возможны незатухающие колебания на одной из собственных частот резонатора.

Описанная выше волновая картина возникновения когерентного излучения при вынужденном испускании света находящимися в разных местах атомами активной среды имеет следующую квантовую интерпретацию. Одна из мод резонатора может возбудиться в результате спонтанного испускания фотона каким-либо возбужденным атомом. В дальнейшем каждый акт вынужденного испускания приводит к увеличению на единицу числа фотонов в этой моде. Фотоны одной моды тождественны, несмотря на то что они были испущены находившимися в разных местах атомами активной среды в разные моменты времени. Этим объясняется когерентность результирующего излучения.

Падающая на полупрозрачное зеркало волна частично выходит из резонатора. Это и есть генерируемый лазером пучок. Кроме того, существуют и другие (бесполезные) потери энергии излучения, обусловленные рассеянием света на неоднородностях заполняющей резонатор среды, дифракцией на краях зеркал и т. п.

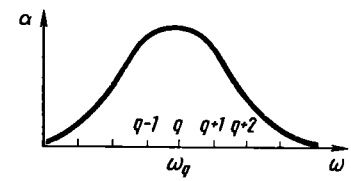
Очевидно, что генерация возможна только тогда, когда падающая на полупрозрачное зеркало после очередного прохода волна имеет энергию не меньшую, чем при предыдущем падении. Это значит, что усиление света в активной среде должно быть достаточно большим, превышающим некоторое значение, называемое

пороговым. Если усиление на двойном проходе резонатора больше суммарных потерь, то с каждым проходом интенсивность волны возрастает. Это возрастание не может, конечно, продолжаться беспредельно. При фиксированной мощности источника накачки инверсия населенностей рабочих уровней будет убывать с увеличением плотности энергии излучения в резонаторе, что в соответствии с (9.37) приведет к уменьшению коэффициента усиления. В результате этого нелинейного *эффекта насыщения* (см. § 10.1) в лазерах непрерывного действия устанавливается стационарный режим генерации, когда суммарные потери энергии в точности компенсируются усилением в активной среде. При этом мощность генерируемого излучения пропорциональна превышению ненасыщенного коэффициента усиления над его пороговым значением, т. е. линейно зависит от мощности источника накачки, поддерживающего в активной среде инверсную населенность.

Спектральная линия, соответствующая переходу между рабочими уровнями атомов активной среды, имеет конечную ширину. Возможные причины уширения были рассмотрены в § 1.8. Помимо радиационного затухания вклад в ширину линии дают столкновения и тепловое движение атомов (в газовой среде), а также возмущение энергетических уровней атомов под влиянием окружения (полей заряженных частиц в газовом разряде, кристаллических полей в твердых телах и т. п.). При однородном уширении контур спектральной линии $F(\omega)$ с хорошей точностью описывается лоренцевской функцией, при неоднородном — гауссовой. Ширина линии $\Delta\omega$ много меньше частоты ω_0, соответствующей центру линии, поэтому спектральная зависимость коэффициента усиления $\alpha(\omega)$ (9.37) повторяет ход функции формы линии $F(\omega)$.

Обычно в пределах контура усиления умещается несколько собственных частот резонатора (рис. 9.4). Если постепенно увеличивать степень возбуждения активной среды, т. е. инверсию населенностей уровней рабочего перехода, то первой достигает порога мода, частота которой расположена ближе к центру спектральной линии. На ней и начинается генерация. Дальнейшее увеличение уровня возбуждения активной среды приводит к достижению порога другими модами. Тогда становится возможной генерация излучения одновременно на нескольких частотах.

В случае однородно уширенной линии перехода различные моды получают энергию от одних и тех же атомов, что приводит к эффектам *нелинейного взаимодействия* (*конкуренции*) *мод*. Насыщение усиления до уровня потерь происходит быстрее для той моды, частота которой лежит дальше других от центра линии усиления. Продолжающийся рост поля других мод вызывает уменьшение усиления и приводит к затуханию этой



9.4

Спектральный контур линии усиления

моды. В результате возможно затухание всех мод, кроме одной, частота которой наиболее близка к центру линии усиления.

При неоднородном уширении линии перехода каждая мода получает энергию от тех атомов, центральные частоты которых совпадают с ее частотой. Поэтому эффекты насыщения приводят к образованию в контуре усиления «провалов» на частотах соответствующих мод резонатора. Это способствует многомодовой генерации, так как становится возможным почти независимое усиление мод, отстоящих друг от друга по частоте больше чем на ширину линии отдельного атома.

В условиях стационарной генерации потери в резонаторе компенсируются за счет вынужденного излучения в активной среде. Отсутствие затухания означает, что добротность, сопоставляемая соответствующей «активной» моде, бесконечно велика, а спектральная линия излучения лазера на этой моде бесконечно узка. Другими словами, усиление в активной среде поддерживает постоянную амплитуду гармонического выходного сигнала, препятствуя его затуханию, но при этом не вносит возмущений в его фазу. Спектр такого излучения состоит из единственной частоты. Фактически же излучение одномодового лазера, хотя и обладает очень высокой степенью монохроматичности, все же по ряду причин имеет конечную спектральную ширину. До сих пор не учитывалось спонтанное испускание возбужденными атомами активной среды, которое также способствует вводу энергии в резонатор. Так как вводимая за счет спонтанных процессов энергия не скоррелирована с полем излучения генерируемой моды, спонтанное излучение вносит случайные флуктуации в его фазу и амплитуду, что приводит к конечной ширине спектра выходного излучения. Иначе можно сказать, что в стационарном режиме при учете спонтанного испускания потери в резонаторе не полностью компенсируются вынужденным излучением (какая-то часть потерь компенсируется спонтанным излучением). Поэтому добротность активной моды уже не обращается в бесконечность, хотя и может быть очень большой. Чем больше выходная мощность, тем уже спектральная линия излучения, так как с ростом плотности энергии поля в резонаторе вынужденное испускание все в большей степени преобладает над спонтанным в механизме передачи энергии от активной среды в моду резонатора. Как показывают расчеты, естественная ширина линии лазерной генерации, обусловленная спонтанным излучением, для гелий-неонового лазера ($\lambda = 632{,}8$ нм) с выходной мощностью 1 мВт при пропускании зеркал 1% и длине резонатора 1 м составляет $\Delta\nu \sim 5 \cdot 10^{-4}$ Гц. Это фантастически малая величина. Соответствующее ей время когерентности $\tau \approx 1/\Delta\nu \sim 2 \cdot 10^{3}$ с, а длина когерентности $L = c\tau \sim 6 \cdot 10^{11}$ м больше диаметра земной орбиты!

На опыте реализовать столь узкую линию излучения не удается из-за технических причин уширения спектра. Согласно (9.39),

частота моды ω_q зависит от длины резонатора и от показателя преломления среды n. Это обстоятельство позволяет плавно перестраивать частоту генерации, например перемещением одного из зеркал. Но изменения длины резонатора и показателя преломления могут происходить и случайным, неконтролируемым образом из-за вибраций, колебаний температуры и давления и т. п. Ценой значительного усложнения конструкции техническое уширение линии излучения может быть существенно уменьшено и относительную ширину $\Delta\omega/\omega$ удается довести до значений 10^{-14}—10^{-16}.

Наряду со столь высокой *временно́й когерентностью,* недостижимой никакими другими способами, лазерное излучение характеризуется также практически полной *пространственной когерентностью.* Это легко продемонстрировать, раздвигая щели в опыте Юнга (без первой входной щели) до самых краев поперечного сечения лазерого пучка. Видность интерференционной картины при этом не уменьшается. Количественные измерения показывают, что для излучения гелий-неонового лазера (λ=632,8 нм) степень пространственной когерентности γ_{12} (см. § 5.5) отличается от единицы менее чем на 10^{-3} даже для тех точек поперечного сечения пучка, где интенсивность составляет всего 0,1% от интенсивности на оси пучка.

Пространственная структура лазерного пучка зависит от геометрии оптического резонатора. От других известных типов резонаторов (например, микроволновых) оптический отличается тем, что его размеры велики по сравнению с длиной волны $[L \sim (10^4 \div 10^7)\,\lambda]$, поэтому он обладает большим числом мод. Однако это «открытый» резонатор, образованный двумя далеко разнесенными зеркалами, и большинство мод характеризуется сильным затуханием из-за ухода излучения за его пределы. Моды с малыми потерями должны (в приближении геометрической оптики) соответствовать такому направлению распространения излучения, чтобы после повторных проходов и отражений излучение не выходило из резонатора. Требование существования таких мод налагает ограничения на соотношение между длиной резонатора и радиусами кривизны его зеркал, известные как *условия устойчивости* (неустойчивый резонатор может использоваться только в системах с очень высоким уровнем усиления в активной среде). Из-за ограниченного размера зеркал распространение света в резонаторе сопровождается дифракционными явлениями, и в общем случае задача расчета поля в резонаторе оказывается довольно сложной.

Электромагнитное поле в резонаторе должно иметь такое распределение амплитуды по поперечному сечению пучка, которое воспроизводит себя на протяжении одного цикла. Для резонатора, образованного сферическими зеркалами, таким свойством обладает *гауссов пучок* (см. § 6.4), характеризуемый быстрым спаданием интенсивности от оси к краям по закону $\exp[-(x^2+y^2)/w^2]$. Распространяющиеся навстречу гауссовы пучки образуют стоячую вол-

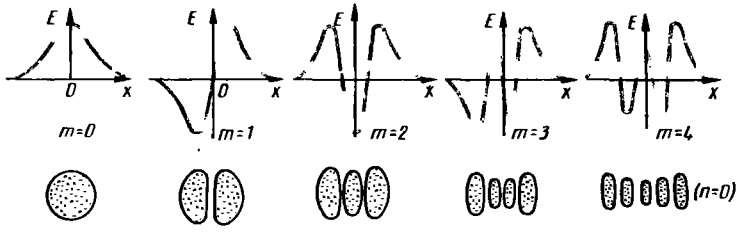
ну при условии, что на длине резонатора укладывается целое число полуволн. Такие моды имеют эквидистантный частотный спектр (9.39). Почти вся энергия излучения в них сосредоточена вблизи оси резонатора в области радиусом $\sim w$. Этот радиус не зависит от апертуры зеркала. Увеличение апертуры зеркала приводит лишь к уменьшению дифракционных потерь и не влияет на поперечный размер пучка. Например, для конфокального резонатора гелий-неонового лазера ($\lambda=632{,}8$ нм) при $L=1$ м радиус пучка на зеркалах $w=0{,}32$ мм.

Помимо обладающего осевой симметрией гауссова пучка возможны моды с более сложным распределением амплитуды по поперечному сечению, описываемым (в случае прямоугольной апертуры зеркал) функциями вида

$$H_n(x/w)\,H_m(y/w)\exp[-(x^2+y^2)/(2w^2)],$$

где $H_n(x)$ — полиномы Эрмита. При n, $m=0$, 1, 2 они имеют вид $H_0(x)=1$, $H_1(x)=2x$, $H_2(x)=4x^2-2$. Такие моды называют *поперечными*. На рис. 9.5 показано распределение амплитуды напряженности поля вдоль направления x для некоторых поперечных мод низкого порядка. Отрицательные значения $E(x)$ свидетельствуют об обращении фазы: при переходе от одного пятна к другому фаза колебаний в данной моде изменяется на противоположную. Это легко продемонстрировать, помещая в пучок экран с двумя отверстиями. Когда отверстия располагаются в соседних пятнах, светлые и темные полосы интерференционной картины меняются местами по сравнению со случаем, когда оба отверстия находятся в пределах одного пятна. Дифракционные потери поперечных мод с m, $n\neq 0$ выше, чем у основной моды ($m=n=0$), и для их возбуждения требуется более высокое усиление в активной среде.

При заданной длине резонатора поперечный размер пучка будет наименьшим для конфокального резонатора. При предельном переходе к плоским зеркалам поперечный радиус w пучка становится очень большим. Фактически это означает, что в резонаторе с плоскими зеркалами гауссов пучок сформироваться не может. В этом случае поле в резонаторе представляет собой стоячую волну с почти плоскими волновыми поверхностями, а поперечное распре-



Распределение амплитуды поля и вид картины, наблюдаемой при освещении экрана лазерным пучком, для некоторых поперечных мод низкого порядка

деление амплитуды для прямоугольной апертуры зеркал $a \times b$ приближенно описывается произведением гармонических функций, которые обращаются в нуль на краях зеркал:

$$\sin[(m+1)\pi x/a]\sin[(n+1)\pi y/b],$$

где $m, n=0, 1, 2, \ldots$ характеризуют число узлов, т. е. линий нулевой амплитуды светового поля на зеркалах в соответствующей поперечной моде (начало координат в плоскости xy совмещено с одной из вершин зеркала). Стоячую волну с таким поперечным распределением поля можно представить как суперпозицию бегущих плоских волн, направления которых образуют с осью резонатора углы $\varphi_m=(m+1)\lambda/(2a)$, $\varphi_n=(n+1)\lambda/(2b)$. Отсюда ясно, что минимальной угловой расходимостью обладает лазерный пучок при генерации на основной моде $m=n=0$. При заданных длине волны λ и поперечном размере a эта расходимость имеет дифракционный характер, т. е. приближается к наименьшей возможной, допускаемой волновой природой света ($\theta \sim \lambda/a$).

Таким образом, как для сферических, так и для плоских зеркал каждой *продольной моде*, т. е. стоячей волне с определенным значением q (числа полуволн, укладывающихся на длине резонатора), имеющей в соответствии с (9.39) частоту $\omega_q=q[\pi c/(nL)]$, соответствует набор *поперечных мод* с различными значениями m и n, которые обычно отстоят по частоте на расстоянии (от 0,2 до 5 МГц), малом по сравнению с интервалом $\Delta\omega=\pi c/(nL)$ (≈ 150 МГц при $L=1$ м) между соседними продольными модами.

В резонаторах с плоскими зеркалами интенсивность поля при удалении от оси спадает медленнее, чем при сферических зеркалах. Поэтому дифракционные потери здесь больше (около 0,1% при $a=b=1$ см, $L=1$ м, $\lambda=0{,}63$ мкм) и больше должна быть площадь поперечного сечения активной среды. С этим обстоятельством, а также с более жесткими требованиями к юстировке плоских зеркал (их параллельность должна быть выдержана с точностью до угловых секунд) связано широкое распространение резонаторов со сферическими зеркалами.

Рассмотрим кратко особенности конструкции и работы некоторых основных типов лазеров. В качестве примера твердотельных лазеров возьмем *лазер на кристалле рубина*. Рубин — это кристалл оксида алюминия Al_2O_3 (корунд), в котором небольшая часть ионов алюминия ($\approx 0{,}05\%$) при выращивании замещена ионами хрома Cr^{3+}. Сам корунд в видимой области прозрачен, и основную роль в работе лазера играют ионы хрома. Упрощенная схема энергетических уровней иона хрома приведена на рис. 9.6, *а*. Инверсия населенностей создается между основным состоянием ε_1 и уровнем ε_2. Этому переходу соответствует длина волны $\lambda=694{,}3$ нм в красной области спектра. Выше уровня ε_2 лежат широкие полосы энергетических уровней ε_3 и ε_3'. Переходам в них из основного состояния соответствуют две широкие полосы поглощения в зеленой и синей областях

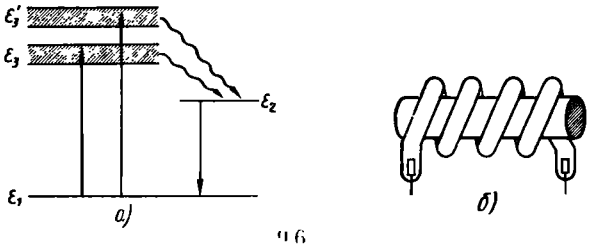


Рис. 9.6

Энергетические уровни иона хрома в рубине (*а*) и рубиновый стержень со спиральной лампой-вспышкой (*б*)

спектра. С этим поглощением связана розовая окраска рубина. Цилиндрический стержень рубина длиной несколько сантиметров и диаметром около 1 см освещается мощным импульсом белого света от лампы-вспышки, обвивающей его в виде спирали (рис. 9.6, *б*). Длительность импульса около 1 мс. При достаточной энергии вспышки большая часть ионов хрома, поглощая свет, переходит в состояния ε_3 и ε_3'. Затем ионы хрома за время порядка 10^{-8} с безызлучательно переходят на уровень ε_2, передавая избыток энергии колебаниям кристаллической решетки. Время жизни ионов хрома в возбужденном состоянии ε_2 составляет несколько миллисекунд, что на несколько порядков величины превышает типичные времена жизни возбужденных состояний (10^{-7}—10^{-8} с). Возбужденные уровни со столь большим временем жизни называют *метастабильными*. При недостаточной энергии вспышки на уровне ε_2 окажется меньше половины всех ионов хрома. Тогда обратный их переход в основное состояние происходит спонтанно (за время ≈ 1 мс) и сопровождается люминисценцией на длине волны 694,3 нм.

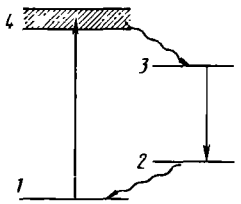
Если же в результате поглощения света вспышки и последующих безызлучательных переходов на уровне ε_2 накопится более половины всех ионов, то между уровнями ε_1 и ε_2 возникает инверсия населенностей ($N_2 > N_1$). Такой механизм ее образования называют *оптической накачкой*. Если рубиновый стержень помещен во внешний оптический резонатор или имеет посеребренные плоскопараллельные торцы, в нем возникает короткий импульс лазерной генерации на длине волны 694,3 нм. Лазерный импульс имеет сложную временну́ю структуру и состоит из нерегулярной последовательности отдельных импульсов длительностью около 1 мкс. Из-за малой длительности импульса (порядка 1 мс) мощность рубинового лазера в импульсе достигает нескольких киловатт при сравнительно небольшой энергии (несколько джоулей).

Мощность излучения в импульсе можно повысить, если добиться сокращения его длительности. Для этого используют *режим модулированной добротности*, сущность которого заключается в следующем. Генерация в лазере начинается, когда инверсия населенностей превысит пороговое значение, зависящее от потерь в резонаторе. Можно задержать начало развития генерации и получить более

высокую концентрацию возбужденных ионов хрома в освещенном кристалле рубина, если на первом этапе возбуждения повысить порог искусственным увеличением потерь, т. е. выключением обратной связи. Для этого одно из зеркал резонатора заменяют быстровращающейся призмой полного отражения. Включение обратной связи и снижение порога генерации происходит только при определенном положении призмы. К этому моменту лампа-вспышка обеспечивает перевод значительной части ионов хрома в возбужденное состояние, и в результате развивается очень короткий (до 10^{-7} с) импульс генерации с мощностью до 10^7 Вт. Еще более короткие импульсы можно получить, если модулировать добротность резонатора с помощью оптического затвора, основанного на эффекте Керра (см. § 4.5). Малая инерционность эффекта Керра позволяет получить импульсы генерации длительностью до 10^{-8} с. Используя последующие каскады усилителей света, можно еще больше сократить длительность импульса, так как при достаточной мощности уже начальная часть задающего импульса вызывает вынужденное высвечивание всех возбужденных ионов в усилителях. Таким путем получают наносекундные *гигантские импульсы* света мощностью 10^4 МВт и выше.

Расходимость излучения рубинового лазера не достигает предела, определяемого дифракцией на выходной апертуре. Это связано с оптической неоднородностью кристалла, а также с многомодовым характером излучения.

Наряду с рубиновым лазером, работающим по *трехуровневой схеме*, широкое распространение получили *четырехуровневые лазеры* на ионах редкоземельных элементов (неодим, самарий), внедренных в кристаллическую (флюорит кальция) или стеклянную матрицу. Эти ионы обладают чрезвычайно подходящей для лазеров структурой энергетических уровней (рис. 9.7). При комнатной температуре разность $\varepsilon_2 - \varepsilon_1$ в несколько раз превосходит kT, поэтому уровень *2* практически пуст. С помощью сильного света лампы накачки ионы неодима возбуждаются в широкую полосу уровней *4*. Время жизни иона в состоянии *4* очень мало, так как с большой вероятностью происходит безызлучательный переход в метастабильное состояние *3*. Инверсия населенностей между уровнями *2* и *3* лазерного перехода достигается очень легко, поскольку нижний уровень *2* практически не заселен.



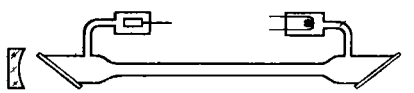
9.7
Энергетическая диаграмма четырехуровневого лазера

Лазер на неодимовом стекле генерирует излучение на длине волны 1,06 мкм в виде импульсов с очень большими энергиями (около 1 кДж). По четырехуровневой схеме работает лазер непрерывного действия на кристалле иттрий-алюминиевого граната (YAG) с примесью неодима с выходной мощностью до 1 кВт.

Газовые лазеры по многим характеристикам превосходят лазеры других типов. Они перекрывают широкий спектральный диапазон (от субмиллиметрового до ультрафиолетового). Среди газовых лазеров всегда можно найти лазер, обладающий по крайней мере одним из следующих свойств: высокой степенью монохроматичности излучения, малой расходимостью излучения (вплоть до дифракционного предела), предельными значениями мощности в непрерывном режиме, высоким к. п. д.

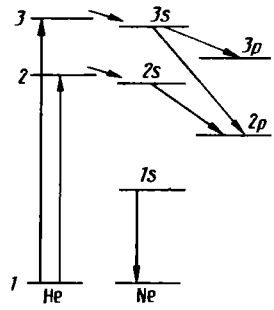
Устройство наиболее распространенного гелий-неонового лазера схематически показано на рис. 9.8. Газоразрядная трубка с внутренним диаметром 1—10 мм и длиной от нескольких десятков сантиметров до 1,5—3 м имеет торцовые плоскопараллельные стеклянные или кварцевые окна, установленные под углом Брюстера к ее оси. Для линейно поляризованного излучения с электрическим вектором в плоскости падения коэффициент отражения от них равен нулю. Поэтому брюстеровские окна обеспечивают линейную поляризацию излучения лазера и исключают потери энергии при распространении света из активной среды к зеркалам и обратно. Трубка помещена в резонатор, образованный зеркалами с многослойными диэлектрическими покрытиями (см. § 5.7). Такие зеркала имеют очень высокий коэффициент отражения в нужном спектральном интервале и почти не поглощают свет. Пропускание зеркала, через которое выводится излучение, выбирается обычно около 1—2%, другого — менее 1%. Особенно удобен резонатор, близкий к конфокальному, так как он вносит малые дифракционные потери и легко поддается юстировке.

Разрядная трубка заполнена смесью гелия и неона в молярном отношении 5:1 при давлении 10^2—10^3 Па. К электродам подключен источник с напряжением в несколько киловольт. Типичная сила тока в разряде — десятки миллиампер. Упрощенная схема энергетических уровней атомов неона и гелия приведена на рис. 9.9. Тлеющий разряд создает условия для возникновения инверсии населенностей уровней в неоне. Гелий служит лишь для резонансного возбуждения неона. Атом гелия обладает двумя метастабильными состояниями *2* и *3*. При столкновениях с электронами плазмы атомы гелия возбуждаются в эти состояния, и из-за большого времени их жизни концентрация атомов в разряде, возбужденных в метастабильные состояния, очень велика. Энергии состояний *2* и *3* близки к энергиям возбужденных состояний 2*s* и 3*s* атома неона, что благоприятно сказывается на передаче энергии возбуждения от атома гелия атому неона при столкновении. В результате таких процессов населенность уровней 2*s* и 3*s* неона возрастает и возникает инверсия населенностей на пе-



9.8
Разрядная трубка газового лазера с ок
иами, установлениыми под углом Брюс
тера

реходах $3s-3p$ ($\lambda = 3392$ нм), $2s-2p$ ($\lambda = 1153$ нм) и $3s-2p$ ($\lambda = 632{,}8$ нм). На этих переходах и возможна генерация. Для получения генерации на одной из указанных длин волн используют зеркала с высоким коэффициентом отражения в соответствующей спектральной области. Высокая оптическая однородность газовой активной среды позволяет получать излучение с очень высокой степенью временно́й и пространственной когерентности.



Упрощенная схема энергетических уровней гелия и неона

амечательной чертой лазеров, тесно связанной с когерентностью их излучения, является исключительная способность к *концентрации световой энергии*: *в спектре* — очень узкая спектральная линия излучения; *во времени* — возможность получения сверхкоротких импульсов света; *в пространстве* и *по направлению распространения* — возможность получения направленного пучка с предельно малой (дифракционной) расходимостью и фокусирования всего излучения в малой области с размерами порядка длины волны.

Исключительные свойства лазерного излучения открывают широкие перспективы для использования лазеров в различных областях науки и техники: монохроматичность и когерентность — в голографии, при обработке информации, в измерительной технике; высокая мощность — в лазерной технологии и энергетике, в нелинейной оптике; малая расходимость излучения — в лазерной связи, локации, геодезии, строительстве и т. д.

Контрольные вопросы

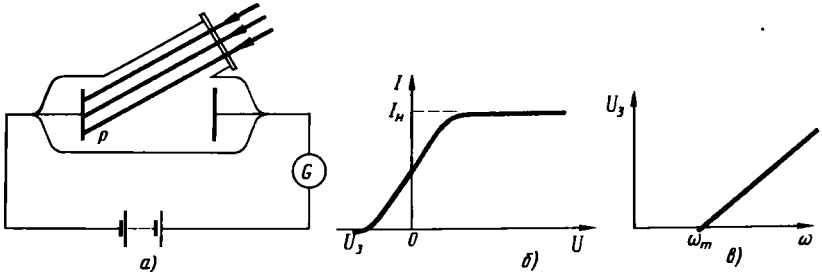
Какие функции в лазере выполняют активная среда и резонатор?
Какими факторами определяется частота, на которой происходит лазерная генерация?
Какими условиями определяется пороговое усиление?
Какие процессы ведут к установлению стационарного режима генерации в случае превышения коэффициента усиления над его пороговым значением?
□ При каких условиях возможна многомодовая генерация?
□ В чем заключаются естественные и технические причины уширения линий генерации?
□ Как на опыте можно продемонстрировать пространственную когерентность лазерного излучения?
Почему дифракционные потери в резонаторе с плоскими зеркалами больше, чем в конфокальном резонаторе таких же размеров?
Поясните способы получения инверсии населенностей в трехуровневых и четырехуровневых лазерах.
□ В чем заключается режим модулированной добротности?
□ Назовите основные области применения лазеров.

9.5. Фотоэлектрический эффект

Корпускулярные свойства излучения наиболее отчетливо обнаруживаются в явлении освобождения электронов из вещества под действием света — фотоэффекта.

Влияние света на электрические процессы было открыто Герцем, заметившим, что проскакивание искры между находящимися под напряжением цинковыми электродами облегчается при освещении их ультрафиолетовым излучением. Первое обстоятельное исследование фотоэффекта было выполнено в 1888—1890 гг. А. Г. Столетовым. Им было установлено, что под действием света тело теряет отрицательный заряд, что действие пропорционально световому потоку и вызывается преимущественно ультрафиолетовыми лучами, что явление протекает практически безынерционно. Впоследствии Ленардом и Томсоном было измерено отношение заряда испускаемых частиц к массе с помощью опытов по отклонению в электрическом и магнитном полях. Эти измерения показали, что под действием света освобождаются электроны.

Для исследования закономерностей фотоэффекта используют установку, схематически показанную на рис. 9.10, *а*. В сосуде поддерживается высокий вакуум. При освещении металлической пластины P через кварцевое окно в цепи возникает ток (*фототок*), измеряемый гальванометром G. Явление в сильной степени зависит от чистоты освещаемой поверхности, поэтому в точных опытах используют свежие поверхности (срезы или поверхности, напыленные в вакууме). Для зависимости силы фототока от приложенного напряжения при неизменном световом потоке (рис. 9.10, *б*) характерно существование участка тока насыщения $I_н$, когда все освобождаемые светом электроны (*фотоэлектроны*) достигают анода, и участка нарастания, начинающегося при некотором значении задерживающего напряжения $U_з$. По измерениям $U_з$ можно определить максимальную скорость v освобождаемых светом электронов с помощью соотношения $eU_з = {}^1/_2 mv^2$. Многочислен-



9.10

Схема опыта для исследования фотоэффекта (*а*), вольт-амперная характеристика (*б*) и зависимость задерживающего напряжения $U_з$ от частоты (*в*)

ными экспериментами установлены следующие основные закономерности фотоэффекта:

1. Сила тока насыщения прямо пропорциональна падающему световому потоку (при неизменном спектральном составе). Этот закон проверен для очень широкого интервала интенсивностей. Отсюда следует, что число электронов, освобождаемых светом в 1 с, пропорционально мощности падающего излучения.

2. Для каждого металла существует максимальная длина волны света (минимальная частота), при которой еще происходит освобождение электронов. Если длина волны превышает эту так называемую *красную границу фотоэффекта*, то эмиссия электронов отсутствует даже при сравнительно большой интенсивности облучающего света*.

3. Максимальная энергия фотоэлектронов линейно зависит от частоты ω падающего света (рис. 9.10, *в*) и не зависит от его интенсивности.

С точки зрения классических волновых представлений о природе излучения сам факт освобождения электронов из металла неудивителен, так как падающая на поверхность электромагнитная волна вызывает вынужденные колебания электронов в металле. Поглощая энергию волны, электрон может накопить ее в количестве, достаточном для преодоления потенциального барьера, удерживающего электрон в металле (т. е. для совершения работы выхода). Если эта картина верна, то энергия фотоэлектрона должна находиться в прямой связи с интенсивностью падающего света. Но опыт показывает, что энергия фотоэлектронов совершенно не зависит от интенсивности света. Увеличение интенсивности приводит лишь к пропорциональному увеличению числа фотоэлектронов. Энергия же отдельного фотоэлектрона зависит только от частоты падающего света.

Более того, даже при очень малой интенсивности фотоэлектроны появляются практически сразу после начала освещения (безынерционно), хотя, по классическим представлениям, в таких условиях требуется конечное время, чтобы электрон мог накопить необходимую энергию.

В 1905 г. Эйнштейн объяснил экспериментальные закономерности фотоэффекта на основе гипотезы световых квантов. Качественная картина с этой точки зрения выглядит следующим образом. Падающее монохроматическое излучение рассматривается как поток фотонов, энергия которых связана с частотой соотношением $\varepsilon = \hbar\omega$. При поглощении фотона его энергия целиком передается одному электрону, и если эта энергия достаточна для того, чтобы

---

* При очень высоких интенсивностях (сфокусированное лазерное излучение) возможен многофотонный фотоэффект (см. § 10.1), для которого красная граница исчезает.

освободить электрон от удерживающих его связей, то он может выйти за пределы поверхности металла. Вероятность одновременного поглощения двух фотонов одним электроном мала, поэтому каждый фотоэлектрон получает энергию от одного фотона*. (Отметим, что, вообще говоря, не каждый поглощенный фотон приводит к освобождению электрона, т. е. квантовый выход — отношение числа фотоэлектронов к числу поглощенных фотонов — обычно меньше единицы.)

По квантовым представлениям, полное число освобожденных электронов пропорционально числу поглощенных фотонов, т. е. сила тока насыщения пропорциональна интенсивности. Но энергия отдельного фотоэлектрона определяется энергией поглощенного фотона $\varepsilon=\hbar\omega$. Отсюда ясно, почему энергия фотоэлектронов линейно зависит от частоты падающего света и совсем не зависит от его интенсивности (т. е. числа падающих фотонов).

Приобретаемая электроном энергия $\varepsilon=\hbar\omega$ частично затрачивается на освобождение из металла. Ее излишек остается в форме кинетической энергии освобожденного электрона. Минимальную энергию A, необходимую для освобождения электрона из металла, называют *работой выхода*. Таким образом, для фотоэлектронов, имеющих максимальную скорость, закон сохранения энергии в элементарном акте поглощения фотона (*уравнение Эйнштейна*) можно записать в виде

$$\hbar\omega={}^1/_2 mv^2+A. \tag{9.40}$$

Очевидно, что при $\hbar\omega<A$ электрон не может выйти из металла. Это значит, что существует некоторая минимальная частота излучения $\omega_m=A/\hbar$, при которой еще возможен фотоэффект. При меньших частотах $\omega<\omega_m$ фотоэффект не наблюдается. Мы видим, что уравнение Эйнштейна (9.40) сразу объясняет существование красной границы фотоэффекта. Для различных металлов работа выхода A и, следовательно, граничная частота ω_m имеют разные значения. Кроме того, на работу выхода существенное влияние оказывают состояние и чистота поверхности металла, в особенности наличие пленки адсорбированного газа. Для большинства металлов красная граница фотоэффекта приходится на ультрафиолетовую область спектра (в опытах Столетова с освещением цинковой пластинки фотоэффект пропадал при переходе от ультрафиолетовых к видимым лучам). Только у щелочных металлов красная граница попадает в область видимого света. Поэтому они используются для покрытия поверхности фотокатода у фотоэлементов,

---

* Это утверждение справедливо при не слишком большой интенсивности падающего света. В 1967 г. был обнаружен многофотонный (нелинейный) фотоэффект в металлах под действием сверхкоротких импульсов лазерного излучения.

предназначенных для работы с видимым светом. Значения работы выхода A, определяемые по красной границе фотоэффекта, находятся в хорошем согласии со значениями, полученными при изучении термоэлектронной эмиссии.

Максимальная кинетическая энергия фотоэлектронов $^1/_2mv^2$ измеряется на опыте по задерживающему напряжению U_3, которое необходимо приложить между фотокатодом и анодом для полного прекращения фототока: $^1/_2mv^2=eU_3$. Уравнение Эйнштейна (9.40) предсказывает линейную зависимость U_3 от частоты падающего света:

$$eU_3=\hbar(\omega-\omega_m). \qquad (9.41)$$

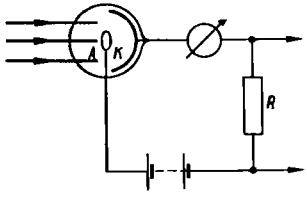
Соотношение (9.41) было подтверждено в тщательных опытах Милликена (1916), создавшего прибор, в котором поверхности исследуемых металлов подвергались очистке в вакууме. По наклону прямой, изображающей зависимость задерживающего напряжения U_3 от частоты ω, находится отношение $\hbar/e$. Этот наклон одинаков для всех металлов. Полученное таким методом значение универсальной постоянной Планка $\hbar$ находится в хорошем согласии со значениями, найденными другими методами (по излучению черного тела, см. § 9.2; по коротковолновой границе сплошного рентгеновского спектра, см. § 9.6). Точка пересечения графика (9.41) с осью абсцисс $\omega=\omega_m$ позволяет определить значение работы выхода $A=\hbar\omega_m$ для исследуемого металла.

Принято считать, что фотоэффект дает наиболее прямое экспериментальное доказательство квантовой природы излучения. Квантовая гипотеза и в самом деле позволяет непринужденно объяснить все основные экспериментальные закономерности фотоэффекта. Но тем не менее следует отметить, что эти закономерности получают исчерпывающее объяснение и в полуклассической теории взаимодействия излучения с веществом, рассматривающей вещество квантово-механически, а излучение — как классическое электромагнитное поле. Это показал Г. Вентцель в 1927 г. С аналогичным положением вещей мы сталкиваемся и в проблеме равновесного излучения. Спектральное распределение энергии (формулу Планка) можно получить, рассматривая нормальные колебания электромагнитного поля в полости как набор квантовых осцилляторов, т. е. как идеальный газ частиц излучения — фотонов (см. § 9.3). Но формулу Планка можно получить и иначе, рассматривая излучение как классическое электромагнитное поле и применяя квантовую гипотезу лишь к находящемуся в равновесии с ним веществу (осцилляторам). Именно так и поступал Планк (см. § 9.2). Полуклассическая теория взаимодействия света с веществом, не привлекая понятия фотона, дает количественное объяснение большинству наблюдаемых явлений. Квантование электромагнитного поля принципиально необходимо для правильного описания некоторых явлений, включающих его флуктуации: спонтанного излучения, лэмбовского сдвига, аномального магнитного момента электрона.

Остановимся теперь на практических применениях фотоэффекта. Фотоэлектрические приемники излучения, преобразующие световой сигнал в электрический, разработаны практически для всего оптического диапазона и широко используются в науке и технике. Устройство простейшего вакуумного фотоэлемента показано на рис. 9.11. Половина небольшого откачанного стеклянного или кварцевого баллона покрыта изнутри тонким слоем металла, который служит фотокатодом. Для видимой и близкой ультрафиолетовой области применяют сурьмяно-цезиевые фотокатоды, имеющие максимум чувствительности в сине-фиолетовой области спектра ($\lambda \approx 450$ нм), для красной и ближней инфракрасной — серебряно-кислородно-цезиевые с максимумом в области $\lambda \approx 800$ нм. Анод в виде небольшого кольца или сетки находится в центре баллона. При освещении фотокатода в цепи возникает ток. Приложенное напряжение должно быть достаточным для насыщения фототока. Применение фотоэлемента для фотометрических измерений основано на строгой пропорциональности силы тока насыщения и потока излучения. Иногда используют прерыватель света, преобразующий постоянный сигнал в переменный, который затем усиливается обычными радиотехническими методами.

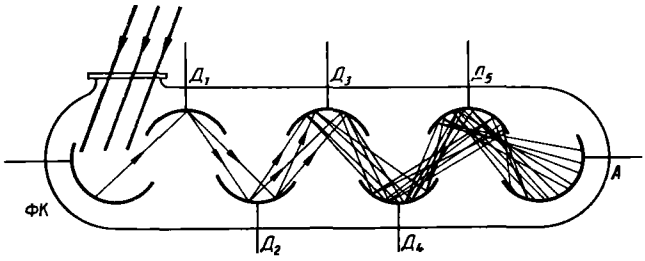
Множество разных задач, решаемых с помощью фотоэлементов, вызвало к жизни большое разнообразие типов фотоэлементов с различными техническими характеристиками.

Существенный прогресс в фотоэлектрических измерениях был достигнут в 40—50-е годы, когда в практику начали широко внедряться *фотоэлектронные умножители* (ФЭУ). Основной недостаток вакуумных фотоэлементов — малая величина вырабатываемых ими электрических сигналов — преодолевается в ФЭУ усилением фототока с помощью вторичной электронной эмиссии. Это явление заключается в освобождении электронов из металла или полупроводника при бомбардировке поверхности пучком быстрых электронов. Отношение числа освобождаемых электронов к числу падающих на поверхность, называемое коэффициентом вторичной эмиссии, зависит от скорости и угла падения пучка электронов, вида и состояния поверхности и для некоторых веществ может достигать весьма больших значений — 10 и выше. Значительное усиление потока электронов получается при использовании в качестве материала эмиттеров сплава сурьмы и цезия.



9.11
Схема простейшего фотоэлемента

Схема устройства фотоумножителя приведена на рис. 9.12. Фотоэлектроны, эмиттируемые при освещении поверхности фотокатода *ФК*, ускоряются электрическим полем и попадают на первый промежуточный электрод (динод) $Д_1$, вызывая эмиссию вторичных электронов. Кон-



9.12
Принципиальная схема фотоумножителя

фигурация и расположение фотокатода и динодов выбраны так, что создаваемые ими электрические поля обеспечивают попадание большинства фотоэлектронов на дииод $Д_1$, а большинство вторичных электронов после ускорения попадает на следующий динод $Д_2$, где процесс умножения повторяется, и т. д. Между последовательными динодами должно быть приложено определенное напряжение. Оно подается через делители напряжения от высоковольтного стабилизированного источника. Вторичные электроны с последнего из динодов (их может быть 10—15) собираются на анод (коллектор) A. Общий коэффициент усиления в такой системе достигает 10^6—10^8, а интегральная чувствительность — тысяч ампер на люмен, что позволяет измерять очень малые световые потоки.

Однако предельная чувствительность фотоэлектрических приемников излучения (как и других физических измерительных приборов) лимитируется не достижимым усилением сигнала, а *собственными шумами* измерительного устройства. Собственными шумами называют случайные, нерегулярные изменения (флуктуации) сигнала, вносимые самим прибором. Ясно, что одновременное усиление сигнала и шума не даст никакой выгоды. Поэтому реальные возможности регистрации и измерения слабых световых потоков определяются *отношением полезного сигнала к шуму*. В тех случаях, когда измеряемый поток значительно превосходит предел чувствительности, шумами определяется погрешность измерений.

Некоторые виды шумов связаны с техническим несовершенством прибора. Например, небольшие непостоянные утечки зарядов по поверхности стекла колбы фотоумножителя могут вызвать нерегулярные движения стрелки измерительного прибора. Такие *технические флуктуации* можно в значительной степени устранить, если приемник выполнить достаточно тщательно и принять необходимые меры предосторожности при работе с ним. Но существуют принципиально неустранимые флуктуации, связанные с атомной структурой вещества (пример таких флуктуаций дает хорошо известное броуновское движение), с конечной величиной элементарного заряда и с квантовой природой взаимодействия излучения с веществом.

Обусловленные собственными флуктуациями шумы в измерительных устройствах рассчитываются на основе статистической механики.

В вакуумном фотоэлементе (рис. 9.11) одна из наиболее очевидных причин возникновения флуктуаций фототока связана с тем, что элементарные акты испускания фотоэлектронов происходят в *случайные* моменты времени аналогично актам термоэлектронной эмиссии из накаленного катода в вакуумных электронных лампах. Обусловленный этой причиной шум получил образное название *дробового*. При постоянной интенсивности падающего излучения, когда можно считать, что за некоторый промежуток времени τ на катод падает вполне определенное число фотонов N_Φ, среднее число испускаемых электронов за время τ составит $\langle N\rangle = \eta N_\Phi$, где η — квантовый выход фотокатода. Однако от измерения к измерению число N фотоэлектронов только *в среднем* остается неизменным, испытывая флуктуации около среднего значения $\langle N\rangle$. Вероятность $P(N)$ испускания за время τ определенного числа N фотоэлектронов в отдельном измерении дается распределением Пуассона: $P(N) = e^{-\alpha}\alpha^N/N!$, где $\alpha = \langle N\rangle$. Мерой флуктуаций числа фотоэлектронов служит величина $\delta N = \langle (N - \langle N\rangle)^2\rangle^{1/2}$, называемая среднеквадратичной флуктуацией. Можно показать, что для пуассоновского случайного процесса, описываемого распределением $P(N)$, среднеквадратичная флуктуация равна корню из среднего числа событий: $\delta N = \sqrt{\langle N\rangle}$. Поэтому флуктуация силы фототока $I = eN/\tau$ равна $\delta I = e\delta N/\tau = e\sqrt{\langle N\rangle}/\tau = \sqrt{eI/\tau}$. В результате дробовой эффект приводит к появлению на нагрузочном сопротивлении R (см. рис. 9.11) флуктуирующего напряжения $U_{др}$, средний квадрат которого равен $\langle U_{др}^2\rangle = (\delta I)^2R^2 = eIR^2/\tau$. При измерении изменяющихся во времени (модулированных) световых потоков в это выражение вместо времени измерения τ (т. е. постоянной времени прибора) войдет величина $1/(2\Delta f)$, где Δf — *полоса пропускания* последующей регистрирующей системы, т. е. интервал частот, в пределах которого происходит дальнейшее усиление переменного выходного напряжения:

$$\langle U_{др}^2\rangle = 2eIR^2\Delta f. \qquad (9.42)$$

Фотоэлемент дает некоторый ток даже при отсутствии падающего излучения. В этот *темновой ток* наряду с утечками изоляции электродов, которые в совершенных приборах малы и постоянны, дает вклад термоэлектронная эмиссия с фотокатода. Термоэмиссия электронов, как и фотоэмиссия, подвержена флуктуациям, приводящим к дробовому шуму. Мощности шумов от статистически независимых источников флуктуаций просто складываются, поэтому полный дробовой шум определяется тем же выражением (9.42), если в нем под I понимать полный ток с поверхности фотокатода, т. е. сумму фототока I_Φ и темнового тока $I_т$: $I = I_\Phi + I_т$. При измерении слабых световых потоков, когда $I_\Phi < I_т$, основной вклад в дробовой шум дают флуктуации темнового тока. Для увеличения пороговой чувствительности иногда прибегают к охлаждению фотокатода и уменьшению его размеров (до размеров фокусируемого на нем светового пятна). Так, при охлаждении сурьмяно-цезиевого катода до температуры жидкого азота темновой ток уменьшается в 10^4 раз, что в соответствии с (9.42) приводит к снижению шумового напряжения и увеличению пороговой чувствительности в $\sqrt{10^4} = 100$ раз. Конечно, охлажденне фотокатода, как и уменьшение его размеров, целесообразно лишь тогда, когда ток термоэмиссии $I_т$ больше фототока I_Φ, так как в противном случае шумы темнового тока тонут в дробовых шумах фотоэмиссии.

Другая причина флуктуаций выходного напряжения фотоэлемента (см. рис. 9.11) связана с хаотическим тепловым движением электронов в нагрузочном сопротивлении R. Тепловой шум в проводниках, интенсивность которого (т. е. средний квадрат хаотического напряжения) растет линейно с увеличением температуры T и сопротивления R, был обнаружен Джонсоном в 1927 г. Спектральная плотность джонсоновского шума в области частот $\hbar\omega \ll kT$ постоянна, и средний квадрат напряжения тепловых шумов определяется формулой Найквиста

$$\langle U_{тепл}^2\rangle = 4kTR\Delta f.$$

Таким образом, для отношения сигнал/шум, определяющего теоретические предельные характеристики вакуумного фотоэлемента, можно написать следующее выражение:

$$\rho=\frac{U_{\text{сигн}}}{\sqrt{\langle U^2_{\text{др}}\rangle+\langle U^2_{\text{тепл}}\rangle}}=\frac{I_\Phi}{\sqrt{[e(I_\Phi+I_\text{т})+2kT/R]2\Delta f}}. \tag{9.43}$$

Повышение чувствительности и точности фотоэлектрических измерений может быть достигнуто уменьшением полосы пропускания Δf. Узкополосный фильтр, настроенный на частоту модуляции переменного светового сигнала, или фильтр нижних частот в случае постоянного сигнала позволяет отфильтровать бо́льшую часть шума, энергия которого в отличие от полезного сигнала равномерно распределена по всем частотам. Сужение полосы пропускания Δf ведет к неизбежному увеличению времени регистрации сигнала ($\tau\sim 1/\Delta f$). При увеличении τ в n раз отношение сигнал/шум, как видно из (9.43), возрастает в $\sqrt{n}$ раз.

Предел чувствительности фотоэлемента с последующим усилителем (см. рис. 9.11) определяется обычно джонсоновскими шумами в нагрузочном сопротивлении R, которые в дальнейшем усиливаются вместе с полезным сигналом и дробовым шумом. Это ограничение устраняется в фотоумножителе, где ток с фотокатода, т. е. сигнал вместе с дробовым шумом, многократно усиливается уже внутри самого ФЭУ без заметного добавления каких-либо новых шумов (при достаточно высоком коэффициенте вторичной эмиссии (>4) ее шумы малосущественны). Если сигнал ФЭУ оказывается все же недостаточным для измерительного прибора, используют, как и в схеме с фотоэлементом, внешний усилитель. Но тепловой шум от нагрузочного сопротивления здесь уже обычно значительно меньше дробового шума на выходе фотоумножителя. В этом главное преимущество фотоумножителя перед простым фотоэлементом. Для отношения сигнал/шум вместо (9.43) теперь получим

$$\rho=\frac{U_{\text{сигн}}}{\sqrt{\langle U^2_{\text{др}}\rangle}}=\frac{I_\Phi}{\sqrt{2e(I_\Phi+I_\text{т})\Delta f}}, \tag{9.44}$$

где I_Φ и $I_\text{т}$ — фототок и темновой ток фотокатода ФЭУ.

Оценим пороговую чувствительность фотоумножителя. При регистрации очень слабых сигналов темновой ток больше фототока и, согласно (9.44), $\rho=I_\Phi/\sqrt{2eI_\text{т}\Delta f}$. Число фотонов, достигающих фотокатода в 1 с, равно $\Phi/(\hbar\omega)$, где Φ — падающий поток излучения (в энергетических единицах). Среднее число фотоэлектронов за 1 с составит $\eta\Phi/(\hbar\omega)$ (η — квантовый выход). Умножая это число на заряд e электрона, получим фототок $I_\Phi=e\eta\Phi/(\hbar\omega)$. Для уверенного обнаружения сигнала должно быть $\rho\geqslant 1$. Условно можно принять, что пороговый сигнал $\Phi_{\text{пор}}$ соответствует $\rho=1$. Тогда $I_\Phi=\sqrt{2eI_\text{т}\Delta f}$ и для пороговой чувствительности получаем

$$\Phi_{\text{пор}}/(\hbar\omega)=(1/\eta)\sqrt{2I_\text{т}\Delta f/e}. \tag{9.45}$$

Отсюда следует, что основные параметры, определяющие предельную чувствительность ФЭУ, — это квантовый выход и темновой ток фотокатода. В лучших образцах ФЭУ при охлаждении число термоэлектронов с 1 см^2 площади катода составляет 5—10 в 1 с. При $\eta=0{,}01$ по формуле (9.45) получаем для пороговой чувствительности $\Phi_{\text{пор}}/\hbar\omega$ величину порядка нескольких сотен фотонов в секунду при ширине полосы пропускания $\Delta f=1$ Гц. Минимальный регистри-

руемый поток можно уменьшить сужением полосы пропускания, т. е. увеличением времени регистрации («накоплением») сигнала.

При низком уровне собственных шумов для измерения слабых световых потоков наилучшие результаты дает работа ФЭУ в *режиме счета фотонов*. Каждый фотоэлектрон после размножения в динодной системе создает на выходе ФЭУ импульс тока. Эти импульсы регистрируются хорошо разработанными в ядерной физике методами. Таким способом можно решать задачи, недоступные другим методам регистрации световых сигналов, в частности задачи непосредственного измерения времен жизни возбужденных состояний атомов.

При использовании фотоумножителя для обнаружения слабого модулированного сигнала в присутствии значительно большего постоянного потока темновой ток может быть малым по сравнению с фототоком. В этом случае точность измерений лимитируется дробовым шумом фототока. Формула (9.44) при $I_т < I_ф$ принимает вид $\rho = \sqrt{I_ф/(2e\Delta f)} = \sqrt{\eta\Phi/(2\Delta f\hbar\omega)}$. Величина, обратная отношению сигнал/шум, характеризует относительную погрешность таких измерений:

$$\Delta\Phi/\Phi = 1/\rho = \sqrt{2\Delta f\hbar\omega/(\eta\Phi)}. \qquad (9.46)$$

При $\lambda = 0{,}5$ мкм, $\Phi = 10^{-7}$ Вт, $\eta = 10^{-1}$ и ширине полосы $\Delta f = 1$ Гц это приводит к относительной погрешности $\Delta\Phi/\Phi \approx 10^{-5}$. Технические флуктуации обычно превосходят этот уровень, так что на практике достижимая точность фотоэлектрических измерений составляет 10^{-3}—10^{-4}.

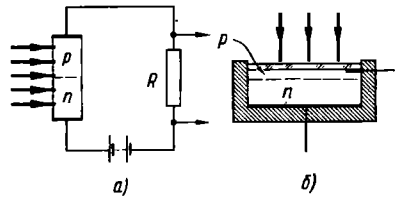
В проведенном выше рассмотрении шумов фотоэлектрических приемников излучения предполагалось, что источник света имеет постоянную интенсивность, т. е. число фотонов, падающих на фотокатод за время наблюдения τ, не испытывает флуктуаций. Такая модель применима к излучению стабилизированного одномодового лазера непрерывного действия. Статистической моделью излучения нелазерных источников может служить хаотически модулированное колебание, амплитуда и фаза которого представляют собой случайные функции времени (см. § 5.4). Характерный временной масштаб изменения амплитуды и фазы — это время когерентности излучения $\tau_{ког}$. Можно показать, что интенсивность такого излучения также хаотически изменяется со временем, испытывая случайные отклонения от среднего значения, причем флуктуации интенсивности равны средней интенсивности. Несмотря на столь большие флуктуации интенсивности, обнаружить их проявление в эксперименте трудно. Объясняется это тем, что временной масштаб флуктуаций интенсивности равен времени когерентности $\tau_{ког}$, которое обратно пропорционально ширине спектрального распределения излучения $\Delta\nu$: $\tau_{ког} \sim 1/\Delta\nu$. Даже для наиболее узких спектральных линий время когерентности мало по сравнению с достижимыми значениями времени наблюдения τ. (Для фотоумножителей постоянная времени τ достигает 10^{-8}—10^{-9} с.) При $\tau_{ког} \ll \tau$ флуктуации интенсивности не влияют на число фотоэлектронов N, эмиттируемых за время τ. Закон распределения числа фотоэлектронов $P(N)$ по-прежнему определяется только случайным характером фотоэмиссии и будет пуассоновским независимо от статистического распределения интенсивности излучения. Сведения о флуктуациях интенсивности сохраняются в статистике фотоэлектронов, когда $\tau_{ког} \gg \tau$. Это условие может выполняться для многомодового лазерного излучения или для лазерного света, рассеянного в турбулентной среде. В подобных случаях статистические свойства излучения могут быть найдены из экспериментальных измерений распределения фотоотсчетов.

Красная граница фотоэффекта ограничивает возможности применения фотоумножителей в инфракрасной области спектра. Для уменьшения работы выхода изготовляют сложные фотокатоды. Нанесение на поверхность серебра кислородно-цезиевого монослоя позволяет расширить спектральную область чувствительности приблизительно до 1,3 мкм, но у таких «длинноволновых» фотокатодов очень мал квантовый выход ($\eta_{макс}=0{,}5\%$), что ограничивает возможности их применения.

Более широкой областью спектральной чувствительности характеризуются фотоэлектрические приемники излучения на основе *внутреннего фотоэффекта* в полупроводниках. Поглощение фотона с энергией, превышающей энергетический интервал между заполненной валентной зоной и свободной зоной проводимости, приводит к образованию пары неравновесных носителей тока — электрона и дырки. Красная граница внутреннего фотоэффекта определяется шириной запрещенной зоны. Она зависит от природы полупроводника и может лежать в области значительно более длинных волн, чем у приемников с внешним фотоэффектом.

Увеличение числа свободных носителей под действием света ведет к возрастанию электропроводимости (фотопроводимость). Основанные на этом явлении приемники излучения называются *фотосопротивлениями*. Фотосопротивления в виде тонких пленок на основе InAs, PbS, PbTe и других полупроводниковых соединений обладают чувствительностью вплоть до длин волн 7—8 мкм. Для видимой и ближней инфракрасной области спектра большей эффективностью обладает другой тип фотоприемника — *фотодиод*. Фотодиод представляет собой полупроводниковую пластинку, внутри которой благодаря различным примесям имеются две области с электронной и дырочной проводимостями (рис. 9.13, *а*). Напряжение от внешнего источника приложено к *n-p*-переходу в запирающем направлении, для которого сопротивление перехода велико. При освещении области *n-p*-перехода в ней появляются неосновные носители (электроны в *p*-области и дырки в *n*-области), что приводит к возникновению тока в цепи. Обычно фотодиод изготавливается так, чтобы свет падал перпендикулярно границе раздела, проходя через тонкий *p*-слой (рис. 9.13, *б*).

Фотодиоды характеризуются широкой спектральной чувствительностью (от ультрафиолетовой области до 1,8 мкм у германиевых фотодиодов) и высоким квантовым выходом ($\eta_{max} \approx 0{,}9$), их постоянная времени $\tau \approx 1$ мкс. Характеристики современных фотодиодов позволяют предпочесть их другим приемникам излучения для решения многих практических задач.



9.13
Принципиальная схема (*а*) и схема устройства (*б*) фотодиода

Фотодиод без внешнего источника э.д.с., называемый *вентильным фотоэлементом*, осуществляет непосредственное преобразование энергии падающего излучения в электрическую энергию. Неравновесные электроны и дырки, образующиеся при поглощении света, пространственно разделяются в переходном слое (фото-э.д.с.), что приводит к возникновению тока во внешней цепи. Кремниевые фотоэлементы такого типа имеют высокий к.п.д. (14—16%) и используются в качестве источника энергии в солнечных батареях космических аппаратов.

Контрольные вопросы

- ☐ Приведите схему экспериментальной установки для изучения фотоэффекта. Каким способом измеряется максимальная скорость фотоэлектронов?
- ☐ Назовите основные закономерности фотоэффекта. Как объяснить их на основе квантовых представлений?
- ☐ Как из опытов по фотоэффекту определить постоянную Планка?
- ☐ Объясните принцип действия фотоэлектронного умножителя.
- ☐ Какие причины вызывают появление собственных шумов вакуумных фотоэмиссионных приемников излучения?
- ☐ Какие приемы используют для увеличения отношения сигнал/шум?
- ☐ Какие факторы определяют пороговую чувствительность фотоумножителя и относительную погрешность при измерении больших световых потоков?
- ☐ Почему большие флуктуации интенсивности излучения нелазерных источников не влияют на статистическое распределение фотоотсчетов?
- ☐ Какими преимуществами обладают приемники излучения на основе внутреннего фотоэффекта?

9.6. Энергия и импульс фотона. Дуализм света

Электромагнитная теория, рассматривающая свет как классические электромагнитные волны (т. е. как возможные решения уравнений Максвелла), исчерпывающе описывает распространение света в пустоте, интерференцию и дифракцию. Будучи дополненной электронной теорией, она оказывается в состоянии охватить и широкий круг вопросов, относящихся к взаимодействию света с веществом, таких, как рассеяние, поглощение, преломление, дисперсия. Но применимость классической теории в вопросах взаимодействия света с веществом имеет предел. Так, например, в фотоэффекте проявляется внезапная пространственная локализация конечной порции энергии света на одном электроне. Такое поведение несовместимо с классической волновой картиной, но получает естественное объяснение в корпускулярной картине, трактующей свет как поток фотонов. Ведь именно частица может доставить конечную порцию энергии в определенную точку.

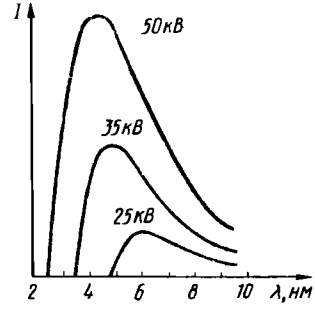
Локализация кванта энергии излучения в пространстве особенно ярко проявляется в фотоэффекте на отдельных атомах или молекулах (фотоионизация). В классической электромагнитной волне энергия непрерывно распределена по всему волновому фронту.

Исходя из известного значения площади эффективного сечения, можно оценить, сколько времени должно пройти с момента начала облучения для того, чтобы атом мог накопить достаточную для вырывания электрона (ионизации) энергию E_i. В условиях реального эксперимента это могут быть недели или месяцы. Однако опыт показывает, что фотоэлектроны с энергией $\hbar\omega - E_i$ появляются практически сразу после начала облучения. Значит, классическое непрерывное распределение энергии по фронту характеризует перенос энергии излучения только в среднем, но не для элементарных актов взаимодействия с веществом, свидетельствующих о пространственной локализации переносимой энергии, что характерно для потока частиц.

Волновой и корпускулярный аспекты излучения связаны друг с другом: атрибут волновой картины — частота ω — входит в соотношение $\varepsilon = \hbar\omega$, определяющее энергию световой частицы — фотона. В опытах по фотоэффекту это соотношение проверяется измерением энергии фотоэлектронов, образуемых фотонами монохроматического излучения известной частоты. Другое подтверждение ему дает обратный процесс — испускание фотонов быстрыми электронами при торможении. Это происходит, например, в антикатоде рентгеновской трубки. Опыт показывает, что при данном ускоряющем напряжении U в рентгеновском спектре отсутствует излучение с длинами волн, меньшими некоторого значения λ_{min}. Коротковолновая граница тормозного излучения λ_{min} определяется ускоряющим напряжением U и не зависит от материала мишени (рис. 9.14). В целом процесс излучения при торможении электрона в электрических полях, создаваемых атомами мишени, весьма сложен, но коротковолновая граница в корпускулярной картине получает очень простое объяснение. Как и в уравнении Эйнштейна (9.39) для фотоэффекта, можно записать закон сохранения энергии в элементарном акте испускания кванта излучения. Фотон получает наибольшую энергию в том случае, когда электрон полностью останавливается при столкновении с ядром атома мишени. Из-за большой массы ядра такой процесс не противоречит закону сохранения импульса. Значит, максимальная энергия испускаемых фотонов $\hbar\omega_{max}$ равна кинетической энергии электронов $^1/_2 mv^2 = eU$:

$$\hbar\omega_{max} = eU, \text{ или } \lambda_{min} = ch/eU. \qquad (9.47)$$

Существование граничной частоты ω_{max} демонстрирует квантовый характер испускания рентгеновского излучения. По измерению ее зависимости от ускоряющего напряжения можно с высокой точностью определить значение постоянной



9.11

Коротковолновая граница тормозного рентгеновского спектра

Планка $\hbar$, при этом получается хорошее согласие со значениями, найденными на основе черного излучения и фотоэффекта, что экспериментально свидетельствует о выполнении соотношения $\varepsilon = \hbar\omega$ между энергией фотона и частотой для очень широкой области спектра и указывает на его всеобщий характер.

Согласно классической электромагнитной теории, бегущая электромагнитная волна обладает импульсом $\mathbf{p}$, направленным вдоль волнового вектора $\mathbf{k}$ и пропорциональным энергии волны: $p = W/c$ (см. § 3.5). Можно ожидать, что такое соотношение между энергией и импульсом выполняется и для элементарного кванта излучения — фотона: $p = \hbar\omega/c$. Учитывая, что отношение ω/c равно волновому числу k, запишем его в векторном виде (вместе с соотношением между энергией фотона и частотой):

$$\varepsilon = \hbar\omega, \quad \mathbf{p} = \hbar\mathbf{k}. \tag{9.48}$$

Частота ω и волновой вектор $\mathbf{k}$ характеризуют волновые свойства монохроматического излучения, а энергия ε и импульс $\mathbf{p}$ — корпускулярные. Второе соотношение (9.48), связывающее импульс фотона с волновым вектором, неизбежно следует из первого, связывающего энергию с частотой, если обратиться к требованию равноправия всех инерциальных систем отсчета, т. е. к принципу относительности. В самом деле, энергия (деленная на постоянный множитель c) и импульс частицы образуют четырехмерный вектор $(\varepsilon/c, \mathbf{p})$, а частота (деленная на c) и волновой вектор образуют четырехмерный волновой вектор $(\omega/c, \mathbf{k})$ монохроматической волны. При переходе от одной инерциальной системы отсчета к другой пространственные и временны́е компоненты 4-векторов в соответствии с преобразованиями Лоренца (8.7) «перемешиваются» друг с другом. Фундаментальное соотношение $\varepsilon = \hbar\omega$ между временны́ми компонентами 4-векторов $(\varepsilon/c, \mathbf{p})$ и $(\omega/c, \mathbf{k})$ будет удовлетворять требованию релятивистской инвариантности, т. е. выполняться одновременно во всех системах отсчета, тогда и только тогда, когда такое же соотношение $\mathbf{p} = \hbar\mathbf{k}$ имеет место и между их пространственными компонентами.

Обращаясь к релятивистской формуле $(\varepsilon/c)^2 - p^2 = m_0^2 c^2$, связывающей энергию и импульс частицы*, и учитывая, что для фотона $p = \varepsilon/c$, видим, что масса покоя фотона $m_0 = 0$. Если мы склонны рассматривать фотон как частицу, этот результат на первый взгляд может показаться странным. Однако это говорит лишь о том, что, обладая некоторыми свойствами частицы, фотон не во всех отношениях похож на обычную частицу: он в любой системе отсчета имеет скорость c, и не существует системы отсчета, в которой бы

* Напомним, что разность квадратов временно́й и пространственной компонент любого 4-вектора одинакова во всех инерциальных системах отсчета (релятивистский инвариант). Приведенное соотношение получается, если эту разность квадратов для 4-вектора энергии-импульса приравнять ее значению в системе отсчета, где частица покоится, и воспользоваться формулой $\varepsilon_0 = m_0 c^2$.

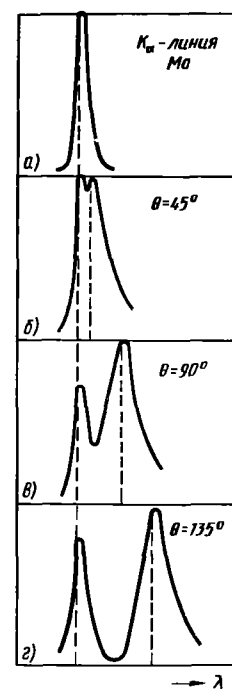
он покоился. Равенство нулю массы покоя означает, что фотон в состоянии покоя — понятие, лишенное смысла.

Таким образом, квант монохроматического электромагнитного поля во взаимодействии с веществом проявляет себя как частица с энергией и импульсом, определяемыми соотношениями (9.48). Взаимодействие света с веществом можно рассматривать как совокупность элементарных актов поглощения, испускания и рассеяния фотонов, в каждом из которых выполняются законы сохранения энергии и импульса. В рассмотренных выше явлениях фотоэффекта и тормозного излучения мы учитывали только закон сохранения энергии при поглощении или испускании фотона, так как массивный катод мог, не участвуя в энергетическом балансе, «принять на себя» любой импульс и этим обеспечить выполнение закона его сохранения. Но существуют явления, в которых импульс фотона обнаруживает себя явно и соотношение $\mathbf{p}=\hbar\mathbf{k}$ допускает экспериментальную проверку. В качестве примера рассмотрим рассеяние рентгеновского излучения электронами, впервые количественно исследованное Комптоном в 1923 г.

Опыты Комптона показали, что при рассеянии пучка монохроматических рентгеновских лучей на мишени из вещества с небольшим атомным номером в рассеянном излучении наряду с неизменной длиной волны появляется спектральная компонента, смещенная в сторону длинных волн. Наблюдаемое изменение длины волны $\Delta\lambda=\lambda'-\lambda$ не зависит от материала мишени, а определяется лишь углом θ между направлением падающего пучка и направлением рассеянного излучения. С увеличением угла θ интенсивность несмещенной компоненты падает, а смещенной — растет (рис. 9.15).

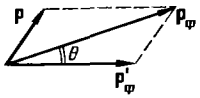
В атомах легких элементов энергия связи электрона (энергия ионизации) порядка 10 эВ, что примерно в тысячу раз меньше энергии рентгеновского фотона ($\hbar\omega\approx 10$ кэВ). Поэтому электроны в этих опытах можно считать практически свободными. Энергия покоя электрона $mc^2\approx 0{,}5$ Мэв, что много больше энергии фотона. Поэтому электрон, который до столкновения покоился или двигался в атоме с нерелятивистской скоростью, и после столкновения с таким фотоном останется нерелятивистским. Пренебрегая начальной энергией электрона, запишем закон сохранения энергии

$$\hbar\omega=\hbar\omega'+p^2/(2m). \qquad (9.49)$$



9.15

Спектр рассеянного на графите излучения K_α-линии молибдена



9.16

Сохранение импульса в эффекте Комптона

Здесь p — импульс электрона после столкновения с фотоном. Закон сохранения импульса

$$\mathbf{p}_\phi = \mathbf{p}'_\phi + \mathbf{p}, \tag{9.50}$$

где $\mathbf{p}_\phi$ и $\mathbf{p}'_\phi$ — импульс фотона до и после рассеяния, запишем с помощью теоремы косинусов (рис. 9.16) в следующем виде:

$$p^2 = p_\phi^2 + p_\phi'^2 - 2p_\phi p'_\phi \cos\theta. \tag{9.51}$$

Подставим в это равенство выражение для импульса фотона через его частоту $p_\phi = \hbar\omega/c$ и квадрат импульса электрона из закона сохранения энергии (9.49):

$$2m\hbar(\omega - \omega') = (\omega^2 + \omega'^2 - 2\omega\omega'\cos\theta)\hbar^2/c^2.$$

Переписав это уравнение в виде

$$\omega - \omega' = [\hbar\omega/(2mc^2)](\omega + \omega'^2/\omega - 2\omega'\cos\theta) \tag{9.52}$$

и учитывая, что $\hbar\omega/(mc^2) \ll 1$, видим, что изменение частоты $\Delta\omega = \omega' - \omega$ мало по сравнению с самой частотой. Поэтому в правой части (9.52) можно заменить ω' на ω. Тогда для относительного сдвига частоты $\Delta\omega/\omega$ при рассеянии получаем $\Delta\omega/\omega = -[\hbar\omega/(mc^2)] \times (1 - \cos\theta) = -[2\hbar\omega/(mc^2)]\sin^2(\theta/2)$. Знак минус показывает, что частота излучения при рассеянии уменьшается. Это и понятно, так как часть своей энергии фотон передает электрону. Переходя теперь от частот к длинам волн $\lambda = 2\pi c/\omega$, $\Delta\lambda/\lambda = -\Delta\omega/\omega$, окончательно получаем

$$\Delta\lambda = [2h/(mc)]\sin^2(\theta/2) = 2\Lambda\sin^2(\theta/2). \tag{9.53}$$

Соотношение (9.53) очень хорошо описывает наблюдаемую на опыте зависимость увеличения длины волны от угла рассеяния (см. рис. 9.15). Входящая в (9.53) константа Λ, называемая *комптоновской длиной волны* электрона, выражается в виде комбинации трех универсальных постоянных: $\Lambda = h/(mc)$. Подставляя их числовые значения, получаем $\Lambda = 0{,}24$ нм, что совпадает с величиной, найденной из измерений смещения длины волны при рассеянии.

Происхождение несмещенной компоненты в спектре рассеянного излучения обусловлено внутренними электронами атомов мишени. Их энергия связи, особенно в тяжелых атомах, сравнима с энергией рентгеновских фотонов, и, следовательно, такие электроны уже нельзя рассматривать как свободные. Поэтому в акте рассеяния фотон обменивается энергией и импульсом с атомом в целом, а так как масса атома велика, то по закону сохранения импульса фотон практически не передает ему своей энергии. Фотоны, рассеянные внутренними электронами, образуют несмещенную компоненту. Из приведенных рассуждений ясно, почему эффект Комптона нельзя наблюдать в видимой области спектра. Энергия фотона видимого света составляет лишь несколько электрон-вольт. При этом даже внешние электроны нельзя считать свободными.

При рассеянии фотона часть его энергии передается электрону. Электроны отдачи играют важную роль в процессе ионизации газов рентгеновским излучением. Используя камеру Вильсона, помещенную в магнитное поле, можно найти импульс и энергию элек-

тронов отдачи, что позволяет еще более полно проверить выполнение законов сохранения. Измерения показали, что в каждом элементарном акте рассеяния электрон приобретает как раз такие энергию и импульс, какие теряет фотон.

Опыты Комптона ярко продемонстрировали, что энергия и импульс фотона действительно выражаются формулами (9.48) и что законы сохранения энергии и импульса выполняются в элементарных актах рассеяния.

Таким образом, опытные факты говорят о том, что в ряде явлений свет обнаруживает корпускулярные свойства: взаимодействие с веществом имеет характер дискретных процессов, в которых поглощается, испускается или рассеивается целый квант. Но представление о свете как потоке классических корпускул несовместимо с классической картиной электромагнитных волн, которая, в свою очередь, находит опытное подтверждение в явлениях интерференции и дифракции.

Впрочем, вопрос о природе света не всегда стоит так категорично: или частицы, или волны. Существует ряд явлений, допускающих корректное объяснение с любой из этих точек зрения. В качестве примеров можно привести эффект Доплера и давление света.

Покажем, что типично волновое на первый взгляд явление Доплера можно объяснить с корпускулярной точки зрения, рассматривая свет как поток фотонов. Пусть «закрепленный» неподвижный атом при переходе в состояние с меньшей энергией испускает фотон с энергией $\hbar\omega$. Разность энергий рассматриваемых состояний атома равна $\hbar\omega$ и не зависит от того, покоится атом или движется. При испускании фотона незакрепленным, свободно движущимся атомом импульс атома изменяется, поскольку испущенный фотон обладает импульсом. Изменяется и кинетическая энергия атома. На основании закона сохранения энергии

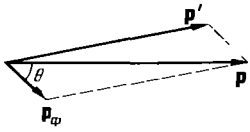
$$\hbar\omega + p^2/(2M) = \hbar\omega' + p'^2/(2M), \qquad (9.54)$$

где $\hbar\omega'$ — энергия испущенного фотона, p и p' — импульс атома до и после испускания фотона, M — масса атома. Начальный и конечный импульсы атома можно связать с импульсом $\mathbf{p}_\phi$ испускаемого фотона с помощью закона сохранения импульса: $\mathbf{p} = \mathbf{p}' + \mathbf{p}_\phi$ (рис. 9.17). Перенесем здесь $\mathbf{p}_\phi$ в левую часть и возведем полученное равенство в квадрат. Учитывая, что импульс фотона крайне мал по сравнению с импульсом движущегося излучающего атома ($p_\phi \ll p$), получаем*

$$p^2 - 2pp_\phi \cos\theta \approx p'^2.$$

Используя это, перепишем закон сохранения энергии (9.54) в следующем виде:

* Пренебрегая квадратом импульса фотона, мы теряем сдвиг частоты из-за эффекта отдачи атома, который может быть существенным при излучении γ-квантов атомным ядром (см. задачи 1, 2). В оптической области эффект отдачи приводит к ничтожному сдвигу частоты.



9.17
Эффект Доплера при испускании фотона

$$\hbar\omega' - \hbar\omega = (pp_\phi/M)\cos\theta.$$

Подставляя сюда импульс испущенного фотона $p_\phi = \hbar\omega'/c$ и учитывая, что p/M есть скорость v движения излучающего атома, находим для относительного сдвига частоты

$$\Delta\omega/\omega = (v/c)\cos\theta, \qquad (9.55)$$

что совпадает с обычным нерелятивистским выражением для эффекта Доплера. Более точную (релятивистскую) формулу можно получить, если в закон сохранения энергии подставлять релятивистское выражение для энергии атома (см. задачу 2).

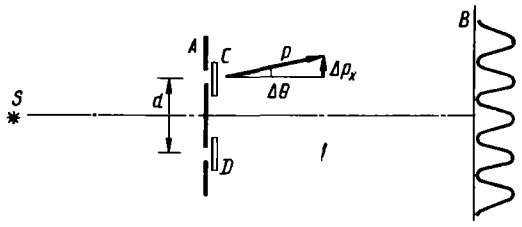
Впрочем, возможность как волновой, так и корпускулярной трактовки эффекта Доплера сразу видна из релятивистски инвариантного характера соотношений (9.48). Преобразование энергии фотона ε при переходе от одной инерциальной системы отсчета к другой происходит по тем же формулам, что и преобразование частоты ω электромагнитной волны, поскольку ε и ω во всех системах отсчета пропорциональны друг другу (коэффициент пропорциональности — универсальная постоянная $\hbar$) и представляют собой временны́е компоненты соответствующих четырехмерных векторов.

Давление света, предсказываемое электромагнитной теорией Максвелла и впервые обнаруженное в опытах П. Н. Лебедева, по корпускулярным представлениям интерпретируется как результат передачи импульса фотонов поглощающей или отражающей стенке. При нормальном падении каждый поглощенный фотон передает стенке импульс $\hbar\omega/c$, отраженный — $2\hbar\omega/c$. На единицу площади стенки в единицу времени падает энергия излучения Uc, где U — объемная плотность энергии. Поэтому число падающих фотонов $N = Uc/\hbar\omega$. Из него NR фотонов отразится, $N(1-R)$ — поглотится (R — энергетический коэффициент отражения поверхности). Давление p света равно полному импульсу, передаваемому падающими фотонами на 1 м^2 поверхности в 1 с, т. е.

$$p = NR2\hbar\omega/c + N(1-R)\hbar\omega/c = U(1+R), \qquad (9.56)$$

что совпадает с выражением (см. § 3.5), полученным на основе электромагнитной теории. Изотропное излучение, распространяющееся по всем направлениям, оказывает на стенку давление $p = U/3$. Такое выражение можно получить, рассматривая излучение в замкнутой полости как идеальный газ фотонов. Учет светового давления необходим при применении термодинамических методов к проблеме равновесного теплового излучения.

Рассмотренные примеры показывают, что при анализе конкретных явлений двойственность света не приводит к противоречиям. Противоречие возникает только тогда, когда мы пытаемся составить общее представление о свете. В самом деле, соотношения (9.48) связыва-



9.18

Интерференционный опыт с двумя щелями

ют волновые и корпускулярные свойства света: правые части содержат частоту ω и волновое число k, определяемые из интерференционных явлений, а левые части ε и p характеризуют фотон как частицу. Но именно сосуществование этих свойств и не может быть логически непротиворечиво объяснено классической физикой, с точки зрения которой понятия волны и частицы исключают друг друга, так как описывают полярно противоположные формы движения. Каким образом частица может иметь некоторую частоту или длину волны?

Для иллюстрации возникающих логических трудностей и того, как они преодолеваются квантовой теорией, рассмотрим простой интерференционный опыт. Свет от точечного источника S падает на непрозрачный экран A, в котором прорезаны две узкие параллельные щели C и D (рис. 9.18). Расстояние между экранами A и B велико по сравнению с расстоянием d между щелями, которое, в свою очередь, много больше длины световой волны. На экране B возникают интерференционные полосы. Каждый фотон, попадая на экран B, ведет себя как частица, вызывая в определенной точке почернение фотопластинки при фотографической записи или вырывание фотоэлектрона при фотоэлектрической регистрации. Но распределение большого числа таких точек регистрации фотонов описывается классической картиной интерференции волн, приходящих от двух щелей.

Придерживаясь корпускулярных представлений, можно было бы пытаться объяснить интерференционные полосы каким-то взаимодействием между фотонами, проходящими через разные щели. Но тогда при очень слабом источнике S, когда в каждый момент в установке не может находиться более одного фотона, интерференционные полосы должны исчезнуть. Прямые опыты такого типа (Тэйлор, 1909 г.) показывают, что никаких изменений в интерференционной картине при уменьшении интенсивности источника не происходит. Это значит, что интерференция представляет собой статистическое свойство отдельного фотона.

Если закрыть одну из щелей, то интерференционные полосы пропадают: распределение интенсивности на экране становится таким же, как при дифракции на одной щели, т. е. при достаточно узкой щели практически равномерным. Это вынуждает нас считать,

что на пути от источника через щели к экрану B излучение ведет себя как волна, ибо появление полос наблюдается лишь когда свет проходит через обе щели. Невозможно представить себе классическую частицу, которая проходила бы сразу через обе щели. Но если фотон проходит через одну щель, то каким образом другая щель, через которую он не проходит, не позволяет ему попасть в те места экрана, куда он мог бы попасть, если бы эта щель была закрыта?

Такая постановка вопроса основана на естественном для классической физики допущении, что частица в любой момент времени имеет определенные координаты, доступные измерению, и, следовательно, движется по определенной траектории. Современная квантовая теория отказывается от этого допущения, утверждая, что говорить об определенном положении частицы имеет смысл лишь в том случае, если при постановке опыта мы позаботимся об измерении ее координаты. Значит, если мы хотим считать, что каждый фотон, подобно частице, проходит только через одну из щелей, то мы должны поставить какой-либо измерительный прибор, который бы фиксировал нам факт прохождения фотона через определенную щель. Можно, например, попытаться с помощью счетчиков C и D фиксировать, через какое отверстие проходит каждый фотон. При этом фотоны демонстрируют действительно корпускулярное поведение: счетчики никогда не срабатывают одновременно, т. е. отдельный фотон сразу через две щели не проходит. Но оказывается, что такое «вмешательство» счетчиков выразится в полном размытии интерференционных полос на экране B!

В самом деле, сохранение интерференционных полос в такой ситуации противоречило бы соотношениям неопределенностей Гейзенберга. Чтобы установить факт прохождения частицы через определенную щель, счетчик должен определять координату x фотона (в направлении, перпендикулярном щелям) с погрешностью Δx, не превышающей половины расстояния между щелями: $\Delta x < d/2$. Такое измерение в соответствии с основными положениями квантовой теории неизбежно вносит неконтролируемое изменение x-составляющей импульса фотона $\Delta p_x \gtrsim h/\Delta x$, т. е. неопределенность $\Delta\theta$ в направлении его распространения (рис. 9.21): $\Delta\theta = \Delta p_x/p \gtrsim h/(p\Delta x)$. Так как $\Delta x < d/2$, а импульс фотона $p = \hbar k = h/\lambda$, то $\Delta\theta > 2\lambda/d$. Эта неопределенность в направлении фотона, вносимая определением щели, через которую он прошел, превышает угловую ширину отдельной интерференционной полосы λ/d, т. е. приводит к полному размытию интерференционной картины. Установив, через какую щель проходит фотон, мы утрачиваем интерференционную картину и не можем говорить о проявлении фотонами волновых свойств.

Подведем некоторые итоги. Попадая на экран B, фотоны обнаруживают корпускулярное поведение, заключающееся в том, что они взаимодействуют с веществом экрана только в отдельных точках. Если при этом не пытаться экспериментально наблюдать траектории фо-

тонов до попадания их на экран (т. е. не устанавливать, через какую щель проходит каждый из них), то, пропустив большое число фотонов, на экране B мы получим интерференционную картину, которую нельзя объяснить иначе как на основе волновых представлений. Предсказать, в какое место экрана попадет отдельный фотон, нельзя. Это можно сделать только в статистическом смысле: вероятность зарегистрировать фотон в области максимума интерференционной картины велика, а в области минимума — мала. Плотность энергии электромагнитного поля, вычисляемая в классической волновой теории на основе уравнений Максвелла, определяет вероятность того, что фотон будет зарегистрирован в данном месте, т. е. дает среднее пространственное распределение, наблюдаемое с очень большим числом фотонов. В этом смысле классическая волновая теория вполне справедлива. Если же мы пытаемся наблюдать отдельные фотоны, например, фиксируем их траектории, как в рассмотренном выше опыте, то фотоны совсем не проявляют волновых свойств. Квантовая теория не отвергает полностью классическую волновую картину, а указывает на ее ограниченность.

Итак, что же такое свет — частицы или волны? Квантовая теория отвечает на этот вопрос так: ни то, ни другое. Когда мы описываем поведение фотона как поведение частицы или волны, мы навязываем классическое описание этому объекту, имеющему существенно неклассическую природу. Свет может характеризоваться только с той его стороны (корпускулярной или волновой), проявление которой обусловлено внешними условиями, создаваемыми экспериментальными *средствами наблюдения*. Такая постановка вопроса позволяет рассматривать и тот случай, когда один и тот же исследуемый объект (свет) обладает несовместимыми по классическим представлениям корпускулярными и волновыми свойствами. В квантовой теории эти свойства не исключают, а *дополняют* друг друга, так как в чистом виде они могут проявиться лишь в разных опытах, производимых при *взаимоисключающих условиях*. Этим и объясняется отсутствие логического противоречия в понятии «корпускулярно-волновой дуализм». Действительно, нет необходимости пытаться представить себе, как это фотон может быть сразу и волной, и частицей. Свет обладает *потенциальной возможностью* проявлять и волновые, и корпускулярные свойства, но эти дополнительные свойства в чистом виде проявляются лишь при взаимоисключающих условиях эксперимента. Адекватный способ описания света определяется выбранным способом наблюдения, а вопрос о том, что же существует «на самом деле» — волна или частица, — лишен содержания.

Важно подчеркнуть, что изложенные здесь теоретические идеи не могут быть чисто логически выведены из каких-то общих принципов. Они возникли на основе анализа большой совокупности экспериментальных фактов. Заметим, что корпускулярно-волновой дуализм присущ не только фотонам, но и любым другим микрообъектам — электронам, протонам, нейтронам и т. п.

Математический аппарат современной квантовой теории (изложение которого выходит за рамки данной книги) органически включает в себя возможность проявления изучаемым микрообъектом (фотоном, электроном) как волновых, так и корпускулярных свойств. Этот аппарат свободен от абстракций, присущих классическому описанию и заключающихся в абсолютизации физического процесса и принципиальной возможности неограниченной его детализации. В его основе лежит явный учет реальных возможностей измерений, проводимых над микрообъектами (относительность к средствам наблюдения). Результат взаимодействия микрообъекта с классически описываемым прибором при заданных внешних условиях характеризуется некоторой вероятностью. Вероятности в квантовой физике имеют совсем иной характер, нежели в классической, где они отражают неполноту наших знаний о физической системе. Здесь они принципиально необходимы, так как отражают объективно существующие при данных условиях потенциальные возможности. Математический аппарат квантовой теории позволяет находить как возможные значения физических величин, так и вероятности получения на опыте тех или иных значений этих величин при измерении в определенных условиях.

Контрольные вопросы

- □ В каких явлениях обнаруживаются волновые и в каких — корпускулярные свойства света?
- □ Как объясняется происхождение коротковолновой границы спектра тормозного рентгеновского излучения?
- □ Какие явления свидетельствуют о том, что фотон обладает импульсом? Как связаны импульс фотона и волновой вектор?
- □ Как объясняется изменение длины волны при рассеянии рентгеновских лучей свободными электронами?
- □ Приведите примеры явлений, допускающих как волновое, так и корпускулярное объяснение.
- □ Объясните, в чем заключается происхождение противоречий, возникающих при попытках составить общее представление о свете, проявляющем то волновые, то корпускулярные свойства. Как эти логические трудности преодолеваются квантовой теорией?

Задачи

1. Первоначально покоившийся атом (или ядро) массы M испускает фотон (γ-квант) при переходе между энергетическими уровнями, разность которых равна $\hbar\omega_0$. Найти частоту ω испущенного излучения. Оцените $\Delta\omega/\omega_0$ для желтой линии натрия и для γ-линии 113 кэВ ядра изотопа гафния $^{177}_{72}\mathrm{Hf}$.

◇ Ответ: $\omega=\omega_0[1-\hbar\omega_0/(2Mc^2)]$. Эта формула описывает *эффект отдачи* при испускании фотона, аналогичный откату ствола орудия при выстреле. Частота (и энергия) фотона всегда меньше той, которую он имел бы при бесконечно большой массе ядра. Для оптических фотонов эффект отдачи

ничтожен (сдвиг частоты много меньше естественной ширины линии). ◇

2. Используя законы сохранения энергии и импульса, получить релятивистскую формулу для частоты фотона, испускаемого движущимся атомом (с учетом эффекта отдачи).

◇ Ответ: $\omega=\omega_0\left(1-\dfrac{\hbar\omega_0}{2Mc^2}\right)\dfrac{\sqrt{1-v^2/c^2}}{1-(v/c)\cos\theta}$.

Здесь θ — угол между направлением испущенного фотона и скоростью атома v. При $v=0$ полученное выражение совпадает с ответом предыдущей задачи. При $\hbar\omega_0 \ll Mc^2$ получаем формулу для релятивистского эффекта Доплера. ◇

9.

10

ОСНОВЫ НЕЛИНЕЙНОЙ ОПТИКИ

● Вскоре после создания лазеров началось бурное развитие нелинейной оптики, объединяющей круг явлений, обусловленных зависимостью оптических параметров среды от интенсивности света.

● Такие оптические характеристики среды, как показатель преломления и коэффициент поглощения, не зависят от интенсивности света, если реакция среды на электрическое поле световой волны описывается линейным материальным уравнением $\mathbf{P} = \varepsilon_0 \chi \mathbf{E}$, т. е. поляризованность пропорциональна напряженности поля, а коэффициент пропорциональности — восприимчивость χ — зависит только от свойств среды. Но это материальное уравнение приближенно: оно справедливо лишь при напряженностях E электрического поля световой волны, малых по сравнению с напряженностями E_a внутриатомных электрических полей. Для оценки E_a можно считать, что по порядку величины $E_a \sim e/(4\pi\varepsilon a^2)$, где a — размер атома. Так как $a \sim 10^{-8}$ см, то $E_a \sim 10^8 - 10^9$ В/см. В пучках света от нелазерных источников достижимые значения напряженности поля не превосходят 0,1—10 В/см и линейное материальное уравнение выполняется с большой точностью.

● Но в интенсивных лазерных пучках напряженность E поля достигает 10^6—10^7 В/см. В таких полях модель гармонического осциллятора для описания поведения оптического электрона атома уже неприменима и связь поляризованности среды с напряженностью поля световой волны становится нелинейной. В результате возникает зависимость оптических характеристик среды от интенсивности излучения, которая приводит не к каким-либо малым поправкам, а к принципиально новым эффектам, не существующим в линейной оптике. Нелинейная оптика существенно расширяет наши представления о взаимодействии света с веществом.

10.1. Некогерентные нелинейные эффекты

Первый нелинейный эффект был открыт С. И. Вавиловым и В. Л. Левшиным еще в 1925 г. Несмотря на трудности опытов с нелазерными источниками, им удалось наблюдать небольшое уменьшение (до 1,5%) коэффициента поглощения уранового стекла с увеличением интенсивности света. Возникновение этого нелинейного *эффекта насыщения* обусловлено выравниванием населенностей двух уровней энергии, между которыми происходят квантовые переходы с поглощением и испусканием света (см. задачу). При большой интенсивности падающего света вероятность вынужденных переходов может приблизиться к вероятности релаксационных спонтанных переходов, посредством которых возбужденные атомы или молекулы избавляются от избытка энергии и возвращаются в основное состояние. С увеличением интенсивности рост поглощаемой мощности замедляется и она приближается к некоторому пределу, определяемому скоростью, с которой возбужденные атомы отдают избыток энергии окружающей среде (скоростью релаксации).

Уменьшение доли поглощаемой мощности с увеличением интенсивности света, т. е. *просветление* нелинейной поглощающей среды при прохождении сильных световых пучков, нашло применение в лазерной технике, где оно используется для модуляции добротности оптических резонаторов с целью генерации сверхкоротких мощных импульсов (см. § 9.4). Ячейка с нелинейной поглощающей средой представляет собой автоматический затвор, открывающийся под действием мощного светового пучка. Очень важно, что такой затвор обладает малой инерционностью, так как после прохождения мощного светового импульса среда быстро становится снова непрозрачной для слабого света.

В активной среде с инверсией населенностей эффект насыщения приводит к уменьшению коэффициента усиления при увеличении интенсивности света и тем самым к установлению стационарного режима генерации в лазерах.

Если непрозрачная для слабого света среда за счет эффекта насыщения становится прозрачной в сильных световых пучках, то прямо противоположная ситуация наблюдается в области оптической прозрачности среды. Здесь в результате *многофотонного поглощения* интенсивный свет может поглощаться гораздо сильнее, чем слабый. При больших плотностях излучения система с уровнями энергии ε_1 и ε_2 может поглотить в элементарном акте два фотона с частотами ω_1 и ω_2 такими, что $\hbar\omega_1+\hbar\omega_2=\varepsilon_2-\varepsilon_1$ (в частном случае $\omega_1=\omega_2=$ $=1/2(\varepsilon_2-\varepsilon_1)/\hbar$). Вероятность поглощения двух фотонов пропорциональна произведению интенсивностей пучков с частотами ω_1 и ω_2 (или квадрату интенсивности при $\omega_1=\omega_2$). Возможно также одновременное поглощение трех фотонов и более. Многофотонное поглощение находит применение в нелинейной лазерной спектроскопии и позволяет получать информацию об энергетических уров-

нях квантовых систем, недоступную для традиционных методов спектроскопии.

К нелинейным эффектам поглощения примыкает и *многофотонный фотоэффект.* В экспериментах с фокусируемыми лазерными пучками достигаются столь высокие плотности световой энергии, что становятся доступными наблюдению процессы, в которых атом одновременно поглощает до 7—8 фотонов. В результате может произойти фотоионизация атома светом малой частоты, т. е. в интенсивных световых пучках исчезает красная граница фотоэффекта на отдельном атоме. Интересно, что Эйнштейн в работе 1905 г., содержащей вывод основного уравнения фотоэффекта, не исключал принципиальной возможности процессов с участием более чем одного фотона.

Рассмотренные выше нелинейные оптические эффекты относятся к числу *некогерентных*, так как в различных участках среды они развиваются в значительной мере независимо и не связаны с передачей нелинейного возбуждения от одной точки среды к другой. Преимущество лазерного излучения для их наблюдения обусловлено его большой мощностью, а роль высокой пространственной когерентности сводится лишь к возможности фокусировки в область малого размера для получения больших плотностей энергии.

Высокие монохроматичность и направленность лазерного излучения играют принципиально важную роль для наблюдения *когерентных* нелинейных оптических эффектов, таких, как генерация оптических гармоник и параметрическое взаимодействие волн. Для них важны фазовые соотношения взаимодействующих волн и характерна возможность получения в определенных условиях пространственного накопления нелинейных эффектов по мере распространения света в среде. Когерентным эффектам уделено основное внимание в дальнейшем изложении.

Контрольные вопросы

Объясните возникновение эффекта насыщения, используя простейшую модель поглощающей среды. Чем объясняется замедление роста поглощаемой мощности при увеличении интенсивности?
Какую роль играют эффекты насыщения в лазерах?
В чем состоят основные отличия закономерностей многофотонного поглощения и многофотонного фотоэффекта от законов соответствующих однофотонных явлений?
Почему рассмотренные здесь нелинейные оптические эффекты называют некогерентными?

Задача

Атомы среды могут находиться в двух состояниях с энергиями ε_1 и ε_2. Рассмотреть распространение в такой среде света с частотой ω, соответствующей переходу между этими уровнями ($\hbar\omega = \varepsilon_2 - \varepsilon_1$). Показать, что

при больших интенсивностях света коэффициент поглощения уменьшается (насыщение).

◇ Пусть N — концентрация атомов, из них N_1 находятся на нижнем уровне (в основном состоянии), N_2 — на верхнем. Изменение N_1 и N_2 происходит из-за спонтанных переходов с верхнего уровня со скоростью $A_{21}N_2$, определяемой коэффициентом Эйнштейна $A_{21}=1/\tau$ (τ — время жизни атома в возбужденном состоянии), и из-за вынужденных переходов с поглощением и испусканием, вероятность которых пропорциональна интенсивности света. Учитывая все три процесса, скорость изменения N_1 представим в виде

$$dN_1/dt = A_{21}N_2 - I(B_{12}N_1 - B_{21}N_2)/c.$$

Учитывая, что $B_{12}=B_{21}\equiv B$, и вводя для краткости обозначения $A\equiv A_{21}$, $W\equiv IB/c$, получаем уравнение $dN_1/dt = AN_2 - W(N_1 - N_2)$, которое вместе с условием $N_1+N_2=N$ позволяет найти населенности N_1 и N_2. Для стационарного состояния ($dN_1/dt=0$) решение имеет вид

$$N_1 = N(A+W)/(A+2W),\quad N_2 = NW/(A+2W).$$

Поглощаемая в единичном объеме мощность P и коэффициент поглощения α [см. (9.36), (9.37)] выражаются через разность населенностей $N_1 - N_2 = NA/(A+2W)$:

$$P=\hbar\omega W(N_1-N_2),\quad \alpha=\hbar\omega B(N_1-N_2)/c.$$

При малой интенсивности падающего излучения ($W=IB/c\ll A$) почти все атомы находятся в основном состоянии: $N_1\approx N$, $N_2\approx 0$, а коэффициент поглощения не зависит от интенсивности и равен $\alpha_0=\hbar\omega BN/c$. С увеличением интенсивности происходит выравнивание населенностей: при $W\to\infty$ N_1 и N_2 стремятся к $N/2$, а коэффициент поглощения α стремится к нулю (просветление среды). Если ввести значение интенсивности $I_{нас}$, при котором α уменьшается вдвое, зависимость α от интенсивности принимает вид

$$\alpha(I) = \frac{\alpha_0}{1+I/I_{нас}}.$$

Поглощаемая мощность P при увеличении интенсивности растет сначала линейно, затем этот рост замедляется и $P(I)$ при $I\to\infty$ стремится к конечному значению $P_{макс}=\hbar\omega NA/2=\hbar\omega N/(2\tau)$. Таким образом, при насыщении поглощаемая мощность определяется временем жизни τ атома в возбужденном состоянии: среда поглощает столько энергии из падающего света, сколько ее могут спонтанно излучить возбуждаемые атомы. ◇

10.2. Материальные уравнения для нелинейных сред

Действующее на молекулу или атом электрическое поле световой волны вызывает смещение зарядов (деформацию электронного облака), что приводит к появлению у молекулы индуцированного дипольного момента. Этот дипольный момент пропорционален напряженности поля волны при не слишком больших ее значениях. В таком приближении поляризуемость молекулы (или атома) и, следовательно, все оптические характеристики среды не зависят от интенсивности световой волны. Но в интенсивных лазерных световых пучках напряженность электрического поля нельзя считать пренебрежимо малой по сравнению с внутриатомными электрическими полями, и связь индуцированного дипольного момента с напряженностью поля уже не будет линейной.

Чтобы составить представление о нелинейной поляризуемости молекулы, будем исходить из простой модели, лежащей в основе классической электронной теории дисперсии (см. § 2.3). Согласно этой модели, смещение $x(t)$ оптического электрона из положения равновесия в поле световой волны $E(t)$ описывается уравнением $m\ddot{x}=eE(t)+F$, где F — сила, удерживающая электрон в положении равновесия. В первом приближении, соответствующем линейной оптике, предполагается квазиупругий характер этой силы, т. е. ее пропорциональность смещению из равновесия $F=-kx$, что отвечает квадратичной зависимости потенциальной энергии электрона от его смещения $U(x)={}^1/_2kx^2$. В следующих приближениях нужно учесть члены более высокой степени при разложении $U(x)$ в ряд по степеням смещения из равновесия:

$$U(x)={}^1/_2kx^2-{}^1/_3m\beta x^3-{}^1/_4m\delta x^4+\ldots. \tag{10.1}$$

Коэффициенты β и δ, как и k, определяются строением молекулы. Нечетные степени в разложении $U(x)$ могут быть только у систем без центра симметрии. В такой усовершенствованной модели уравнение движения электрона в пренебрежении затуханием принимает вид

$$\ddot{x}+\omega_0^2x=(e/m)E(t)+\beta x^2+\delta x^3+\ldots, \tag{10.2}$$

где $\omega_0=\sqrt{k/m}$ — частота собственных гармонических колебаний при небольших амплитудах. Когда ангармонические члены βx^2, δx^3, ... имеют характер небольших поправок, уравнение (10.2) можно решать методом последовательных приближений. В нулевом приближении ангармонические члены отбрасываются и (10.2) сводится к основному уравнению (2.30) классической теории дисперсии линейной оптики. В монохроматическом поле $E(t)=E_0\cos\omega t$ его решение имеет вид $x_0(t)=(e/m)E_0/(\omega_0^2-\omega^2)\cos\omega t$, что дает обычное выражение для линейной поляризуемости $\alpha_0=(e^2/m\varepsilon_0)/(\omega_0^2-\omega^2)$. Решение нулевого приближения $x_0(t)$ подставляется в ангармонические члены уравнения (10.2) вместо $x(t)$, и в следующем приближении нужно решать линейное уравнение, в котором играющая роль вынуждающей силы правая часть содержит кроме $\cos\omega t$ гармонические слагаемые на удвоенной и утроенной частотах:

$$\cos^2\omega t={}^1/_2(1+\cos 2\omega t),\quad \cos^3\omega t={}^3/_4\cos\omega t+{}^1/_4\cos 3\omega t.$$

Поэтому и частное решение неоднородного уравнения, описывающее установившиеся вынужденные колебания, кроме слагаемого $x_0(t)$ на основной частоте ω содержит слагаемые с частотами 2ω, 3ω, ...:

$$x(t)=\frac{e/m}{\omega_0^2-\omega^2}E_0\cos\omega t+\frac{\beta}{2}\left[\frac{(e/m)E_0}{\omega_0^2-\omega^2}\right]^2\left[\frac{1}{\omega_0^2}+\frac{\cos 2\omega t}{\omega_0^2-(2\omega)^2}\right]+$$

$$+\frac{\delta}{4}\left[\frac{(e/m)E_0}{\omega_0^2-\omega^2}\right]^3\left[\frac{3\cos\omega t}{\omega_0^2-\omega^2}+\frac{\cos 3\omega t}{\omega_0^2-(3\omega)^2}\right]+\ldots. \tag{10.3}$$

Совершающий колебания оптический электрон становится источником вторичных волн. Вынужденные колебания электрона под действием падающей волны на удвоенной и утроенной частотах приводят к генерации в среде световых волн с частотами, кратными частоте падающей на среду волны. Это значит, что в сильных оптических полях нарушается один из основных принципов линейной оптики, заключающийся в неизменности частоты света при переходе из одной среды в другую. Считалось, что частота задается процессами в источнике света, а именно частотой колебаний оптического электрона в излучающем атоме, и не меняется при изменении условий распространения. Но при выполнении определенных условий (см. § 10.3) даже небольшой ангармонизм оптических электронов прозрачной среды позволяет получить удвоение частоты проходящего света.

Кроме того, кубичная ангармоничность [слагаемое δx^3 в уравнении (10.2)] вызывает изменение поляризуемости молекулы на основной частоте ω. Объединяя члены с $\cos \omega t$ в (10.3), находим, что эта поляризуемость равна

$$\alpha = \alpha_0 + \alpha_2 E_0^2, \quad \alpha_0 = \frac{e^2/m\varepsilon_0}{\omega_0^2 - \omega^2}, \quad \alpha_2 = \frac{3}{4}\,\delta\alpha_0\,\frac{(e/m)^2}{(\omega_0^2 - \omega^2)^3}. \tag{10.4}$$

Таким образом, в поляризуемости молекулы появляется член, пропорциональный интенсивности световой волны. Он приводит к зависимости показателя преломления от интенсивности волны, проявляющейся в эффектах самовоздействия.

Если ангармонический осциллятор подвержен одновременному действию двух монохроматических полей с частотами ω_1 и ω_2, то в спектре его вынужденных колебаний помимо основных и кратных частот присутствуют комбинационные (суммарные и разностные) частоты. Этим объясняется эффект взаимодействия волн в нелинейной среде, ведущий к генерации волн на суммарной и разностной частотах.

Вынужденное движение электронов среды в поле световой волны макроскопически проявляется в возникновении поляризованности среды, которая складывается из индуцированных электрическим полем волны дипольных моментов отдельных молекул. Из-за ангармонических членов в потенциальной энергии электрона (10.1) эти дипольные моменты помимо линейного члена, пропорционального напряженности электрического поля световой волны, содержат члены, пропорциональные более высоким степеням напряженности поля. Поэтому в сильных световых полях материальное уравнение, связывающее поляризованность среды с напряженностью электрического поля, становится нелинейным. Для однородной анизотропной среды такое материальное уравнение в общем случае можно

записать в виде

$$P_i = \varepsilon_0 \Big[\sum_k \varkappa_{ik} E_k + \sum_{k,\,l} \varkappa_{ikl} E_k E_l + \sum_{k,\,l,\,m} \varkappa_{iklm} E_k E_l E_m + \ldots \Big], \qquad (10.5)$$

где i, k, l, m пробегают значения декартовых индексов x, y, z. Тензор второго ранга $\varkappa_{ik}$ представляет собой обычную или *линейную восприимчивость* среды (см. § 2.2), а тензоры высших рангов $\varkappa_{ikl}$, $\varkappa_{iklm}$ называются соответственно *квадратичной и кубичной восприимчивостями.* Их значения могут быть выражены через соответствующие физические характеристики среды (нелинейные поляризуемости молекул и концентрацию молекул). Но чтобы не ограничивать применимость материального уравнения (10.5) рамками выбранной модели среды, линейную и нелинейные восприимчивости можно рассматривать как феноменологические параметры среды. Поле E в (10.5) предполагается монохроматическим, а восприимчивости $\varkappa_{ik}$, $\varkappa_{ikl}$, ... зависят от его частоты ω.

Выше на примере простой классической модели было показано, что нелинейные восприимчивости появляются за счет ангармонических членов в потенциальной энергии оптического электрона. Электронный механизм возникновения нелинейности преобладает в твердых телах. Но зависимость оптических характеристик среды от интенсивности световой волны может быть обусловлена не только влиянием поля волны на поляризуемость молекулы (ее внутренние степени свободы), но и воздействием на концентрацию и ориентацию молекул, т. е. на внешние степени свободы. Эти факторы играют главную роль в жидкостях.

Одна из причин локальных изменений концентрации молекул в световой волне связана с электрострикцией. Электрическое поле в диэлектрике создает всестороннее давление, пропорциональное квадрату напряженности поля. В результате сжатия возникают локальные изменения плотности среды (т. е. концентрации молекул), а тем самым и ее оптических характеристик.

Если молекулы среды анизотропны, но при отсутствии внешнего поля все ориентации молекул равновероятны (что характерно для газов и жидкостей), то среда в целом изотропна. При большой интенсивности излучения электрическое поле волны оказывает ориентирующее действие на анизотропные молекулы (высокочастотный эффект Керра, см. § 4.5). В результате среда оказывается двоякопреломляющей, причем добавки к показателям преломления n_o и n_e пропорциональны интенсивности света.

Феноменологическое материальное уравнение (10.5) учитывает все возможные механизмы возникновения нелинейных восприимчивостей. Однако следует иметь в виду, что оно справедливо только в установившемся режиме, т. е. не описывает переходных процессов. В то же время разные причины зависимости оптических характеристик среды от интенсивности обладают разной степенью инерционности. При электрострикции время установления стационарного распределения плотности по порядку величины равно отношению размера a поперечного сечения светового пучка к скорости $v_{зв}$

упругих волн (звука). При $a \sim 1$ мм и $v_{зв} \sim 1$ км/с это время порядка 10^{-6} с. Инерционность ориентационного механизма возникновения нелинейной восприимчивости определяется временем поворота молекулы, равным по порядку величины 10^{-12} с. Поэтому в случае коротких лазерных импульсов (короче 10^{-6} с) основной вклад дает эффект Керра, тогда как в импульсах большой длительности вклад электрострикции в некоторых веществах становится преобладающим.

В анизотропной среде направление вектора **P** в общем случае не совпадает с направлением напряженности **E** электрического поля. Поэтому материальное уравнение (10.5) имеет тензорный характер. Если среда обладает центром симметрии, то в (10.5) все тензоры $\varkappa$ нечетных рангов обращаются в нуль. Так будет, например, в изотропной среде или в кубическом кристалле. Поэтому в них невозможны нелинейные эффекты, обусловленные квадратичной восприимчивостью $\varkappa_{ikl}$, например генерация второй гармоники. Тем не менее при качественном изучении таких явлений можно воспользоваться упрощенной изотропной моделью нелинейной среды, считая поляризованность **P** параллельной напряженности **E** и полагая в материальном уравнении (10.5) восприимчивости всех рангов скалярами:

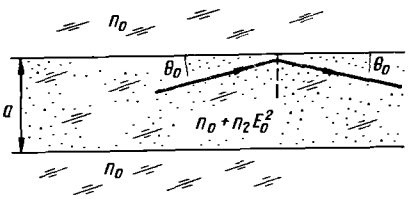
$$\mathbf{P} = \varepsilon_0 (\varkappa \mathbf{E} + \varkappa_2 E \mathbf{E} + \varkappa_3 E^2 \mathbf{E} + \ldots). \qquad (10.6)$$

Используя эту упрощенную модель, следует помнить, что в кристалле в каждом данном направлении могут распространяться волны только с определенными поляризациями. В материальном уравнении (10.6) восприимчивости $\varkappa$, $\varkappa_2$, $\varkappa_3$ для каждой из волн имеют свои значения.

Остановимся кратко на эффектах самовоздействия, возникающих при распространении в нелинейной среде интенсивного светового пучка, имеющего ограниченное поперечное сечение. Будем рассматривать только ту часть электромагнитного поля в среде, которая изменяется с частотой ω падающей световой волны. В феноменологическом материальном уравнении (10.6) вклад в поляризованность среды на частоте ω дают первый и третий члены, содержащие линейную и кубичную восприимчивости. Наличие члена с $\varkappa_3$ в поляризованности на частоте ω эквивалентно появлению в диэлектрической проницаемости $\varepsilon(\omega)$ и, следовательно, в показателе преломления $n(\omega)$ слагаемого, пропорционального квадрату амплитуды E_0, т. е. интенсивности световой волны:

$$n(\omega) = n_0 + n_2 E_0^2, \qquad (10.7)$$

где n_0 — показатель преломления, фигурирующий в линейной оптике, т. е. при малых интенсивностях света, а $n_2(\omega)$ — зависящий от свойств среды коэффициент, выражающийся через кубичную восприимчивость $\varkappa_3$. Из (10.7) следует, что ограниченный интенсивный пучок света делает среду оптически неоднородной. В пределах пучка, где $E_0 \neq 0$, показатель преломления будет иным, чем вне его,



10.1
К оценке порога самофокусировки

где $E_0=0$. Как известно (см. § 7.1), луч в неоднородной среде отклоняется от прямолинейного направления в сторону бо́льших значений показателя преломления. В результате возникает явление *самофокусировки* (при $n_2>0$) и дефокусировки (при $n_2<0$), предсказанное теоретически Г. А. Аскарьяном в 1962 г. и впервые наблюдавшееся Н. Ф. Пилипецким и А. Р. Рустамовым в 1965 г.

Сущность явления самофокусировки проще всего понять, рассматривая входящий в однородную среду пучок с одинаковой по всему сечению амплитудой (рис. 10.1). Тогда в среде под воздействием пучка образуется как бы стержень из вещества с более высоким (при $n_2>0$) коэффициентом преломления. Лучи, распространяющиеся внутри такого «стержня» под небольшим углом к его оси, испытывают полное отражение. Предельный угол луча с осью пучка θ_0, при котором происходит полное отражение, определяется соотношением $(n_0+n_2E_0^2)\cos\theta_0=n_0$. При малых значениях этого угла $\cos\theta_0\approx 1-\theta_0^2/2$, поэтому $\theta_0^2\approx 2n_2E_0^2/n_0$. Наклоненные к оси пучка лучи возникают в результате дифракции при ограничении диафрагмой его поперечных размеров, причем максимальный угол отклонения $\theta_{\text{диф}}$ по порядку величины равен $\lambda/a=\lambda_0/(n_0a)$, где a — поперечный размер пучка, λ_0 — длина волны в вакууме. При $\theta_{\text{диф}}>\theta_0$ пучок света по мере распространения расширяется из-за дифракции, но это происходит медленнее, чем в линейной среде. При $\theta_{\text{диф}}=\theta_0$ полное отражение полностью компенсирует дифракцию и площадь сечения пучка остается неизменной, т. е. пучок создает в среде своеобразный световод, в котором свет распространяется без дифракционной расходимости. Такой режим называется *самоканализацией* светового пучка. Приравнивая выражения для $\theta_{\text{диф}}$ и θ_0, находим пороговое значение амплитуды $E_{0\,\min}^2=\lambda_0^2/(2n_0n_2a^2)$. Отсюда по известному значению n_2 для данной нелинейной среды можно оценить минимальную мощность светового пучка, необходимую для наблюдения этого явления. В случае сероуглерода и рубинового лазера ($\lambda_0=694{,}3$ нм) $P_{\min}\approx 20$ кВт. Для некоторых сортов стекла $P_{\min}\sim 1$ Вт, что позволяет наблюдать явление даже в малоинтенсивных пучках лазеров непрерывного действия.

В реальных световых пучках интенсивность и, следовательно, показатель преломления возрастают от краев к оси пучка. Поэтому при превышении пороговой мощности лучи искривляются в сторону оси и концентрируются в области большей интенсивности, т. е. нелинейная среда ведет себя подобно объемной собирающей линзе, фокус которой находится на некотором расстоянии от входа пучка в среду, пропорциональном начальному радиусу пучка и обратно пропорциональном квадратному корню из его интенсивности. По ме-

ре приближения к фокусу лучи все более искривляются, т. е. самофокусировка имеет лавинообразный характер. В мощном световом пучке вслед за первым фокусом может появиться ряд последующих, причем с ростом мощности их число увеличивается и они приближаются к месту входа пучка в нелинейную среду. В случае коротких световых импульсов фокусы могут двигаться с околосветовыми скоростями.

В уравнения Максвелла (2.3) — (2.4), описывающие распространение света в веществе, поляризованность $\mathbf{P}$ среды входит в качестве источников в правую часть. Когда в материальном уравнении (10.6), связывающем $\mathbf{P}$ с $\mathbf{E}$, квадратичные и кубичные по степеням $\mathbf{E}$ члены существенны, подстановка $\mathbf{P}$ в уравнения Максвелла приводит к системе нелинейных уравнений для векторов $\mathbf{E}$ и $\mathbf{B}$ световой волны. Нелинейность уравнений означает нарушение принципа суперпозиции, согласно которому распространение световой волны в среде никак не сказывается на распространении других световых волн. Таким образом, справедливость принципа суперпозиции для света в веществе ограничивается приближением линейной оптики.

Для описания нелинейных эффектов удобно разделить поляризованность среды на линейную и нелинейную части: $\mathbf{P}=\mathbf{P}^{л}+\mathbf{P}^{нл}$. Влияние обусловленных линейной частью $\mathbf{P}$ зарядов и токов (т. е. источников полей в уравнениях Максвелла) удобно учесть, вводя вектор электрической индукции

$$\mathbf{D}=\varepsilon_0\mathbf{E}+\mathbf{P}^{л},$$

который связан с напряженностью $\mathbf{E}$ электрического поля линейным материальным уравнением (2.13):

$$\mathbf{D}=\varepsilon_0\varepsilon(\omega)\mathbf{E},\ \varepsilon(\omega)=1+\varkappa(\omega). \tag{10.8}$$

Входящая в (10.8) диэлектрическая проницаемость выражается только через линейную восприимчивость, т. е. имеет тот же смысл, что и в линейной оптике. Если в линейной оптике таким путем для векторов $\mathbf{E}$, $\mathbf{D}$ и $\mathbf{B}$ в среде получалась система однородных уравнений (2.6)—(2.9), аналогичная уравнениям поля в вакууме, то теперь в правой части остаются члены, связанные с нелинейной частью поляризованности:

$$\nabla\mathbf{D}=-\nabla\mathbf{P}^{нл}, \tag{10.9}$$

$$\varepsilon_0c^2\nabla\times\mathbf{B}-\frac{\partial\mathbf{D}}{\partial t}=\frac{\partial\mathbf{P}^{нл}}{\partial t}, \tag{10.10}$$

$$\nabla\mathbf{B}=0, \tag{10.11}$$

$$\nabla\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}=0, \tag{10.12}$$

Эту систему, как и уравнение (10.2) для ангармонического осциллятора, можно решать методом последовательных приближений. В нулевом приближении правые части отбрасываются, и (10.9)—

(10.12) превращаются в обычные однородные уравнения линейной оптики. Возможное их решение — плоская монохроматическая волна $\mathbf{E}(\mathbf{r}, t)=\mathbf{E}_0\cos(\mathbf{kr}-\omega t)$, волновой вектор которой удовлетворяет соотношению (2.23): $k^2=\varepsilon(\omega)\omega^2/c^2$. Для нахождения следующего приближения в правые части уравнений (10.9) и (10.10) подставим нелинейную поляризованность $\mathbf{P}^{нл}=\varepsilon_0(\varkappa_2 E\mathbf{E}+\varkappa_3 E^2\mathbf{E}+\ldots)$, в которой $\mathbf{E}(\mathbf{r}, t)$ заменено выражением из нулевого приближения. В результате получаются линейные, но неоднородные уравнения с известными правыми частями. Эти правые части можно рассматривать как дополнительные источники волн, обусловленные нелинейной поляризованностью среды. Каждый элемент объема среды dV излучает как осциллирующий диполь с дипольным моментом $\mathbf{P}^{нл}dV$. Если в нулевом приближении есть только одна монохроматическая волна частотой ω, то эти осцилляции происходят на частотах ω, 2ω, 3ω, ... Соответствующие решения уравнений (10.9)—(10.12) нужно искать в виде распространяющихся в среде волн с такими же частотами. При этом для волны каждой частоты в левых частях уравнений (10.9)—(10.10) при переходе от $\mathbf{D}$ к $\mathbf{E}$ следует брать значение диэлектрической проницаемости $\varepsilon(\omega)$ при той же частоте, т. е. $\varepsilon(\omega)$, $\varepsilon(2\omega)$, $\varepsilon(3\omega)$, ...

Контрольные вопросы

- ☐ Как нужно видоизменить модель, используемую в классической электронной теории дисперсии, чтобы объяснить нелинейную поляризуемость молекул?
- ☐ Объясните с помощью модели ангармонического осциллятора возникновение вторичных волн с кратными частотами.
- ☐ Какими причинами может быть обусловлена зависимость показателя преломления от интенсивности света?
- ☐ Какой вид имеет феноменологическое материальное уравнение для нелинейной среды?
- ☐ В чем заключается метод последовательных приближений для решения уравнений Максвелла в случае нелинейной среды?

10.3. Генерация второй гармоники

Рассмотрим нелинейные эффекты, обусловленные квадратичной восприимчивостью $\varkappa_2$ в (10.6). В поле монохроматической волны частотой ω нелинейная часть поляризованности имеет вид

$$\mathbf{P}^{нл}=\varepsilon_0\varkappa_2 E_0\mathbf{E}_0\cos^2(\mathbf{kr}-\omega t)={}^1/_2\varepsilon_0\varkappa_2 E_0\mathbf{E}_0+{}^1/_2\varepsilon_0\varkappa_2 E_0\mathbf{E}_0\cos 2(\mathbf{kr}-\omega t). \qquad (10.13)$$

Первый член здесь не зависит от времени. Это значит, что в нелинейной среде при прохождении через нее интенсивной световой волны возникает статическая поляризованность, создающая постоянное однородное электрическое поле, как в плоском конденсаторе. Такой эффект получил название *оптического детектирования*, так

как он вполне аналогичен выпрямлению переменного электрического тока и находит применение в детекторах мощных световых пучков. В детекторе «проходного» типа измеряется напряжение на обкладках конденсатора, нанесенных на боковые грани нелинейного кристалла, возникающее при прохождении сквозь него лазерного импульса. Преимущество такого детектора заключается в том, что кристалл почти не поглощает измеряемый пучок света.

Второй член в правой части (10.13) изменяется со временем с удвоенной частотой основной волны: $\omega_2=2\omega$. С ним связана генерация второй гармоники в нелинейной среде. Представим его как вещественную часть комплексного выражения $^1/_2\varepsilon_0\varkappa_2 E_0\mathbf{E}_0\exp[2i(\mathbf{kr}-\omega t)]$ и подставим в правые части уравнений Максвелла (10.9)—(10.10). Частное решение этих уравнений, соответствующее вынужденным колебаниям на частоте 2ω, можно искать в виде $\mathbf{E}(t)=\mathbf{E}_1\exp[2i\times\times(\mathbf{kr}-\omega t)]$, $\mathbf{B}(t)=\mathbf{B}_1\exp[2i(\mathbf{kr}-\omega t)]$. Используя также материальное уравнение (10.8), из уравнений Максвелла получаем:

$$\mathbf{kE}_1=0, \tag{10.14}$$

$$c^2\mathbf{k}\times\mathbf{B}_1+\varepsilon(2\omega)\omega\mathbf{E}_1=-^1/_2\varkappa_2\omega E_0\mathbf{E}_0, \tag{10.15}$$

$$\mathbf{kB}_1=0, \tag{10.16}$$

$$\mathbf{k}\times\mathbf{E}_1-\omega\mathbf{B}_1=0. \tag{10.17}$$

Из (10.14) и (10.16) видно, что векторы $\mathbf{E}_1$ и $\mathbf{B}_1$ ортогональны вектору $\mathbf{k}$, а из (10.17) — что они взаимно перпендикулярны, причем векторы $\mathbf{E}_1$, $\mathbf{B}_1$ и $\mathbf{k}$ образуют правую тройку, как орты $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. Выразим с помощью (10.17) $\mathbf{B}_1$ через $\mathbf{E}_1$ и подставим в уравнение (10.15):

$$c^2\mathbf{k}\times(\mathbf{k}\times\mathbf{E}_1)+\varepsilon(2\omega)\omega^2\mathbf{E}_1=-^1/_2\varkappa_2\omega^2 E_0\mathbf{E}_0. \tag{10.18}$$

Преобразуем двойное векторное произведение $\mathbf{k}\times(\mathbf{k}\times\mathbf{E}_1)=\mathbf{k}(\mathbf{kE}_1)-k^2\mathbf{E}_1$ и учтем, что $\mathbf{kE}_1=0$, а квадрат волнового вектора волны нулевого приближения удовлетворяет соотношению $k^2=\varepsilon(\omega)\omega^2/c^2$. Тогда из (10.18) получим амплитуду напряженности $\mathbf{E}_1$ электрического поля вынужденной волны на удвоенной частоте:

$$\mathbf{E}_1=\frac{^1/_2\varkappa_2 E_0\mathbf{E}_0}{\varepsilon(\omega)-\varepsilon(2\omega)}. \tag{10.19}$$

Выберем ось z вдоль направления входящей в нелинейную среду интенсивной волны частотой ω. На границе среды (при $z=0$) амплитуда второй гармоники равна нулю. Чтобы удовлетворить этому условию, к найденному выше частному решению нужно прибавить общее решение однородной системы уравнений, соответствующей (10.9)—(10.12). Такое решение представляет собой свободно распространяющуюся в среде волну $\mathbf{E}(\mathbf{r},t)=\mathbf{E}_2\cos(\mathbf{k}_2\mathbf{r}-\omega_2 t)$ с частотой $\omega_2=2\omega$ и волновым вектором $\mathbf{k}_2$, удовлетворяющим соотношению $k_2^2=\varepsilon(\omega_2)\omega_2^2/c^2$. Ее направление распространения и амплитуду $\mathbf{E}_2$ следует выбрать так, чтобы на всей плоскости $z=0$

суммарная напряженность поля вынужденной и свободной волн частоты 2ω обращалась в нуль. В результате для второй гармоники получаем следующее выражение:

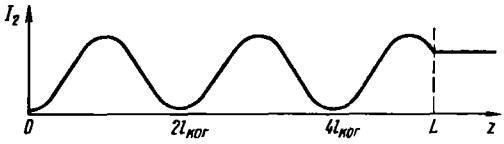
$$\mathbf{E}(z,\ t)=\frac{{}^1/_2\varkappa_2 E_0\mathbf{E}_0}{\varepsilon(\omega)-\varepsilon(2\omega)}[\cos(2kz-2\omega t)-\cos(k_2 z-2\omega t)]. \qquad (10.20)$$

Здесь первое слагаемое представляет собой вынужденную волну, фазовая скорость которой совпадает со скоростью распространения поляризованности, т. е. со скоростью v_1 исходной волны частотой ω: $v_1=c/n(\omega)$. Второе слагаемое описывает свободное распространение в среде вторичных волн частотой 2ω, которое происходит с иной фазовой скоростью $v_2=c/n(2\omega)$. Сложение волн одинаковой частоты 2ω, но распространяющихся с различными фазовыми скоростями, дает волну (10.20), амплитуда напряженности которой периодически зависит от z, т. е. от глубины проникновения в нелинейную среду:

$$\mathbf{E}(z,\ t)=\frac{\varkappa_2 E_0\mathbf{E}_0}{\varepsilon(\omega)-\varepsilon(2\omega)}\sin\left(\frac{k_2-2k}{2}z\right)\sin\left(\frac{k_2+2k}{2}z-2\omega t\right). \qquad (10.21)$$

Зависимость интенсивности второй гармоники от z показана на рис. 10.2. Интенсивность обращается в нуль при $(k_2/2-k)z=m\pi$, где m — целое число. В таких точках вторичные волны частотой 2ω, испущенные разными точками среды, гасят друг друга в результате интерференции. Так как $k_2/2-k=[n(2\omega)-n(\omega)]\omega/c$, то это условие можно записать в виде $z=2ml_{\text{ког}}$, где расстояние $l_{\text{ког}}=(\lambda/4)/[n(2\omega)-n(\omega)]$ называют когерентной длиной. На интервале от 0 до $l_{\text{ког}}$ фазовые соотношения таковы, что энергия от исходной волны передается второй гармонике, а на интервале от $l_{\text{ког}}$ до $2l_{\text{ког}}$ — возвращается в исходную волну. Этот процесс периодически повторяется по мере распространения исходной волны в нелинейной среде.

Впервые генерация второй гармоники была осуществлена Франкеном с сотрудниками в 1961 г. при прохождении импульса излучения рубинового лазера ($\lambda=694$ нм) через пластинку кварца. Излучение второй гармоники ($\lambda=347$ нм) соответствует ближней ультрафиолетовой области. Измерения показали, что ее интенсивность испытывает осцилляции по мере наклона пластинки к оси лазерного пучка. Эти осцилляции объясняются изменением длины пути пучка в пластинке: когда эта длина равна нечетному числу длин $l_{\text{ког}}$, интенсивность выходящей из пластинки второй гармоники максимальна (рис. 10.2). Полученное из периода осцилляций значение когерентной длины



10.2

Зависимость интенсивности второй гармоники от глубины проникновения в нелинейную среду

хорошо согласуется с известными значениями показателя преломления для $\lambda = 694$ нм и $\lambda = 347$ нм.

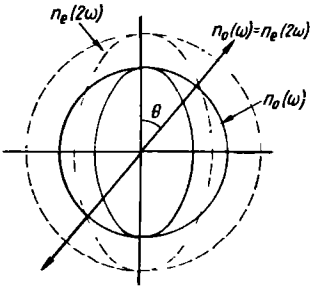
Однако коэффициент преобразования энергии основного пучка в энергию второй гармоники в описываемых опытах был чрезвычайно мал ($\sim 10^{-8}$). Для кварца в случае $\lambda = 694$ нм $n(2\omega) - n(\omega) = 0{,}025$ и $l_{\text{ког}} = 10\lambda = 0{,}7 \cdot 10^{-2}$ мм, т. е. эффективное расстояние, на котором энергия основного пучка передается второй гармонике, оказывается очень малым — всего лишь несколько длин волн. Эту трудность в 1962 г. преодолели Джордмейн, Мейкер и Терхьюн, предложив простой и весьма остроумный метод согласования показателей преломления для основного пучка и второй гармоники.

Когда показатели преломления на частотах ω и 2ω одинаковы, когерентная длина $l_{\text{ког}}$ обращается в бесконечность. В этом случае на протяжении всего пути в нелинейной среде наблюдается переход энергии от исходной волны ко второй гармонике. Так происходит потому, что при $n(2\omega) = n(\omega)$, т. е. $k_2 = 2k$, исходная волна, создающая нелинейную поляризованность среды на частоте 2ω, и испускаемые средой на этой частоте вторичные волны распространяются с одинаковой фазовой скоростью и фазовые соотношения между ними всюду одинаковы. Вся нелинейная среда действует как объемная сфазированная решетка элементарных диполей с максимумом излучения в направлении распространения. Условие $n(2\omega) = n(\omega)$ или $\varepsilon(2\omega) = \varepsilon(\omega)$ называется *фазовым* или *пространственным синхронизмом* исходной волны и ее второй гармоники. При его выполнении выражение (10.21) становится неопределенным. Чтобы раскрыть неопределенность при $\varepsilon(2\omega) = \varepsilon(\omega)$, преобразуем в нем знаменатель: $\varepsilon(2\omega) - \varepsilon(\omega) = n^2(2\omega) - n^2(\omega) = [n(2\omega) + n(\omega)][n(2\omega) - n(\omega)] \approx 2n\Delta n$, где $\Delta n = n(2\omega) - n(\omega)$. Учитывая, что $(k_2 - 2k)/2 = \Delta n\omega/c$, получаем из (10.21) следующее выражение для интенсивности второй гармоники:

$$I_2 \sim \varkappa_2^2 I_0^2 z^2 (\sin u/u)^2, \quad u = \Delta n \omega z/c. \tag{10.22}$$

При выполнении условия фазового синхронизма $u = 0$ для всех z и отношение $(\sin u)/u$ принимает максимальное значение, равное 1. Происходит пространственное накопление эффекта обмена энергией, и интенсивность второй гармоники пропорциональна квадрату толщины нелинейной среды. Конечно, ее энергия черпается из исходной волны, и формула (10.22) верна лишь для тех z, пока $I_2 \ll I_0$. В противном случае метод последовательных приближений, которым она получена, становится неприменимым, так как в нем интенсивность исходной волны полагается заданной и неизменной. Но вывод о том, что для эффективного преобразования энергии исходной волны в энергию второй гармоники необходимо выполнение условия фазового синхронизма, остается, конечно, в силе.

Для изотропной среды выполнение условия $n(2\omega) = n(\omega)$ невозможно, так как в области ее прозрачности показатель преломления монотонно изменяется с частотой (нормальная дисперсия). Однако



10.3
Сечение поверхностей показателей преломления n_o и n_e в кристалле KDP и направление фазового синхронизма

фазовый синхронизм можно осуществить между обыкновенной и необыкновенной волнами в некоторых кристаллах. На рис. 10.3 приведены сечения поверхностей, изображающих зависимость показателей преломления n_o и n_e от направления волновой нормали для отрицательного одноосного кристалла дигидрофосфата калия KH_2PO_4 (KDP). Сплошные линии соответствуют частоте ω, штриховые — удвоенной частоте 2ω. Видно, что поверхность $n_o(\omega)$ пересекается с поверхностью $n_e(2\omega)$. Поэтому для обыкновенной волны с частотой ω и необыкновенной волны с частотой 2ω, распространяющихся под определенным углом θ к оптической оси, соответствующим направлению из центра на точку пересечения $n_o(\omega)$ и $n_e(2\omega)$, выполняется условие пространственного синхронизма. Несмотря на различное направление поляризации этих волн, оказывается возможным их нелинейное взаимодействие, так как в анизотропной среде квадратичная восприимчивость представляет собой тензор $\varkappa_{ikl}$. Электрическое поле исходной обыкновенной волны, напряженность которого направлена перпендикулярно оптической оси, вызывает смещение зарядов на удвоенной частоте и в других направлениях, порождая волну второй гармоники, поляризованную в плоскости главного сечения. При правильном выборе направления исходной волны таким путем удается большую часть ее энергии перевести во вторую гармонику. Отклонение от направления синхронизма приводит к уменьшению интенсивности второй гармоники в соответствии с множителем $(\sin^2 u)/u^2$ в (10.22). Для инфракрасного излучения гелий-неонового лазера ($\lambda = 1{,}15$ мкм) в кристалле KDP направление синхронизма образует угол $\theta = 41{,}5°$ с оптической осью. Интенсивность второй гармоники спадает почти до нуля при отклонении направления исходной волны в кристалле от угла θ всего на 3′.

Генерация второй гармоники в нелинейных кристаллах используется для преобразования инфракрасного излучения мощных лазеров в видимое. Например, вторая гармоника лазера на иттрий-алюминиевом гранате, активированном неодимом, представляет собой зеленый свет, весьма удобный для разнообразных оптических исследований. При выполнении условия фазового синхронизма удается получить к.п.д. преобразования около 20—30%. Однако более эффективными оказались системы, в которых нелинейный кристалл помещается внутри лазерного резонатора. Как видно из (10.22), интенсивность второй гармоники пропорциональна квадрату интенсивности исходного излучения, которое внутри резонатора имеет значительно большую мощность, чем на выходе лазера. Зеркала

резонатора с многослойными диэлектрическими покрытиями обеспечивают очень высокий коэффициент отражения на основной частоте (99,8—99,9%) и пропускание свыше 90% для второй гармоники. На торцовые поверхности нелинейного кристалла наносится просветляющее покрытие, препятствующее отражению основного излучения. Кристалл преобразует энергию основного пучка во вторую гармонику, которая практически полностью покидает резонатор. При оптимальном согласовании оптических элементов можно обеспечить выходное излучение лазера только на частоте второй гармоники.

Генерация второй гармоники с точки зрения квантовой теории представляет собой процесс с участием трех фотонов: два фотона исходного излучения частоты ω превращаются в один фотон второй гармоники. Удвоению частоты $\omega_2=2\omega$ соответствует сохранение энергии в элементарном акте, так как энергия фотона второй гармоники $\hbar\omega_2$ равна суммарной энергии двух исходных фотонов $2\hbar\omega$. Условию пространственного синхронизма $k_2=2k$ соответствует сохранение импульса: $\hbar k_2=2\hbar k$.

Генерацию третьей гармоники в нелинейной среде можно получить за счет кубичной восприимчивости $\varkappa_3$ в (10.6). Исходное излучение частотой ω создает в нелинейной среде поляризованность, осциллирующую на утроенной частоте 3ω. Элементарные вторичные волны третьей гармоники, испускаемые разными элементами объема среды, будут иметь всюду одинаковое фазовое соотношение с возбуждающей их волной поляризованности при совпадении показателей преломления на частотах ω и 3ω. Дисперсия среды на интервале $(\omega, 3\omega)$ еще больше, чем в случае второй гармоники. Это ограничивает выбор кристаллов, в которых возможно выполнение условия пространственного синхронизма $n(3\omega)=n(\omega)$, так как двупреломление должно быть настолько большим, чтобы поверхности $n_o(\omega)$ и $n_e(3\omega)$ еще пересекались. Но главная трудность связана с малым значением кубичной восприимчивости, что вынуждает применять очень интенсивное исходное излучение. Интенсивность третьей гармоники пропорциональна кубу его интенсивности.

Контрольные вопросы

- В чем заключается эффект оптического детектирования?
- Как изменяется интенсивность второй гармоники в зависимости от расстояния, пройденного волной в нелинейной среде?
- Что называется когерентной длиной и каков ее физический смысл?
- Какое условие должно выполняться для эффективной передачи энергии исходного пучка во вторую гармонику?
- Как можно обеспечить пространственный синхронизм исходной волны и второй гармоники?
- В чем преимущество внутрирезонаторного метода удвоения частоты лазерного излучения?

10.4. Параметрическое преобразование частоты

В § 10.2 говорилось о том, что при распространении в среде двух монохроматических волн с частотами ω_1 и ω_2 в спектре вынужденных колебаний оптических электронов из-за квадратичного ангармонизма помимо основных и кратных частот появляются суммарная и разностная частоты. Такое нелинейное взаимодействие волн можно трактовать как следствие изменения оптических параметров среды под действием сильного поля одной из волн на ее основной частоте ω_1. В результате возникает модуляция с частотой ω_1 фазы второй волны, имеющей частоту ω_2, что эквивалентно появлению волн на боковых частотах $\omega_1+\omega_2$ и $\omega_1-\omega_2$. Эти процессы аналогичны рассмотренной в § 10.3 генерации второй гармоники. Они также обусловлены квадратичной восприимчивостью среды в (10.6), которая может быть отлична от нуля в кристаллах без центра симметрии. По существу, генерация второй гармоники представляет собой частный случай процесса генерации суммарной частоты, когда складываемые частоты ω_1 и ω_2 одинаковы и берутся от одного источника.

Практический интерес процессов генерации суммарных и разностных частот обусловлен тем, что, смешивая излучение двух лазеров в нелинейной среде, можно получить когерентное излучение в области спектра, отличной от исходной, т. е. расширить спектральный диапазон доступных источников когерентного излучения. Другое важное применение процесса сложения частот связано с возможностью создания чувствительных и малоинерционных детекторов инфракрасного излучения. Если в видимой области ($\lambda\approx 500$ нм) фотоумножители позволяют регистрировать потоки порядка ста фотонов в 1 с, то в области ~ 10 мкм для надежной регистрации существующими приемниками необходимы потоки порядка 10^8 фотонов в 1 с. Поэтому возможность преобразования инфракрасного излучения в видимое даже с относительно невысокой эффективностью представляется чрезвычайно привлекательной.

Преобразование «вверх» частоты ω_1 сигнальной волны от слабого источника возможно при ее взаимодействии в нелинейной среде с сильной «волной накачки» частотой ω_2. Нелинейная поляризованность среды $P_i^{\text{нл}}=\varepsilon_0\sum_{k,l}\varkappa_{ikl}E_{1k}\cos(\mathbf{k}_1\mathbf{r}-\omega_1 t)E_{2l}\cos(\mathbf{k}_2\mathbf{r}-\omega_2 t)$ содержит слагаемое на частоте $\omega_3=\omega_1+\omega_2$. Полагая для простоты направления напряженностей электрических полей обеих волн одинаковыми, это слагаемое запишем в виде $P^{\text{нл}}=\frac{1}{2}\varepsilon_0\varkappa_2 E_1E_2\cos[(\mathbf{k}_1+\mathbf{k}_2)\mathbf{r}-(\omega_1+\omega_2)t]$. В результате в среде возникает вторичная волна с суммарной частотой $\omega_3=\omega_1+\omega_2$. Интенсивность этой волны нарастает по мере распространения, если ее фазовая скорость и направление такие же, как у возбуждающей ее волны поляризованности $P^{\text{нл}}$. Другими словами, для эффективного перехода энергии сигнальной волны и волны накачки к волне с суммарной частотой ω_3 должно выполняться следующее векторное условие простран-

ственного синхронизма: $\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2$. В изотропной прозрачной среде в области нормальной дисперсии удовлетворить этому условию одновременно с условием $\omega_3 = \omega_1 + \omega_2$ невозможно (см. задачу). Пространственный синхронизм, как и в случае генерации второй гармоники, можно получить в некоторых кристаллах, если использовать обыкновенные и необыкновенные волны. Условия синхронизма для разных процессов (удвоения частоты, генерации суммарной и разностной частот) не совпадают, что позволяет реализовать на опыте какой-либо один нелинейный процесс при полном подавлении остальных.

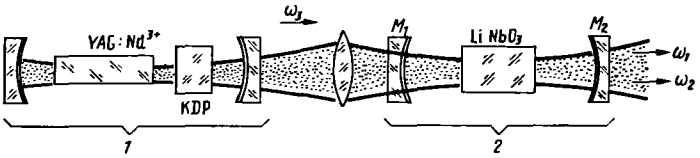
Процесс параметрического взаимодействия в нелинейной среде трех волн, часто́ты которых связаны соотношением $\omega_1 + \omega_2 = \omega_3$, можно использовать для усиления слабой волны частотой ω_1 или ω_2 за счет энергии мощной волны накачки с большей частотой ω_3 (рис. 10.4). Принцип *параметрического усиления света* был предложен С. А. Ахмановым и Р. В. Хохловым в 1962 г. и осуществлен Джордмейном и Миллером в 1965 г.

Взаимодействие сильной волны накачки с частотой ω_3 и слабой сигнальной волны с частотой ω_1 за счет квадратичной восприимчивости $\varkappa_2$ приводит к появлению в нелинейной поляризованности осцилляций на разностной частоте $\omega_2 = \omega_3 - \omega_1$. При выполнении векторного условия пространственного синхронизма $\mathbf{k}_2 = \mathbf{k}_3 - \mathbf{k}_1$ распространяющиеся в направлении вектора $\mathbf{k}_2$ вторичные волны частотой ω_2, испускаемые во всем объеме нелинейной среды, складываются синфазно. В результате энергия сильной волны накачки частоты ω_3 эффективно передается «холостой» волне с разностной частотой $\omega_2 = \omega_3 - \omega_1$ и сигнальной волне частотой ω_1, вызывая ее усиление.

На квантовом языке параметрическое усиление можно рассматривать как процесс вынужденного распада фотона волны накачки с энергией $\hbar\omega_3$ на два фотона с энергиями $\hbar\omega_1$ и $\hbar\omega_2$. Поэтому при прохождении некоторого расстояния в нелинейной среде увеличение плотности потока фотонов в сигнальной и холостой волнах одинаково и равно уменьшению плотности потока фотонов волны накачки. Соотношение $\omega_3 = \omega_1 + \omega_2$ выражает закон сохранения энергии в таком процессе.

Условие пространственного синхронизма $\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2$ можно рассматривать как закон сохранения импульса $\hbar\mathbf{k}_3 = \hbar\mathbf{k}_1 + \hbar\mathbf{k}_2$ для элементарного акта распада.

Наряду с рассмотренным выше *вынужденным* процессом распада фотона волны накачки на два фотона, вероятность которого пропорциональна интенсивности сигнальной волны, в нелинейной среде возможен и соответствующий *спонтанный* процесс распада, происходящий и при отсутствии сигнальной волны. Такое явление было обнаружено на опыте в 1967 г. и называется *параметрической люминесценцией* или параметрическим рассеянием. Процессы



10.4
Схема параметрического генератора света:
1 — лазер накачки с удвоителем частоты; *2* — нелинейный кристалл в оптическом резонаторе

параметрического усиления и люминесценции находятся в такой же связи, как вынужденное и спонтанное испускание фотона квантовой системой при переходе из возбужденного состояния в основное.

Введением положительной обратной связи параметрический усилитель можно превратить в генератор. Для этого нелинейную среду, как в лазерах, помещают в оптический резонатор, образованный двумя зеркалами (рис. 10.4). Нелинейный кристалл ориентируется так, что для волн, распространяющихся в одном направлении перпендикулярно зеркалам, выполняется условие пространственного синхронизма $\mathbf{k}_1^o + \mathbf{k}_2^o = \mathbf{k}_3^e$ либо $\mathbf{k}_1^o + \mathbf{k}_2^e = \mathbf{k}_3^e$. Зеркала M_1 и M_2 имеют высокие коэффициенты отражения для частот ω_1 и ω_2, так что сигнальная и холостая частоты ω_1 и ω_2 соответствуют высокодобротным модам резонатора. Зеркало M_1 одновременно должно быть прозрачно для частоты ω_3 излучения накачки. При достаточно большой мощности волны накачки параметрическое усиление одной из волн ω_1 или ω_2 на длине нелинейного кристалла превысит суммарные потери за проход, возникающие из-за неполного отражения от зеркал, поглощения, рассеяния и других причин. Тогда происходит самовозбуждение генератора (с затравкой из-за параметрической люминесценции) и возникает когерентное излучение на частотах ω_1 и ω_2.

В перпендикулярном зеркалам резонатора направлении, соответствующем его максимальной добротности, условие пространственного синхронизма $\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2$ при фиксированной частоте накачки ω_3 выполняется для вполне определенных частот ω_1 и ω_2, связанных соотношением $\omega_1 + \omega_2 = \omega_3$. На них и происходит параметрическая генерация. При изменении ориентации кристалла или его температуры или при наложении постоянного электрического поля часто́ты ω_1 и ω_2 изменяются. Эти обстоятельства можно использовать для плавной перестройки частоты параметрического генератора. Полоса перестройки частот генерации определяется диапазоном частот, для которого в данном нелинейном кристалле возможен пространственный синхронизм. Весьма жесткие требования предъявляются к источнику накачки. Его излучение должно иметь высокую монохроматичность и малую угловую расходимость. Недостаточная временна́я и пространственная когерентность излуче-

ния накачки нарушает условие пространственного синхронизма и поэтому увеличивает пороговую мощность и снижает к. п. д.

Обычно в качестве излучения накачки используют вторую гармонику неодимового лазера на иттрий-алюминиевом гранате (YAG:Nd). Созданные на основе высококачественных нелинейных кристаллов ниобата лития $LiNbO_3$, ниобата бария-натрия $Ba_2NaNb_5O_{15}$, дигидрофосфата калия KDP и аммония ADP параметрические генераторы света позволяют получать когерентное излучение с плавной перестройкой частоты во всем видимом и инфракрасном диапазоне спектра.

Контрольные вопросы

- □ В чем заключается процесс параметрического преобразования частоты света в нелинейной среде?
- □ Какие перспективы связаны с возможностью параметрического преобразования «вверх» частоты слабого инфракрасного излучения?
- □ Объясните принцип параметрического усиления света в нелинейной среде. Какую роль играет здесь условие пространственного синхронизма?
- □ В чем состоит параметрическое усиление света с точки зрения квантовых представлений?
- □ Объясните принцип действия параметрического генератора света.
- □ Каким способом можно осуществлять плавную перестройку частоты генерируемого излучения?

Задача

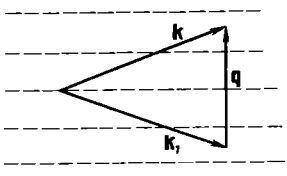
Показать, что в изотропной среде в области нормальной дисперсии невозможно выполнение условия пространственного синхронизма $\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2$ для волн, частоты которых связаны соотношением $\omega_3 = \omega_1 + \omega_2$.

◇ Так как $k = \omega n(\omega)/c$, то $k_3 = \omega_3 n(\omega_3)/c = (\omega_1 + \omega_2) n(\omega_3)/c = \omega_1 n(\omega_3)/c + \omega_2 n(\omega_3)/c$. Но $\omega_3 > \omega_1$ и $\omega_3 > \omega_2$, поэтому в области нормальной дисперсии $n(\omega_3) > n(\omega_1)$ и $n(\omega_3) > n(\omega_2)$. Таким образом, волновое число $k_3 > \omega_1 n(\omega_1)/c + \omega_2 n(\omega_2)/c = k_1 + k_2$. Поэтому условие $\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2$ не может выполняться даже тогда, когда все три волны распространяются в одном направлении. Это тем более невозможно при различных направлениях векторов $\mathbf{k}_1$, $\mathbf{k}_2$ и $\mathbf{k}_3$, так как в противном случае в образованном ими треугольнике одна сторона $\mathbf{k}_3$ была бы длиннее суммы длин двух других сторон.

10.5. Вынужденное рассеяние Мандельштама — Бриллюэна

Параметрические процессы преобразования частоты (см. § 10.4) происходят при взаимодействии трех электромагнитных волн в нелинейной среде. Взаимное влияние этих волн через нелинейную восприимчивость приводит к обмену энергией между ними и делает возможным усиление слабой волны за счет энергии мощной когерентной волны накачки. Здесь мы рассмотрим аналогичный нелинейный процесс взаимодействия трех волн в среде, из которых две электромагнитные, а одна упругая.

Пусть в среде распространяется упругая плоская монохроматическая волна с частотой Ω и волновым вектором $\mathbf{q}$. Ее фазовая скорость Ω/q — это скорость звука v, зависящая от упругих характеристик и средней плотности среды. Происходящие в упругой



10.5
Плоскости равных значений плотности в упругой волне с волновым вектором q

волне периодические изменения плотности $\delta\varrho = a\cos(\mathbf{qr} - \Omega t)$ приводят к пространственной модуляции показателя преломления среды $\delta n \sim \delta\varrho$, т. е. создают своеобразную объемную прозрачную (фазовую) синусоидальную дифракционную решетку. Эта решетка ничем не отличается от фазовой объемной голограммы (см. § 7.7), на которой зарегистрирована интерференционная картина от опорной волны $E_0\cos(\mathbf{kr} - \omega t)$ и плоской предметной волны $E_1\cos(\mathbf{k}_1\mathbf{r} - \omega_1 t)$ (рис. 10.5). Пусть для определенности угол между векторами $\mathbf{q}$ и $\mathbf{k}$ острый. Отражающие поверхности голограммы будут иметь такую же ориентацию и интервал, как и плоскости равных значений показателей преломления в упругой волне, если волновые векторы опорной и предметной волн удовлетворяют условию $\mathbf{k} - \mathbf{k}_1 = \mathbf{q}$ (см. задачу). Чтобы эти поверхности перемещались вдоль направления $\mathbf{q}$ со скоростью звука, частота предметной волны ω_1 должна быть меньше ω на частоту Ω упругой волны: $\omega - \omega_1 = \Omega$. Так как $\Omega \ll \omega$, то векторы $\mathbf{k}$ и $\mathbf{k}_1$ имеют почти одинаковую длину $k = \omega n/c$ и образуют равнобедренный треугольник (рис. 10.5).

При падении на такую голограмму только одной опорной волны $E_0\cos(\mathbf{kr} - \omega t)$ в результате дифракции произойдет восстановление предметной волны $E_1\cos(\mathbf{k}_1\mathbf{r} - \omega_1 t)$. Поэтому дифракция монохроматической световой волны $E_0\cos(\mathbf{kr} - \omega t)$ на упругой волне $\delta\varrho = a\cos(\mathbf{qr} - \Omega t)$ должна приводить к возникновению как раз такой дифрагировавшей волны $E_1\cos(\mathbf{k}_1\mathbf{r} - \omega_1 t)$ с волновым вектором $\mathbf{k}_1 = \mathbf{k} - \mathbf{q}$ и частотой $\omega_1 = \omega - \Omega$. Когда же угол между $\mathbf{q}$ и $\mathbf{k}$ тупой, дифрагировавшая волна имеет волновой вектор $\mathbf{k}_1 = \mathbf{k} + \mathbf{q}$ и частоту $\omega + \Omega$.

В конденсированных средах тепловое движение проявляется в колебаниях атомов и молекул. Из-за сильной связи атомов друг с другом колебание одного из них неизбежно передается соседним атомам. В целом тепловое движение в твердых телах можно рассматривать как совокупность всевозможных упругих волн, соответствующих различным нормальным колебаниям. Их спектр простирается от низких звуковых частот до гиперзвуковых ($\sim 10^9 - 10^{11}$ Гц). На основе такой модели теплового движения Дебай объяснил наблюдаемую на опыте температурную зависимость теплоемкости твердых тел с простыми кристаллическими решетками. Рассеяние света на упругих волнах, обусловленных тепловым движением, называется *рассеянием Мандельштама — Бриллюэна* (см. § 2.10).

С квантовой точки зрения рассеяние Мандельштама—Бриллюэна представляет собой рассеяние фотона исходного светового пучка с испусканием или поглощением кванта упругих колебаний среды — фонона, т. е. квазичастицы с энергией $\hbar\Omega$ и импульсом $\hbar\mathbf{q}$. В элементарном акте рассеяния выполняются законы сохранения энергии

$\hbar\omega \pm \hbar\Omega = \hbar\omega_1$ и импульса $\hbar\mathbf{k} \pm \hbar\mathbf{q} = \hbar\mathbf{k}_1$, соответствующие изменению частоты $\omega \pm \Omega = \omega_1$ и направления света $\mathbf{k} \pm \mathbf{q} = \mathbf{k}_1$ в классической волновой картине рассеяния. Вероятность испускания фонона и появления нового рассеянного фотона не зависит от наличия таких фотонов, т. е. от интенсивности рассеянного света, и в этом смысле рассматриваемое рассеяние следует считать *спонтанным* процессом.

Существует аналогичный процесс *вынужденного* рассеяния фотона с испусканием фонона, вероятность которого пропорциональна интенсивности рассеянного света. Вынужденное рассеяние Мандельштама—Бриллюэна (ВРМБ) было открыто на опыте в 1964 г. Чиао, Таунсом и Стойчевым. Они обнаружили, что мощное лазерное излучение частоты ω вызывает в кристалле появление когерентной упругой волны частоты Ω с одновременным испусканием света на частоте $\omega - \Omega$. Явление наблюдается лишь тогда, когда мощность лазерного излучения превышает некоторое пороговое значение.

С точки зрения классической электромагнитной теории ВРМБ можно рассматривать как процесс параметрического усиления упругой волны с частотой Ω и холостой электромагнитной волны с частотой $\omega - \Omega$ за счет энергии мощной электромагнитной волны накачки с частотой ω. Поясним это. При больших значениях напряженности электрического поля световой волны становится существенным не только влияние создаваемых упругой волной оптических неоднородностей на распространение света, но и влияние света на оптические параметры среды. Такое влияние обусловлено, в частности, явлением электрострикции: в электрическом поле в диэлектрике возникает дополнительное давление, пропорциональное квадрату напряженности электрического поля $p \sim E^2$. Пусть, например, в результате рассеяния интенсивной световой волны $E_0 \cos(\mathbf{kr} - \omega t)$ на упругой волне $a\cos(\mathbf{qr} - \Omega t)$ появилась световая волна $E_1 \cos[(\mathbf{k} - \mathbf{q})\mathbf{r} - (\omega - \Omega)t]$. Рассмотрим суммарную напряженность электрического поля падающей и рассеянной волн. Для нахождения создаваемого ими давления нужно сумму напряженностей возвести в квадрат, что после элементарных тригонометрических преобразований даст слагаемые, изменяющиеся на удвоенных световых частотах, и одно слагаемое с частотой Ω. Упругие волны с оптическими частотами существовать не могут, поэтому соответствующие слагаемые в p можно не учитывать и возбуждение упругих волн в среде под действием света будет описываться выражением

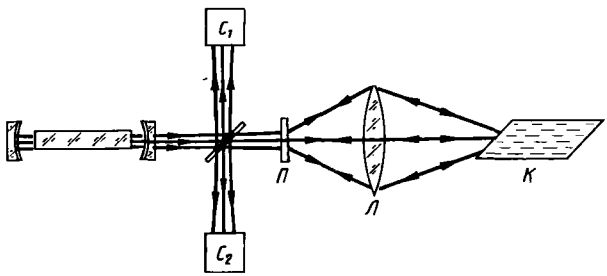
$$p \sim E_0 E_1 \cos(\mathbf{qr} - \Omega t). \qquad (10.23)$$

Эта волна создаваемого светом давления распространяется в том же направлении и с той же скоростью, что и первичная упругая волна, благодаря которой возникла рассеянная световая волна. Поэтому будет происходить параметрическое усиление первоначально слабой тепловой упругой волны, как и рассеянного на ней света. При достаточно большой интенсивности исходного

светового пучка, когда усиление превысит потери, амплитуды упругой волны и рассеянного света будут нарастать по мере распространения в нелинейной среде. Их энергия черпается из волны накачки, поэтому нарастание продолжается до тех пор, пока интенсивность рассеянного света не станет сравнима с ее интенсивностью. В отличие от некогерентного спонтанного рассеяния на тепловых флуктуациях плотности возникающее при ВРМБ излучение когерентно.

При ограниченных размерах нелинейной среды и поперечного сечения светового пучка накачки наиболее интересен случай рассеяния назад, когда усиливаемые упругая и световая волны распространяются навстречу и каждая из них обеспечивает положительную обратную связь для процесса параметрического усиления другой. Если когерентный падающий пучок пространственно неоднороден, т. е. его интенсивность не постоянна по поперечному сечению, то при ВРМБ происходит интереснейшее явление *обращения волнового фронта*, не имеющее аналога в классической оптике. Схема эксперимента по его наблюдению приведена на рис. 10.6. Волновой фронт интенсивного лазерного пучка, имеющего высокую направленность, существенно искажается поставленной на его пути фазовой пластинкой *П* со случайными неоднородностями. Расходимость пучка возрастает при этом в десятки раз. Затем линза *Л* с большой апертурой, достаточной для того, чтобы перехватить весь расширенный пучок, направляет свет в кювету *К*, заполненную сероуглеродом или метаном при высоком давлении. Небольшая часть лазерного пучка отражается плоскопараллельной пластинкой, и его угловое распределение в дальней зоне регистрируется измерительной системой C_1. Аналогичная система C_2 регистрирует рассеянный назад свет, также прошедший через линзу *Л* и фазовую матовую пластинку *П*.

Эксперименты показали, что первоначально широкий пучок рассеянного света после прохождения через те же оптические элементы в обратной последовательности становится столь же направ-



10.6
Схема эксперимента по наблюдению обращения волнового фронта

ленным, как и в первоначальный лазерный пучок. Так происходит потому, что волновой фронт излучения, рассеянного назад в нелинейной среде, в точности воспроизводит сколь угодно сложную структуру волнового фронта падающей волны, отличаясь лишь противоположным направлением распространения (небольшое уменьшение частоты можно не принимать во внимание, так как его относительная величина $\Delta\omega/\omega = \Omega/\omega = 2nv/c$ имеет порядок отношения скорости звука к скорости света). По выражению академика Р. В. Хохлова, «кювету с рассеивающим веществом можно рассматривать как волшебное зеркало, изменяющее знак времени».

Эффект обращения волнового фронта, как и голография, уточняет наши представления о необратимых и обратимых оптических явлениях. Информация о первоначальной структуре когерентного светового пучка не теряется при его прохождении через непоглощающую матовую пластинку. Механизм вынужденного рассеяния обращает искаженный волновой фронт с сохранением этой информации. При прохождении в обратном направлении через ту же пластинку все внесенные ею искажения волнового фронта полностью компенсируются и пучок восстанавливает свою структуру, т. е. возвращается в исходное состояние, но обращенное во времени. Конечно, абсолютная обратимость здесь все же не достигается как в отношении мощности, так и в отношении частоты, которая получает небольшой сдвиг из-за затраты энергии на возбуждение упругих волн в нелинейной среде.

Можно провести аналогию рассматриваемых явлений с голографическим восстановлением волнового фронта (см. § 7.7). Если при восстановлении объемной голограммы использовать волну противоположного направления по сравнению с опорной волной, то происходит восстановление действительного изображения предмета, которое расположено там, где находился предмет при записи. Это значит, что такое восстановление голограммы также приводит к обращению волнового фронта предметной волны. Восстановление формы волнового фронта при повторном прохождении через фазовую пластинку после его обращения в нелинейной среде можно сопоставить со свойствами голографических интерферограмм с двойной экспозицией. Наличие фазовых неоднородностей на пути от объекта к голограмме, искажающих волновой фронт предметной волны, не сказывается на виде интерференционных полос, так как все искажения, создаваемые статическими неоднородностями, компенсируются при второй экспозиции.

Поясним, почему для обращения волнового фронта при ВРМБ исходный когерентный пучок должен быть пространственно неоднородным. Выделим из общего падающего пучка с плохой направленностью две плоские волны, распространяющиеся под некоторым углом друг к другу. В результате их интерференции распределение интенсивности в каждом сечении имеет вид дифракционной решетки с

чередующимися максимумами и минимумами. Такое же пространственное распределение имеет и вероятность вынужденного рассеяния. Поэтому интерференция рассеянных волн должна дать точно такую же картину, что и возбуждающие исходные волны, так как при этом положение максимумов интенсивности рассеянного света совпадает с положением максимумов вероятности рассеяния. Чтобы волновой фронт рассеянной волны в деталях воспроизводил фронт падающей, в ограниченном поперечном сечении должно умещаться достаточное число полос этой картины, т. е. две выделенные волны исходного пучка должны иметь заметно различные направления. Если обе волны распространяются почти параллельно, то в среде конечных размеров никакой «решетки» распределения интенсивности не возникает, т. е. не фиксируется информация о том, какие направления энергетически выгодны для образования рассеянных волн. Эксперимент подтверждает этот вывод: если исходный пучок имеет предельно высокую направленность, то никакого обращения волнового фронта не происходит, так как пучок вынужденно рассеянного света имеет существенно большую расходимость, чем исходный.

С помощью обращения волнового фронта в нелинейной среде возможно решение актуальной проблемы получения мощного излучения с предельно малой расходимостью от лазеров с оптически неоднородными усиливающими средами. Случайные неоднородности искажают волновой фронт и ухудшают направленность лазерного излучения. Улучшение направленности излучения мощных лазерных систем может быть достигнуто по следующей схеме. Пучок света с дифракционной расходимостью от маломощного задающего лазера проходит через мощный усилитель с активной средой низкого оптического качества, значительно увеличивая при этом свою расходимость. После обращения волнового фронта в кювете с нелинейной средой излучение вновь проходит через тот же усилитель в противоположном направлении. При этом одновременно с дальнейшим увеличением интенсивности происходит компенсация искажений волнового фронта и выходящее излучение имеет близкую к предельной направленность. Таким методом можно решить задачу фокусировки мощного излучения в предельно малом объеме.

Контрольные вопросы

- ☐ Проведите аналогию между рассеянием когерентного света на упругой волне и восстановлением фазовой объемной голограммы.
- ☐ Какими элементарными процессами описывается рассеяние Мандельштама — Бриллюэна в квантовой теории?
- ☐ В чем заключается вынужденное рассеяние Маидельштама — Бриллюэна?
- ☐ Поясните физический механизм параметрического усиления упругой и рассеянной волн за счет энергии волны накачки при ВРМБ.
- ☐ Почему для обращения волнового фронта с помощью ВРМБ световой пучок должен быть пространственно неоднородным?

Задача

Показать, что в фазовой объемной голограмме, получаемой с опорной волной $E_0\cos(\mathbf{kr}-\omega t)$ и плоской предметной волной $E_1\cos(\mathbf{k}_1\mathbf{r}-\omega_1 t)$, частота которой ω_1 близка к ω, пространственное распределение показателя преломления имеет вид $\delta n(\mathbf{r},t)=b\cos(\mathbf{qr}-\Omega t)$, где $\mathbf{q}=\mathbf{k}-\mathbf{k}_1$ и $\Omega=\omega-\omega_1\ll\omega$.

◇ Представим амплитуду опорной волны в виде суммы $E_0=E_1+(E_0-E_1)$. Интерференционный узор на голограмме обусловлен только слагаемым с амплитудой E_1. Складывая его с предметной волной, находим

$$E_1[\cos(\mathbf{kr}-\omega t)+\cos(\mathbf{k}_1\mathbf{r}-\omega_1 t)]=$$
$$=2E_1\cos\left(\frac{\mathbf{k}-\mathbf{k}_1}{2}\mathbf{r}-\frac{\omega-\omega_1}{2}t\right)\cos\left(\frac{\mathbf{k}+\mathbf{k}_1}{2}\mathbf{r}-\frac{\omega+\omega_1}{2}t\right).$$

Пусть для определенности $\omega_1<\omega$, так что $\omega-\omega_1=\Omega>0$. По условию, частота ω_1 мало отличается от ω, поэтому первый сомножитель в полученном выражении можно рассматривать как медленно меняющуюся амплитуду $A(\mathbf{r},t)=2E_1\cos^1/_2(\mathbf{qr}-\Omega t)$. Вариации показателя преломления в фазовой голограмме пропорциональны интенсивности в интерференционной картине, т. е. квадрату амплитуды $A^2(\mathbf{r},t)=2E_1^2[1+\cos(\mathbf{qr}-\Omega t)]$. Отсюда $\delta n(\mathbf{r},t)=b\cos(\mathbf{qr}-\Omega t)$. Если же $\omega_1>\omega$, то $\Omega=\omega_1-\omega$ и $\mathbf{q}=\mathbf{k}_1-\mathbf{k}$. ◇

10.6 Вынужденное комбинационное рассеяние

При распространении монохроматического излучения в среде, молекулы которой могут совершать собственные колебания на частоте Ω, в спектре рассеянного света появляются линии, сдвинутые относительно частоты ω падающего излучения на $\pm\Omega$. Происхождение этих линий *комбинационного рассеяния* (см. § 2.10) объясняется тем, что при колебаниях атомов в молекуле, когда нормальная координата x, описывающая отклонение атомов от равновесной конфигурации, изменяется по закону $x(t)=a\cos\Omega t$, электронная поляризуемость молекулы α изменяется на частоте Ω:

$$\alpha(t)=\alpha_0+(\partial\alpha/\partial x)_0 x(t)=\alpha_0+\mu a\cos\Omega t, \qquad (10.24)$$

где α_0 — поляризуемость в равновесной конфигурации, а коэффициент $\mu=(\partial\alpha/\partial x)_0$ характеризует изменение поляризуемости при смещении атомов из равновесных положений. Индуцируемые падающим излучением осцилляции дипольного момента молекулы

$$p=\varepsilon_0\alpha E=\varepsilon_0[\alpha_0+(\partial\alpha/\partial x)_0 x]E=$$
$$=\varepsilon_0(\alpha_0+\mu a\cos\Omega t)E_0\cos\omega t \qquad (10.25)$$

оказываются промодулированными и в их спектре присутствуют составляющие с частотами $\omega\pm\Omega$ [см. (1.79)], которыми и обусловлено появление соответствующих спектральных компонент в рассеянном излучении. По квантовым представлениям, стоксова компонента с частотой $(\omega-\Omega)$ возникает, когда энергия фотона $\hbar\omega$ уменьшается на величину $\hbar\Omega$, равную энергии возбуждаемого при рассеянии кванта колебаний молекулы, антистоксова $(\omega+\Omega)$ — когда квант $\hbar\Omega$ первоначально возбужденной молекулы передается излучению. Интенсивность антистоксовой компоненты, пропорциональная числу

возбужденных молекул, много меньше интенсивности стоксовой, так как отношение числа возбужденных и невозбужденных молекул $\exp[-\hbar\Omega/(kT)] \ll 1$.

Явление *вынужденного комбинационного рассеяния* (ВКР), соответствующее описанному выше спонтанному процессу, было открыто на опыте Вудбери и Нг в 1962 г. ВКР также заключается в испускании спектральных компонент, сдвинутых относительно возбуждающего излучения на частоту внутримолекулярных колебаний, но вероятность этого процесса зависит от интенсивностей падающего и рассеянного излучений. ВКР возникает только при интенсивности падающего пучка, превышающей некоторую пороговую величину. В отличие от спонтанного рассеяния, интенсивность которого очень мала (10^{-7} — 10^{-8} часть возбуждающего потока), при ВКР доля рассеянного потока достигает десятков процентов. Помимо линий с частотами $(\omega \pm \Omega)$ появляются линии более высоких порядков $(\omega \pm 2\Omega)$, ..., и рассеяние имеет четко выраженный направленный характер.

Некоторые из особенностей ВКР можно понять с помощью простой классической модели явления. При объяснении спонтанного комбинационного рассеяния принималось во внимание только влияние колебаний молекулы на рассеиваемые ею световые волны. Но поле световой волны оказывает и обратное влияние на молекулу, раскачивая ее колебания. Покажем это. Энергия диполя с моментом p, индуцированным полем с напряженностью E, равна $-pE/2 = -\varepsilon_0 \alpha E^2/2$. Поэтому сила, действующая на диполь со стороны поля, равна $F_x = -\partial U/\partial x' = \varepsilon_0 \mu E^2/2$. Эту силу нужно включить в правую часть уравнения, описывающего колебания атомов в молекуле:

$$\ddot{x}' + 2\Gamma\dot{x}' + \Omega^2 x' = \varepsilon_0 \mu E^2/(2M). \qquad (10.26)$$

Здесь M — приведенная масса атомов, константа Γ характеризует затухание собственных колебаний. Ограничимся в (10.26), кроме напряженности поля возбуждающего излучения $E_0 \cos \omega t$, вкладом только первой стоксовой компоненты $E_c = \cos(\omega - \Omega)t$, и из квадрата их суммы учтем лишь член, изменяющийся с частотой Ω, т. е. в резонансе с собственными колебаниями молекулы:

$$\ddot{x}' + 2\Gamma\dot{x}' + \Omega^2 x' = {}^1/_2(\varepsilon_0 \mu E_0 E_c/M)\cos \Omega t. \qquad (10.27)$$

Решение этого уравнения, описывающее вынужденные колебания, в случае слабого затухания $(\Gamma \ll \Omega)$ имеет вид

$$x(t) = x_0' \sin \Omega t, \quad x_0' = \varepsilon_0 \mu E_0 E_c/(4M\Gamma\Omega). \qquad (10.28)$$

Таким образом, поля возбуждающего и стоксова излучений приводят к резонансной раскачке колебаний молекулы с амплитудой, пропорциональной $E_0 E_c$. Эти индуцированные колебания, в свою очередь, вызывают еще большую модуляцию колебаний дипольного момента, происходящих под действием возбуждающего излучения, и тем самым приводят к усилению стоксова излучения и к возникновению у дипольного момента *новых спектральных компонент*. Чтобы

увидеть это, подставим $x(t)$ из (10.28) в выражение $p = \varepsilon_0(\alpha_0 + \mu x)[E_0 \cos \omega t + E_c \cos(\omega - \Omega)t]$. После простых преобразований получим:

$$p = \varepsilon_0 \alpha_0 E(t) - {}^1/_2 \varepsilon_0 \mu x_0' E_0 \sin(\omega - \Omega)t + {}^1/_2 \varepsilon_0 \mu x_0' E_c \sin \omega t + {}^1/_2 \varepsilon_0 \mu x_0' E_0 \sin(\omega + \Omega)t - {}^1/_2 \varepsilon_0 \mu x_0' E_c \sin(\omega - 2\Omega)t. \quad (10.29)$$

Здесь первый член соответствует линейной поляризованности среды, которая определяет индукцию $D = \varepsilon_0 E + P^{\text{л}}$. Второй член с частотой $(\omega - \Omega)$ описывает усиление стоксова излучения. В самом деле, совершаемая над молекулой в 1 с работа поля стоксова излучения

$$P_c = -{}^1/_4 \varepsilon_0 \mu x_0' E_0 E_c(\omega - \Omega) = -(\varepsilon_0 \mu E_0 E_c)^2(\omega - \Omega)/(16 M \Gamma \Omega)$$

отрицательна, т. е. происходит усиление излучения, пропорциональное E_0^2 и E_c^2. С третьим членом в (10.29) связана работа поля возбуждающего излучения над молекулой $P = {}^1/_4 \varepsilon_0 \mu x_0' E_0 E_c \omega = (\varepsilon_0 \mu E_0 E_c)^2 \omega/(16 M \Gamma \Omega) > |P_c|$, которая идет на усиление колебаний молекулы и стоксовой компоненты рассеянного излучения.

Возбуждающее излучение и испытавшая большое усиление стоксова компонента рассеянного излучения создают в среде, как видно из четвертого и пятого членов (10.29), когерентный ансамбль диполей, излучающих на антистоксовой частоте $(\omega + \Omega)$ и стоксовой частоте второго порядка $(\omega - 2\Omega)$. Излучение на второй стоксовой частоте возникает еще и потому, что первая стоксова компонента $(\omega - \Omega)$, достигая большой интенсивности, сама начинает играть роль возбуждающего излучения и испытывает вынужденное рассеяние с уменьшением частоты еще на Ω. Процесс увеличения числа спектральных компонент рассеянного излучения ограничивается из-за конечного запаса мощности исходного возбуждающего пучка.

Направленный характер ВКР обусловлен интерференцией когерентных вторичных волн, испускаемых диполями в различных точках рассеивающей среды. Результат интерференции зависит от фазовых соотношений между этими волнами и от геометрических условий эксперимента. Для стоксова излучения условия фазового синхронизма, обеспечивающие пространственное накопление эффекта преобразования энергии возбуждающего излучения в рассеянное, выполняются в любом направлении. На опыте обычно возбуждается только одна стоксова мода с наименьшими потерями. Если рассеивающая среда находится вне резонатора, наиболее благоприятные условия для преобразования энергии узкого возбуждающего пучка в стоксово излучение соответствуют его распространению по ходу пучка, так как при параллельном расположении пучков объем среды, где происходит их взаимодействие, намного больше, чем для внеосевого направления.

Для антистоксова рассеяния в среде с нормальной дисперсией условия пространственного синхронизма выполняются в направле-

ниях, образующих небольшой угол (3° в нитробензоле для излучения рубинового лазера) с возбуждающим пучком, поэтому излучение с частотой $(\omega + \Omega)$ распространяется вдоль конической поверхности, ось которой совпадает с возбуждающим лазерным пучком.

На опыте ВКР сопровождается обычно другими нелинейными явлениями, в частности самофокусировкой (см. § 10.2) возбуждающего лазерного пучка, благодаря которой облегчается достижение порогового значения интенсивности.

Контрольные вопросы

- ☐ Почему интенсивность антистоксовой компоненты спонтанного комбинационного рассеяния много меньше, чем стоксовой?
- ☐ В чем заключаются основные отличия ВКР от спонтанного?
- ☐ Как на основе классической модели объясняется возникновение вынужденного когереитного стоксова и антистоксова рассеяния?
- ☐ Чем объясняется направленный характер рассеянного излучения при ВКР?

10.

Заключение

Вот и подошло к концу изучение очередного раздела курса физики. Пройден долгий путь, на котором пришлось преодолеть немало трудностей. Для многнх нз вас оптнка лишь составная часть общего физнческого образовання, нужная главным образом для успешного занятия другнми наукамн. Для других это основа будущей профессии, необходнмая не только для формирования соаременного научного мышления, но н для успешного решения конкретных научных и технпчеcких проблем, выдзигаемых практикой. Хочется надеяться, что в любом случае изучение этой кннгн доставило чнтателю не только муки преодоления трудиостей, но н радости открытий н постиженнй, и не отбило акус к дальнейшему продвижению по избранному пути.

Очень важно сознавать, что наука как род человеческой деятельности — это отнюдь не то, что включено в учебннки. Неизбежное для учебника «прнчесанное» изложение науки создает у читателя нллюзню завершенностн: кажется, что все уже сделано, осталось, быть может, уточнить лишь некоторые детали. Исторня науки постоянно опровергает такое заблуждение. Учебнкам полувековой давностн тоже присуща подобная завершенность, но сколько замечательных научных открытий принесли нам последние десятилетия! Пример оптикн показывает, что открытня совершаются не только в таких новых областях наукн, как физика высоких энергий и космологня, где основные принципы еще не установлены и необходим выход за пределы уже известных представлений. Физическнй фундамент для создания лазеров, открытия голографин и многих эффектов нелинейной оптикн был заложеи очень давно, и тем не менее потребовалнсь долгие годы, чтобы настало время для этих открытий. Опыт прошлого, в том числе и совсем недавнего, говорит о том, что оптика еще не раз вознаградит усилия исследователей новымн открытиями, преображающими жизнь человека н раскрывающнми красоту и гармонию окружающего мнра.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Аберрация геометрическая 335
— звездиая 393, 406, 407, 410
— сферическая 353
— хроматическая 335, 358
Абсорбция 101
Анализатор 20
Анастигмат 357
Аиизотропия 195
Апертура 348
— интерференции 238
— полной поляризации 192
— числовая 363, 370
Аподизация 299
Астигматизм 356

Билинза Бийе 209
Бипризма Френели 209, 239

Вектор волновой 16, 79
— Пойнтинга 31, 40
Видность полос интерференционных 69, 220, 225
Виньетирование 351
Волны гауссовы 18
— линейно поляризованные 19
— необыкиовенные 184
— иеоднородные 15, 79
— обыкновенные 184
— однородные 80
— поляризованные 22
— стоячие 25
— сферические 35
— электромагнитные 12, 15, 75, 180
Время жизни атома 42
— затухания радиационного 42, 438
— когерентности 222
— релаксации 101

Гипотеза статистическая 55
Гироскоп лазерный 416
Гиротропия 108
Глубина резкости 348
Голография 377
Граница фотоэффекта красная 465

Давление световое 167, 170
Детектирование 488
Детекторы 34
Диафрагма апертурная 348
— поли зрения 350
Дисперсия 80, 83, 94, 98
— аномальная 89, 91, 132
— временная 112
— нормальная 86, 131
— орнентационная 98, 101
— отрнцательная 93
— пространственная 112
— релаксационная 101
— спектральная прибора 262, 308, 315
— угловая 308, 315
Дисторсия 357
Дифракция Фраунгофера 277, 284, 291, 305
— Френеля 277, 279
Длина волны комптоновская 470
— когерентности 222
— пути оптическая 334

Закон Брюстера 148
— Бугера 81, 443

— Вина 423, 424
— Гаусса 12
— индукции электромагнитной 13
— Кирхгофа 422
— Ламберта 68
— равнораспределения энергии 428
— Рэлея 119
— смещения 424
— сохранения импульса 470
— — энергии 31, 471, 472
— Стефана — Больцмана 423, 424
— Хаббла 409
Закон отражения 142, 144
— преломления 142, 144
Затенение 351
Затухание радиационное 41, 84
Зеркала лазерные 303
— Ллойда 240
— Френеля 208, 239
Зона волновая 39
Зоны Френеля 271

Излучение Вавилова — Черенкова 136, 138
— вынужденное 434
— дипольное 38, 40
— квазимонохроматическое 52
— лазерное 337
— неполяризованное 60, 61
— переходное 139
— равновесное 419, 422
— спонтанное 437
— тормозное 467
Изображение оптическое 334
— стигматическое 343
Иммерсия 371
Импульс волны электромагнитной 168
Инверсия населенностей 93, 443
Индикатриса рассеяния 117
Интенсивность света 32, 67, 120
Интерференция 202, 218
— многолучевая 255
Интерферометр Жамена 248—250
— Майкельсона 213, 253, 396
— — звездный 245, 369
— Рождественского 249
— Рэлея 248
— Фабри — Перо 256, 260, 262, 264,

Источник протяженный 68, 234
— точечный 67

Катодолюминесценция 418
Когерентность 116
— временная 222, 228
— полная 229
— пространственная 234
— частичная 229
Коллектив 351
Кольца Ньютона 217
Кома 354
Контур лоренцевский 53, 58, 90
Коэффициент отражения 149
— поглощения 81, 443
— увлечения эфира 394, 406
Коэффициенты Эйнштейна 438
Критерий Рэлея 321, 367

Лазеры 453, 454
Линза 340
— ахроматическая 359
— полевая 351
Локализация потока энергии 33
Лучепреломление двойное 175, 194, 195, 197
Люминесценция 495

Метод контраста фазового 374
— крюков Рождественского 250
— призм скрещенных 91
Мираж 331
Модуляция 123, 377
Момент дипольный 73, 99, 106
— импульса 172
Мультиплеты 65

Накачка оптическая 452
Намагниченность 75
Населенность инверсная 93, 443
Насыщение 447, 479, 481

Опалесценция критическая 120
Опыт Винера 28
— Кундта 92
— Кундта — Вуда 92
— Майкельсона 396
— Поля 211, 240
Опыт Физо 395, 406
— Юнга 202, 235, 449
Освещенность 67, 69
Осциллятор 38, 83, 433
Отверстие относительное 362
Отражение полное 95, 153, 155, 156
— света 161

Параллакс звезд 393
Параллелепипед Френеля 159
Пластинка зонная 274
Плотность излучення спектральная 419, 425
— энергии поли магнитного 32
Поглощение света 91, 94, 165
Полярнзаторы 20, 109
Поляризация волн 18, 108
— ионная 99
— света квазимонохроматическая 54, 177
Поляризуемость атомная 85, 93
Порядок спектра 306
Постоянная Верде 105, 107
— затухания 52
— интерферометра 263
— решетки 307
Потери гауссовы 297, 345
— днфракционные 300
Потеря полуволны 146
Поток излучения 67
— световой 361
Призма Волластона 193
— Николя 191
— полярнзационная 191
— Френеля 110
— Фуко 192
Прнпцип Гюйгенса 268
— Гюйгенса — Френеля 268, 278, 282
— суперпозиции 45
— Ферма 333, 334

Радиоинтерферометры 245
Радиус расходнмости дифракционный 299
— электрона класснческий 122
Распределенне Максвелла 59
Рассеянне комбинационное 123, 503
— Мандельштама — Брнллюэна 497
— Ми 119
— молекулярное 119
— рэлеевское 123
— света 116
Расходимость пучков 287
Резкость полос 259
Резонатор концентричесний 301
— неустойчивый 301
— оптический 300, 449
— открытый 264
— софокусный 301
Рефрактометр Аббе 154
Рефракцня удельная 86
Решетка дифракционная 305, 310, 312

Самофокуснровка 486
Сателлнты 123
Свет естественный 123, 151
— квазнмонохроматнческий 55
— поляризованный 61, 151
— — частично 61, 151
— рассеянный 123
Светимость 69
Свойство поперечности волн 16
Сила оптнческаи 339, 342
— осцнллятора 93
— разрешающая 321, 367
— света 67
Скин-эффект 165
Скорость атома наивероятнейшая 59
— групповаи 128, 131
— лучевая 181
— распространения энергин 133
— света 127, 129
— снгнала 86, 135
— фазовая 16, 79, 128, 132
— фронта 135
Спектр волны 47
— излучення источника 57

Спираль Корню 280
Способность инструментов разрешающая 367
Среда идеально однородная 117
— мутная 117
Степень когерентности временной 243
— — пространственной 243
— поляризации 151
Сечение рассеяния 121

Тело черное 68
Ток поляризационный 74

Увеличение нормальное 363, 369
— прибора 363
— угловое 344
Увлечение эфира 394, 406
Угол Брюстера 147
— полной поляризации 148
— поля зрения 351
Умножители фотоэлектронные 460
Уравнение волновое 14, 15
— непрерывности 13, 31
— эйконала 329
— Эйнштейна 459
Уравнения Максвелла 13, 72, 78
— материальные 75, 481, 482
Услоане излучения 143
— синусов Аббе 354
— цикличности 300
Условия граничные 142

Фаза волны 16
Фильтры оптические 265
Флуктуации плотности 119
Флуоресценция резонансная 122
Формула дисперсионная 100
— Зельмейера 251
— Планка 430
— Рэлея 131, 132
— Рэлея — Джинса 428
— Френеля 146
— Эйри 258
Формулы Крамерса — Кронига 78
Фотодиод 466, 467
Фотолюминесценция 418
Фотосопротивление 465
Фотоэффект внутренний 465
— многофотонный 480
Фронт волновой 206
Функция автокорреляционная 228
— аппаратная 316, 368
— корреляционная 242
Фурье-спектрометр 253, 254

Хемилюминесценция 418

Частота ларморовская 64, 102
— комбинационная 124
— критическая 95
— модуляции 124
— плазменная 94
Число волновое 16
Чувствительность пороговая 463

Ширина линий спектральных доплеровская 58
— — — естественная 53
— — — неоднородная 58
— — — однородная 58
— — — столкновительная 57

Экспозиция 67
Электролюминесценция 418
Энергия волн электромагнитных 30
— осциллятора 42
Эффект Доплера 58,
— Зеемана 62, 64
— — аномальный 65
— — нормальный 65
— — обратный 102
— Керра 195, 198
— Комптона 171, 470

— Коттона — Мутона 199
— насыщения 479
— отдачи 171
— Поккельса
— Саньяка 413
— Тиндаля 1
— туннельный 156
— Фарадея 102, 103, 107
— фотоэлектрический 456
— Штарка 66

Яркость 68, 70, 361

Учебное издание

Евгений Иванович Бутиков

ОПТИКА

Зав. редакцией литературы по физике
и математике **Е. С. Гридасова**
Редактор **Г. Н. Чернышева**
Мл. редакторы **С. А. Доровских, Н. П. Майкова**
Оформление художника **Ю. Д. Федичкина**
Художественный редактор **В. И. Пономаренко**
Технический редактор **З. В. Нуждина**
Корректор **Г. И. Кострикова**

ИБ № 5421

Изд. № ФМ-796. Сдано в набор 21.03.86. Подп. в печать 03.11.86. Т—17589. Формат $60\times90^1/_{16}$. Бум. офс. № 1. Гарнитура «Таймс». Печать офсетная. Объем 32 усл. печ. л. + + 0,25 усл. печ. л. форзацы. 65,5 усл. кр.-отт. 34,93 уч.-изд. л. + 0,44 уч.-изд. л. форзацы. Тираж 23 000 экз. Зак. № 1205. Цена 1 р. 70 к.

Издательство «Высшая школа», 101430, Москва, ГСП-4, Неглинная ул., д. 29/14. Ярославский полиграфкомбинат Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 150014, Ярославль, ул. Свободы, 97.