Б. САЛЕХ, М. ТЕЙХ

ОПТИКА И ФОТОНИКА
ПРИНЦИПЫ И ПРИМЕНЕНИЯ

Том 1

Перевод с английского
В.Л. Дербова

Издательский Дом
ИНТЕЛЛЕКТ

ДОЛГОПРУДНЫЙ
2012

Б. Салех, М. Тейх

Оптика и фотоника. Принципы и применения. Пер. с англ.: Учебное пособие. В 2 т. Т. 1 / Б. Салех, М. Тейх – Долгопрудный: Издательский Дом «Интеллект», 2012. – 760 с.: цв. вкл.

ISBN 978-5-91559-038-9

Сравнительно новый термин «фотоника» возник по аналогии с хорошо известным термином «электроника». Это современное состояние науки о взаимодействии света и вещества и многочисленных технологических приложениях.

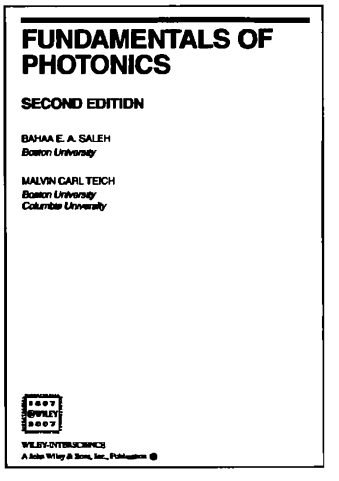
Этот термин отражает квантовую (фотонную) природу света и включает широкий круг физических явлений, методов и устройств, используемых для генерации света, управления его свойствами, передачи, регистрации, воздействия светом на вещество и оптической диагностики материальных сред. В учебной литературе на русском языке, рассчитанной на студентов физических и технических специальностей, в настоящее время отсутствует книга, объединяющая указанный круг проблем. Данный пробел призван восполнить перевод на русский язык второго издания книги известных американских специалистов.

Содержание книги охватывает оптику лучей, волн и пучков, фурье-оптику, электромагнитную теорию света, поляризационную оптику, оптику фотонных кристаллов, волноводов и резонаторов, элементы статистической и квантовой оптики, взаимодействие фотонов с атомами, лазерные усилители и лазеры, оптику полупроводников, полупроводниковые источники и приемники фотонов, акусто- и электрооптику, основы нелинейной оптики, включая оптику ультракоротких импульсов света, а также основные сведения об оптических системах связи и их элементах – оптических соединителях и переключателях.

Начиная с элементарных основ оптики, авторы достаточно быстро подводят читателя к самым современным научным достижениям и техническим решениям. Математический аппарат изложен лаконично, но достаточно строго, наглядность обеспечивается большим количеством иллюстраций.

Каждый раздел книги снабжен хорошо продуманным набором задач, что делает ее весьма полезной как для преподавателей, так и для самостоятельной работы студентов.

Огромный объём материала, охватывающего все разделы оптики, потребовал выпустить книгу на русском языке в виде двухтомника.



ISBN 978-5-91559-038-9
ISBN 978-0-4713-5832-9 (англ.)

© 2007, John Wiley & Sons
© 2012, ООО Издательский Дом «Интеллект», перевод на русский язык, оригинал-макет, оформление

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ К ВТОРОМУ ИЗДАНИЮ 10
ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ 18

Глава 1
ОПТИКА ЛУЧЕЙ 21

1.1. Постулаты лучевой оптики 23
1.2. Простые оптические элементы 27
1.2.1. Зеркала 27
1.2.2. Плоские границы 31
1.2.3. Сферические границы и линзы 34
1.2.4. Световоды 38
1.3. Оптика сред с градиентным показателем преломления 40
1.3.1. Уравнение луча 41
1.3.2. Оптические элементы с градиентным показателем преломления 43
*1.3.3. Уравнение эйконала 48
1.4. Матричная оптика 50
1.4.1. Матрица передачи луча 51
1.4.2. Матрицы простых оптических элементов 52
1.4.3. Матрицы каскада оптических элементов 54
1.4.4. Периодические оптические системы 57

Глава 2
ОПТИКА ВОЛН 69

2.1. Постулаты волновой оптики 70
2.2. Монохроматические волны 72
2.2.1. Комплексное представление и уравнение Гельмгольца 73
2.2.2. Элементарные волны 75
2.2.3. Параксиальные волны 80
*2.3. Связь между волновой и лучевой оптикой 82
2.4. Простые оптические элементы 84
2.4.1. Отражение и преломление 84

В конце каждой главы приведены литература, рекомендуемая авторами, и задачи. — *Прим. издательства.*

2.4.2. Прохождение через оптические элементы ... 85
2.4.3. Оптические элементы с градиентным показателем преломления ... 93
2.5. Интерференция ... 94
2.5.1. Интерференция двух волн ... 94
2.5.2. Многоволновая интерференция ... 101
2.6. Полихроматический и импульсный свет ... 105
2.6.1. Временно́е и спектральное описание ... 106
2.6.2. Световые биения ... 110

Глава 3
ОПТИКА ПУЧКОВ 116

3.1. Гауссов пучок 117
3.1.1. Комплексная амплитуда ... 117
3.1.2. Свойства ... 119
3.1.3. Качество пучка ... 129
3.2. Прохождение через оптические элементы ... 130
3.2.1. Прохождение через тонкую линзу 130
3.2.2. Формирование пучка ... 133
3.2.3. Отражение от сферического зеркала ... 137
*3.2.4. Прохождение через произвольную оптическую систему ... 138
3.3. Пучки Эрмита—Гаусса ... 141
3.4. Пучки Лагерра—Гаусса и Бесселя ... 145

Глава 4
ФУРЬЕ-ОПТИКА ... 151

4.1. Распространение света в свободном пространстве ... 154
4.4.1. Пространственные гармоники и плоские волны 154
4.1.2. Передаточная функция свободного пространства ... 164
4.1.3. Функция отклика на импульсное воздействие для свободного пространства ... 168
4.1.4. Принцип Гюйгенса—Френеля ... 170
4.2. Оптическое преобразование Фурье ... 170
4.2.1. Преобразование Фурье в дальней зоне ... 171
4.2.2. Преобразование Фурье с помощью линзы 174
4.3. Дифракция света ... 178
4.3.1. Дифракция Фраунгофера ... 180
4.3.2. Дифракция Френеля ... 184
4.4. Формирование изображения 189
4.4.1. Лучевая оптика однолинзовой изображающей системы ... 189
4.4.2. Волновая оптика формирования изображения в 4*f*-системе 191
4.4.3. Волновая оптика однолинзовой изображающей системы ... 195
4.4.4. Формирование изображения в ближнем поле ... 200
4.5. Голография ... 203

Глава 5
ЭЛЕКТРОМАГНИТНАЯ ОПТИКА 221

5.1. Электромагнитная теория света 223
5.2. Электромагнитные волны в диэлектрических средах 227
5.2.1. Линейные, недиспергирующие, однородные и изотропные среды 228
5.2.2. Нелинейные, диспергирующие, неоднородные, или неизотропные, среды 230
5.3. Монохроматические электромагнитные волны 236
5.4. Элементарные электромагнитные волны 239
5.4.1. Плоские, сферические и гауссовы электромагнитные волны 239
5.4.2. Связь между электромагнитной и скалярной волновой оптикой 245
5.4.3. Векторные пучки 246
5.5. Поглощение и дисперсия 247
5.5.1. Поглощение 247
5.5.2. Дисперсия 251
5.5.3. Резонансная среда 253
5.5.4. Оптика проводящих сред 260
5.6. Распространение импульсов в средах с дисперсией 265
5.7. Оптика магнитных материалов и метаматериалов 274

Глава 6
ПОЛЯРИЗАЦИОННАЯ ОПТИКА 283

6.1. Поляризация света 285
6.1.1. Поляризация 285
6.1.2. Матричное представление 290
6.2. Отражение и преломление 298
6.3. Оптика анизотропных сред 306
6.3.1. Показатели преломления 307
6.3.2. Распространение вдоль главной оси 310
6.3.3. Распространение в произвольном направлении 312
6.3.4. Дисперсионные соотношения, лучи, волновые фронты и перенос энергии 314
6.3.5. Двулучепреломление 319
6.4. Оптическая активность и магнитооптика 322
6.4.1. Оптическая активность 322
6.4.2. Магнитооптика: эффект Фарадея 326
6.5. Оптика жидких кристаллов 328
6.6. Поляризационные устройства 332
6.6.1. Поляризаторы 332
6.6.2. Фазовые пластинки 333
6.6.3. Вращатели плоскости поляризации 335
6.6.4. Невзаимные поляризационные устройства 336

Глава 7
ОПТИКА ФОТОННЫХ КРИСТАЛЛОВ 343

7.1. Оптика многослойных диэлектрических сред 346
7.1.1. Матричная теория многослойной оптики 346
7.1.2. Эталон Фабри—Перо 357
7.1.3. Решетка Брэгга 361
7.2. Одномерные фотонные кристаллы 372
7.2.1. Моды Блоха 373
7.2.2. Матричная оптика периодических сред 377
7.2.3. Фурье-оптика периодических сред 386
7.2.4. Границы между периодическими и однородными средами 389
7.3. Двумерные и трехмерные фотонные кристаллы 392
7.3.1. Двумерные фотонные кристаллы 393
7.3.2. Трехмерные фотонные кристаллы 396

Глава 8
ОПТИКА ВОЛНОВОДОВ 406

8.1. Планарные зеркальные волноводы 408
8.2. Планарные диэлектрические волноводы 418
8.2.1. Волноводные моды 419
8.2.2. Распределения полей 423
8.2.3. Дисперсионные соотношения и групповые скорости 426
8.3. Двумерные волноводы 429
8.4. Фотонно-кристаллические волноводы 434
8.5. Оптическая связь в волноводах 436
8.5.1. Устройства ввода 436
8.5.2. Связанные волноводы 439
8.5.3. Периодические волноводы 446
8.6. Металлические волноводы с размерами меньше длины волны (плазмоника) 447

Глава 9
ВОЛОКОННАЯ ОПТИКА 452

9.1. Направляемые лучи 454
9.1.1. Волокна со ступенчатым профилем показателя преломления 454
9.1.2. Градиентные волокна 458
9.2. Направляемые волны 460
9.2.1. Волокна со ступенчатым профилем показателя преломления 461
9.2.2. Одномодовые волокна 471
9.2.3. Квазиплоские волны в волокнах со ступенчатым и градиентным профилем показателя преломления 474

9.3. Затухание и дисперсия ... 481
9.3.1. Затухание 482
9.3.2. Дисперсия ... 485
9.4. Микроструктурные и фотонно-кристаллические волокна ... 496

Глава 10
ОПТИКА РЕЗОНАТОРОВ ... 504

10.1. Резонаторы с плоскими зеркалами ... 507
10.1.1. Моды резонатора ... 507
10.1.2. Внеосевые моды резонатора ... 518
10.2. Резонаторы со сферическими зеркалами ... 519
10.2.1. Удержание лучей ... 520
10.2.2. Гауссовы моды ... 523
10.2.3. Резонансные частоты ... 528
10.2.4. Моды Гаусса—Эрмита ... 529
10.2.5. Конечные апертуры и дифракционные потери ... 531
10.3. Дву- и трехмерные резонаторы ... 533
10.3.1. Двумерные прямоугольные резонаторы ... 533
10.3.2. Круговые резонаторы и моды шепчущей галереи ... 535
10.3.3. Трехмерные резонаторы в виде прямоугольной полости ... 537
10.4. Микрорезонаторы ... 540
10.4.1. Прямоугольные микрорезонаторы ... 541
10.4.2. Резонаторы в виде микростолбиков, микродисков и микроторов ... 542
10.4.3. Микросферические резонаторы ... 543
10.4.4. Фотонно-кристаллические микрорезонаторы ... 545

Глава 11
СТАТИСТИЧЕСКАЯ ОПТИКА ... 551

11.1. Статистические свойства случайного света 553
11.1.1. Оптическая интенсивность ... 553
11.1.2. Временна́я когерентность и спектр 555
11.1.3. Пространственная когерентность ... 562
11.1.4. Продольная когерентность ... 567
11.2. Интерференция частично когерентного света ... 570
11.2.1. Интерференция двух частично когерентных волн ... 570
11.2.2. Интерференция и временна́я когерентность ... 571
11.2.3. Интерференция и пространственная когерентность ... 575
11.3. Прохождение частично когерентного света через оптические системы ... 579
11.3.1. Распространение частично когерентного света ... 580
11.3.2. Формирование изображений в некогерентном свете ... 582
11.3.3. Приобретение пространственной когерентности при распространении ... 586
11.4. Частичная поляризация ... 591

Глава 12
ОПТИКА ФОТОНОВ 602

12.1. Фотон 604
12.1.1. Энергия фотона 606
12.1.2. Поляризация фотона 607
12.1.3. Положение фотона 610
12.1.4. Импульс фотона 612
12.1.5. Интерференция фотона 615
12.1.6. Временна́я локализация фотона 617
12.2. Потоки фотонов 620
12.2.1. Средний поток фотонов 621
12.2.2. Случайность потока фотонов 624
12.2.3. Статистика числа фотонов 626
12.2.4. Случайное разбиение фотонных потоков 634
*12.3. Квантовые состояния света 637
12.3.1. Когерентные состояния света 641
12.3.2. Сжатые состояния света 642

Глава 13
ФОТОНЫ И АТОМЫ 652

13.1. Уровни энергии 653
13.1.1. Атомы 654
13.1.2. Молекулы 660
13.1.3. Твердые тела 663
13.2. Заселенность уровней энергии 674
13.2.1. Распределение Больцмана 674
13.2.2. Распределение Ферми—Дирака 676
13.3. Взаимодействие фотонов с атомами 677
13.3.1. Взаимодействие одномодового света с атомом 677
13.3.2. Спонтанное излучение 682
13.3.3. Вынужденное излучение и поглощение 685
13.3.4. Уширение линий 689
*13.3.5. Усиленное спонтанное излучение 695
*13.3.6. Лазерное охлаждение атомов и лазерные ловушки 696
13.4. Тепловое излучение 698
13.4.1. Тепловое равновесие между фотонами и атомами 698
13.4.2. Спектр излучения черного тела 700
13.5. Люминесценция и рассеяние света 703
13.5.1. Виды люминесценции 703
13.5.2. Фотолюминесценция 705
13.5.3. Рассеяние света 708

Глава 14
ЛАЗЕРНЫЕ УСИЛИТЕЛИ 716

14.1. Теория лазерного усиления 719
14.1.1. Коэффициент и ширина полосы усиления 720
14.1.2. Фазовый сдвиг 724
14.2. Накачка усилителя 725
14.2.1. Скоростные уравнения 726
14.2.2. Схемы накачки 730
14.3. Распространенные лазерные усилители 736
14.3.1. Рубин 736
14.3.2. Стекло с неодимом 738
14.3.3. Кварцевое волокно, легированное эрбием 739
14.3.4. Волоконные ВКР-усилители 742
14.3.5. Таблица избранных лазерных переходов 745
14.4. Нелинейность усилителя 745
14.4.1. Насыщение усиления в однородно-уширенной среде 745
*14.4.2. Насыщение усиления в неоднородно-уширенной среде 751
*14.5. Шум усилителя 754

ПРЕДИСЛОВИЕ К ВТОРОМУ ИЗДАНИЮ

С момента выхода первого издания в 1991 г. книга «Основы фотоники» переиздавалась около 20 раз, была переведена на чешский и японский языки и использовалась по всему миру как учебник и справочник. За этот период продолжалось быстрое развитие фотоники, открывшее путь новым технологиям, таким как телекоммуникации и применения в промышленности и медицине. Второе издание обобщает эти достижения, сохраняя однотомный формат книги в пределах разумного объема. (В русском переводе книга выходит в двух томах. — *прим. пер.*)

В своей новой структуре «Основы фотоники» продолжают служить самодостаточным, стоящим на уровне современных требований учебником вводного уровня, обеспечивающим логическое сочетание теории и приложений. Многие читатели первого издания были удовлетворены его обильными иллюстрациями. Это качество усилено во втором издании за счет введения цветных иллюстраций по всей книге, обеспечивающих наглядность и облегчающих ее чтение.

Притом, что все 22 главы первого издания существенно обновлены, для второго издания принципиально добавление двух новых глав: одна о фотонных кристаллах, а другая — об оптике ультракоротких импульсов. В них отражаются достижения, оказывающие существенное и растущее влияние на фотонику в последние десять лет.

Новая глава об **оптике фотонных кристаллов** даст основу для понимания оптики слоистых сред, включая решетки Брэгга, с помощью матричного подхода. Распространение света в одномерных периодических средах изучается с использованием мод Блоха с помощью матричного метода и метода Фурье. Вводится понятие фотонной запрещенной зоны. Разрабатывается теория распространения света в двух- и трехмерных фотонных кристаллах и получаются соответствующие дисперсионные соотношения и структуры запрещенных зон. Добавлены также разделы о фотонно-кристаллических волноводах, дырчатых волокнах и фотонно-кристаллических резонаторах в соответствующих местах других глав.

Новая глава об **оптике ультракоротких импульсов** содержит разделы о пико- и фемтосекундных оптических импульсах, их описании, формировании и сжатии, а также об их распространении в оптических волокнах в пределах применимости линейной оптики. Разделы о нелинейной оптике таких импульсов включают импульсные параметрические взаимодействия и оптические солито-

ны. Дается обзор методов регистрации ультракоротких оптических импульсов с использованием доступных сравнительно медленных приемников.

Кроме добавления двух новых глав, была полностью переписана глава об **оптических межсоединениях и коммутаторах**, дополненная такими темами, как маршрутизация и переключение по времени и длине волны и новые типы оптических переключателей. Глава об **оптоволоконной связи** также была существенно обновлена и дополнена материалом по сетям, использующим уплотнение с разделением сигналов по длине волны; в ней теперь даются лаконичные обзоры таких проблем, как компенсация дисперсии и управление ею, а также оптическая связь с применением солитонов.

Продолжающееся развитие технологии изготовления оптических устройств породило **нанофотонику**, имеющую дело с оптическими процессами, происходящими на субволновых (нанометровых) пространственных масштабах. Нанофотонные устройства и системы включают квантоворазмерные структуры, такие как квантовые точки, наночастицы и периодические наноструктуры, используемые для синтеза *метаматериалов* с экзотическими оптическими свойствами, такими как отрицательный показатель преломления. К ним также относятся конфигурации системы, в которых свет (или его взаимодействие с веществом) ограничен нанометровыми (а не микрометровыми) областями вблизи границ, как в оптике *поверхностных плазмонов*. Нераспространяющиеся поля, такие как создаваемые вблизи поверхности, где происходит полное внутреннее отражение, также имеют такое ограничение. Нераспространяющиеся поля присутствуют в непосредственной близости от апертур субволновых размеров, таких как открытый конец конического волокна (тейпера). Их применение делает возможным получение изображений с разрешением, превосходящим дифракционный предел, и образует основу *оптики ближнего поля*. Многие из этих развивающихся областей описаны в соответствующих местах во втором издании данной книги.

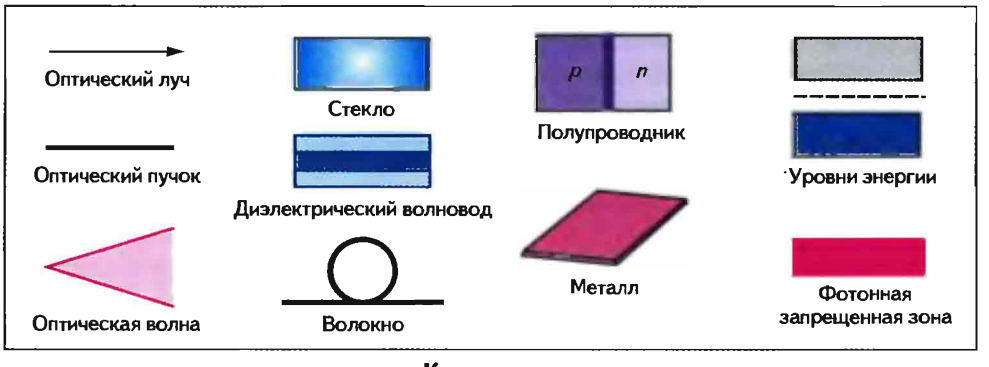
Новые разделы были добавлены в процессе обновления различных глав. К новым темам, добавленным в ранее написанные главы, относятся: пучки Лагерра—Гаусса, получение изображений с помощью ближнего поля, уравнение Зельмейера, быстрый и медленный свет, оптика проводящих сред и плазмоника, дважды отрицательные метаматериалы, сфера Пуанкаре и параметры Стокса, поляризационная модовая дисперсия, моды типа шепчущей галереи, микрорезонаторы, оптическая когерентная томография, орбитальный момент фотона.

В главах по оптике лазеров новые темы включают: волоконные усилители на основе редкоземельных элементов и комбинационного рассеяния, лазеры крайнего ультрафиолетового излучения, рентгеновские лазеры и лазеры на свободных электронах, химические лазеры и лазеры со случайной генерацией. В области оптоэлектроники добавлены: структуры и устройства на основе нитрида галлия, суперлюминесцентные диоды, органические светодиоды и светодиоды белого света, квантово-размерные лазеры, квантово-каскадные лазеры, лазеры с микрорезонаторами, фотонно-кристаллические лазеры, матричные приемники, низкошумящие лавинные фотодиоды, лавинные фотодетекторы одиночных фотонов, инфракрасные фотодетекторы на квантовых ямах.

Глава о нелинейной оптике была дополнена материалом по перестроечным кривым параметрического взаимодействия, устройствам с фазовым квазисинхронизмом, двухволновому смешению и кросс-модуляции фазы, генерации терагерцового излучения, а также другим нелинейно оптическим явлениям, связанным с короткими оптическим импульсами, включая усиление импульса с чирпом и генерацию суперконтинуума. Глава об электрооптике теперь содержит обсуждение модуляторов на основе электропоглощения.

Приложение В о **модах линейных систем** было расширено и теперь предлагает читателю полный обзор концепции мод, появляющихся в книге множество раз. Наконец, добавлены новые задачи и упражнения, причем их нумерация разделена во избежание недоразумений.

В настоящем полноцветном издании в большинстве иллюстраций использован цветовой код, отраженный в нижеприведенной карте. Пучки и распределения света изображены розовым (кроме многоцветных пучков, как, например, в нелинейной оптике). Стекло и стеклянные волокна изображаются светло-голубым. Полупроводники окрашены фиолетовым, причем яркость отражает различную степень легирования, металлы показаны красно-сиреневым. Схемы уровней и зон энергии изображены серым или темно-синим цветом, а фотонные запрещенные зоны — красным.



Карта цветов

Организация книги

В своем новом воплощении книга «Основы фотоники» содержит 24 главы, сгруппированных в шесть частей, как показано на диаграмме ниже. Форма книги модульная, так что она может использоваться читателями, имеющими различные потребности; она также дает возможность преподавателю выбрать темы для различных курсов. Существенный материал одной главы часто кратко повторяется в другой, чтобы сделать каждую главу максимально самодостаточной. Например, в начале гл. 24 «Волоконно-оптические системы связи» дается краткий обзор необходимого материала предшествующих глав, касающийся волокон, источников света, приемников и усилителей. Таким образом, в распоряжение читателя предоставляются сведения об элементной базе

сетей прежде, чем перейти к обсуждению структуры и функционирования всей системы связи, где указанные элементы применяются.

Основы	Распространение волн	Лазерная оптика	Элементная база оптической связи
1. Оптика лучей	7. Оптика фотонных кристаллов	13. Фотоны и атомы	19. Акустооптика
2. Оптика волн	8. Оптика волноводов	14. Лазерные усилители	20. Электрооптика
3. Оптика пучков	9. Волоконная оптика	15. Лазеры	21. Нелинейная оптика
4. Фурье-оптика	10. Оптика резонаторов	16. Оптика полупроводников	22. Оптика сверхбыстрых процессов
5. Электромагнитная оптика	11. Статистическая оптика	17. Полупроводниковые источники фотонов	23. Оптические межсоединения и коммутаторы
6. Поляризационная оптика	12. Оптика фотонов	18. Полупроводниковые детекторы фотонов	24. Волоконно-оптические системы связи
		Оптоэлектроника	**Системы оптической связи**

Осознавая различную степень математической подготовки потенциального читателя, мы решили представлять трудный материал в два приема: на вводном уровне, обеспечивающем понимание физического смысла и мотивации, а затем на уровне более глубокого и строгого анализа. Примером такого подхода является гл. 20 («Электрооптика»), в которой рассмотрение сначала проводится в рамках скалярной модели, а затем — с помощью тензорного формализма.

Мы старались использовать общепринятые символы и обозначения везде, где это возможно. Ввиду широкого круга рассматриваемых проблем большое число символов имеет по нескольку значений: в конце книги прилагается список обозначений и символов, помогающий внести ясность в этот вопрос. По всей книге важные формулы выделены рамками, чтобы облегчить последующее их использование. Разделы повышенной сложности отмечены звездочками и могут быть при желании пропущены. Краткие сводки полученных результатов даются всюду, где этого требует сущность излагаемого материала.

Каждая глава содержит также упражнения, набор задач и список литературы для чтения. Примеры реальных систем включены для того, чтобы подчеркнуть идеи, лежащие в основе актуальных практических применений, а в приложениях приводятся свойства одно- и двумерного преобразования Фурье, теория линейных систем и моды линейных систем.

Типичные курсы

Главы этой книги можно комбинировать различным образом для использования в семестровых или полусеместровых учебных курсах. Типичные примеры таких курсов приведены ниже. Некоторые из них могут быть частью цикла курсов. Возможны и другие варианты выбора, подходящие к конкретным целям преподавателя и студента.

Оптика/фотоника			
1. Оптика лучей	7. Оптика фотонных кристаллов	13. Фотоны и атомы	19. Акустооптика
2. Оптика волн	8. Оптика волноводов	14. Лазерные усилители	20. Электрооптика
3. Оптика пучков	9. Волоконная оптика	15. Лазеры	21. Нелинейная оптика
4. Фурье-оптика	10. Оптика резонаторов	16. Оптика полупроводников	22. Оптика сверхбыстрых процессов
5. Электромагнитная оптика	11. Статистическая оптика	17. Полупроводниковые источники фотонов	23. Оптические межсоединения и коммутаторы
6. Поляризационная оптика	12. Оптика фотонов	18. Полупроводниковые детекторы фотонов	24. Волоконно-оптические системы связи

Первые шесть глав книги подходят для вводного курса «Оптика» или «Фотоника». К ним можно добавить гл. 11 «Статистическая оптика» для введения некогерентного и частично-когерентного света или вводные разделы гл. 8 и 9 «Оптика волноводов» и «Волоконная оптика» для знакомства с приложениями.

	Оптическая обработка информации		
1. Оптика лучей	7. Оптика фотонных кристаллов	13. Фотоны и атомы	19. Акустооптика
2. Оптика волн	8. Оптика волноводов	14. Лазерные усилители	20. Электрооптика
3. Оптика пучков	9. Волоконная оптика	15. Лазеры	21. Нелинейная оптика
4. Фурье-оптика	10. Оптика резонаторов	16. Оптика полупроводников	22. Оптика сверхбыстрых процессов
5. Электромагнитная оптика	11. Статистическая оптика	17. Полупроводниковые источники фотонов	23. Оптические межсоединения и коммутаторы
6. Поляризационная оптика	12. Оптика фотонов	18. Полупроводниковые детекторы фотонов	24. Волоконно-оптические системы связи

Курс «Оптическая обработка информации» можно начать с основ волновой оптики и оптики пучков с включением Фурье-оптики (когерентное формирование и обработка изображений), а также некогерентное и частично-когерентное формирование изображений в статистической оптике. Затем можно включить материал по устройствам, используемым для обработки аналоговых данных (раздел «Акустооптика»), а закончить рассмотрением коммутаторов и логических элементов, используемых при обработке цифровых данных.

			Оптика волноводов
1. Оптика лучей	7. Оптика фотонных кристаллов	13. Фотоны и атомы	19. Акустооптика
2. Оптика волн	8. Оптика волноводов	14. Лазерные усилители	20. Электрооптика
3. Оптика пучков	9. Волоконная оптика	15. Лазеры	21. Нелинейная оптика
4. Фурье-оптика	10. Оптика резонаторов	16. Оптика полупроводников	22. Оптика сверхбыстрых процессов
5. Электромагнитная оптика	11. Статистическая оптика	17. Полупроводниковые источники фотонов	23. Оптические межсоединения и коммутаторы
6. Поляризационная оптика	12. Оптика фотонов	18. Полупроводниковые детекторы фотонов	24. Волоконно-оптические системы связи

Курс «Оптика волноводов» можно начать с введения в распространение волн в слоистых и периодических средах (гл. 7 «Оптика фотонных кристаллов») и продолжить с использованием глав «Оптика волноводов», «Волоконная оптика» и «Оптика резонаторов». Дополнительные темы могут включать «Электрооптику» и «Оптические межсоединения и коммутаторы».

Лазеры			
1. Оптика лучей	7. Оптика фотонных кристаллов	13. Фотоны и атомы	19. Акустооптика
2. Оптика волн	8. Оптика волноводов	14. Лазерные усилители	20. Электрооптика
3. Оптика пучков	9. Волоконная оптика	15. Лазеры	21. Нелинейная оптика
4. Фурье-оптика	10. Оптика резонаторов	16. Оптика полупроводников	22. Оптика сверхбыстрых процессов
5. Электромагнитная оптика	11. Статистическая оптика	17. Полупроводниковые источники фотонов	23. Оптические межсоединения и коммутаторы
6. Поляризационная оптика	12. Оптика фотонов	18. Полупроводниковые детекторы фотонов	24. Волоконно-оптические системы связи

Курс «Лазеры» можно начать с оптики пучков и резонаторов, затем следует теория взаимодействия света с веществом (гл. 13), лазерного усиления и генерации (гл. 14 и 15) с включением полупроводниковых светоизлучающих диодов и лазеров (гл. 16 и 17). Введение в физику фемтосекундных лазеров можно обеспечить путем включения соответствующих разделов из главы «Оптика сверхбыстрых процессов».

	Оптоэлектроника		
1. Оптика лучей	7. Оптика фотонных кристаллов	13. Фотоны и атомы	19. Акустооптика
2. Оптика волн	8. Оптика волноводов	14. Лазерные усилители	20. Электрооптика
3. Оптика пучков	9. Волоконная оптика	15. Лазеры	21. Нелинейная оптика
4. Фурье-оптика	10. Оптика резонаторов	16. Оптика полупроводников	22. Оптика сверхбыстрых процессов
5. Электромагнитная оптика	11. Статистическая оптика	17. Полупроводниковые источники фотонов	23. Оптические межсоединения и коммутаторы
6. Поляризационная оптика	12. Оптика фотонов	18. Полупроводниковые детекторы фотонов	24. Волоконно-оптические системы связи

Три главы, посвященные оптике полупроводников, полупроводниковым источникам, усилителям и детекторам, образуют основу для курса «Оптоэлектроника». Этот материал можно сопроводить основами оптики из предыдущих глав и расширить путем включения таких тем, как устройства на жидких кристаллах (разд. 6.5 и 20.3), полупроводниковые модуляторы на основе электропоглощения (разд. 20.5) и введение в использование фотонных устройств для коммутации и/или связи (гл. 23 и 24, соответственно).

			Фотонные устройства
1. Оптика лучей	7. Оптика фотонных кристаллов	13. Фотоны и атомы	19. Акустооптика
2. Оптика волн	8. Оптика волноводов	14. Лазерные усилители	20. Электрооптика
3. Оптика пучков	9. Волоконная оптика	15. Лазеры	21. Нелинейная оптика
4. Фурье-оптика	10. Оптика резонаторов	16. Оптика полупроводников	22. Оптика сверхбыстрых процессов
5. Электромагнитная оптика	11. Статистическая оптика	17. Полупроводниковые источники фотонов	23. Оптические межсоединения и коммутаторы
6. Поляризационная оптика	12. Оптика фотонов	18. Полупроводниковые детекторы фотонов	24. Волоконно-оптические системы связи

Другая возможная тема курса лекций, совмещающего устройства на основе фотонных кристаллов и волноводов с электрооптическими, акустооптическими и нелинейно-оптическими устройствами, с включением оптики ультракоротких импульсов, оптических средств связи и переключателей — это «Фотонные устройства».

Волоконно-оптическая связь			
1. Оптика лучей	7. Оптика фотонных кристаллов	13. Фотоны и атомы	19. Акустооптика
2. Оптика волн	8. Оптика волноводов	14. Лазерные усилители	20. Электрооптика
3. Оптика пучков	9. Волоконная оптика	15. Лазеры	21. Нелинейная оптика
4. Фурье-оптика	10. Оптика резонаторов	16. Оптика полупроводников	22. Оптика сверхбыстрых процессов
5. Электромагнитная оптика	11. Статистическая оптика	17. Полупроводниковые источники фотонов	23. Оптические межсоединения и коммутаторы
6. Поляризационная оптика	12. Оптика фотонов	18. Полупроводниковые детекторы фотонов	24. Волоконно-оптические системы связи

Курс «Волоконно-оптическая связь» мог бы включать оптические волноводы и волокна, полупроводниковые источники и усилители света (возможно, также подразд. 14.3.3 и 14.3.4 о волоконных и рамановских волоконных усилителях) в качестве базового материала для главы «Волоконно-оптическая связь» (гл. 24). Если требуется выделить волоконно-оптические сети, можно также включить разд. 23.3 о фотонных переключателях.

Бостон, Массачусетс — *Бахаа Е.А. Салех*
19 декабря 2006 г. — *Малвин Карл Тейх*

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

Оптика — старая и уважаемая наука, охватывающая генерацию, распространение и регистрацию света. Три основных достижения способствовали обновлению оптики в последние тридцать лет и ее возрастающей важности в современной технологии: изобретение лазера, изготовление оптических волокон с низкими потерями, появление полупроводниковых лазерных устройств. В результате развития этих достижений появились новые научные дисциплины и вошли в употребление новые термины: электрооптика, оптоэлектроника, квантовая электроника, квантовая оптика, оптическая технология. Хотя нет полного единства в вопросе о точном значении этих терминов, в общем и целом согласие на этот счет существует.

Фотоника

Термин «*электрооптика*» применяется для оптических устройств, в которых играют роль электрические эффекты (например, лазеры, электрооптические модуляторы и переключатели). *Оптоэлектроника* относится к устройствам и системам существенно электронной природы, но включающих в себя свет (примерами являются светодиоды, дисплеи на основе жидких кристаллов, матричные фотоприемники). Термин «*квантовая электроника*» употребляется в связи с устройствами и системами, принципиальную основу которых составляет взаимодействие света с веществом (примерами служат лазеры и нелинейно-оптические устройства, используемые для усиления и смещения волн). Исследования квантовых и когерентных свойств света лежат в области *квантовой оптики*. Термин «*оптическая технология*» применялся для описания приборов и систем, используемых в оптической связи и оптической обработке информации.

В последние годы в употребление вошел термин «**фотоника**». Возникший по аналогии с электроникой, он отражает растущие связи между оптикой и электроникой, усиливаемые возрастающей ролью, которую играют в оптических системах полупроводниковые материалы и устройства. *Электроника* включает в себя управление потоками электрических зарядов (в вакууме или веществе); *фотоника* — управление фотонами (в свободном пространстве или материальной среде). Обе дисциплины очевидно перекрываются, поскольку электроны часто управляют потоком фотонов и, наоборот, фотоны управляют потоками электронов. Термин «*фотоника*» отражает также важность фотонной природы света для описания действия многих оптических устройств.

Тематика книги

Данная книга вводит читателя в основы фотоники. Термин «*фотоника*» широко используется и охватывает все вышеупомянутые области, включая нижеперечисленные.

- *Генерация* когерентного света лазерами и некогерентного света люминесцентными источниками, такими как светодиоды.
- *Передача* света в свободном пространстве, через обычные оптические элементы, такие как линзы, диафрагмы и изображающие системы, а также через волноводы, например оптические волокна.
- *Модуляция*, переключение и сканирование света с использованием электрически, акустически или оптически управляемых приборов.
- *Усиление* и *преобразование частоты* света с использованием взаимодействия волн в нелинейных материалах.
- *Регистрация* света.

Эти области нашли постоянно растущее применение в оптической связи, обработке сигналов, зондировании, отображении информации, печати и передаче энергии света.

Авторский подход и представление материала

Фундаментальные основы фотоники представлены в ряде глав, которые в сжатой форме содержат:

- четыре теории света (каждая из которых является более общей, чем предыдущая): лучевую, волновую, электромагнитную и фотонную оптику;
- теорию взаимодействия света с веществом;
- теорию полупроводниковых материалов и их оптические свойства.

Эти главы служат основным строительным материалом, используемым в других главах для описания *генерации* света (лазерами и светодиодами), *передачи* света (оптическими пучками, при дифракции, формировании изображений, в оптических волноводах и волокнах), *модуляции* и переключении света (с использованием электрооптических, акустооптических и нелинейно-оптических устройств) и *регистрации* света (с помощью фотоприемников). Приводится множество приложений и примеров, так что книга представляет современное состояние как теории, так и практики. Заключительная глава посвящена исследованию волоконно-оптической связи, которая дает особенно богатый ассортимент примеров, в которых генерация, передача, модуляция и прием световых сигналов являются частями единой фотонной системы, предназначенной для передачи информации.

Теории света представлены в порядке возрастания их сложности. Так, свет вначале рассматривается как лучи, затем как скалярные волны, затем как электромагнитные волны и, наконец, как фотоны. Каждое из этих описаний имеет свою область применимости. Наш подход состоит в том, чтобы для каждого приложения использовать простейшую теорию, обеспечивающую его адекватное описание. Поэтому оптика лучей используется для описания систем получения изображений, а также для объяснения их ограничения в волноводах и резонаторах. Скалярная волновая теория обеспечивает описание оптических пуч-

ков, существенное для понимания лазеров, а также Фурье-оптики, полезной для описания когерентных оптических систем и голографии. Электромагнитная теория обеспечивает основу описания поляризации и дисперсии света, оптики направляемых волн, волокон и резонаторов. Оптика фотонов служит для описания взаимодействия света и вещества, объясняя такие процессы, как генерация и детектирование света, а также смешение света в нелинейных средах.

Аудитория, на которую рассчитана книга

Книга «Основы фотоники» адресована:

- студентам, специализирующимся в области электротехники или прикладной физики на старших курсах;
- в качестве пособия, содержащего все необходимые сведения широкому кругу специалистов для самообразования;
- специалистам, желающим продолжить профессиональное образование в соответствии с потребностями промышленности, университетов и профессиональных обществ.

Предполагается, что читатель имеет базовый уровень образования в инженерной или прикладной физике, включая современную общую физику, электричество и магнетизм, движение волн. Знание линейных систем и элементарной квантовой механики полезно, но не существенно. Нашей целью было дать такое введение в фотонику, которое фокусирует внимание на понятиях, составляющих основу наиболее актуальных приложений. Поэтому книгу не следует рассматривать как справочник, охватывающий все фотонные устройства и системы. На самом деле некоторые области фотоники совсем не вошли в нее, а многие отдельные главы легко могли бы быть расширены до самостоятельных монографий.

Задачи, списки литературы и приложения

В конце каждой главы приводится набор задач. Задачи пронумерованы в соответствии с разделами главы, к которым они относятся. Довольно часто задачи посвящены идеям или приложениям, не упоминаемым в тексте, аналитическим выводам и численным расчетам, иллюстрирующим значения важных величин. Задачи, отмеченные звездочкой, имеют повышенную сложность. В тексте каждой главы имеется также множество упражнений, помогающих читателю приобрести более глубокое понимание материала или расширить его.

В приложениях даны свойства одно- и двумерных преобразований Фурье, элементы теории линейных систем и моды линейных систем (важные для поляризационных устройств, оптических волноводов и резонаторов); ссылки на приложения даны в соответствующих местах на протяжении всей книги. Каждая глава заканчивается списком литературы, включающим избранные важные книги, обзорные статьи и небольшое количество классических оригинальных статей особой важности.

Мадисон, Висконсин — *Бахаа Е.А. Салех*
Нью Йорк, 3 апреля 1991 г. — *Малвин Карл Тейх*

ГЛАВА 1 ОПТИКА ЛУЧЕЙ

Сэр Исаак Ньютон (1642—1727) выдвинул оптическую теорию, в которой свет представляет собой набор частиц, движущихся прямолинейно

Пьер Ферма (1601—1665) сформулировал принцип, согласно которому свет выбирает траекторию своего движения

Свет можно описать как электромагнитное волновое явление, подчиняющееся тем же теоретическим принципам, которые управляют всеми другими видами электромагнитного излучения, такими как радиоволны и рентгеновские лучи. Эта концепция света называется **электромагнитной оптикой**. Электромагнитное излучение распространяется в виде двух взаимосвязанных *векторных* волн: волны электрического поля и волны магнитного поля. Однако многие оптические явления можно описать с помощью упрощенной теории скалярных волн, в которой свет описывается одной скалярной волновой функцией. Это приближенный способ рассмотрения света называется скалярной волновой оптикой или просто **волновой оптикой**.

Когда свет распространяется через объекты, размеры которых намного больше, чем его длина волны, или вокруг таких объектов, волновая природа света слабо проявляет себя, и поведение света адекватно описывается с помощью лучей, подчиняющихся набору геометрических правил. Эта модель света называется **лучевой оптикой**. С математической точки зрения лучевая оптика является пределом волновой оптики при длине волны, стремящейся к нулю.

Таким образом, электромагнитная оптика включает волновую оптику, которая, в свою очередь, включает оптику лучей, как показано на рис. 1.1. Лучевая и волновая оптика являются приближенными теориями, справедливость которых определяется успешной аппроксимацией результатов более общей электромагнитной теории.

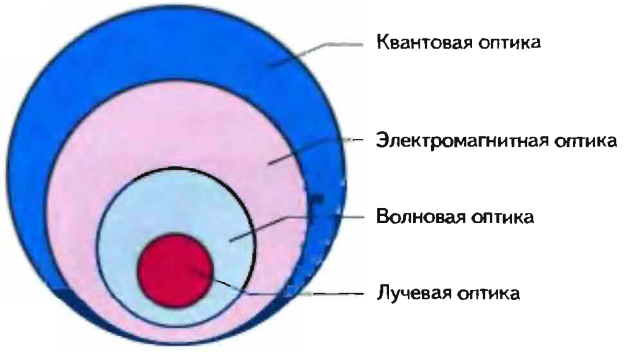


Рис. 1.1. Теория квантовой оптики дает объяснение почти всех оптических явлений. Электромагнитная теория света (электромагнитная оптика) обеспечивает наиболее полное описание света в границах классической оптики. Волновая оптика является скалярным приближением электромагнитной. Лучевая оптика — предел волновой оптики в случае очень малых длин волн

Хотя электромагнитная оптика обеспечивает наиболее полное описание света в границах **классической оптики**, некоторые оптические явления существенно квантовы по своей природе и не могут быть описаны классическим образом. Эти явления описываются квантовой версией электромагнитной теории, известной как **квантовая электродинамика**. Применительно к оптическим явлениям эта теория называется также **квантовой оптикой**.

Исторически оптические теории развивались примерно в такой последовательности: (1) лучевая оптика → (2) волновая оптика → (3) электромагнитная оптика → (4) квантовая оптика. Каждая следующая модель совершеннее и сложнее предыдущей, и ее появление вызвано необходимостью объяснения результатов более тонких и точных экспериментов. Оптимально выбирать простейшую из моделей, описывающих данное частное явление, однако это бывает трудно сделать *a priori* и достигается в результате опыта.

Из педагогических соображений начальные главы настоящей книги следуют исторической последовательности, предоставленной выше. Каждая модель света начинается с набора постулатов, принимаемых без доказательства, из которых выводятся различные конкретные результаты. Показано, что постулаты каждой модели выводятся из модели более высокого уровня в качестве частных результатов. В данной главе мы начинаем с оптики лучей.

О данной главе

Лучевая оптика — простейшая теория света. Свет описывается лучами, распространяющимися в различных оптических средах согласно набору геометрических правил. Лучевая оптика называется также **геометрической оптикой**. Лучевая оптика — приближенная теория. Хотя она адекватно описывает большую часть нашего повседневного опыта, касающегося света,

существует много явлений, которые она не способна описать. Многочисленные подтверждения этого читатель найдет в последующих главах настоящей книги.

Лучевая оптика связана с *положением* и *направлением* лучей. Она, следовательно, полезна при изучении *формирования изображений* — лучи от каждой точки предмета собираются и направляются оптическим элементом в соответствующую точку изображения. Лучевая оптика позволяет определить условия, при которых свет направляется данной средой, такой как стеклянное волокно. В изотропных средах лучи указывают направление, в котором распространяется *энергия света*. Можно построить пучки лучей таким образом, чтобы плотность числа лучей была пропорциональна плотности потока энергии света. Например, если свет изотропно испускается точечным источником, то энергия, связанная с лучами внутри некоторого конуса, пропорциональна охватываемому этим конусом телесному углу. Можно проследить за распространением лучей от источника через некоторую оптическую систему, чтобы определить оптическую энергию, проходящую через данную площадку.

Данная глава начинается с набора постулатов, из которых выводятся простые правила, определяющие распространение лучей в оптических средах. В разд. 1.2 эти правила применены к простым оптическим элементам, таким как зеркала, плоские и сферические границы раздела двух сред. Распространение лучей в неоднородных средах с плавно меняющимся показателем преломления рассмотрено в разд. 1.3. Оптика таких сред является основой технологии, которая стала важной частью современной оптики.

Оптические элементы часто центрированы относительно оптической оси, вдоль которой лучи распространяются с небольшими отклонениями. Такие лучи называются **параксиальными лучами**. Данное предположение составляет основу **параксиальной оптики**. Изменение положения и наклона параксиального луча в процессе его распространения через оптическую систему можно эффективно описать с помощью алгебр матриц 2×2. В разд. 1.4 рассказывается о данном алгебраическом методе, называемом **матричной оптикой**.

1.1. ПОСТУЛАТЫ ЛУЧЕВОЙ ОПТИКИ

Постулаты лучевой оптики

- Свет распространяется в виде лучей. Лучи испускаются источниками света и могут наблюдаться, когда они достигают оптического приемника.
- Оптическая среда характеризуется величиной $n \geq 1$, называемой **показателем преломления**. Показатель преломления равен $n = c_0/c$, где c_0 — скорость света в свободном пространстве; c — скорость света в среде. Таким образом, время, затрачиваемое светом на прохождение расстояния d, есть $d/c = nd/c_0$. Оно пропорционально произведению nd, известному как **оптическая длина пути**.

• В неоднородной среде показатель преломления $n(\boldsymbol{r})$ является функцией координат $\boldsymbol{r} = (x, y, z)$. Следовательно, оптическая длина пути вдоль данной кривой между двумя точками A и B есть

$$\text{Оптическая длина пути} = \int_A^B n(\boldsymbol{r})\,\mathrm{d}s, \tag{1.1}$$

где $\mathrm{d}s$ — элемент длины кривой. Время, затрачиваемое светом на проход от A до B, пропорционально оптической длине пути.

• **Принцип Ферма.** Оптический луч распространяется между двумя точками A и B по линии, вдоль которой затрачиваемое время (или оптическая длина пути) имеет экстремум по отношению к соседним кривым. Математически это выражается как

$$\delta \int_A^B n(\boldsymbol{r})\,\mathrm{d}s = 0, \tag{1.2}$$

где символ δ читается как «вариация величины». Условие (1.2) означает, что оптическая длина пути на указанной кривой имеет максимум, минимум или точку перегиба. Обычно это минимум, т. е.

Луч света распространяется по линии наименьшего времени.

Иногда минимальность времени распространения обеспечивается несколькими путями, вдоль которых свет идет одновременно. Пример максимального времени распространения рассмотрен в задаче 1 к разд. 1.2.

В данной главе мы используем постулаты лучевой оптики для вывода правил, определяющих распространение световых лучей, их отражение и преломление на границах между различными средами и их прохождение через различные элементы оптических систем. Изобилие результатов, применимых к многочисленным оптическим системам, получается без необходимости каких-либо других предположений или правил, касающихся природы света.

Распространение в однородной среде

В однородной среде показатель преломления везде одинаков, как и скорость света. Линия наименьшего времени, требуемая принципом Ферма, следовательно, есть линия минимальной длины. Принцип *пути наименьшей длины* известен как **принцип Герона**. Линия наименьшей длины между двумя точками есть прямая, так что *в однородной среде лучи света распространяются вдоль прямых линий* (рис. 1.2).

Отражение от зеркала

Зеркала изготовляются из металла с хорошо отполированной поверхностью или металлических или диэлектрических пленок, нанесенных на

такую подложку, как стекло. Свет отражается от зеркала в соответствии с законом отражения:

Отраженный луч лежит в плоскости падения; угол отражения равен углу падения.

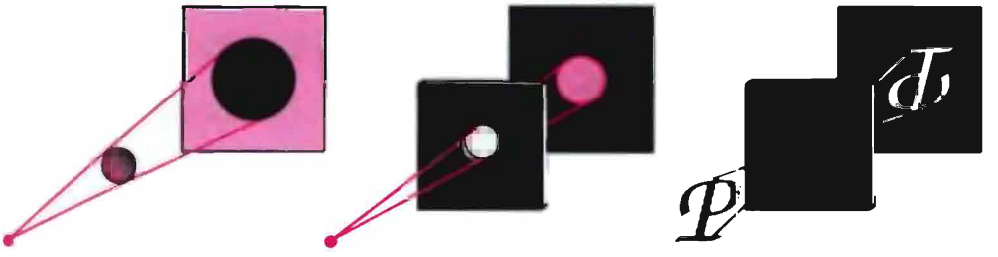


Рис. 1.2. Лучи света распространяются по прямым линиям. Тени представляют собой точное изображение препятствий

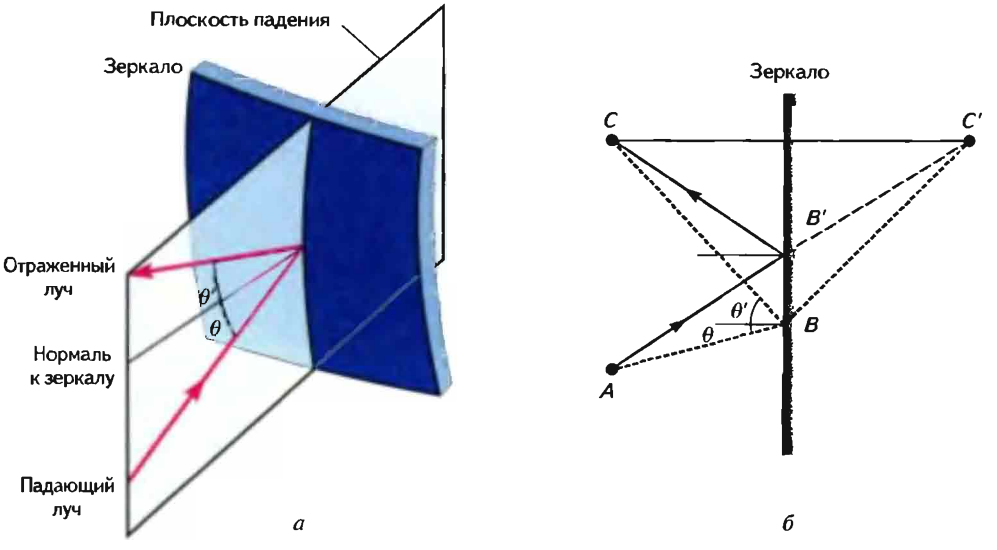


Рис. 1.3. Отражение от поверхности искривленного зеркала (*а*); геометрическое построение для доказательства закона отражения (*б*)

Плоскость падения — это плоскость, образованная падающим лучом и нормалью к поверхности зеркала в точке падения. Углы падения и отражения определяются рис. 1.3, *а*. Для доказательства закона отражения используем принцип Герона. Рассмотрим луч, попадающий из точки A в точку C в результате отражения от плоского зеркала, как показано на рис. 1.3, *б*. По принципу Герона для зеркала бесконечно малой толщины расстояние $\overline{AB} + \overline{BC}$ должно быть минимальным. Если C' — зеркальное изображение точки C, то $\overline{BC} = \overline{BC'}$, так что $\overline{AB} + \overline{BC'}$ должно быть минимальным. Это происходит, если линия $\overline{ABC'}$ прямая, т. е. если B совпадает с B', так что $\theta = \theta'$

Отражение и преломление на границе между двумя средами

На границе между двумя средами с показателями преломления n_1 и n_2 падающий луч расщепляется на два — отраженный и преломленный (прошедший) лучи (рис. 1.4). Отраженный луч подчиняется закону отражения. Преломленный луч подчиняется закону преломления:

Преломленный луч лежит в плоскости падения, угол преломления θ_2 связан с углом падения θ_1 законом Снелла

$$\boxed{n_1 \sin\theta_1 = n_2 \sin\theta_2.} \qquad (1.3)$$

Закон Снелла

Пропорция, в которой луч отражается и преломляется, не описывается лучевой оптикой.

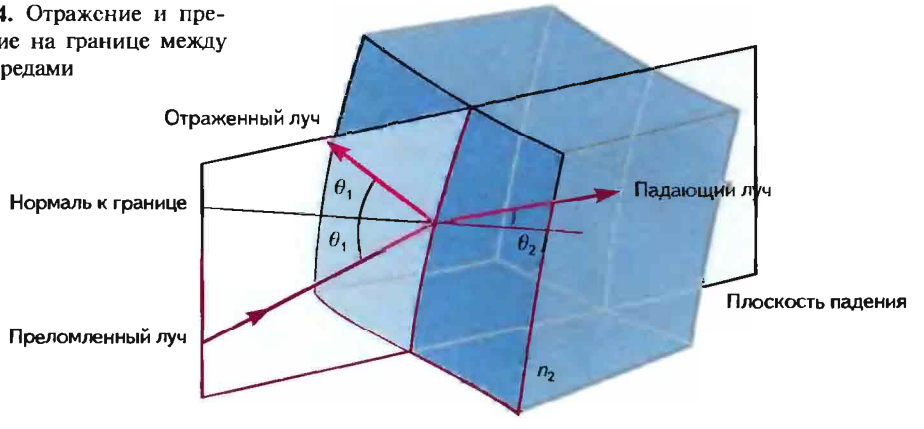


Рис. 1.4. Отражение и преломление на границе между двумя средами

Упражнение 1.1

Доказательство закона Снелла

Доказательство закона Снелла — упражнение на использование принципа Ферма. Для лучей, показанных на рис. 1.5, найдем минимум оптической длины пути $n_1\overline{AB} + n_2\overline{BC}$ между точками A и C. Таким образом, имеем задачу оптимизации: минимизировать $n_1 d_1 \sec\theta_1 + n_2 d_2 \sec\theta_2$ по отношению к углам θ_1 и θ_2 при условии, что $d_1 \operatorname{tg}\theta_1 + d_2 \operatorname{tg}\theta_2 = d$. Покажите, что решение этой задачи минимизации с ограничением дает закон Снелла.

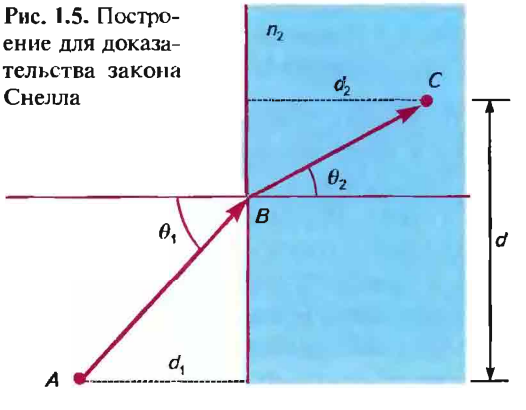


Рис. 1.5. Построение для доказательства закона Снелла

Три простых правила — распространение по прямой и законы отражения и преломления — применяются в разд. 1.2 к нескольким геометрическим конфигурациям зеркал и прозрачных оптических элементов без необходимости возвращаться к принципу Ферма.

1.2. ПРОСТЫЕ ОПТИЧЕСКИЕ ЭЛЕМЕНТЫ

1.2.1. Зеркала

Плоские зеркала

Плоское зеркало отражает лучи, выходящие из точки P_1 так, что отраженные лучи выглядят выходящими из точки P_2 за зеркалом, называемой изображением (рис. 1.6).

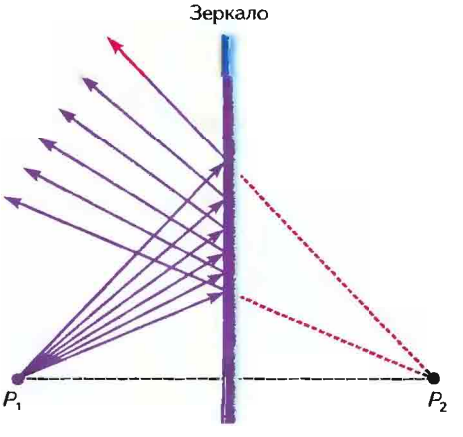


Рис. 1.6. Отражение света от плоского зеркала

Параболические зеркала

Поверхность параболического зеркала — параболоид вращения. Он имеет полезное свойство фокусировать все лучи, параллельные его оси, в одну точку, называемую фокусом. Расстояние $\overline{PF} = f$, показанное на рис. 1.7, называется фокусным расстоянием. Параболические зеркала часто используются в качестве элементов, собирающих свет в телескопах. Они также служат для получения параллельных пучков света от источников, таких как лампа-вспышка.

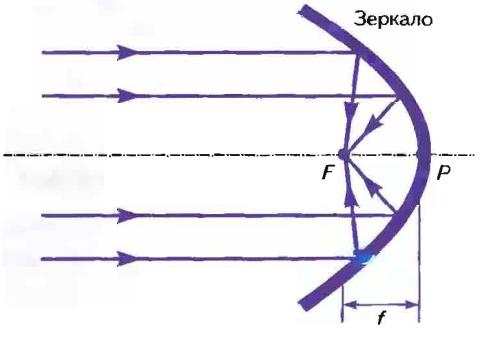


Рис. 1.7. Фокусировка света параболическим зеркалом

Эллиптические зеркала

Эллиптическое зеркало отражает все лучи, выходящие из одного из двух его фокусов, например P_1, и собирает (отображает) их в другой фокус P_2 (рис. 1.8). В соответствии с принципом Герона расстояния, проходимые светом между точками P_1 и P_2, одинаковы.

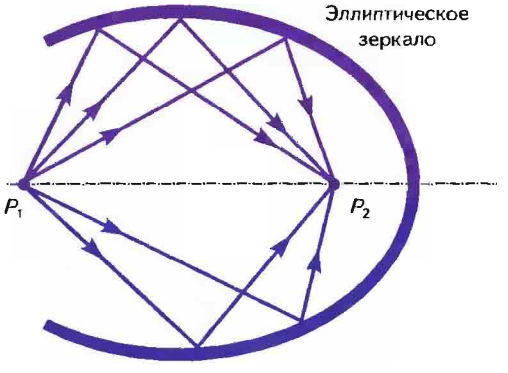


Рис. 1.8. Отражение от эллиптического зеркала

Сферические зеркала

Сферическое зеркало легче изготовить, чем параболическое и эллиптическое. Однако оно не обладает ни фокусирующим свойством параболического зеркала, ни отображающим свойством эллиптического зеркала. Как показано на рис. 1.9, параллельные лучи после отражения пересекают ось в различных точках, их огибающая (штриховая линия) называется каустической кривой. Однако параллельные лучи, близкие к оси, приблизительно фокусируются в одну точку F на расстоянии $(-R)/2$ от центра зеркала C. По соглашению, R считается отрицательным для вогнутого зеркала и положительным для выпуклого.

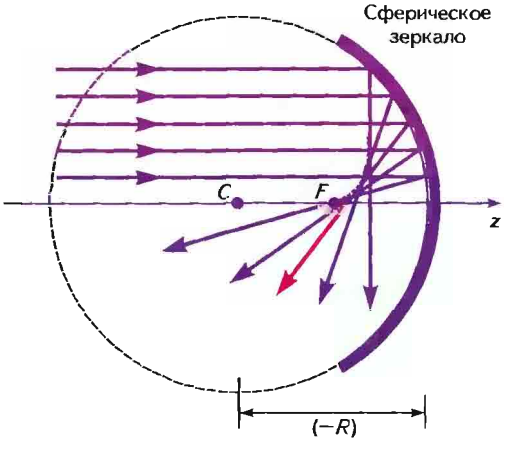


Рис. 1.9. Отражение параллельных лучей от вогнутого сферического зеркала

Параксиальные лучи, отраженные от сферических зеркал

Лучи, образующие малые углы (такие, что $\sin\theta \approx \theta$) с осью зеркала, называются **параксиальными лучами**. В **параксиальном приближении**, где рассматриваются только параксиальные лучи, сферическое зеркало обладает фокусирующим свойством, как параболическое, и отображающим свойством, как эл-

линтическое. Совокупность правил, вытекающих из этого приближения, образует **параксиальную оптику**, называемую также оптикой первого порядка, или гауссовой оптикой.

Сферическое зеркало радиусом R, таким образом, ведет себя как параболическое зеркало с фокусным расстоянием $f = R/2$. В самом деле понятно, так как в точках, близких к оси, параболу можно аппроксимировать окружностью, радиус которой равен радиусу кривизны параболы (рис. 1.10).

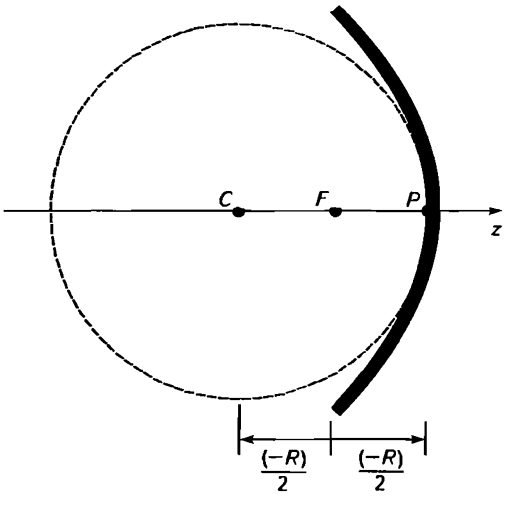


Рис. 1.10. Сферическое зеркало аппроксимирует параболическое зеркало для параксиальных лучей

Все параксиальные лучи, выходящие из любой точки, лежащей на оси сферического зеркала, отражаются и фокусируются в единственную соответственную точку оси. Это можно увидеть (рис. 1.11), рассматривая луч, испущенный под углом θ_1 из точки P_1, расположенной на расстоянии z_1 от вогнутого зеркала радиуса R, и отражающийся под углом $(-\theta_2)$, чтобы затем пересечь ось в точке P_2, лежащей на расстоянии z_2 от зеркала. Угол θ_2 отрицательный, поскольку луч отклоняется вниз. Поскольку сумма углов треугольника равна 180°, имеем $\theta_1 = \theta_0 - \theta$ и $(-\theta_2) = \theta_0 + \theta$, так что $(-\theta_2) + \theta_1 = 2\theta_0$. Если угол θ_0 достаточно мал, справедливо приближенное равенство $\operatorname{tg}\theta_0 \approx \theta_0$, так что $\theta_0 \approx y/(-R)$, откуда

$$(-\theta_2) + \theta_1 \approx \frac{2y}{(-R)}, \tag{1.4}$$

где y — высота точки, в которой происходит отражение. Напомним, что R отрицательно, поскольку зеркало вогнутое. Аналогично, если θ_1 и θ_2 малы, то

$$\theta_1 = \frac{y}{z_1}; \quad (-\theta_2) = \frac{y}{z_2},$$

так что (1.4) дает

$$\frac{y}{z_1} + \frac{y}{z_2} \approx \frac{2y}{(-R)},$$

т. е. окончательно

$$\frac{1}{z_1} + \frac{1}{z_2} \approx \frac{2}{(-R)}. \tag{1.5}$$

Это соотношение не сдержит y (т. е. не зависит от θ_1) до тех пор, пока справедливо параксиальное приближение. Это означает, что все параксиальные лучи, начинающиеся в точке P_1, приходят в точку P_2. Расстояния z_1 и z_2 измеряются в системе координат с осью z, направленной влево. Точки с отрицательными z, следовательно, лежат правее зеркала.

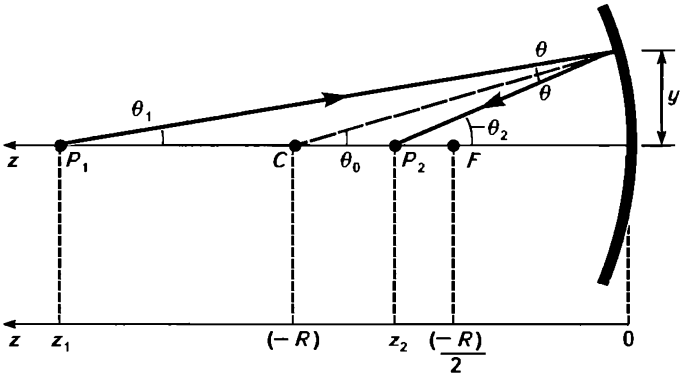


Рис. 1.11. Отражение параксиальных лучей от вогнутого сферического зеркала радиусом $R < 0$

Согласно (1.5), лучи, испущенные из точек, удаленных от зеркала на большое расстояние ($z = \infty$), фокусируются в точку F на расстоянии $z_2 = (-R)/2$. Это означает, что в рамках параксиального приближения все лучи, идущие из бесконечности (параллельные оси зеркала), фокусируются в точку на расстоянии f от зеркала, называемом фокусным расстоянием:

$$\boxed{f = \frac{(-R)}{2}.} \qquad (1.6)$$

Фокусное расстояние
Сферическое зеркало

Уравнение (1.5) обычно записывается в виде

$$\boxed{\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f},} \qquad (1.7)$$

Уравнение изображения
(параксиальное приближение)

которое известно как уравнение изображения. Для справедливости этого уравнения как падающий, так и отраженный лучи должны быть параксиальными.

Упражнение 1.2

Формирование изображения сферическим зеркалом

Покажите, что в рамках параксиального приближения лучи, исходящие из точки $P_1 = (x_y, y_z)$, отражаются в точку $P_2 = (x_y, y_z)$, где z_1 и z_2 удовлетворяют (1.7) и $y_2 = -y_1 z_2 / z_1$ (рис. 1.12). Это означает, что лучи из каждой точки плоско-

сти $z = z_1$ пересекаются в единственной соответствующей точке на плоскости $z = z_2$, так что зеркало ведет себя как система, формирующая изображение с увеличением $-z_2/z_1$. Отрицательное увеличение означает, что изображение перевернутое.

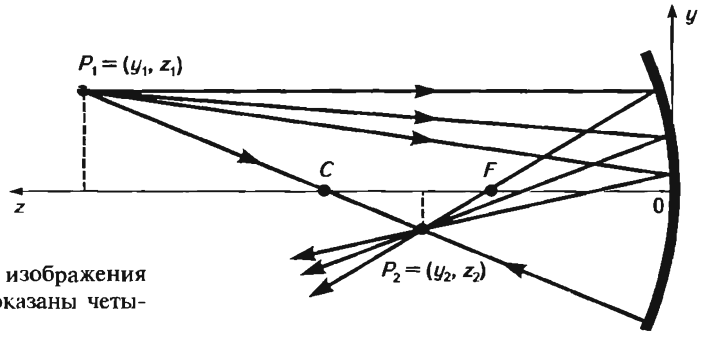
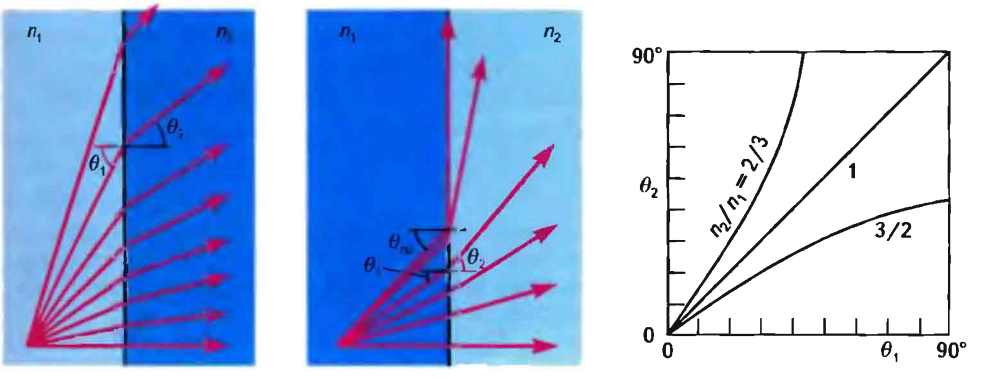


Рис. 1.12. Формирование изображения сферическим зеркалом. Показаны четыре отдельных луча

1.2.2. Плоские границы

Связь между углом падения θ_1 и углом отражения θ_2 на плоской границе раздела между двумя средами с показателями преломления n_1 и n_2 определяется законом Снелла (1.3). Эта связь показа на рис. 1.13 для двух случаев.

- *Внешнее преломление* ($n_1 < n_2$). Когда луч падает из среды с меньшим показателем преломления, $\theta_2 < \theta_1$, и преломленный луч отклоняется в направлении от границы.
- *Внутреннее преломление* ($n_1 > n_2$). Если падающий луч находится в среде с большим показателем преломления, $\theta_2 > \theta_1$, и преломленный луч отклоняется в сторону границы.



Внешнее преломление Внутреннее преломление

Рис. 1.13. Связь между углом преломления и углом падения

Преломленные лучи отклоняются таким образом, чтобы минимизировать оптическую длину пути, т. е. увеличить расстояние, проходимое в среде с более

низким показателем преломления за счет уменьшения расстояния, проходимого в среде с большим показателем преломления. В обоих случаях, когда углы малы (т. е. лучи параксиальны), соотношение между углами θ_1 и θ_2 приблизительно линейно, $n_1\theta_1 \approx n_2\theta_2$, или $\theta_2 \approx (n_1/n_2)\theta_1$.

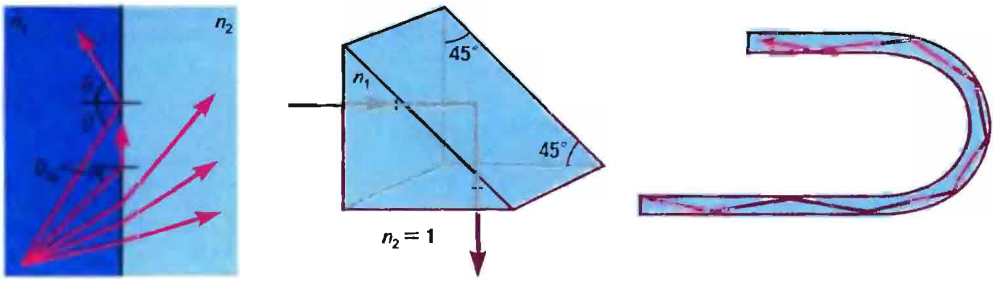
Полное внутреннее отражение

Для внутреннего преломления ($n_1 > n_2$) угол преломления больше угла падения, $\theta_2 > \theta_1$, так что с ростом θ_1 угол θ_2 раньше достигает 90° (см. рис. 1.13). Это происходит при $\theta_1 = \theta_{пр}$ (**предельный угол**), при этом $n_1 \sin\theta_{пр} = n_2 \sin(\pi/2) = n_2$, так что

$$\theta_{пр} = \sin^{-1}\frac{n_2}{n_1}. \qquad (1.8)$$

Предельный угол

Когда $\theta_1 > \theta_{пр}$, закон Снелла (1.3) не выполняется, и преломление не происходит. Падающий луч полностью отражается, как если бы поверхность раздела была идеальным зеркалом (рис. 1.14, *а*). На основе явления полного внутреннего отражения создано много оптических устройств и систем, таких как отражающие призмы (рис 1.14, *б*) и оптические волокна (см. подразд. 1.2.4). С использованием электромагнитной оптики можно показать (формулы Френеля в гл. 6), что вся энергия передается отраженному свету, поэтому процесс полного внутреннего отражения высокоэффективен.



а *б* *в*

Рис. 1.14. Полное внутреннее отражение на плоской границе (*а*). Отражательная призма (*б*). Если $n_1 > \sqrt{2}$ и $n_2 = 1$ (воздух), то $\theta_{пр} < 45°$; поскольку $\theta_i = 45°$, луч полностью отражается. Лучи направляются за счет полного внутреннего отражения от внутренней поверхности оптического волокна (*в*)

Призмы

Призма с показателем преломления n и углом α при вершине (рис. 1.15) отклоняет луч, падающий под углом θ на угол

$$\theta_d = \theta - \alpha + \arcsin\left[\sqrt{n^2 + \sin^2\theta}\,\sin\alpha - \sin\theta\cos\alpha\right]. \qquad (1.9)$$

Это можно показать, дважды используя закон Снелла на двух преломляющих поверхностях призмы. Если угол α очень мал (тонкая призма) и угол θ также очень мал (параксиальное приближение), то (1.9) приближенно дает

$$\theta_d \approx (n-1)\alpha. \qquad (1.10)$$

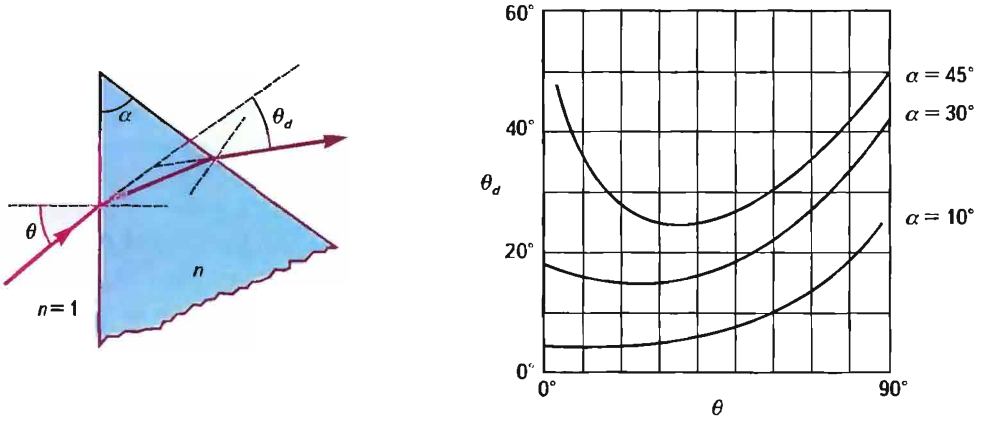


Рис. 1.15. Отклонение луча призмой. Угол отклонения θ_d показан как функция угла падения θ при различных фиксированных углах при вершине призмы α для $n = 1{,}5$. Когда α и θ малы, $\theta_d \approx (n-1)\alpha$ и приближенно не зависит от θ. Когда $\alpha = 45°$ либо $\theta = 0°$, происходит полное внутреннее отражение, как показано на рис. 1.14, *б*

Светоделители

Светоделитель — это оптическое устройство, которое расщепляет падающий пучок на отраженный и проходящий пучки, как показано на рис. 1.16. Светоделители также часто используются для сведения двух пучков света в один пучок (рис. 1.16, *в*). Часто светоделители изготавливаются путем нанесения тонкой полупрозрачной металлической или диэлектрической пленки на стеклянную подложку. Тонкая стеклянная пластинка или призма также могут служить светоделителями.

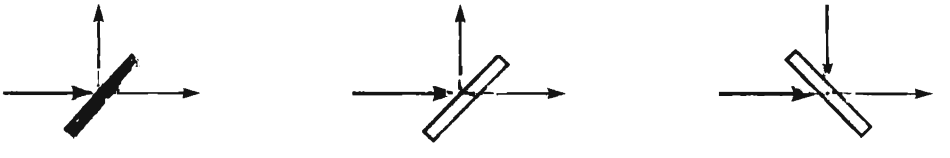


Рис. 1.16. Устройства для разделения и сведения пучков

1.2.3. Сферические границы и линзы

Теперь изучим преломление лучей на сферической границе радиуса R между двумя средами с показателями преломления n_1 и n_2. По соглашению R считается положительным для выпуклой и отрицательным для вогнутой границы. Применим закон Снелла, связывающий углы падения и преломления относительно нормали к поверхности, определяемой радиусом-вектором, проведенным в точку преломления из центра сферы C. Эти углы следует отличать от углов θ_1 и θ_2, определяемых относительно оси z. Если рассматривать только параксиальные лучи, образующие малые углы с осью системы, такие что $\sin\theta \approx \theta$ и $\operatorname{tg}\theta \approx \theta$, можно обнаружить следующие свойства:

• луч, образующий угол θ_1 с осью z и пересекающий границу на высоте y, где он образует угол θ_0 с радиусом-вектором (рис. 1.17, *а*), изменяет направление таким образом, что преломленный луч образует угол θ_2 с осью z и угол θ_3 с радиусом-вектором. Угол падения, таким образом, равен $\theta_1 + \theta_2$, в то время как угол преломления равен θ_3, так что

$$\theta_2 \approx \frac{n_1}{n_2}\theta_1 - \frac{n_2 - n_1}{n_2}\frac{y}{R}; \tag{1.11}$$

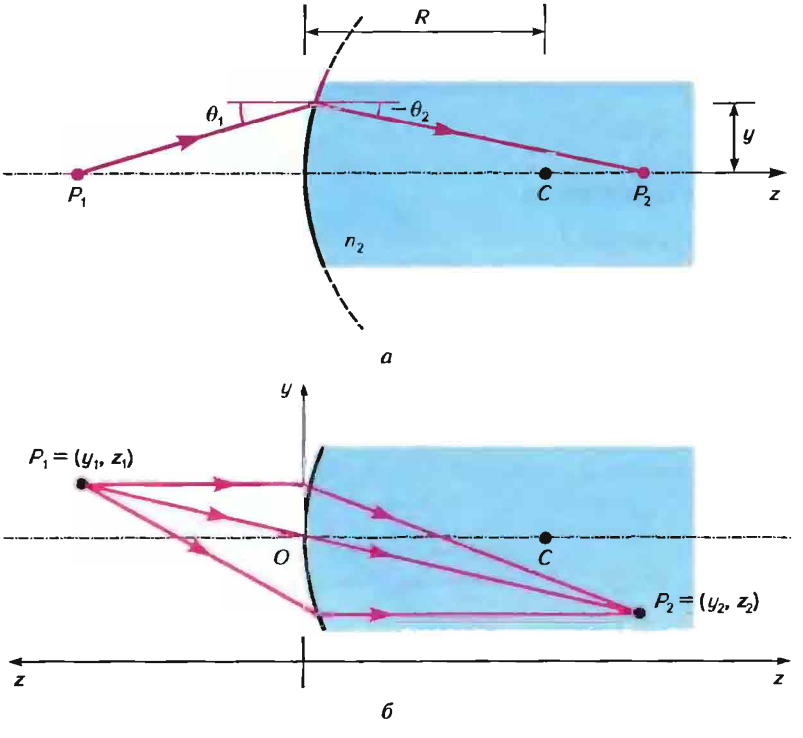


Рис. 1.17. Преломление на выпуклой сферической границе ($R > 0$)

• все параксиальные лучи, исходящие из точки $P_1 = (y_1, z_1)$ в плоскости $z = z_1$, пересекаются в точке $P_2 = (y_2, z_2)$ в плоскости $z = z_2$, где

$$\frac{n_1}{z_1} + \frac{n_2}{z_2} \approx \frac{n_2 - n_1}{R}; \tag{1.12}$$

$$y_2 = -\frac{n_1}{n_2}\frac{z_2}{z_1} y_1. \tag{1.13}$$

Плоскости $z = z_1$ и $z = z_2$ называются сопряженными плоскостями. Каждая точка первой плоскости отображается в соответствующую точку (изображение) второй плоскости с увеличением $-(n_1/n_2)(z_2/z_1)$. Как и раньше, отрицательное увеличение означает, что изображение перевернуто. По соглашению положение точки P_1 определяется в системе координат с осью, направленной влево, а P_2 — в системе координат с осью, направленной вправо (т. е. если точка P_2 лежит левее границы, то z_2 отрицательно).

Сходство перечисленных свойств со свойствами сферического зеркала очевидно. Важно помнить, что описанные свойства, позволяющие формировать изображения, являются приближенными. Они имеют место только для параксиальных лучей. Лучи, образующие большие углы, не подчиняются этим параксиальным законам; отклонение от них приводит к искажению изображения, называемому **аберрацией**.

Упражнение 1.3

Формирование изображения

Выведите (1.11). Докажите, что параксиальные лучи, исходящие из P_1, проходят через P_2, когда выполняются (1.12) и (1.13).

Упражнение 1.4

Изображающая поверхность, свободная от аберраций

Найдите уравнение выпуклой асферической (несферической) поверхности раздела между двумя средами с показателями преломления n_1 и n_2, такой, что все лучи (не обязательно параксиальные), исходящие из точки P_1, лежащей на оси на расстоянии z_1 слева от поверхности, пересекаются в точке P_2, лежащей на оси на расстоянии z_2 справа (см. рис. 1.17, *а*).

Примечание. В соответствии с принципом Ферма оптические длины пути между двумя точками равны для всех лучей.

Линзы

Сферическая линза ограничена двумя сферическими поверхностями. Таким образом, она полностью описывается радиусами поверхностей R_1 и R_2, толщиной Δ и показателем преломления n материала (рис. 1.18). Стеклянная линза в воздухе может быть рассмотрена как комбинация двух сферических границ раздела: воздух—стекло и стекло—воздух.

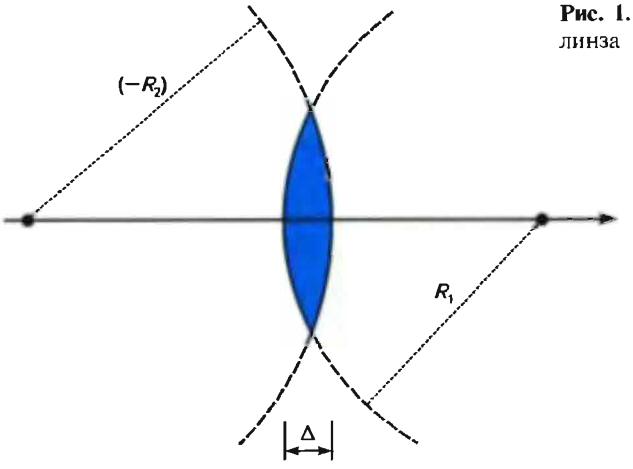


Рис. 1.18. Двояковыпуклая сферическая линза

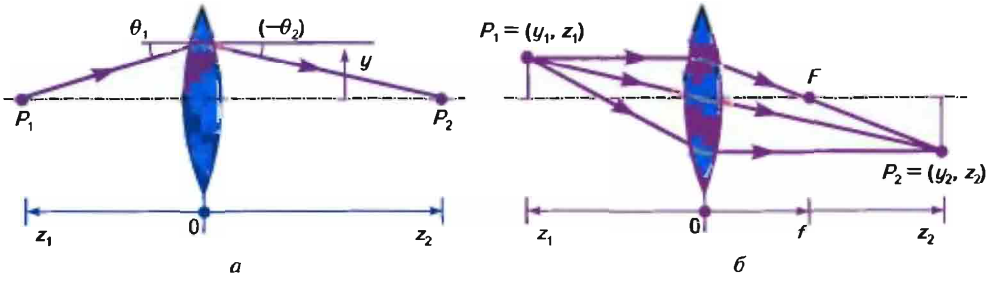


Рис. 1.19. Лучи, преломляемые тонкой линзой (*а*); формирование изображения тонкой линзой (*б*)

Рассмотрим луч, образующий с осью z угол θ_1 и пересекающий первую поверхность на высоте y (рис. 1.19, *а*). Применим (1.11) к первой поверхности, что даст угол наклона θ для преломленного луча, который мы продолжим до пересечения со второй поверхностью. Затем еще раз применим (1.11) с заменой θ_1 на θ, что даст угол наклона θ_2 луча на выходе из линзы. Результат, вообще говоря, имеет сложный вид. Однако, если линза тонкая, можно считать, что падающий луч выходит из линзы приблизительно на той же высоте y, на которой он входит в нее. При этом предположении получаем следующие соотношения:

- углы наклона преломленного и падающего лучей связаны соотношением

$$\theta_2 = \theta_1 - \frac{y}{f}, \tag{1.14}$$

где f, называемое фокусным расстоянием, определяется формулой

$$\boxed{\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right);} \tag{1.15}$$

Фокусное расстояние
Тонкая сферическая линза

• все лучи, исходящие из точки $P_1 = (y_1, z_1)$, пересекаются в точке $P_2 = (y_2, z_2)$ (рис. 1.19, *б*), где

$$\boxed{\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f};} \qquad (1.16)$$

Уравнение изображения

$$\boxed{y_2 = -\frac{z_2}{z_1} y_1.} \qquad (1.17)$$

Увеличение

Эти результаты идентичны результатам для сферического зеркала [см. (1.7) и упражнение 1.2].

Полученные уравнения отражают тот факт, что каждая точка плоскости $z = z_1$ отображается в соответственную точку плоскости $z = z_2$ с фактором увеличения $-z_2/z_1$. Увеличение равно единице при $z_1 = z_2 = 2f$. Таким образом, фокусное расстояние линзы f полностью определяет ее действие на параксиальные лучи. Как отмечалось выше, положение точек P_1 и P_2 задается в системах координат с осями, направленными влево и вправо, соответственно, а радиусы кривизны положительны для выпуклых и отрицательны для вогнутых поверхностей. Для двояковыпуклой линзы, показанной на рис. 1.18, R_1 положительно, а R_2 отрицательно, так что оба члена в (1.15) складываются и дают положительное f.

Упражнение 1.5

Доказательство формул тонкой линзы

Используя (1.11) вместе с определением фокусного расстояния (1.15), докажите (1.14) и (1.16).

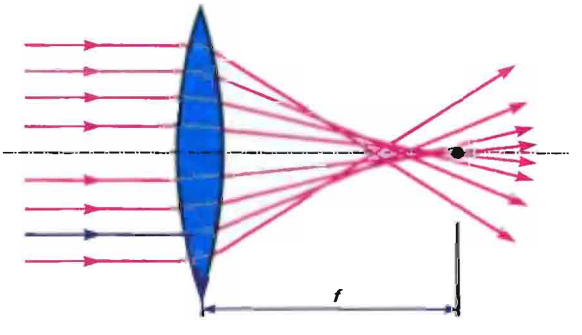


Рис. 1.20. Непараксиальные лучи не пересекаются в параксиальном фокусе

Подчеркнем, что все вытекающие отсюда выражения справедливы только для параксиальных лучей. Присутствие непараксиальных лучей приводит к аберрациям, как показано на рис. 1.20.

1.2.4. Световоды

Свет можно направлять из одного места в другое с помощью набора линз или зеркал, как схематически показано на рис. 1.21. Поскольку преломляющие элементы (такие как линзы) обычно частично отражают свет, а зеркала частично его поглощают, накапливающиеся потери оптической мощности будут значительны, когда число направляющих элементов велико. Можно изготовить элементы, в которых эти эффекты минимизированы (например, линзы с противоотражательным покрытием), но система получается сложной и дорогой.

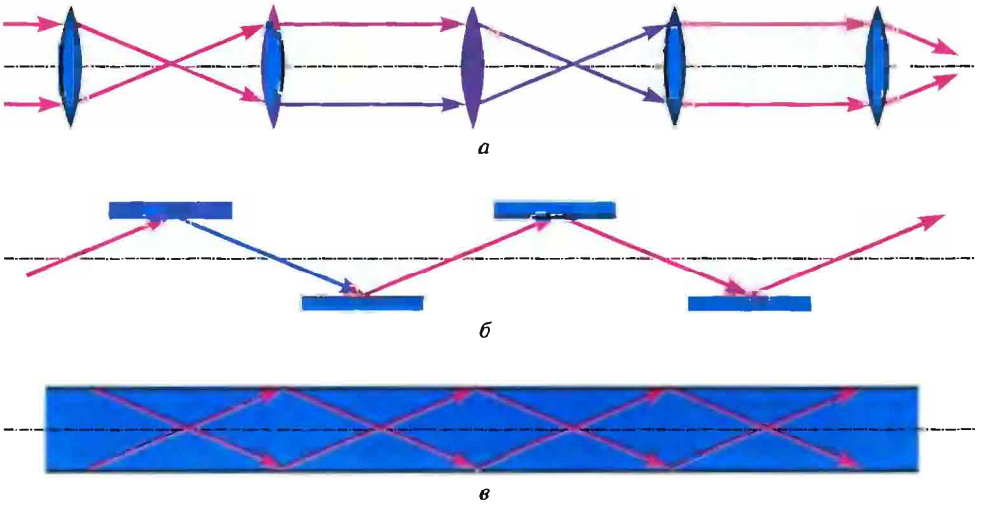


Рис. 1.21. Направление света линзами (*а*), зеркалами (*б*), с помощью полного внутреннего отражения (*в*)

Идеальным механизмом для направления света является полное внутреннее отражение на границе между двумя средами с различными показателями преломления. Лучи многократно отражаются, не испытывая преломления. Стеклянные волокна высокой химической чистоты используются для направленного распространения света на расстояния в десятки километров со сравнительно низкими потерями оптической мощности.

Оптическое волокно представляет собой «трубопровод» для света, состоящий из двух концентрических стеклянных (или пластиковых) цилиндров (рис. 1.22). Внутренний, называемый сердцевиной, имеет показатель преломления n_1, а наружный, называемый оболочкой, — несколько меньший показатель преломления $n_2 < n_1$. Лучи света, распространяющегося в сердцевине, полностью отражаются от оболочки, если угол их падения больше критического угла

$$\bar{\theta} > \theta_{\text{крит}} = \arcsin\left(\frac{n_2}{n_2}\right).$$

Следовательно, лучи, образующие угол $\theta = 90° - \bar{\theta}$ с оптической осью, не выходят из волновода, если $\theta < \bar{\theta}_{крит}$, где

$$\bar{\theta}_{крит} = 90° - \theta_{крит} = \arccos\left(\frac{n_2}{n_1}\right).$$

Оптические волокна используются в оптических системах связи (см. гл. 9 и 24). Некоторые важные свойства оптических волокон предлагается вывести в упражнении 1.6.

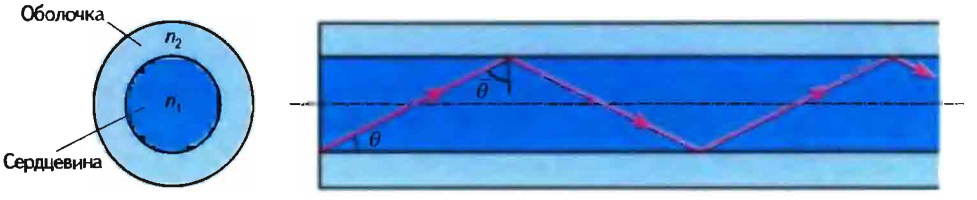


Рис. 1.22. Оптическое волокно. Лучи света направляются за счет многократного полного внутреннего отражения. Здесь θ — угол наклона луча по отношению к оси волокна, так что $\bar{\theta} = 90° - \theta$ — угол падения на границу раздела двух диэлектриков

Упражнение 1.6

Числовая апертура и угол приема оптического волокна

Оптическое волокно освещается светом от источника [например, светоизлучающего диода (СИД)]. Показатели преломления сердцевины и оболочки равны n_1 и n_2 соответственно, показатель преломления воздуха равен 1 (рис. 1.23). Покажите, что половина угла при вершине θ_a конуса лучей, принимаемых волокном и распространяющихся в нем без преломления на границе между сердцевиной и оболочкой, дается выражением

$$\boxed{\text{NA} = \sin\theta_a = \sqrt{n_1^2 - n_2^2}.} \qquad (1.18)$$

Числовая апертура
Оптическое волокно

Угол θ_a называется углом приема, а параметр $\text{NA} = \sin\theta_a$ — **числовой апертурой** волокна. Вычислите числовую апертуру и угол приема для волокна из кварцевого стекла с $n_1 = 1{,}475$ и $n_2 = 1{,}460$.

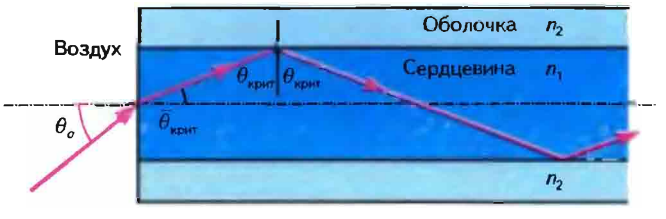


Рис. 1.23. Угол приема оптического волокна

Пленение света в средах с высоким показателем преломления

Свет, образующийся внутри среды с высоким показателем преломления, часто не может выйти наружу, особенно если поверхности среды параллельны друг другу. Это происходит потому, что некоторые лучи претерпевают многократное полное внутреннее отражение без преломления на границе с воздухом. Принцип иллюстрируется упражнением 1.7.

Упражнение 1.7

Свет, плененный в светоизлучающем диоде

1. Предположим, что свет генерируется во всех направлениях внутри материала с показателем преломления n, вырезанном в форме параллелепипеда (рис. 1.24). Материал окружен воздухом с единичным показателем преломления. Этот процесс происходит в светоизлучающем диоде (см. гл. 17). Чему равен угол конуса внутри материала, внутри которого лучи света могут выйти наружу с каждой стороны? Что происходит с другими лучами? Каково численное значение этого угла для GaAs ($n = 3{,}6$)?

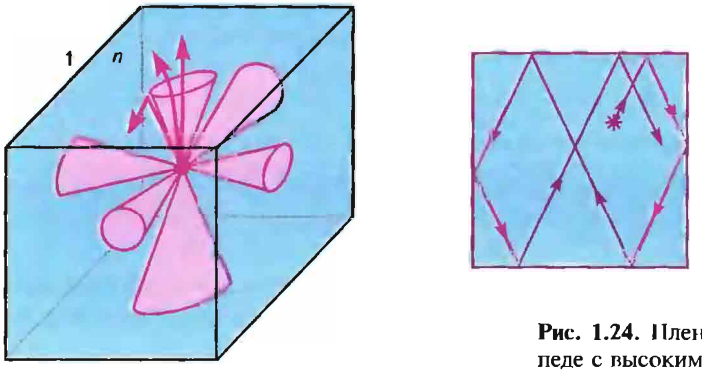


Рис. 1.24. Пленение света в параллелепипеде с высоким показателем преломления

2. Предположим, что когда свет генерируется изотропно, величина оптической мощности, связанная с лучами внутри данного конуса, пропорциональна соответствующему телесному углу. Покажите, что отношение оптической мощности, выходящей наружу к полной мощности света, генерируемого внутри материала, равно $3\left(1-\sqrt{1-1/n^2}\right)$ при условии, что $n > \sqrt{2}$. Чему равно численное значение этого отношения для GaAs?

1.3. ОПТИКА СРЕД С ГРАДИЕНТНЫМ ПОКАЗАТЕЛЕМ ПРЕЛОМЛЕНИЯ

Материал с градиентным показателем преломления (ГПП) имеет показатель преломления, который меняется от точки к точке как непрерывная функция $n(\boldsymbol{r})$. Такие материалы часто изготавливаются путем добавления при-

месей (легирующих добавок) в регулируемых концентрациях. В среде с ГПП оптические лучи следуют искривленным, а не прямолинейным траекториям. При надлежащем выборе $n(\boldsymbol{r})$ плоскопараллельная пластинка с ГПП может оказывать на лучи света такое же действие, как и обычные оптические элементы, такие как линза или призма.

1.3.1. Уравнение луча

Для определения траектории светового луча в неоднородной среде с показателем преломления $n(\boldsymbol{r})$ используем принцип Ферма

$$\delta\int_A^B n(\boldsymbol{r})\,\mathrm{d}s = 0, \tag{1.19}$$

где $\mathrm{d}s$ — бесконечно малый элемент длины вдоль траектории луча между точками A и B. Если траектория описывается функциями $x(s)$, $y(s)$ и $z(s)$, где s — длина вдоль траектории (рис. 1.25), то, используя вариационное исчисление, можно показать[1], что $x(s)$, $y(s)$ и $z(s)$ должны удовлетворять трем дифференциальным уравнениям в частных производных

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(n\frac{\mathrm{d}x}{\mathrm{d}s}\right)=\frac{\partial n}{\partial x};\quad \frac{\mathrm{d}}{\mathrm{d}s}\left(n\frac{\mathrm{d}y}{\mathrm{d}s}\right)=\frac{\partial n}{\partial y};\quad \frac{\mathrm{d}}{\mathrm{d}s}\left(n\frac{\mathrm{d}z}{\mathrm{d}s}\right)=\frac{\partial n}{\partial z}. \tag{1.20}$$

Вводя вектор $\boldsymbol{r}(s)$, компонентами которого являются $x(s)$, $y(s)$ и $z(s)$, уравнения (1.20) можно записать в компактном векторном виде

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}s}\left(n\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s}\right)=\nabla n,} \tag{1.21}$$

Уравнение луча

где ∇n — градиент n, т. е. вектор с декартовыми компонентами $\partial n/\partial x$, $\partial n/\partial y$ и $\partial n/\partial z$. Уравнение (1.21) называется **уравнением луча**.

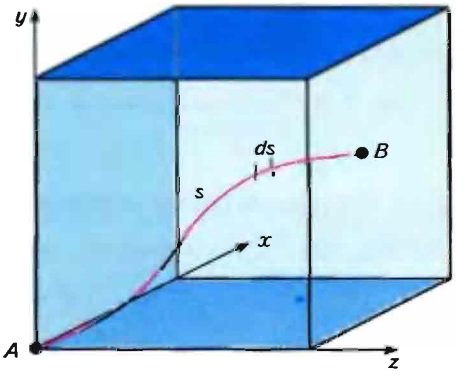


Рис. 1.25. Траектория луча описывается либо параметрически тремя функциями $x(s)$, $y(s)$ и $z(s)$, либо двумя функциями $x(z)$, $y(z)$

[1] Этот вывод выходит за рамки настоящей книги. Подробнее см., например, *R. Weinstock.* Calculus of Variations, Dover, 1974.

Один из подходов к решению уравнения луча состоит в описании траектории двумя функциями $x(z)$, $y(z)$, тогда элемент длины равен

$$ds = dz\sqrt{1+\left(\frac{dx}{dz}\right)^2+\left(\frac{dy}{dz}\right)^2}$$

Подставляя это выражение в (1.21), приходим к двум дифференциальным уравнениям в частных производных для функций $x(z)$, $y(z)$. В общем случае соответствующие вычисления достаточно сложны, однако они сильно упрощаются в параксиальном приближении.

Параксиальное уравнение луча

В параксиальном приближении траектории лучей почти параллельны оси z, так что $ds \approx dz$ (рис. 1.26). Уравнения луча (1.21) в этом случае упрощаются

$$\boxed{\frac{d}{dz}\left(n\frac{dx}{dz}\right)\approx\frac{\partial n}{\partial x};\quad \frac{d}{dz}\left(n\frac{dy}{dz}\right)\approx\frac{\partial n}{\partial y}.} \qquad (1.22)$$

Параксиальные уравнения луча

При заданном $n(x, y, z)$ эти два дифференциальных уравнения в частных производных можно решить относительно определяющих траекторию функций $x(z)$, $y(z)$.

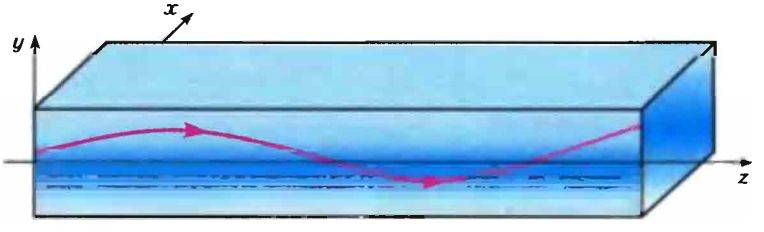


Рис. 1.26. Траектория параксиального луча в среде с градиентным показателем преломления

В предельном случае однородной среды, когда n не зависит от x, y, z, уравнения (1.22) дают

$$\frac{d^2x}{dz^2}=0;\quad \frac{d^2y}{dz^2}=0,$$

откуда следует, что x и y — линейные функции z, так что траектории лучей представляют собой прямые линии. Более интересные случаи будут рассмотрены далее.

1.3.2. Оптические элементы с градиентным показателем преломления

Слой с градиентным показателем преломления

Рассмотрим плоский слой материала, показатель преломления которого $n = n(y)$ не зависит от x и z и непрерывно меняется в направлении y (рис. 1.27). Траектории параксиальных лучей в плоскости y—z описываются параксиальным уравнением луча

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(n\frac{\mathrm{d}y}{\mathrm{d}z}\right)=\frac{\mathrm{d}n}{\mathrm{d}y}, \tag{1.23}$$

из которого следует

$$\frac{\mathrm{d}^2y}{\mathrm{d}z^2}=\frac{1}{n(y)}\frac{\mathrm{d}n(y)}{\mathrm{d}y}. \tag{1.24}$$

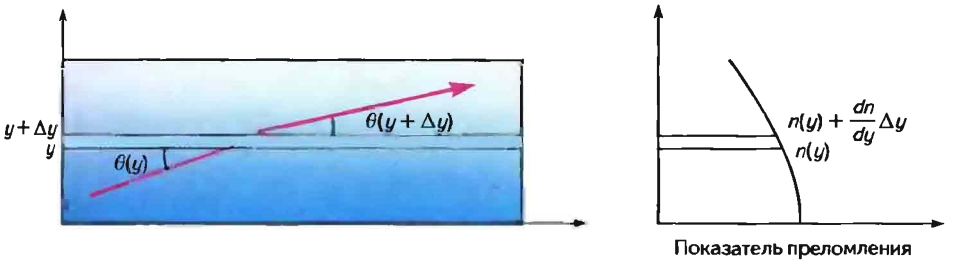


Рис. 1.27. Рефракция в слое с градиентным показателем преломления

При заданном $n(y)$ и начальных условиях (y и $\mathrm{d}y/\mathrm{d}z$ при $z = 0$) уравнение (1.24) можно решить относительно функции $y(z)$, которая описывает траекторию луча.

Вывод параксиального уравнения луча в слое с ГПП с использованием закона Снелла

Уравнение (1.24) можно вывести также непосредственно из закона Снелла (см. рис. 1.27). Пусть

$$\theta(y)\approx\frac{\mathrm{d}y}{\mathrm{d}z}$$

есть угол наклона луча относительно оси z в точке (y, z). После прохождения слоя толщиной Δy этот угол меняется и становится равным $\theta(y + \Delta y)$. Эти два угла связаны законом Снелла, где θ, как видно из рис. 1.27, есть угол, дополнительный к углу падения (преломления)

$$\begin{aligned} &n(y)\cos\theta(y)=n(y+\Delta y)\cos\theta(y+\Delta y)= \\ &=\left[n(y)+\frac{\mathrm{d}n}{\mathrm{d}y}\Delta y\right]\left[\cos\theta(y)-\frac{\mathrm{d}\theta}{\mathrm{d}y}\Delta y\sin\theta(y)\right], \end{aligned} \tag{1.25}$$

где мы применили разложение

$$f(y+\Delta y) = f(y) + \frac{df}{dy}\Delta y$$

к функциям $f(y) = n(y)$ и $f(y) = \cos\theta(y)$. В пределе $\Delta y \to 0$ после исключения члена с $(\Delta y)^2$ получаем дифференциальное уравнение

$$\frac{dn}{dy} = n\frac{d\theta}{dy}\operatorname{tg}\theta. \tag{1.26}$$

Для параксиальных лучей θ очень мало, так что $\operatorname{tg}\theta \approx \theta$. Подставляя $\theta = dy/dz$ в (1.26), получаем (1.24).

Пример 1.1

Слой с параболическим профилем показателя преломления

Важным частным случаем распределения показателя преломления в слое является

$$n^2(y) = n_0^2\left(1 - \alpha^2 y^2\right). \tag{1.27}$$

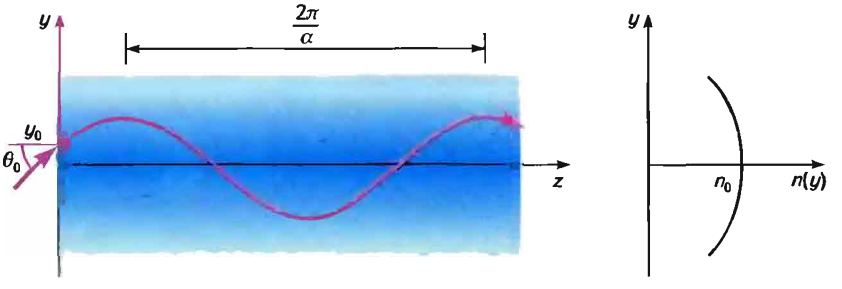


Рис. 1.28. Траектория луча в слое с ГПП параболического профиля (SELFOC)

Это симметричная функция y, имеющая максимальное значение при $y = 0$ (рис. 1.28). Стеклянный слой с таким профилем известен под коммерческим названием SELFOC. Обычно α выбирается достаточно малым, так что $\alpha^2 y^2 \ll 1$ для всех представляющих интерес значений y. При этом условии

$$n(y) = n_0\sqrt{1-\alpha^2 y^2} \approx n_0\left(1 - \frac{1}{2}\alpha^2 y^2\right),$$

т. е. $n(y)$ представляет собой параболическое распределение. Поскольку $n(y) - n_0 = n_0$, относительное изменение показателя преломления очень мало. Если вычислить производную от (1.27), правая часть (1.24) принимает вид

$$\frac{1}{n}\frac{dn}{dy} = -\left(\frac{n_0}{n}\right)^2 \alpha^2 y \approx -\alpha^2 y,$$

так что (1.24) сводится к уравнению

$$\frac{d^2y}{dz^2} \approx -\alpha^2 y. \tag{1.28}$$

Решения этого уравнения — гармонические функции с периодом $2\pi/\alpha$. Предполагая, что луч имел начальное положение $y(0) = y_0$ и начальный наклон $dy/dz = \theta_0$ при $z = 0$ в среде с ГПП, находим траекторию луча

$$y(z) = y_0 \cos \alpha z + \frac{\theta_0}{\alpha} \sin \alpha z, \tag{1.29}$$

для которой наклон ведет себя как

$$\theta(z) = \frac{dy}{dz} = -y_0 \alpha \sin \alpha z + \theta_0 \cos \alpha z. \tag{1.30}$$

Луч колеблется вокруг центра слоя с периодом по координате $2\pi/\alpha$, который называется **шагом**, как показано на рис. 1.28.

Максимальное отклонение луча

$$y_{max} = \sqrt{y_0^2 + \left(\frac{\theta_0}{\alpha}\right)^2}$$

а максимальный угол наклона луча $\theta_{max} = \alpha y_{max}$. Справедливость данного приближенного анализа гарантирована, если $\theta_{max} \ll 1$. Если $2y_{max}$ меньше, чем толщина слоя, то луч остается внутри слоя, который служит световодом. На рис. 1.29 показаны траектории множества лучей, передаваемых по слою SELFOC. Обратите внимание, что все лучи имеют один и тот же шаг. Такой слой с ГПП можно использовать как линзу, что демонстрируется в упражнении 1.8.

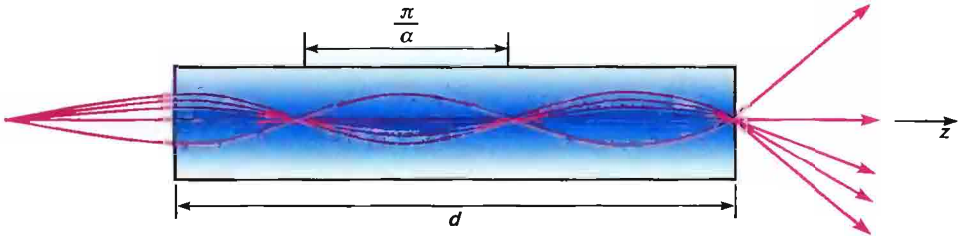


Рис. 1.29. Траектории лучей от внешнего точечного источника в слое SELFOC

Упражнение 1.8

Слой с ГПП как линза

Покажите, что слой SELFOC длиной $d < \pi/2\alpha$ с показателем преломления (1.27) ведет себя как цилиндрическая линза (линза, фокусирующая лучи в плоскости y—z) с фокусным расстоянием

$$f \approx \frac{1}{n_0 d\alpha \sin\alpha}. \tag{1.31}$$

Покажите, что главная точка (ее определение видно из рис. 1.30) лежит на расстоянии

$$\overline{AH} \approx \frac{1}{n_0\alpha} \operatorname{tg} \frac{\alpha d}{2}$$

от торца слоя. Нарисуйте траектории лучей в частных случаях $d = \pi/\alpha$ и $d = \pi/2\alpha$.

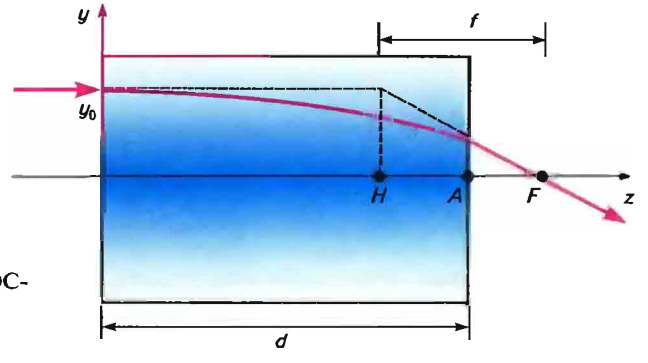


Рис. 1.30. Использование SELFOC-слоя в качестве линзы:
F — фокус; H — главная точка

Волокна с градиентным показателем преломления

Волокно с градиентным показателем преломления представляет собой стеклянный цилиндр, показатель преломления которого непрерывно зависит от расстояния до его оси. В параксиальном приближении траектории лучей определяются параксиальным уравнением лучей (1.22). Рассмотрим, например, распределение показателя преломления

$$n^2 = n_0^2\left[1 - \alpha^2\left(x^2 + y^2\right)\right]. \tag{1.32}$$

Подставляя (1.32) в (1.22) и предполагая, что $\alpha^2(x^2 + y^2) \ll 1$, для всех представляющих интерес x и y, получим

$$\frac{\mathrm{d}^2x}{\mathrm{d}z^2} \approx -\alpha^2 x; \quad \frac{\mathrm{d}^2y}{\mathrm{d}z^2} \approx -\alpha^2 y. \tag{1.33}$$

Таким образом, x и y являются гармоническими функциями z с периодом $2\pi/\alpha$. Начальное положение (x_0, y_0) и наклон луча

$$\theta_{x_0} = \frac{dx}{dz}; \quad \theta_{y_0} = \frac{dy}{dz}$$

при $z = 0$ определяют амплитуды и фазы этих гармонических функций. Ввиду осевой симметрии можно без ограничения общности положить $x_0 = 0$. Тогда решение уравнений (1.33) имеет вид

$$x(z) = \frac{\theta_{x_0}}{\alpha} \sin \alpha z;$$
$$y(z) = \frac{\theta_{y_0}}{\alpha} \sin \alpha z + y_0 \cos \alpha z. \tag{1.34}$$

Если $\theta_{x_0} = 0$, т. е. падающий луч лежит в меридиональной плоскости (плоскости, проходящей через ось цилиндра, в данном случае это плоскость y—z), то луч останется в этой плоскости и будет распространяться по синусоидальной траектории, так же как в слое с ГПП параболического профиля (рис. 1.31, *а*).

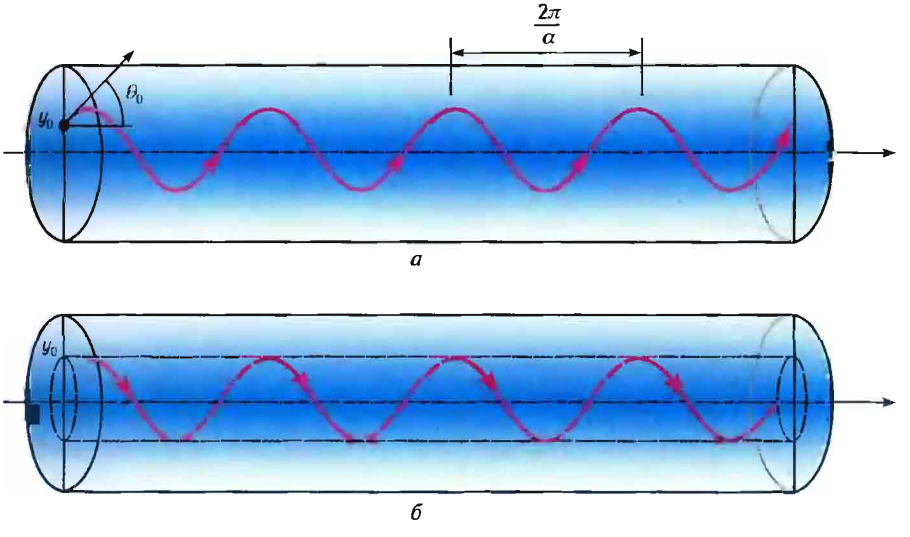


Рис. 1.31. Меридиональный (*а*) и винтовой (*б*) лучи в волокне с градиентным показателем преломления параболического профиля

С другой стороны, если $\theta_{y_0} = 0$, а $\theta_{x_0} = \alpha y_0$, то

$$x(z) = y_0 \sin \alpha z;$$
$$y(z) = y_0 \cos \alpha z, \tag{1.35}$$

так что луч распространяется по винтовой линии, лежащей на поверхности цилиндра радиуса y_0 (рис. 1.31, *б*). В обоих случаях луч остается внутри волокна, которое служит световодом. Для разных падающих лучей получаются различные винтовые траектории.

Волокна с градиентным показателем преломления и их использование в оптических линиях связи обсуждаются в гл. 9 и 24.

Упражнение 1.9

Числовая апертура волокна с градиентным показателем преломления

Рассмотрим волокно радиусом a с градиентным показателем преломления, профиль которого задан формулой (1.32). Луч падает из воздуха в центр торца волокна, образуя внутри него угол θ_0 с его осью (рис. 1.32). В параксиальном приближении покажите, что числовая апертура равна

$$\boxed{\mathrm{NA} \equiv \sin\theta_a \approx n_0 a\alpha,} \qquad (1.36)$$

Числовая апертура
Волокно с градиентным показателем преломления

где θ_a — максимальный угол приема, для которого траектория луча остается внутри волокна. Сравните результат с числовой апертурой волокна со ступенчатым показателем преломления, рассмотренным в упражнении 1.6. Для правильного сравнения возьмите показатели преломления сердцевины и оболочки равными, соответственно, $n_1 = n_0$ и

$$n_2 = n_0\sqrt{1-\alpha^2 a^2} \approx n_0\left(1-\frac{1}{2}\alpha^2 a^2\right).$$

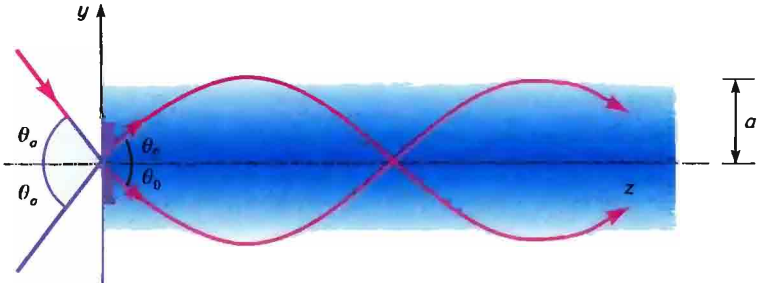


Рис. 1.32. Угол приема оптического волокна с градиентным показателем преломления

*1.3.3. Уравнение эйконала

Траектории лучей часто характеризуются поверхностями, по отношению к которым они нормальны. Пусть $S(\boldsymbol{r})$ — скалярная функция, такая, что поверхности ее равных значений $S(\boldsymbol{r})$ = const всюду нормальны к лучам

(рис. 1.33). Если функция $S(\boldsymbol{r})$ известна, то траектории лучей можно легко построить, поскольку нормалью к поверхности равных значений функции в точке $\boldsymbol{r}$ является направление градиента $\nabla S(\boldsymbol{r})$. Функция $S(\boldsymbol{r})$, называемая эйконалом, аналогична потенциальной функции $V(\boldsymbol{r})$ в электростатике; роль оптических лучей в этом случае играют силовые линии электрического поля $\boldsymbol{E} = -\nabla V$.

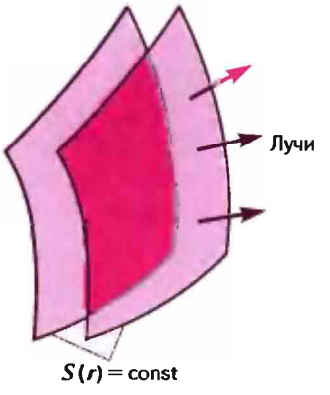


Рис. 1.33. Траектории лучей нормальны к поверхностям равных значений $S(\boldsymbol{r})$

Чтобы удовлетворить принципу Ферма (который является главным постулатом лучевой оптики), эйконал $S(\boldsymbol{r})$ должен удовлетворять дифференциальному уравнению в частных производных, известному как **уравнение эйконала**:

$$\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 + \left(\frac{\partial S}{\partial z}\right)^2 = n^2, \tag{1.37}$$

которое обычно записывают в векторной форме

$$\boxed{|\nabla S|^2 = n^2,} \tag{1.38}$$

Уравнение эйконала

где $|\nabla S|^2 = \nabla S \cdot \nabla S$. Вывод уравнения эйконала из принципа Ферма — математическое упражнение, выходящее за рамки настоящей книги[2]. Можно также показать, что принцип Ферма (и уравнение луча) следует из уравнения эйконала. Поэтому в качестве основного постулата лучевой оптики можно брать либо принцип Ферма, либо уравнение эйконала.

Интегрируя уравнение эйконала (1.38) вдоль траектории луча между точками A и B, получим

$$S(\boldsymbol{r}_B) - S(\boldsymbol{r}_A) = \int_A^B |\nabla S|\,\mathrm{d}s = \int_A^B n\,\mathrm{d}s = \text{оптическая длина пути между } A \text{ и } B. \tag{1.39}$$

Это означает, что разность $S(\boldsymbol{r}_B) - S(\boldsymbol{r}_A)$ представляет собой оптическую длину пути между A и B. В электростатическом аналоге роль оптической длины пути играет разность потенциалов.

---

[2] См., например, *Борн М., Вольф Е.* Основы оптики. М.: Наука, 1973.

Чтобы определить траектории лучей в неоднородной среде с показателем преломления $n(\boldsymbol{r})$, мы можем либо решить уравнение луча (1.21), как мы делали выше, либо решить уравнение эйконала для $S(\boldsymbol{r})$, а затем вычислить градиент ∇S.

Если среда однородна, т. е. $n(\boldsymbol{r})$ постоянно, величина ∇S также постоянна, так что нормали к волновому фронту (лучи) должны быть прямыми линиями. Поверхности $S(\boldsymbol{r}) = \text{const}$ могут быть параллельными плоскостями или концентрическими сферами, как показано на рис. 1.34.

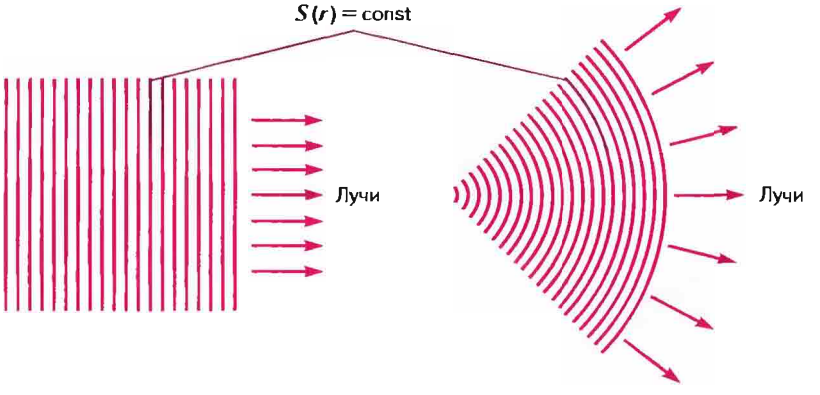


Рис. 1.34. Лучи и поверхности постоянного $S(\boldsymbol{r})$ в однородной среде

Уравнение эйконала вновь рассматривается с точки зрения соотношения между лучевой и волновой оптикой в разд. 2.3.

1.4. МАТРИЧНАЯ ОПТИКА

Матричная оптика представляет собой метод для расчета хода параксиальных лучей. Предполагается, что лучи распространяются в одной плоскости, поэтому данный формализм применим к системам с планарной геометрией и к меридиональным лучам в системах с аксиальной симметрией.

Луч описывается положением и углом наклона по отношению к оптической оси. Эти переменные меняются по мере распространения луча через оптическую систему. В параксиальном приближении положение и угол наклона луча в плоскости входа и выхода связаны двумя *линейными* алгебраическими уравнениями. Таким образом, оптическая система описывается матрицей 2×2, которая называется матрицей передачи луча.

Удобство матричных методов состоит в том, что матрица передачи луча для каскада элементов (или систем) есть произведение матриц передачи луча отдельных элементов (или систем). Таким образом, матричная оптика обеспечивает удобный формализм для описания оптических систем в параксиальном приближении.

1.4.1. Матрица передачи луча

Рассмотрим аксиально-симметричную оптическую систему, образованную последовательностью преломляющих и отражающих поверхностей, центрированных на одной оси (оптической оси). Выберем ось z вдоль оптической оси в направлении общего хода лучей. Рассмотрим лучи, лежащие в плоскости, проходящей через оптическую ось, например, в плоскости y—z. Будем следить за ходом луча через систему, т. е. за тем, как он пересекает каждую поверхность на том или ином расстоянии от оси. Луч, пересекающий поперечную плоскость при данном z, полностью характеризуется координатой пересечения y и углом наклона θ (рис. 1.35).

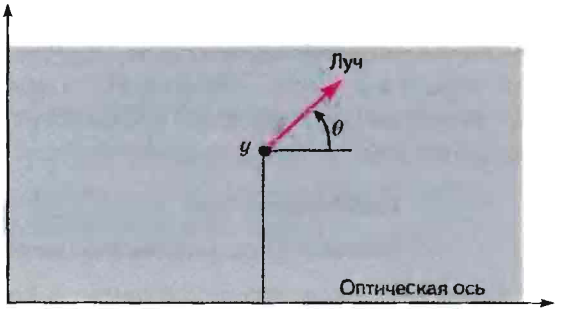


Рис. 1.35. Луч характеризуется своей координатой y и углом наклона θ

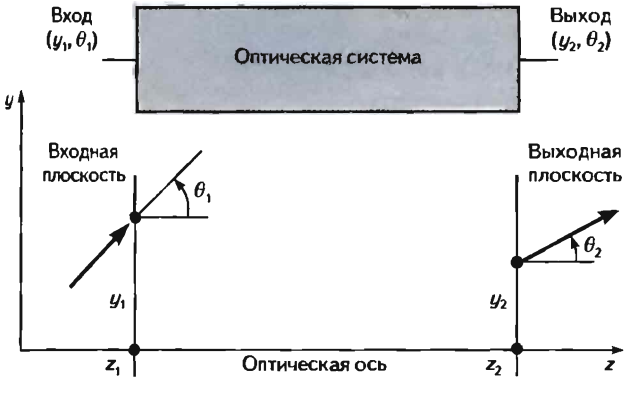


Рис. 1.36. Луч входит в оптическую систему в точке z_1 на расстоянии y_1 от оси и с углом наклона θ_1 и выходит в точке z_2 на расстоянии y_2 от оси и с углом наклона θ_2

Оптическая система представляет собой совокупность оптических элементов, расположенных между двумя поперечными плоскостями при z_1 и z_2, называемых входной и выходной плоскостью соответственно. Система полностью характеризуется ее действием на входящий луч с произвольным положением и наклоном (y_1, θ_1). Система направляет луч таким образом, что на выходе он имеет новое положение и наклон (y_2, θ_2) в выходной плоскости (рис. 1.36).

В параксиальном приближении, когда все углы достаточно малы, так что $\sin\theta \approx \theta$, соотношение между (y_2, θ_2) и (y_1, θ_1) линейно и может быть записано в общем виде

$$y_2 = Ay_1 + B\theta_1; \tag{1.40}$$

$$\theta_2 = Cy_1 + D\theta_1, \tag{1.41}$$

где A, B, C и D — действительные числа. Уравнения (1.40) и (1.41) удобно записать в матричном виде

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix}. \tag{1.42}$$

Матрица **M** с элементами A, B, C и D полностью характеризует оптическую систему, поскольку она позволяет выразить (y_2, θ_2) для любых (y_1, θ_1). Она называется **матрицей передачи луча**. Как будет видно на примерах, приведенных в подразд. 1.4.2, углы, получающиеся отрицательными, указывают на наклон луча вниз по отношению к оси z. Отрицательные радиусы соответствуют вогнутым поверхностям, а положительные радиусы — выпуклым.

Упражнение 1.10

Частные виды матриц передачи луча

Рассмотрите следующие ситуации, в которых один из элементов матрицы передачи луча исчезает.

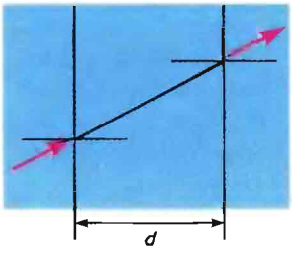
1. Покажите, что если $A = 0$, то все лучи, входящие в систему под одним и тем же углом, выходят из нее в одной и той же точке, т. е. параллельные лучи на входе фокусируются в одну точку на выходе.

2. Каковы особые свойства систем, у которых $B = 0$, $C = 0$ или $D = 0$?

1.4.2. Матрицы простых оптических элементов

Распространение в свободном пространстве

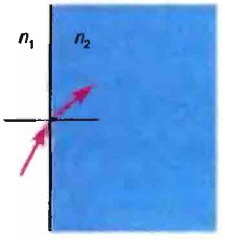
Поскольку лучи распространяются вдоль прямых линий в средах с однородным показателем преломления, таких как свободное пространство, луч, прошедший расстояние d, преобразуется в соответствии с формулами $y_2 = y_1 + \theta_1 d$ и $\theta_2 = \theta_1$. Матрица передачи луча, соответственно, имеет вид



$$\mathbf{M} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}. \tag{1.43}$$

Преломление на плоской границе

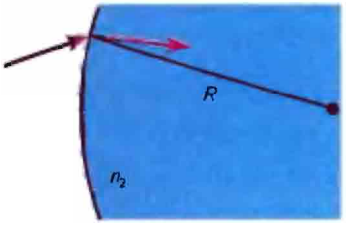
На плоской границе раздела двух сред с показателями преломления n_1 и n_2 угол наклона луча изменяется в соответствии с законом Снелла $n_1 \sin\theta_1 = n_2 \sin\theta_2$. В параксиальном приближении $n_1\theta_1 \approx n_2\theta_2$. Положение луча не меняется, $y_1 = y_2$. Матрица передачи луча имеет вид



$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{n_1}{n_2} \end{bmatrix}. \qquad (1.44)$$

Преломление на сферической границе

Связь между углами θ_1 и θ_2 для параксиальных лучей, преломляющихся на сферической границе раздела двух сред, дается соотношением (1.11). Высота луча не меняется, $y_1 \approx y_2$. Матрица передачи луча есть

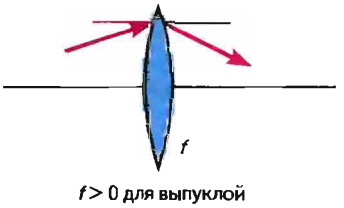


$R > 0$ для выпуклой
и $R < 0$ для вогнутой поверхности

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -\dfrac{(n_2 - n_1)}{n_2 R} & \dfrac{n_1}{n_2} \end{bmatrix}. \qquad (1.45)$$

Прохождение через тонкую линзу

Связь между θ_1 и θ_2 для параксиальных лучей, проходящих через тонкую линзу с фокусным расстоянием f, дается соотношением (1.14). Поскольку высота луча остается неизменной ($y_1 = y_2$), имеем



$f > 0$ для выпуклой
и $f < 0$ для вогнутой поверхности

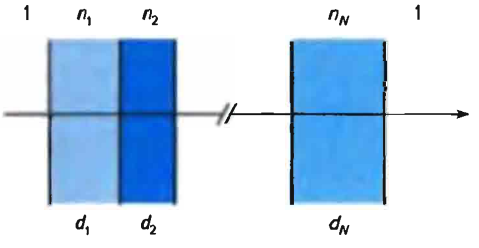
$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -\dfrac{1}{f} & 1 \end{bmatrix}. \qquad (1.46)$$

Обратите внимание на порядок перемножения матриц: справа стоит матрица первого элемента по ходу луча, которая первой преобразует вектор-столбец падающего луча, и так далее. Произведение матриц в общем случае не коммутативно, хотя и ассоциативно.

Упражнение 1.11

Набор параллельных прозрачных пластин

Рассмотрите набор из N параллельных плоских прозрачных пластин с показателями преломления n_1, n_2, ..., n_N и толщинами d_1, d_2, ..., d_N, расположенных в воздухе ($n = 1$) перпендикулярно оси z. Используя метод полной математической индукции, докажите, что матрица передачи луча системы есть



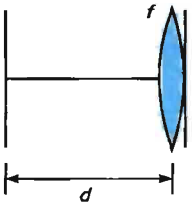
$$\mathbf{M} = \begin{bmatrix} 1 & \sum_{i=1}^{N} \frac{d_i}{n_i} \\ 0 & 1 \end{bmatrix}. \qquad (1.50)$$

Заметьте, что порядок следования пластин не влияет на полную матрицу передачи луча. Какова будет матрица передачи луча для неоднородной прозрачной пластины толщиной d_0 с показателем преломления $n(z)$?

Упражнение 1.12

Участок свободного пространства и линза

Покажите, что матрица передачи луча для участка свободного пространства длиной d, за которым следует линза с фокусным расстоянием f, есть



$$\mathbf{M} = \begin{bmatrix} 1 & d \\ -\frac{1}{f} & 1 - \frac{d}{f} \end{bmatrix}. \qquad (1.51)$$

Упражнение 1.13

Изображение тонкой линзой

Выведите выражение для матрицы передачи луча для оптической системы, состоящей из двух участков свободного пространства с тонкой линзой между ними, как показано на рис. 1.37. Покажите, что при выполнении условия формирования изображения $\left(\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{f}\right)$ все лучи, выходящие из одной точки

входной плоскости, пересекают выходную плоскость в одной точке y_2, независимо от их наклона. Покажите также, что если $d_2 = f$, то все параллельные падающие лучи фокусируются линзой в одну точку в выходной плоскости.

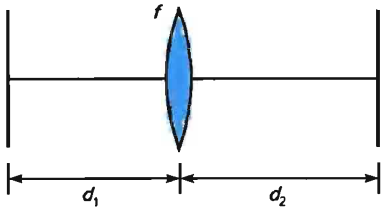


Рис. 1.37. Однолинзовая изображающая система

Упражнение 1.14

Изображение толстой линзой

Рассмотрите стеклянную линзу с показателем преломления n толщиной d с двумя сферическими поверхностями одинакового радиуса R (рис. 1.38). Определите матрицу передачи луча системы между двумя плоскостями на расстоянии d_1 и d_2 от вершин линзы. Линза находится в воздухе (показатель преломления равен 1). Покажите, что система является изображающей (т. е. входная и выходная плоскости сопряжены друг другу), если

$$\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f} \quad \text{или} \quad s_1 s_2 = f^2, \tag{1.52}$$

где

$$z_1 = d_1 + h, \quad s_1 = z_1 - f; \tag{1.53}$$

$$z_2 = d_2 + h, \quad s_2 = z_2 - f \tag{1.54}$$

и

$$h = \frac{(n-1)fd}{nR}; \tag{1.55}$$

$$\frac{1}{f} = \frac{(n-1)}{R}\left[2 - \frac{n-1}{n}\frac{d}{R}\right]. \tag{1.56}$$

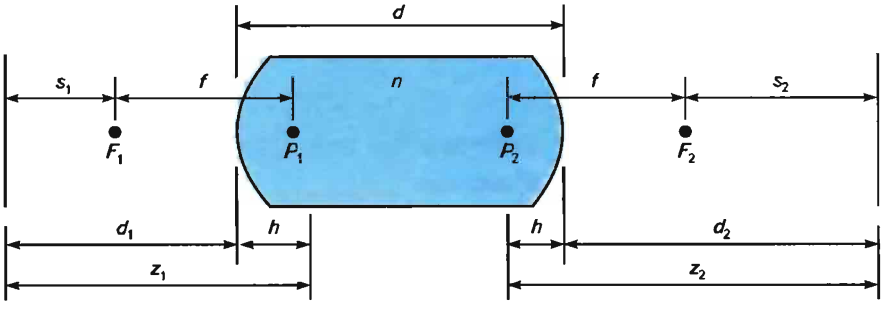


Рис. 1.38. Формирование изображения толстой линзой:
P_1 и P_2 — главные точки; F_1 и F_2 — фокальные точки

Точки F_1 и F_2 называются передней и задней **фокальными точками** системы. Точки P_1 и P_2 называются первой и второй **главными точками** соответственно. Покажите важность этих точек путем построения хода лучей, падающих параллельно оптической оси.

1.4.4. Периодические оптические системы

Периодическая оптическая система представляет собой каскад из одинаковых оптических элементов. Примером является последовательность равноотстоящих идентичных передающих линз, используемых в качестве световода, как показано на рис. 1.21, *а*. Другим примером является отражение света между двумя зеркалами, образующими оптический резонатор (см. подразд. 10.2.1), в этом случае луч снова и снова проходит один и тот же оптический элемент (обход резонатора с отражениями). Даже однородная среда, такая как стеклянное волокно, может рассматриваться как периодическая система, если ее разделить на прилегающие друг к другу идентичные участки одинаковой длины. Перейдем к формулировке общей теории распространения лучей в периодических системах на основе матричных методов.

Разностное уравнение для положения луча

Периодическая система состоит из каскада идентичных элементарных систем (ступеней), каждая из которых имеет матрицу передачи луча (A, B, C, D), как показано на рис. 1.39. Луч входит в систему, имея начальное положение y_0 и наклон θ_0. Для определения положения и наклона луча (y_m, θ_m) на выходе m-й ступени применим матрицу $ABCD$ m раз

$$\begin{bmatrix} y_m \\ \theta_m \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^m \begin{bmatrix} y_0 \\ \theta_0 \end{bmatrix}. \tag{1.57}$$

Можно также повторно применять соотношения

$$y_{m+1} = Ay_m + B\theta_m; \tag{1.58}$$

$$\theta_{m+1} = Cy_m + D\theta_m, \tag{1.59}$$

чтобы по (y_0, θ_0) найти (y_1, θ_1), затем по (y_1, θ_1) найти (y_2, θ_2) и так далее с помощью компьютерной программы.

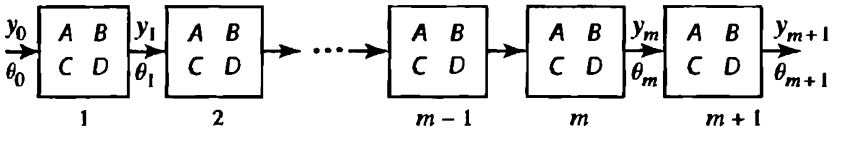


Рис. 1.39. Каскад идентичных оптических систем

Интересно вывести уравнения, которые описывают динамику положения y_m, $m = 0, 1, \ldots$, независимо от угла θ_m. Это достигается исключением θ_m из (1.58) и (1.59). Из (1.58) получаем

$$\theta_m = \frac{y_{m+1} - Ay_m}{B}. \tag{1.60}$$

Замена m на $m + 1$ в (1.60) дает

$$\theta_{m+1} = \frac{y_{m+2} - Ay_{m+1}}{B}. \tag{1.61}$$

Подставляя (1.60) и (1.61) в (1.59), находим

$$\boxed{y_{m+2} = 2by_{m+1} - F^2 y_m,} \tag{1.62}$$

Рекуррентное соотношение для положения луча

где

$$b = \frac{A + D}{2}; \tag{1.63}$$

$$F^2 = AD - BC = \det[\mathrm{M}], \tag{1.64}$$

а $\det[\mathbf{M}]$ — определитель **M**.

Уравнение (1.62) — линейное разностное уравнение, определяющее положение луча y_m. Его можно решить методом итераций, определяя y_2 по y_0 и y_1, затем y_3 по y_1 и y_2 и так далее. Величина y_1 может быть найдена по y_0 и θ_0 с использованием (1.58) с $m = 0$.

Полезно, однако, вывести явное выражение для y_m посредством решения разностного уравнения (1.62). Как и для линейных дифференциальных уравнений, решение, удовлетворяющее линейному разностному уравнению и данным начальным условиям, единственно. Поэтому разумно попытаться угадать вид такого решения для уравнения (1.62). Используем пробное решение в виде геометрической прогрессии

$$y_m = y_0 h^m, \tag{1.65}$$

где h — постоянная. Подставляя (1.65) в (1.62), немедленно убеждаемся, что пробное решение подходит при условии, что h удовлетворяет квадратному алгебраическому уравнению

$$h^2 - 2bh + F^2 = 0, \tag{1.66}$$

из которого

$$h = b \pm j\sqrt{F^2 - b^2} \tag{1.67}$$

Результаты можно представить в более компактном виде, если ввести переменную

$$\varphi = \arccos\left(\frac{b}{F}\right) \tag{1.68}$$

таким образом, чтобы

$$b = F\cos\varphi; \quad \sqrt{F^2 - b^2} = F\sin\varphi$$

и, следовательно,

$$h = F(\cos\varphi \pm j\sin\varphi) = F\exp(\pm j\varphi),$$

после чего (1.65) превращается в

$$y_m = y_0 F^m \exp(\pm jm\varphi).$$

Общее решение можно построить из двух решений с положительным и отрицательным знаком в виде их линейной комбинации. Сумма двух экспоненциальных функций всегда может быть записана в виде гармонической (тригонометрической) функции, так что

$$y_m = y_{\max} F^m \sin(m\varphi + \varphi_0), \tag{1.69}$$

где $y_{\max}$ и φ_0 — постоянные, определяемые из начальных условий y_0 и y_1. В частности, полагая $m = 0$, получим

$$y_{\max} = \frac{y_0}{\sin\varphi_0}.$$

Параметр F связан с определителем матрицы передачи луча для одного элемента соотношением

$$F = \sqrt{\det[\mathbf{M}]}.$$

Можно показать, что, независимо от вида элемента,

$$\det[\mathbf{M}] = \frac{n_1}{n_2},$$

где n_1 и n_2 — показатели преломления начального и конечного сечения элементарной системы. Этот общий результат несложно проверить на всех рассмотренных нами примерах оптических элементов. Поскольку при перемножении матриц их определители перемножаются, полученное соотношение $\det[\mathbf{M}] = n_1/n_2$ применимо к любому каскаду из таких элементов. Например, если

$$\det[\mathbf{M}_1] = \frac{n_1}{n_2}; \quad \det[\mathbf{M}_2] = \frac{n_2}{n_3},$$

то

$$\det[\mathbf{M}_1\mathbf{M}_2] = \frac{n_2}{n_3}\frac{n_1}{n_2} = \frac{n_1}{n_3}.$$

В большинстве приложений первая и последняя ступени каскада — воздух ($n = 1$), $n_1 = n_2$, так что $\det [M] = 1$ и $F = 1$. В этом случае решение уравнения для положения луча есть

$$\boxed{y_m = y_{\max} \sin(m\varphi + \varphi_0).} \qquad (1.70)$$

Положение луча
Периодическая система

Далее мы будем предполагать, что $F = 1$. Соответствующее решение для угла наклона луча получается с использованием соотношения

$$\theta_m = \frac{y_{m+1} - Ay_m}{B},$$

которое следует из (1.58).

Условие гармонической траектории

Для того, чтобы y_m было гармонической, а не гиперболической функцией, величина $\varphi = \arccos b$ должна быть действительной. Для этого требуется

$$\boxed{|b| \le 1 \quad \text{или} \quad \frac{1}{2}|A + D| \le 1.} \qquad (1.71)$$

Условие устойчивости

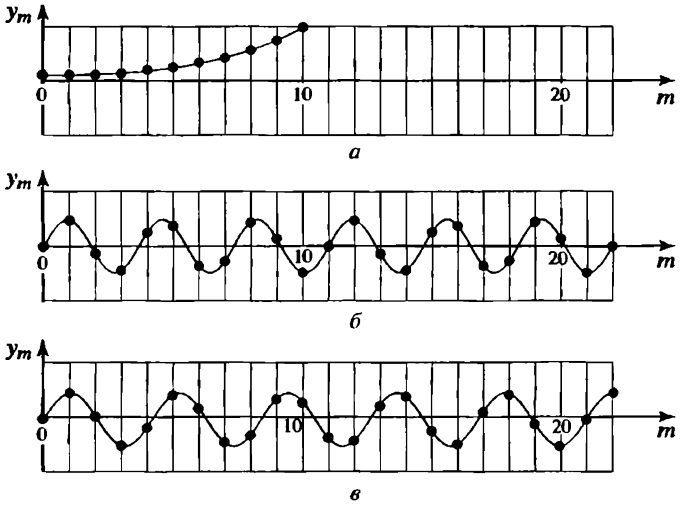


Рис. 1.40. Примеры траекторий в периодических оптических системах:

а — неустойчивая траектория ($b > 1$); *б* — устойчивая периодическая траектория ($\varphi = 6\pi/11$, период равен 11 ступеням); *в* — устойчивая непериодическая траектория ($\varphi = 1{,}5$)

Если, напротив, $|b| > 1$, то φ мнимое, и решение является гиперболической функцией sh или ch, которая неограниченно растет, как показано на рис. 1.40, *а*. Гармоническое решение гарантирует ограниченность y_m для всех m с максимальным значением $y_{\max}$. Таким образом, ограничение $|b| \le 1$ обеспечивает условие **устойчивости** (ограниченности) траектории луча.

Поскольку y_m и y_{m+1} — гармонические функции, то гармонической функцией является и угол наклона, что можно показать, используя (1.70), (1.60) и несложные тригонометрические преобразования. Таким образом,

$$\theta_m = \theta_{\max} \sin(m\varphi + \varphi_1),$$

где постоянные $\theta_{\max}$ и φ_1 определяются из начальных условий. Максимальный угол наклона $\theta_{\max}$ должен быть достаточно малым, чтобы не выходить за пределы параксиального приближения, лежащего в основе данного анализа.

Условие периодичности траектории

Гармоническая функция (1.70) периодична по m, если можно найти такое целое число s, что $y_m = y_{m+s}$ для любого m. Наименьшее из таких чисел есть период. В этом случае ход луча повторяется через каждые s шагов. Данное условие выполняется, если $s\varphi = 2\pi q$, где q — целое. Таким образом, необходимое и достаточное условие периодичности траектории заключается в том, чтобы $\varphi/2\pi$ было рациональным числом q/s. Например, если $\varphi = 6\pi/11$, то $\varphi/2\pi = 3/11$, и траектория периодична с периодом в 11 ступеней. Этот случай иллюстрируется рис. 1.40, *б*. Мы вернемся к рассмотрению периодических оптических систем в гл. 7.

Резюме

Параксиальный луч ($\theta_{\max} \ll 1$), проходя через каскад одинаковых элементарных оптических систем, каждая из которых характеризуется матрицей передачи луча с элементами (A, B, C, D), для которой $AD - BC = 1$, распространяется вдоль гармонической (и, следовательно, ограниченной) траектории, если выполняется условие

$$\frac{|A + D|}{2} \le 1,$$

называемое условием устойчивости. Тогда положение луча на m-й ступени каскада $y_m = y_{\max} \sin(m\varphi + \varphi_0)$, $m = 0, 1, 2, \ldots$, где

$$\varphi = \arccos\frac{|A + D|}{2}.$$

Постоянные $y_{\max}$ и φ_0 определяются начальными положениями y_0 и $y_1 = Ay_0 + B\theta_0$, где θ_0 — начальный наклон луча. Углы наклона луча связаны с его положениями формулой

$$\theta_m = \frac{y_{m+1} - Ay_m}{B}$$

и подчиняются гармоническому закону $\theta_m = \theta_{\max} \sin(m\varphi + \varphi_1)$. Траектория луча периодична с периодом s, если $\varphi/2\pi$ — рациональное число q/s.

Пример 1.2

Последовательность равноотстоящих одинаковых линз

Набор одинаковых линз с фокусным расстоянием f, разделенных расстоянием d, как показано на рис. 1.41, можно использовать для передачи света между двумя пунктами. Элементарная ступень — участок свободного пространства длиной d, за которым следует линза, — имеет матрицу передачи луча (1.51); $A = 1$, $B = d$, $C = -1/f$, $D = 1 - d/f$. Параметр

$$b = \frac{|A + D|}{2} = 1 - \frac{d}{2f}$$

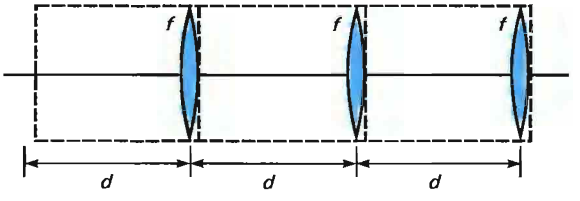


Рис. 1.41. Периодическая последовательность линз

и определитель равен единице. Условие устойчивости траектории луча $|b| \le 1$ или $-1 \le b \le 1$ в этом случае

$$0 \le d \le 4f, \qquad (1.72)$$

так что расстояние между линзами должно быть меньше, чем четыре фокусных расстояния. При этом условии положения параксиальных лучей описываются гармонической функцией

$$y_m = y_{\max} \sin(m\varphi + \varphi_0); \quad \varphi = \arccos\left(1 - \frac{d}{2f}\right). \qquad (1.73)$$

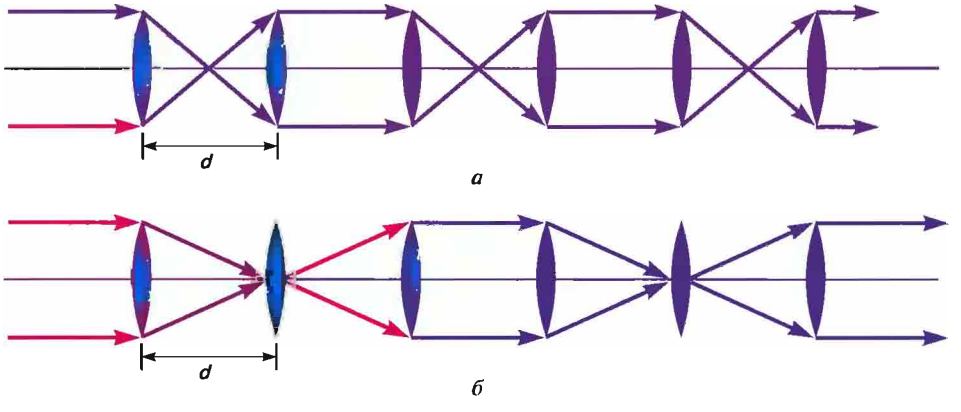


Рис. 1.42. Примеры устойчивых траекторий лучей в периодической системе линз: *а* — $d = 2f$; *б* — $d = f$

Когда $d = 2f$, $\varphi = \pi/2$ и $\varphi/2\pi = 1/4$, так что траектория произвольного луча периодична с одинаковым периодом, равным четырем ступеням. Когда $d = f$, $\varphi = \pi/3$ и $\varphi/2\pi = 1/6$, так что траектория повторяется через каждые шесть ступеней. Эти случаи показаны на рис. 1.42.

Упражнение 1.15

Периодическая последовательность пар различных линз

Исследуйте траектории параксиальных лучей, проходящих через систему пар линз с различными фокусными расстояниями f_1 и f_2, как показано на рис. 1.43. Покажите, что траектория ограничена (устойчива), если

$$0 \le \left(1 - \frac{d}{2f_1}\right)\left(1 - \frac{d}{2f_2}\right) \le 1. \tag{1.74}$$

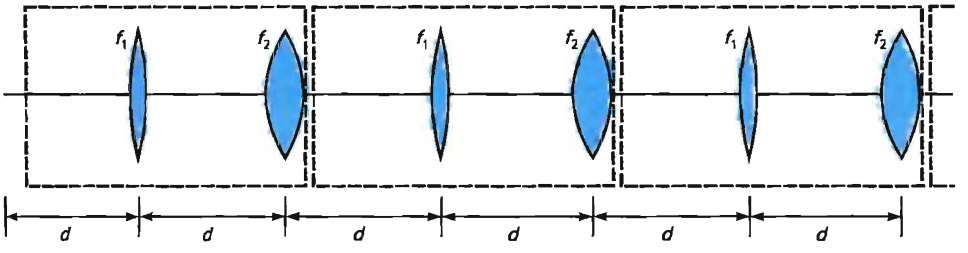


Рис. 1.43. Периодическая последовательность пар линз

Упражнение 1.16

Оптический резонатор

Параксиальные лучи многократно отражаются между двумя сферическими зеркалами с радиусами R_1 и R_2, разделенными расстоянием d (рис. 1.44). Рассматривая эту систему как периодическую, у которой элементарная ступень представляет собой полный обход между зеркалами, определите условие устойчивости траектории луча. Оптические резонаторы подробно рассматриваются в гл. 10.

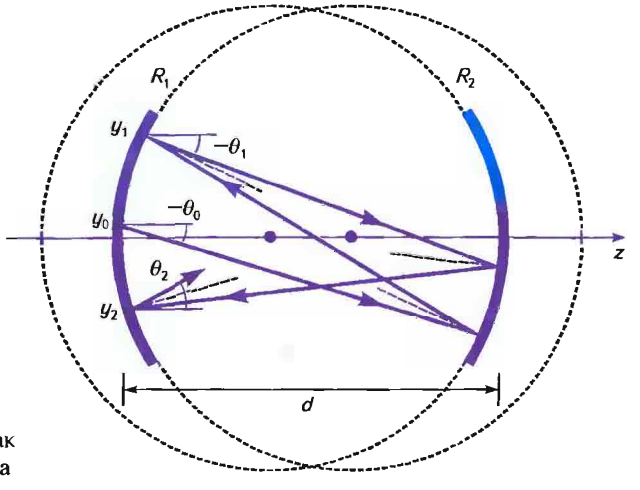


Рис. 1.44. Оптический резонатор как периодическая оптическая система

Рекомендуемая литература

ОБЩАЯ

Pedrotti F.L., *Pedrotti L.M.* and *Pedrotti L.S.* Introduction to Optics. Prentice Hall, 3rd ed. 2006.

Sharma K.K. Optics: Principles and Applications. Academic Press, 2006.

Walther A. The Ray and Wave Theory of Lenses. Cambridge University Press, 1995; paperback ed., 2006.

Moeller K.D. Optics: Learning by Computing with Examples Using Maple, MathCad, Mathematica and MATLAB. Springer-Verlag, 2nd ed. 2006.

Poon T.-C. and *Kim T.* Engineering Optics with MATLAB. World Scientific, 2006.

Siciliano A. Optics: Problems and Solutions. World Scientific, 2006.

Chartier G. Introduction to Optics. Springer-Verlag, 2005.

Strong J. Concepts of Classical Optics. Freeman, 1958; Dover; paperback ed. 2004.

Brooker G. Modern Classical Optics. Oxford University Press, 2003.

Born M., *Wolf E.* Principles of Optics. Cambridge University Press, 7th expanded and corrected ed. 2002.

Hecht E. Optics. Addison-Wesley, 4th ed. 2002.

Mansuripur M. Classical Optics and Its Applications. Cambridge University Press, 2002.

Keating M.P. Geometric, Physical, and Visual Optics. Butterworth-Heinemann, 2nd ed. 2002.

Young M. Optics and Lasers Including Fibers and Optical Waveguides. Springer-Verlag, 1977, 5th ed. 2000.

Meyer-Arendt J.R. Introduction to Classical and Modern Optics. Prentice Hall, 1972, 4th ed. 1995.

Blaker J.W., *Rosenblum W.M.* Optics An Introduction for Students of Engineering. Macmillan, 1993.

Moore D.T., ed. Selected Papers on Gradient-Index Optics. SPIE Optical Engineering Press (Milestone Series. Vol. 67), 1993.

Jenkins F.A. and *White H.E.* Fundamentals of Optics. McGraw-Hill, 1937, 4th revised ed. 1991.

Banerjee P.P., *Poon T.-C.* Principles of Applied Optics. Aksen Associates, 1991.

Guenther R.D. Modern Optics. Wiley, 1990.

Hecht E., *Zajac A.* Optics. Addison-Wesley, 1974; 2nd ed. 1990.

Welford W.T. Optics, Oxford University Press, 1976, 3rd ed. 1988.

Wood R.W. Physical Optics. Macmillan, 3rd ed. 1934; Optical Society of America, 1988.

Klein M.V., *Furtak T.E.* Optics. Wiley, 1982, 2nd ed. 1986.

Marchand E.W. Gradient-Index Optics. Academic Press, 1978.

Carlson F.P. Introduction to Applied Optics for Engineers. Academic Press, 1977.

Ditchburn R.W. Light. Academic Press, 3rd ed. 1976.

Hecht E. Schaum's Outline of Optics. McGraw-Hili, paperback ed. 1974.

Rossi B.B. Optics. Addison-Wesley, 1957; reprinted 1965.

Stone J.M. Radiation and Optics. McGraw-Hili, 1963.

Sommerfeld A. Lectures on Theoretical Physics: Optics. Academic Press, paperback ed. 1954.

ГЕОМЕТРИЧЕСКАЯ ОПТИКА

Kravtsov Yu.A. Geometrical Optics in Engineering Physics. Alpha Science, 2005.

Greivenkamp J.E. Field Guide to Geometrical Optics. SPIE Optical Engineering Press, 2004.

Wolf K.B. Geometric Optics on Phase Space. Springer-Verlag, 2004.

Katz M. Introduction to Geometrical Optics. World Scientific, 2002.

Ditteon R. Modern Geometrical Optics. Wiley, 1998.

Colombini F., Lerner N., eds. Geometrical Optics and Related Topics. Birkhauser, 1997.

Mouroulis P., Macdonald J. Geometrical Optics and Optical Design. Oxford University Press, 1997.

Loshin D.S. The Geometrical Optics Workbook. Butterworth-Heinemann, 1991.

Fry G.A. Geometrical Optics. Chilton, 1969; reprinted 1981.

Welford W.T., Winston R. The Optics of Nonimaging Concentrators. Academic Press, 1978.

Stavroudis O.N. The Optics of Rays, Wavefronts, and Caustics. Academic Press, 1972.

Zimmer H.-G. Geometrical Optics. Springer-Verlag, 1970.

Nussbaum A. Geometric Optics: An Introduction. Addison-Wesley, 1968.

Luneburg R.K. and *Herzberger M.* Mathematical Theory of Optics. University of California Press, 1964; reprinted 1966.

РАСЧЕТ ОПТИЧЕСКИХ СИСТЕМ

Gross H., ed. Handbook of Optical Systems. Wiley, 2005.

Malacara D. and *Malacara Z.* Handbook of Optical Design. Marcel Dekker, 1994; 2nd ed. 2004.

Fischer R.E., Tadic-Galeb B. Optical System Design. McGraw-Hill, 2000.

Smith W.J. Modern Optical Engineering: The Design of Optical Systems. McGraw-Hill, 1966; 3rd ed.2000.

Nussbaum A. Optical System Design. Prentice Hall, 1998.

O'Shea D.C. Elements of Modern Optical Design. Wiley, 1985.

Kingslake R. Optical System Design. Academic Press, 1983.

Levi L. Applied Optics: A Guide to Optical System Design. Wiley. Vol. 1, 1968; Vol. 2, 1980.

МАТРИЧНАЯ ОПТИКА

Gerrard A., Burch J.M. Introduction to Matrix Methods in Optics. Wiley, 1975; Dover, paperback ed. 1994.

Blaker J.W. Geometric Optics: The Matrix Theory. Marcel Dekker, 1971.

Brouwer W. Matrix Methods in Optical Instrument Design. Benjamin, 1964.

ИСТОРИЧЕСКАЯ И ПОПУЛЯРНАЯ ЛИТЕРАТУРА

Weiss R.J. A Brief History of Light and Those that Lit the Way. World Scientific, 1996.

Hall A.R. All was Light: An Introduction to Newton's Opticks. Clarendon Press/ Oxford University Press, 1993.

Kingslake R. A History of the Photographic Lens. Academic Press, 1989.

Sobel M.I. Light. University of Chicago Press, 1987.

Sabra A.I. Theories of Light from Descartes to Newton. Cambridge University Press, 1981.

Newton I. Opticks or a Treatise of the Reflections, Refractions, Inflections & Colours of Light. 4th ed. 1704; Dover, reissued 1979.

van Heel A.C.S., Velze C.H.F. What is Light? McGraw-Hill, 1968; reprinted 1978.

Ronchi V. The Nature of Light: An Historical Survey. Harvard University Press, 1970.

Tolansky S. Revolution in Optics. Penguin, 1968.

Tolansky S. Curiosities of Light Rays and Light Waves. Elsevier, 1965.

Bragg W.H. Universe of Light. Dover, paperback ed. 1959.

Rüchardt E. Light, Visible and Invisible. University of Michigan Press, 1958.

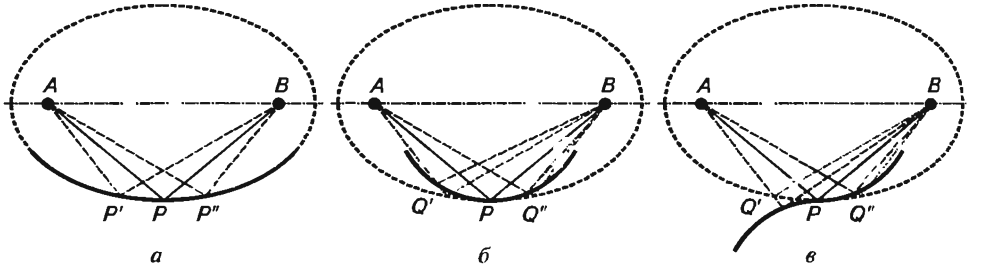
Задачи

К РАЗДЕЛУ 1.1

1. **Принцип Ферма с максимальным временем.** Рассмотрим эллиптическое зеркало, показанное на рис. 1, *а*, фокусы которого обозначены *A* и *B*. Геометрические свойства эллипса таковы, что длина пути $\overline{APB}$ равна длине путей $\overline{AP'B}$ и $\overline{AP''B}$ для соседних точек эллипса.

а. Теперь рассмотрим другое зеркало, радиус кривизны которого меньше, чем у эллиптического зеркала, которое касается его в точке *P*, как показано на рис. 1, *б*. Покажите, что путь луча $\overline{APB}$ между точками *A* и *B* соответствует максимальному времени распространения, т. е. его длина больше, чем у соседних путей $\overline{AQ'B}$ и $\overline{AQ''B}$.

б. Наконец, рассмотрим зеркало, которое пересекает эллипс и при этом касается его в точке *P*, как показано на рис. 1, *в*. Покажите, что возможные пути луча $\overline{AQ'B}$, $\overline{APB}$ и $\overline{AQ''B}$ демонстрируют наличие точки перегиба.



а *б* *в*

Рис. 1. Отражение от эллиптического зеркала (*а*); отражение от вписанного касательного зеркала большей кривизны (*б*); отражение от касательного зеркала с кривизной, меняющейся от выпуклого к вогнутому (*в*)

К РАЗДЕЛУ 1.2

1. **Прохождение через плоскопараллельные пластины.**

а. Используя закон Снелла, покажите, что луч, входящий в плоскопараллельную пластину толщины d с показателем преломления n_1, находящуюся в воздухе ($n \approx 1$), выходит из нее параллельно первоначальному направлению. Луч не обязательно параксиальный. Выведите выражение для поперечного смещения луча в зависимости от угла падения θ. Объясните результат с точки зрения принципа Ферма.

б. Если пластина представляет собой пачку N плоскопараллельных слоев, плотно прилегающих друг к другу и имеющих толщины d_1, d_2, ..., d_N и показатели преломления n_1, n_2, ..., n_N, покажите, что выходящий луч параллелен входящему. Для угла θ_m в слое с номером m покажите, что $n_m \sin \theta_m = n \sin \theta$, $m = 1, 2,$

2. **Линза в воде.** Определите фокусное расстояние f двояковыпуклой линзы с радиусами кривизны 20 см и 30 см и показателем преломления $n = 1{,}5$. Каково будет фокусное расстояние этой линзы, если ее погрузить в воду ($n = 4/3$)?

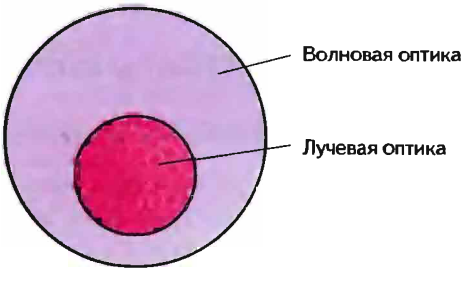
ГЛАВА 2 ОПТИКА ВОЛН

Христиан Гюйгенс (1629—1695) выдвинул несколько новых концепций распространения световых волн

Томас Юнг (1773—1829) отстаивал волновую теорию света и открыл принцип оптической интерференции

Свет распространяется в виде волн. В свободном пространстве световые волны распространяются с постоянной скоростью $c_0 = 3{,}0 \cdot 10^8$ м/с (30 см/нс, или 0,3 мм/пс, или 0,3 мкм/фс). Как показано на рис. 2.1 (см. цв. вклейку), диапазон оптических длин волн включает три области: инфракрасную (от 0,76 до 300 мкм), видимую (от 390 до 760 нм) и ультрафиолетовую (от 10 до 390 нм). Соответствующие частоты простираются от 1 ТГц в дальнем инфракрасной области до $3 \cdot 10^{16}$ Гц в вакуумном ультрафиолете.

Рис. 2.2. Волновая оптика включает в себя оптику лучей. Оптика лучей является пределом волновой оптики в случае очень малых длин волн



Волновая теория света включает в себя лучевую теорию (рис. 2.2). Строго говоря, оптика лучей представляет собой предельный случай волновой опти-

ки, когда длина волны бесконечно мала. Однако, на самом деле длина волны не должна быть равна нулю, чтобы можно было пользоваться лучевой оптикой. Пока световая волна распространяется сквозь объекты или вокруг объектов, размеры которых много больше длины волны, лучевой теории вполне достаточно для описания большинства оптических явлений. Поскольку длина волны видимого света намного меньше размеров обычных предметов, встречающихся в повседневной жизни, проявления волновой природы света неочевидны без внимательного наблюдения.

О данной главе

В данной главе свет описывается скалярной функцией, которая называется волновой функцией и удовлетворяет дифференциальному уравнению второго порядка, известным как волновое уравнение. Обсуждение физического смысла волновой функции отложим до гл. 5, где рассматривается электромагнитная оптика и где мы увидим, что волновая функция представляет одну из компонент электрического или магнитного поля. Волновое уравнение и соотношение между плотностью оптической мощности и волновой функцией являются постулатами скалярной теории, известной как **волновая оптика**. Следствием этих простых постулатов является несметное множество фактов. Волновая оптика дает основу для описания обширного круга оптических явлений, выходящих за рамки оптики лучей, включая интерференцию и дифракцию, как будет показано в этой и следующих двух главах.

Волновая оптика, однако, имеет свои ограничения. Она не может обеспечить полное описание отражения и преломления света на границах раздела диэлектрических сред, а также учесть оптические явления, требующие векторного формализма, такие как поляризационные эффекты. Эти проблемы будут освещены в гл. 5, так же как и условия, при которых скалярная волновая оптика дает хорошее приближение для электромагнитной оптики.

Данная глава начинается с постулатов оптики волн (см. разд. 2.1). В разд. 2.2—2.5 мы рассмотрим монохроматические волны; полихроматический свет обсуждается в разд. 2.6. Элементарные волны, такие как плоские и сферические волны, вводятся в разд. 2.2. В разд. 2.3 показывается, как оптика лучей вытекает из оптики волн. Взаимодействие оптических волн с простыми оптическим элементами, такими так зеркала, призмы, линзы и решетки, изучается в разд. 2.4. Интерференция как важное проявление волновой природы света является предметом разд. 2.5 и 2.6.

2.1. ПОСТУЛАТЫ ВОЛНОВОЙ ОПТИКИ

Волновое уравнение

Свет распространяется в виде волн. В свободном пространстве световые волны распространяются со скоростью c_0. Однородная прозрачная среда, такая как стекло, характеризуется единственной постоянной — показателем

преломления n (≥ 1). В среде с показателем преломления n свет распространяется с уменьшенной скоростью

$$\boxed{c = \frac{c_0}{n}.} \qquad (2.1)$$

Скорость света в среде

Оптическая волна математически описывается действительной функцией координат $\boldsymbol{r} = (x, y, z)$ и времени t, обозначаемой $u(\boldsymbol{r}, t)$ и называемой волновой функцией. Она удовлетворяет дифференциальному уравнению в частных производных, которое называется волновым уравнением

$$\boxed{\nabla^2 u - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0,} \qquad (2.2)$$

Волновое уравнение

где ∇^2 — оператор Лапласа, который в декартовых координатах имеет вид

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Любая функция, удовлетворяющая (2.2), представляет возможную оптическую волну.

Поскольку волновое уравнение линейно, справедлив **принцип суперпозиции**: если $u_1(\boldsymbol{r}, t)$ и $u_2(\boldsymbol{r}, t)$ представляют возможные световые волны, то

$$u(\boldsymbol{r}, t) = u_1(\boldsymbol{r}, t) + u_2(\boldsymbol{r}, t)$$

также представляет возможную световую волну.

На границе раздела между двумя средами волновая функция меняется в зависимости от показателей преломления сред. Однако законы, которым подчиняется это изменение, зависят от физического смысла, приписываемого волновой функции, которая, как будет показано в гл. 5, является компонентой электромагнитного поля. Лежащий в основе физический смысл показателя преломления получается из электромагнитной оптики (см. подразд. 5.5.3).

Волновое уравнение приближенно применимо и к средам с показателем преломления, зависящим от координат, если эта зависимость медленная на пространственных масштабах, сравнимых с длиной волны. В этом случае среда называется локально однородной. Для таких сред n в (2.1) и c в (2.2) просто заменяются функциями координат $n(\boldsymbol{r})$ и $c(\boldsymbol{r})$ соответственно.

Интенсивность, мощность и энергия

Оптическая **интенсивность** $I(\boldsymbol{r}, t)$, определяемая как оптическая мощность, приходящаяся на единицу площади (единица измерения Вт/см^2), пропорциональна среднему квадрату волновой функции

$$\boxed{I(\boldsymbol{r}, t) = 2\langle u^2(\boldsymbol{r}, t)\rangle.} \qquad (2.3)$$

Оптическая интенсивность

Операция $\langle\ \rangle$ означает усреднение по временному интервалу, намного большему оптического периода, но намного меньшему, чем любые другие интересующие нас времена, например длительность светового импульса. Оптические периоды чрезвычайно малы, например для света с длиной волны 600 нм период равен $2 \cdot 10^{-15}$ с = 2 фс. Данное понятие будет уточнено в разд. 2.6.

Физический смысл волновой функции $u(\boldsymbol{r}, t)$ не был введен нами явно, однако важной представляется ее связь (2.3) с физически измеримой величиной — оптической интенсивностью. Существует некоторый произвол в определении волновой функции и ее связи с интенсивностью. Например, формула (2.3) могла быть записана без множителя 2, а в волновую функцию мог быть введен масштабный множитель $\sqrt{2}$, при этом интенсивность осталась бы неизменной. Однако выбор множителя 2 в (2.3) удобен, что будет показано ниже.

Оптическая **мощность** $P(t)$ (единица измерения ватт, Вт), протекающая через поверхность A, нормальную по отношению к направлению распространения света, равна интегралу от интенсивности

$$P(t) = \int_A I(\boldsymbol{r}, t)\,\mathrm{d}A. \tag{2.4}$$

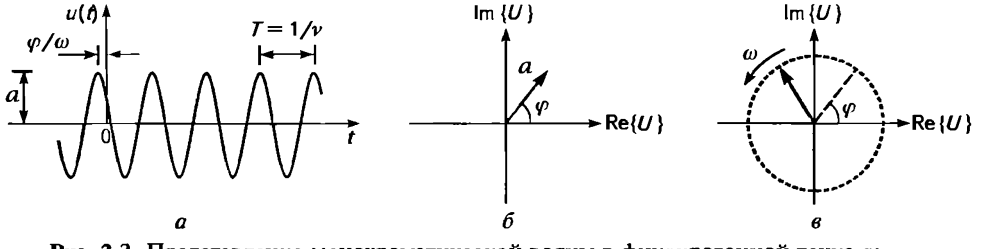
Оптическая **энергия** (единица измерения джоуль, Дж), накопленная за данный промежуток времени, есть интеграл от оптической мощности по указанному промежутку времени.

2.2. МОНОХРОМАТИЧЕСКИЕ ВОЛНЫ

Монохроматическая волна представляется волновой функцией с гармонической зависимостью от времени

$$u(\boldsymbol{r}, t) = a(\boldsymbol{r})\cos[2\pi\nu t + \varphi(\boldsymbol{r})], \tag{2.5}$$

как показано на рис. 2.3, *а*, где $a(\boldsymbol{r})$ — амплитуда; $\varphi(\boldsymbol{r})$ — фаза; ν — частота (колебаний/с или Гц); $\omega = 2\pi\nu$ — угловая частота (радиан/с или $\mathrm{с}^{-1}$); $T = 1/\nu = 2\pi/\omega$ — период (с).



а *б* *в*

Рис. 2.3. Представление монохроматической волны в фиксированной точке $\boldsymbol{r}$:

а — волновая функция $u(t)$ является гармонической функцией времени; *б* — комплексная амплитуда $U = a\exp(j\varphi)$ — постоянный вектор на комплексной плоскости; *в* — комплексная волновая функция $U(t) = U\exp(j2\pi\nu t)$ — вектор на комплексной плоскости, вращающийся с угловой скоростью $\omega = 2\pi\nu$ радиан/с

Как амплитуда, так и фаза в общем случае зависят от координат, однако во всех точках пространства волновая функция является гармонической функцией времени с частотой ν. Оптические волны имеют частоты, лежащие в диапазоне от $3 \cdot 10^{11}$ до $3 \cdot 10^{16}$ Гц, как показано на рис. 2.1.

2.2.1. Комплексное представление и уравнение Гельмгольца

Комплексная волновая функция

Удобно представлять действительную волновую функцию $u(\boldsymbol{r}, t)$, входящую в (2.6), с помощью комплексной функции

$$U(\boldsymbol{r}, t) = a(\boldsymbol{r})\exp[j\varphi(\boldsymbol{r})]\exp(j2\pi\nu t), \tag{2.6}$$

так что

$$u(\boldsymbol{r}, t) = \mathrm{Re}\{U(\boldsymbol{r}, t)\} = \frac{1}{2}[U(\boldsymbol{r}, t) + U^*(\boldsymbol{r}, t)], \tag{2.7}$$

где символ «*» обозначает комплексное сопряжение. Функция $U(\boldsymbol{r}, t)$, называемая **комплексной волновой функцией**, полностью описывает волну; волновая функция $u(\boldsymbol{r}, t)$ является просто ее действительной частью. Как и волновая функция $u(\boldsymbol{r}, t)$, комплексная волновая функция $U(\boldsymbol{r}, t)$ должна удовлетворять волновому уравнению

$$\boxed{\nabla^2 U - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} = 0.} \tag{2.8}$$

Волновое уравнение

Обе функции удовлетворяют одинаковым граничным условиям.

Комплексная амплитуда

Уравнение (2.6) можно записать в виде

$$U(\boldsymbol{r}, t) = U(\boldsymbol{r})\exp(j2\pi\nu t), \tag{2.9}$$

где не зависящий от времени множитель $U(\boldsymbol{r}) = a(\boldsymbol{r})\exp|j\varphi(\boldsymbol{r})|$ называется **комплексной амплитудой** волны. Волновая функция $u(\boldsymbol{r}, t)$, таким образом, связана с комплексной амплитудой как

$$\begin{aligned} u(\boldsymbol{r}, t) &= \mathrm{Re}\{U(\boldsymbol{r})\exp(j2\pi\nu t)\} = \\ &= \frac{1}{2}[U(\boldsymbol{r})\exp(j2\pi\nu t) + U^*(\boldsymbol{r})\exp(-j2\pi\nu t)]. \end{aligned} \tag{2.10}$$

В данной точке $\boldsymbol{r}$ комплексная амплитуда $U(\boldsymbol{r})$ является комплексным числом, изображенным на рис. 2.3, *б*, у которого модуль $|U(\boldsymbol{r})| = a(\boldsymbol{r})$ — амплитуда волны, а аргумент $\arg\{U(\boldsymbol{r})\} = \varphi(\boldsymbol{r})$ — ее фаза. Комплексная волновая функция

$U(\boldsymbol{r}, t)$ графически изображается как вектор на комплексной плоскости, вращающийся с угловой скоростью $\omega = 2\pi\nu$ радиан/с. Его начальное значение при $t = 0$ есть комплексная амплитуда $U(\boldsymbol{r})$.

Уравнение Гельмгольца

Подставляя $U(\boldsymbol{r}, t) = U(\boldsymbol{r})\exp(j2\pi\nu t)$ из (2.9) в волновое уравнение (2.8), получаем дифференциальное уравнение для комплексной амплитуды $U(\boldsymbol{r})$

$$\boxed{\nabla^2 U + k^2 U = 0,} \qquad (2.11)$$

Уравнение Гельмгольца

известное как **уравнение Гельмгольца**, где

$$\boxed{k = \frac{2\pi\nu}{c} = \frac{\omega}{c}} \qquad (2.12)$$

Волновое число

называется **волновым числом**. Различные решения получаются при различных граничных условиях.

Оптическая интенсивность

Оптическая интенсивность определяется подстановкой (2.5) в (2.7):

$$2u^2(\boldsymbol{r}, t) = 2a^2(\boldsymbol{r})\cos^2[2\pi\nu t + \varphi(\boldsymbol{r})] = |U(\boldsymbol{r})|^2\left\{1 + \cos\left(2[2\pi\nu t + \varphi(\boldsymbol{r})]\right)\right\}. \qquad (2.13)$$

При усреднении (2.13) по времени, много большему оптического периода $1/\nu$, второй член обращается в нуль, поэтому

$$\boxed{I(\boldsymbol{r}) = |U(\boldsymbol{r})|^2.} \qquad (2.14)$$

Оптическая интенсивность

Оптическая интенсивность монохроматической волны есть квадрат модуля ее комплексной амплитуды.

Интенсивность монохроматической волны *не зависит* от времени.

Волновые фронты

Волновые фронты — это поверхности равной фазы $\varphi(\boldsymbol{r}) = \text{const}$. Постоянные часто выбираются кратными 2π, так что $\varphi(\boldsymbol{r}) = 2\pi q$, где q — целое. Нормаль к волновому фронту в точке $\boldsymbol{r}$ параллельна вектору градиента $\nabla\varphi(\boldsymbol{r})$ (вектору с компонентами $\frac{\partial\varphi}{\partial x}$, $\frac{\partial\varphi}{\partial y}$, $\frac{\partial\varphi}{\partial z}$ в декартовой системе координат). Она представляет направление, в котором скорость изменения фазы максимальна.

Резюме

• Монохроматическая волна частоты ν описывается *комплексной волновой функцией* $U(\boldsymbol{r}, t) = U(\boldsymbol{r}) \exp(j2\pi\nu t)$, которая удовлетворяет волновому уравнению.

• *Комплексная амплитуда* $U(\boldsymbol{r})$ удовлетворяет уравнению Гельмгольца; ее модуль $|U(\boldsymbol{r})|$ и аргумент $\arg\{U(\boldsymbol{r})\}$ представляют собой *амплитуду* и *фазу* волны соответственно. Оптическая *интенсивность* есть $I(\boldsymbol{r}) = |U(\boldsymbol{r})|^2$. Волновые фронты — поверхности постоянной фазы $\varphi(\boldsymbol{r}) = \arg\{U(\boldsymbol{r})\} = 2\pi q$, где q — целое.

• Волновая функция $u(\boldsymbol{r}, t)$ есть действительная часть комплексной волновой функции, $u(\boldsymbol{r}, t) = \mathrm{Re}\{U(\boldsymbol{r}, t)\}$. Волновая функция также удовлетворяет волновому уравнению.

2.2.2. Элементарные волны

Простейшие решения уравнения Гельмгольца в однородной среде — это плоские и сферические волны.

Плоская волна

Плоская волна имеет комплексную амплитуду

$$U(\boldsymbol{r}) = A \exp(-j\boldsymbol{k} \cdot \boldsymbol{r}) = A \exp\left[-j\left(k_x x + k_y y + k_z z\right)\right], \tag{2.15}$$

где A — комплексная постоянная, называемая **комплексной огибающей**; $\boldsymbol{k} = (k_x, k_y, k_z)$ называется **волновым вектором**. Подставляя (2.15) в уравнение Гельмгольца (2.11), получаем

$$k_x^2 + k_y^2 + k_z^2 = k^2,$$

так что модуль волнового вектора $\boldsymbol{k}$ есть волновое число k.

Поскольку фаза волны есть $\arg\{U(\boldsymbol{r})\} = \arg\{A\} - \boldsymbol{k} \cdot \boldsymbol{r}$, поверхности постоянной фазы (волновые фронты) подчиняются уравнению

$$\boldsymbol{k} \cdot \boldsymbol{r} = k_x x + k_y y + k_z z = 2\pi q + \arg\{A\},$$

где q — целое число. Это уравнение описывает систему параллельных плоскостей, перпендикулярных волновому вектору $\boldsymbol{k}$, откуда и происходит название «плоские волны». Соседние плоскости разделены расстоянием $\lambda = 2\pi/k$, где λ определяется формулой

$$\boxed{\lambda = \frac{c}{\nu}} \tag{2.16}$$

Длина волны

и называется длиной волны. Плоская волна имеет постоянную интенсивность $I(\boldsymbol{r}) = |A|^2$ во всех точках пространства, поэтому она переносит бесконечную

мощность. Ясно, что такая волна является идеализацией, поскольку она существует повсюду и всегда.

Если ось z направить вдоль волнового вектора $\boldsymbol{k}$, то $U(\boldsymbol{r}) = A\exp(-jkz)$ и соответствующая волновая функция, полученная из (2.10), есть

$$u(\boldsymbol{r},\ t) = |A|\cos[2\pi\nu t - kz + \arg\{A\}] = |A|\cos\left[2\pi\nu\left(t - \frac{z}{c}\right) + \arg\{A\}\right]. \quad (2.17)$$

Таким образом, волновая функция периодична во времени с периодом $1/\nu$ и периодична в пространстве с периодом $2\pi/k$, равным длине волны λ (рис. 2.4); c называется **фазовой скоростью** волны.

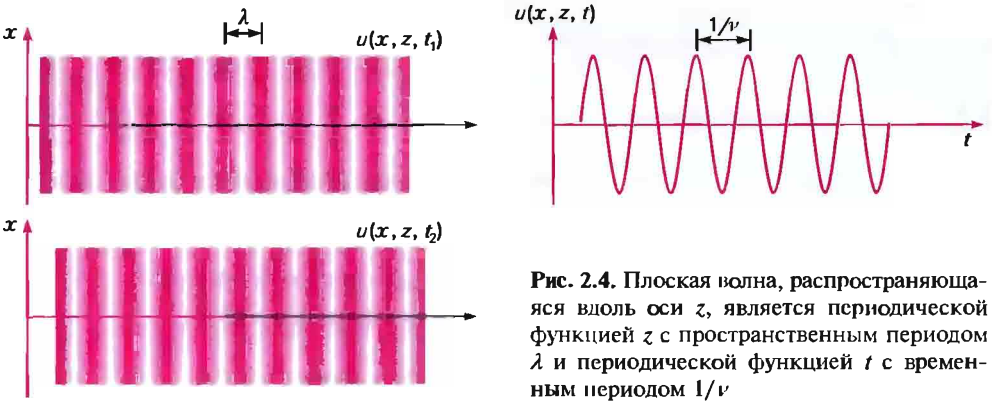


Рис. 2.4. Плоская волна, распространяющаяся вдоль оси z, является периодической функцией z с пространственным периодом λ и периодической функцией t с временным периодом $1/\nu$

В среде с показателем преломления n волна имеет фазовую скорость

$$c = \frac{c_0}{n}$$

и длину волны

$$\lambda = \frac{c}{\nu} = \frac{c_0}{n\nu},$$

так что

$$\lambda = \frac{\lambda_0}{n},$$

где λ_0 — длина волны в свободном пространстве,

$$\lambda_0 = \frac{c_0}{\nu}.$$

Таким образом, при заданной частоте ν длина волны в среде уменьшается в n раз по сравнению с длиной волны в свободном пространстве. Волновое число $k = 2\pi/\lambda$, соответственно, увеличивается по сравнению с его значением в свободном пространстве $k_0 = 2\pi/\lambda_0$ в n раз.

При распространении монохроматической волны через среды с различным показателем преломления ее частота остается неизменной, а скорость, длина волны и волновое число меняются:

$$c = \frac{c_0}{n}; \quad \lambda = \frac{\lambda_0}{n}; \quad k = nk_0. \tag{2.18}$$

Длины волн, показанные на рис. 2.1, относятся к свободному пространству ($n = 1$).

Сферическая волна

Другим простым решением уравнения Гельмгольца (в сферических координатах) является сферическая волна

$$U(r) = \frac{A_0}{r}\exp(-ikr), \tag{2.19}$$

где r — расстояние от начала координат; k — волновое число,

$$k = \frac{2\pi\nu}{c} = \frac{\omega}{c};$$

A_0 — постоянная. Интенсивность

$$I(r) = \frac{|A_0|^2}{r^2}$$

обратно пропорциональна квадрату расстояния. Принимая для простоты, что $\arg\{A_0\} = 0$, получаем, что волновые фронты являются поверхностями $kr = 2\pi q$ или $r = q\lambda$, где q — целое. Это концентрические сферы, разделенные радиальным расстоянием $\lambda = 2\pi/k$, радиус которых растет с фазовой скоростью c (рис. 2.5).

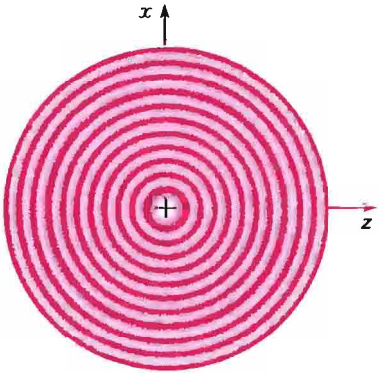


Рис. 2.5. Сечение волновых фронтов сферической волны

Сферическая волна, выходящая из точки r_0, имеет комплексную амплитуду

$$U(r) = \frac{A_0}{|r - r_0|}\exp\left(-jk|r - r_0|\right).$$

Ее волновые фронты — сферы с центром в точке r_0. Волна с комплексной амплитудой

$$U(r) = \frac{A_0}{r}\exp(+jkr)$$

называемая сферической, распространяется не наружу (из центра), а внутрь (по направлению к центру).

Приближение Френеля для сферической волны: параболоидальная волна

Исследуем сферическую волну, исходящую из точки $\boldsymbol{r} = 0$, в точках $\boldsymbol{r} = (x, y, z)$, достаточно близких к оси z, но удаленных от начала координат, так что

$$\sqrt{x^2 + y^2} \ll z.$$

Если бы эти точки были концами лучей, исходящих из начала координат, к ним можно было бы применять параксиальное приближение лучевой оптики (см. разд. 1.2). Введем обозначение

$$\theta^2 = \frac{x^2 + y^2}{z^2} \ll 1$$

и воспользуемся приближением, основанным на разложении Тейлора:

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} = z\sqrt{1 + \theta^2} = \\ &= z\left(1 + \frac{\theta^2}{2} - \frac{\theta^4}{8} + \ldots\right) \approx z\left(1 + \frac{\theta^2}{2}\right) = z + \frac{x^2 + y^2}{2z}. \end{aligned} \tag{2.20}$$

Подставим полученное выражение

$$r \approx z + \frac{x^2 + y^2}{2z}$$

в фазу сферической волны $U(\boldsymbol{r})$, описываемой формулой (2.19). Более грубое приближение $r \approx z$ используем для амплитуды, поскольку она менее чувствительна к ошибке, чем фаза. Результат известен как **приближение Френеля** для сферической волны

$$\boxed{U(\boldsymbol{r}) = \frac{A_0}{z}\exp(-jkz)\exp\left[-jk\frac{x^2 + y^2}{2z}\right].} \tag{2.21}$$

Приближение Френеля для сферической волны

Это приближение играет важную роль в упрощении теории распространения оптических волн через отверстия (**дифракции**), которое обсуждается в гл. 4.

Комплексную амплитуду в (2.21) можно рассматривать как амплитуду плоской волны $A_0 \exp(-jkz)$, модулируемой множителем

$$\frac{1}{z}\exp\left[-jk\frac{x^2 + y^2}{2z}\right],$$

включающим фазу $k(x^2 + y^2)/2z$. Роль этого фазового множителя состоит в искривлении плоских волновых фронтов с превращением их в параболоидальные поверхности (рис. 2.6), поскольку

$$\frac{x^2 + y^2}{z} = \text{const}$$

является уравнением параболоида вращения. В данной области сферическая волна хорошо аппроксимируется параболоидальной волной. Когда z становится очень большим, параболоидальная добавка к фазе в (2.21) стремится к нулю, так что полная фаза становится равной kz. Поскольку величина A_0/z медленно меняется с ростом z, сферическая волна постепенно превращается в плоскую волну $\exp(-jkz)$, как показано на рис. 2.6.

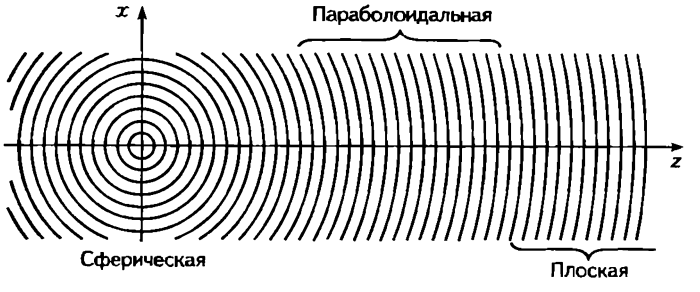


Рис. 2.6. Вблизи оси z и вдали от источника сферическая волна аппроксимируется параболоидальной волной. В точках, значительно удаленных от источника, сферическая волна превращается в плоскую волну

Однако условие применимости приближения Френеля не сводится просто к $\theta^2 \ll 1$. Хотя третий член степенного разложения $\theta^4/8$ может быть очень мал по сравнению с первым и вторым членами, будучи умножен на kz, он может стать сравнимым с π. Поэтому приближение, используемое в дальнейшем, верно при условии

$$kz\frac{\theta^4}{8} \ll \pi \quad \text{или} \quad x^2 + y^2 \ll 4z^3\lambda.$$

Для точек (x, y), лежащих внутри круга радиусом a с центром на оси z, это условие принимает вид $a^4 = 4z^3\lambda$ или

$$\frac{N_F\theta_m^2}{4} \ll 1, \tag{2.22}$$

где $\theta_m = a/z$ — максимальный угол; N_F — **число Френеля**,

$$\boxed{N_F = \frac{a^2}{\lambda z}}. \tag{2.23}$$

Число Френеля

Упражнение 2.1

Применимость приближения Френеля

Найдите радиус круга, внутри которого сферическая волна с длиной волны $\lambda = 633$ нм, источник которой удален на расстояние 1 м, может быть аппроксимирована параболоидальной волной. Определите максимальный угол θ_m и число Френеля N_F.

2.2.3. Параксиальные волны

Волна называется параксиальной, если нормали к ее волновым фронтам представляют собой параксиальные лучи. Один из путей построения параксиальной волны — взять плоскую волну $A\exp(-jkz)$, которую мы будем рассматривать как «несущую», и модифицировать или «модулировать» ее комплексную огибающую A, сделав ее медленно меняющейся функцией координат $A(\boldsymbol{r})$. Комплексная амплитуда получающейся волны есть

$$U(\boldsymbol{r}) = A(\boldsymbol{r})\exp(-jkz). \tag{2.24}$$

Изменения огибающей $A(\boldsymbol{r})$ и ее производной по координате z должно быть малы на расстояниях порядка длины волны $\lambda = 2\pi/k$, так чтобы волна по своей природе оставалась приблизительно плоской.

Волновая функция параксиальной волны

$$u(\boldsymbol{r}, t) = |A(\boldsymbol{r})|\cos[2\pi\nu t - kz + \arg\{A(\boldsymbol{r})\}]$$

изображена на рис 2.7, *а* как функция z при $t = 0$ и $x = y = 0$. Это синусоидальная функция z с амплитудой $|A(0, 0, z)|$ и фазой $\arg\{A(0, 0, z)\}$, обе из которых медленно меняются с ростом z. Поскольку фаза $\arg\{A(x, y, z)\}$ мало меняется на расстоянии порядка длины волны, плоские волновые фронты $kz = 2\pi q$ несущей волны лишь слегка искривляются, так что нормали к ним представляют собой параксиальные лучи (рис. 2.7, *б*).

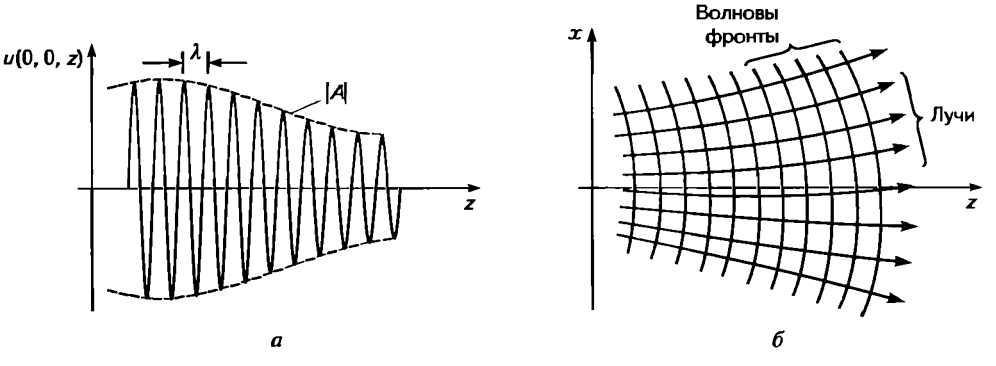


Рис. 2.7. Зависимость волновой функции параксиальной волны от z при $x = y = 0$ (*а*); волновые фронты параксиальной волны и нормали к ним в плоскости x—z (*б*)

Параксиальное уравнение Гельмгольца

Для того, чтобы параксиальная волна (2.24) удовлетворяла уравнению Гельмгольца (2.11), комплексная огибающая $A(\boldsymbol{r})$ должна удовлетворять другому дифференциальному уравнению в частных производных, которое получается путем подстановки (2.24) в (2.11). Предположение, что $A(\boldsymbol{r})$ медленно меняется с ростом z, означает, что на расстоянии $\Delta z = \lambda$ изменение $\Delta A \ll A$. Это

неравенство нужно применять отдельно к действительной и мнимой частям комплексного числа A. Поскольку

$$\Delta A = \frac{\partial A}{\partial z}\Delta z = \frac{\partial A}{\partial z}\lambda,$$

получаем, что

$$\frac{\partial A}{\partial z} \ll \frac{A}{\lambda} = \frac{Ak}{2\pi}$$

и, следовательно,

$$\frac{\partial A}{\partial z} \ll kA. \tag{2.25}$$

Сама производная $\frac{\partial A}{\partial z}$ также должна мало меняться на расстоянии λ, так чтобы

$$\frac{\partial^2 A}{\partial z^2} \ll k\frac{\partial A}{\partial z},$$

что дает

$$\frac{\partial^2 A}{\partial z^2} \ll k^2 A. \tag{2.26}$$

Подставляя (2.24) в (2.11) и пренебрегая $\frac{\partial^2 A}{\partial z^2}$ по сравнению с $k\frac{\partial A}{\partial z}$, получаем дифференциальное уравнение в частных производных относительно комплексной огибающей $A(\boldsymbol{r})$

$$\boxed{\nabla_T^2 A - j2k\frac{\partial A}{\partial z} = 0,} \tag{2.27}$$

Параксиальное уравнение Гельмгольца

где ∇_T^2 — поперечный оператор Лапласа,

$$\nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

Уравнение (2.27) представляет собой уравнение Гельмгольца в **приближении медленно меняющейся амплитуды**. Мы будем называть его просто **параксиальным уравнением Гельмгольца**. Оно имеет некоторое сходство с уравнением Шредингера в квантовой физике [см. (13.1)]. Простейшим решением параксиального уравнения Гельмгольца является параболоидальная волна (упражнение 2.2), которая представляет собой параксиальное приближение сферической волны. Одним из наиболее интересных и полезных решений является **гауссов пучок**, которому посвящена гл. 3.

Упражнение 2.2

Параболоидальная волна и гауссов пучок

Проверьте, что параболоидальная волна с комплексной огибающей

$$A(\boldsymbol{r}) = \frac{A_0}{z}\exp\left[-jk\frac{x^2+y^2}{2z}\right]$$

[см. (2.21)] удовлетворяет параксиальному уравнению Гельмгольца (2.27). Покажите, что волна с комплексной амплитудой

$$A(\boldsymbol{r}) = \frac{A_1}{q(z)}\exp\left[-jk\frac{x^2+y^2}{2q(z)}\right],$$

где $q(z) = z + jz_0$; z_0 — постоянная, также удовлетворяет параксиальному уравнению Гельмгольца. Такая волна, называемая гауссовым пучком, рассматривается в гл. 3. Постройте график распределения интенсивности в гауссовом пучке в плоскости $z = 0$.

*2.3. СВЯЗЬ МЕЖДУ ВОЛНОВОЙ И ЛУЧЕВОЙ ОПТИКОЙ

Покажем теперь, что оптика лучей вытекает из волновой оптики как предельный случай при длине волны, стремящейся к нулю. Рассмотрим монохроматическую волну, длина которой в свободном пространстве равна λ_0, распространяющуюся в среде с показателем преломления $n(\boldsymbol{r})$, который достаточно медленно меняется от точки к точке, так что среду можно считать локально однородной. Запишем комплексную амплитуду в (2.9) в виде

$$U(\boldsymbol{r}) = a(\boldsymbol{r})\exp\left[-jk_0S(\boldsymbol{r})\right], \tag{2.28}$$

где $a(\boldsymbol{r})$ — абсолютная величина; $-k_0S(\boldsymbol{r})$ — фаза; $k_0 = 2\pi/\lambda_0$ — волновое число в свободном пространстве. Предположим, что $a(\boldsymbol{r})$ меняется в зависимости от $\boldsymbol{r}$ достаточно медленно и может считаться постоянной на расстояниях порядка длины волны λ_0.

Волновые фронты представляют собой поверхности $S(\boldsymbol{r}) = \text{const}$, а нормали к ним направлены вдоль вектора градиента ∇S. В окрестности заданной точки $\boldsymbol{r}_0$ волну приближенно можно считать плоской с амплитудой $a(\boldsymbol{r}_0)$ и волновым вектором $\boldsymbol{k}$, абсолютная величина которого $k = n(\boldsymbol{r}_0)k_0$, а направление параллельно вектору градиента ∇S в точке $\boldsymbol{r}_0$. Различным точкам в окрестности данной точки соответствуют локальные плоские волны с различными амплитудами и различными волновыми векторами.

В оптике лучей было показано, что лучи направлены по нормали к поверхностям равных значений функции $S(\boldsymbol{r})$, названной эйконалом (см. подразд. 1.3.3). Свяжем локальные волновые векторы (нормали к волновым фронтам) в волновой оптике с лучами в лучевой оптике, тогда функция $S(\boldsymbol{r})$, пропорциональная

фазе волны, есть не что иное, как эйконал в оптике лучей (рис. 2.8). Как будет вскоре показано, такая связь имеет формальную математическую основу. В рамках данной аналогии оптика лучей может служить для приближенного определения того, как те или иные оптические элементы влияют на нормали к волновым фронтам (см. рис. 2.8).

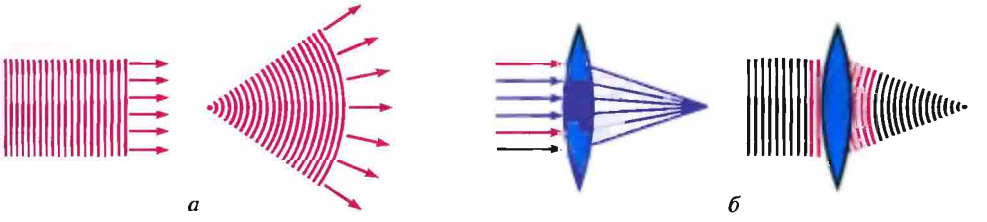


Рис. 2.8. Лучи в лучевой оптике ортогональны к волновым фронтам в волновой оптике (*а*) (см. также рис. 1.34); влияние линзы на лучи и волновые фронты (*б*)

Уравнение эйконала

Подстановка (2.28) в уравнение Гельмгольца (2.11) даст

$$k_0^2\left[n^2 - |\nabla S|^2\right]a + \nabla^2 a - jk_0\left[2\nabla S\cdot\nabla a + a\nabla^2 S\right] = 0, \tag{2.29}$$

где $a = a(\boldsymbol{r})$ и $S = S(\boldsymbol{r})$. Действительная и мнимая части выражения, стоящего в левой части (2.29), должны обращаться в нуль. Приравнивая действительную часть к нулю и используя равенство $k_0 = 2\pi/\lambda_0$, получаем

$$|\nabla S|^2 = n^2 + \left(\frac{\lambda_0}{2\pi}\right)^2 \frac{\nabla^2 a}{a}. \tag{2.30}$$

Предположение о том, что a мало изменяется на расстояниях порядка λ_0, означает, что

$$\frac{\lambda_0^2\nabla^2 a}{a} \ll 1,$$

так что вторым членом в правой части можно пренебречь в пределе $\lambda_0 \to 0$, тогда

$$\boxed{|\nabla S|^2 \approx n^2.} \tag{2.31}$$

Уравнение эйконала

Это уравнение эйконала (1.38), которое можно рассматривать как основной постулат оптики лучей (из этого уравнения можно вывести принцип Ферма и наоборот).

Таким образом, скалярная функция $S(\boldsymbol{r})$, пропорциональная фазе в волновой оптике, есть эйконал в оптике лучей. Это согласуется также с тем фактом, что в оптике лучей $S(\boldsymbol{r}_B) - S(\boldsymbol{r}_A)$ равно длине оптического пути между точками $\boldsymbol{r}_A$ и $\boldsymbol{r}_B$.

Уравнение эйконала является предельным случаем уравнения Гельмгольца при $\lambda_0 \to 0$. При заданном $n(\boldsymbol{r})$ можно использовать уравнение эйконала для определения $S(\boldsymbol{r})$. Приравнивая к нулю мнимую часть (2.29), мы получим связь между a и S, что дает возможность найти волновую функцию.

2.4. ПРОСТЫЕ ОПТИЧЕСКИЕ ЭЛЕМЕНТЫ

В данном разделе мы исследует действие оптических элементов, таких как зеркала, прозрачные пластины, призмы и линзы, на оптические волны.

2.4.1. Отражение и преломление

Отражение от плоского зеркала

Пусть плоская волна с волновым вектором $\boldsymbol{k}_1$ падает на плоское зеркало, расположенное в свободном пространстве в плоскости $z = 0$. Образуется отраженная волна с волновым вектором $\boldsymbol{k}_2$. Углы падения и отражения равны θ_1 и θ_2, как показано на рис. 2.9. Сумма двух волн удовлетворяет уравнению Гельмгольца, если волновые числа одинаковы, т. е. $k_1 = k_2 = k_0$. На поверхности зеркала должны выполняться некоторые граничные условия. Поскольку эти условия одинаковы во всех точках (x, y), необходимо, чтобы волновые фронты обеих волн совпадали, т. е.

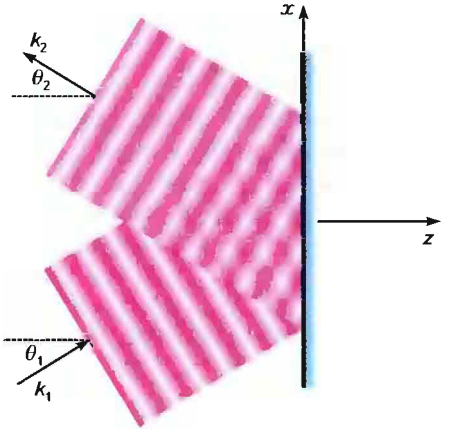


Рис. 2.9. Отражение плоской волны от плоского зеркала. Согласование волновых фронтов на поверхности зеркала требует, чтобы углы падения и отражения были равны

$$\boldsymbol{k}_1 \cdot \boldsymbol{r} = \boldsymbol{k}_2 \cdot \boldsymbol{r} \quad \text{для всех } \boldsymbol{r} = (x, y, 0). \tag{2.32}$$

Подставляя в (2.4) $\boldsymbol{r} = (x, y, 0)$, $\boldsymbol{k}_1 = (k_0 \sin\theta_1, 0, k_0 \cos\theta_1)$ и $\boldsymbol{k}_2 = (k_0 \sin\theta_2, 0, -k_0 \cos\theta_2)$, получаем $k_0 x \sin\theta_1 = k_0 x \sin\theta_2$, откуда $\theta_1 = \theta_2$, так что углы падения и отражения равны. Таким образом, закон отражения в оптике лучей применим к волновым векторам плоских волн.

Отражение и преломление на плоской границе раздела диэлектриков

Рассмотрим теперь плоскую волну с волновым вектором $\boldsymbol{k}_1$, падающую на плоскую границу раздела между двумя однородными средами с показателями преломления n_1 и n_2. Граница лежит в плоскости $z = 0$ (рис. 2.10).

Возникают преломленная и отраженная волны с волновыми векторами $\boldsymbol{k}_2$ и $\boldsymbol{k}_3$ соответственно. Комбинация трех волн удовлетворяет уравнению Гельмгольца во всех точках, если каждая из волн имеет соответствующее волновое число в среде, где она распространяется ($k_1 = k_2 = nk_0$ и $k_2 = n_2 k_0$).

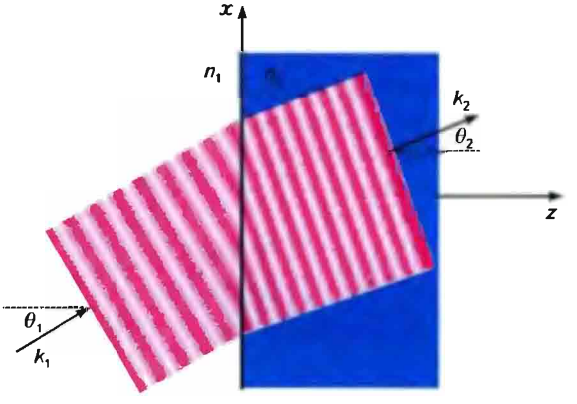


Рис. 2.10. Преломление плоской волны на границе раздела диэлектриков. Согласование волновых фронтов на границе: расстояние $\overline{P_1P_2}$ для падающей волны,

$$\frac{\lambda_1}{\sin\theta_1} = \frac{\lambda_0}{n_1 \sin\theta_1},$$

равно таковому для преломленной волны

$$\frac{\lambda_2}{\sin\theta_2} = \frac{\lambda_0}{n_2 \sin\theta_2},$$

откуда следует закон Снелла

Так как граничные условия инвариантны по отношению к x и y, необходимо, чтобы волновые фронты всех трех волн совпадали, т. е.

$$\boldsymbol{k}_1 \cdot \boldsymbol{r} = \boldsymbol{k}_2 \cdot \boldsymbol{r} = \boldsymbol{k}_3 \cdot \boldsymbol{r} \quad \text{для всех } \boldsymbol{r} = (x, y, 0). \tag{2.33}$$

Поскольку

$$\boldsymbol{k}_1 = (n_1 k_0 \sin\theta_1,\ 0,\ n_1 k_0 \cos\theta_1);$$

$$\boldsymbol{k}_3 = (n_1 k_0 \sin\theta_3,\ 0,\ -n_1 k_0 \cos\theta_3);$$

$$\boldsymbol{k}_2 = (n_2 k_0 \sin\theta_2,\ 0,\ n_2 k_0 \cos\theta_2),$$

где θ_1, θ_2 и θ_3 — углы падения, преломления и отражения соответственно, из (2.33) следует, что $\theta_1 = \theta_3$ и $n_1 \sin\theta_1 = n_2 \sin\theta_2$. Это закон отражения и закон преломления (закон Снелла) из оптики лучей, которые теперь применимы к волновым векторам.

В рамках скалярной волновой оптики невозможно определить амплитуды отраженной и преломленной волн, поскольку граничные условия не полностью определены в этой теории. Это будет сделано в разд. 6.2 в рамках электромагнитной оптики (см. гл. 5 и 6).

2.4.2. Прохождение через оптические элементы

Перейдем к изучению прохождения оптических волн через прозрачные оптические элементы, такие как пластины, призмы и линзы. Мы будем пренебрегать отражением от поверхностей этих элементов, поскольку его нельзя должным образом учесть в рамках скалярной волновой теории света. Не будем мы учитывать и поглощение материала, оставив это до разд. 5.5. Прин-

ципиальный акцент здесь будет сделан на фазовом сдвиге, вносимом указанными элементами, и соответствующем искривлении волнового фронта.

Прохождение через прозрачную пластину

Рассмотрим вначале прохождение плоской волны через прозрачную пластину с показателем преломления n и толщиной d, окруженную свободным пространством. Пусть поверхности пластины — плоскости $z = 0$ и $z = d$, а падающая волна распространяется вдоль оси z (рис. 2.11). Пусть $U(x, y, z)$ — комплексная амплитуда волны. Поскольку внешние и внутренние отражения не учитываются, функция $U(x, y, z)$ предполагается непрерывной на границах. Тогда отношение

$$t(x, y) = \frac{U(x, y, d)}{U(x, y, 0)}$$

представляет собой **комплексный амплитудный коэффициент пропускания** пластины; он позволяет определить $U(x, y, d)$ для произвольного $U(x, y, 0)$ на входе. Влияние отражения рассмотрено в разд. 6.2, а эффект многократного отражения внутри пластины — в разд. 10.1.

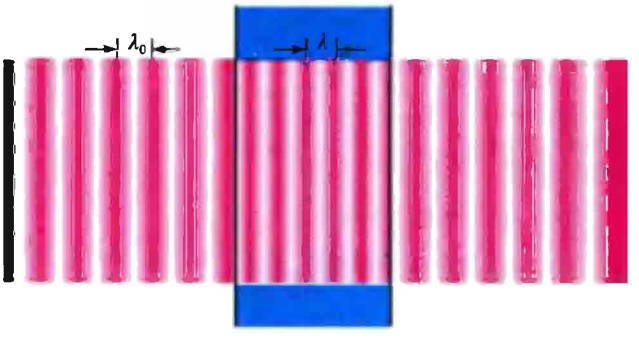


Рис. 2.11. Прохождение плоской волны через прозрачную пластину

Поскольку внутри пластины волна продолжает распространяться как плоская вола с волновым числом nk_0, величина $U(x, y, z)$ пропорциональна $\exp(-jnk_0z)$. Таким образом, отношение

$$\frac{U(x, y, d)}{U(x, y, 0)} = \exp(-jnk_0d),$$

так что

$$\boxed{t(x, y) = \exp(-jnk_0d).} \qquad (2.34)$$

Пропускание
Прозрачная пластина

Видно, что пластина вносит фазовый сдвиг

$$nk_0d = 2\pi\frac{d}{\lambda}.$$

Если плоская падающая волна образует угол θ с осью z и имеет волновой вектор $\boldsymbol{k}$ (рис. 2.12), то преломленная и прошедшая волны также плоские с волновыми векторами $\boldsymbol{k}_1$ и $\boldsymbol{k}$ и углами θ_1 и θ соответственно, причем θ_1 и θ связаны законом Снелла $\sin\theta = n\sin\theta_1$. Комплексная амплитуда $U(x, y, z)$ внутри пластины теперь пропорциональна

$$\exp(-j\boldsymbol{k}_1 \cdot \boldsymbol{r}) = \exp[-jnk_0(z\cos\theta_1 + x\sin\theta_1)],$$

так что комплексный амплитудный коэффициент пропускания пластины $U(x, y, d)/U(x, y, 0)$ есть

$$t(x, y) = \exp(-jnk_0 d\cos\theta_1). \tag{2.35}$$

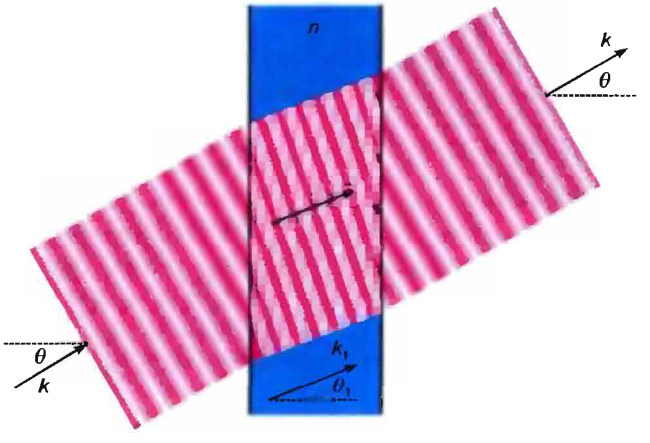


Рис. 2.12. Прохождение наклонной плоской волны через тонкую прозрачную пластину

Если угол падения θ мал (т. е. падающая волна *параксиальная*), то $\theta_1 \approx \theta/n$ также мало, и приближение

$$\cos\theta_1 \approx 1 - \frac{\theta_1^2}{2}$$

дает

$$t(x, y) = \exp(-jnk_0 d)\exp\left(jk_0\frac{\theta^2 d}{2n}\right).$$

Если пластина *достаточно тонкая*, а угол θ *достаточно мал*, так что

$$k_0\frac{\theta^2 d}{2n} \ll 2 \quad \text{или} \quad \frac{d}{\lambda_0}\frac{\theta^2}{2n} \ll 1,$$

то коэффициент пропускания пластины приближенно определяется формулой (2.34). При этих условиях коэффициент пропускания пластины практически не зависит от угла θ.

Тонкая прозрачная пластина переменной толщины

Найдем теперь амплитудный коэффициент пропускания для тонкой прозрачной пластины, толщина которой $d(x, y)$ плавно меняется в зависимости от x и y, на которую падает произвольная *параксиальная* волна. Пластина находится между плоскостями $z = 0$ и $z = d_0$, которые будем рассматривать как границы, заключающие в себе оптический элемент (рис 2.13).

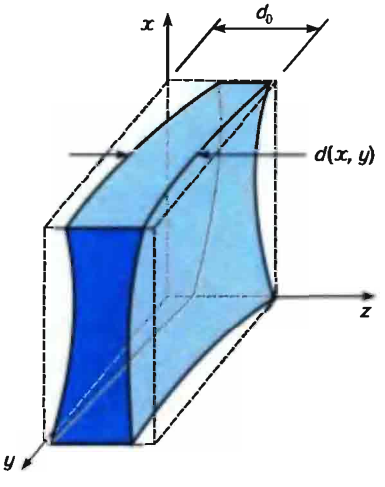


Рис. 2.13. Прозрачная пластина переменной толщины

В окрестности точки $(x, y, 0)$ падающая параксиальная волна может рассматриваться как локально плоская волна, распространяющаяся под небольшим углом к оси z. Она проходит через тонкую пластину толщиной $d(x, y)$, окруженную с обеих сторон тонкими слоями воздуха, полная толщина которых $d_0 - d(x, y)$. В соответствии с приближенным соотношением (2.34) локальный коэффициент пропускания является произведением коэффициентов пропускания тонкого слоя воздуха толщиной $d_0 - d(x, y)$ и тонкого слоя материала толщиной $d(x, y)$, так что

$$t(x, y) \approx \exp[-jnk_0 d(x, y)] \exp\{jk_0[d_0 - d(x, y)]\},$$

откуда

$$\boxed{t(x, y) \approx h_0 \exp[-j(n-1)k_0 d(x, y)]}, \qquad (2.36)$$

Коэффициент пропускания
Пластина переменной толщины

где $h_0 = \exp(-jk_0 d_0)$ — постоянный фазовый множитель. Это соотношение справедливо в параксиальном приближении (при малых углах θ), где толщина d_0 достаточно мала, так что

$$\frac{(d_0/\lambda_0)\theta^2}{2n} \ll 1.$$

Упражнение 2.3

Прохождение через призму

С помощью (2.36) покажите, что комплексный амплитудный коэффициент пропускания тонкой перевернутой призмы толщиной d_0 с углом $\alpha \ll 1$ при вершине (рис. 2.14) есть

$$t(x, y) = h_0 \exp[-j(n-1)k_0 \alpha x],$$

где $h_0 = \exp(-jk_0 d_0)$. Какое действие оказывает призма на падающую плоскую волну, распространяющуюся вдоль оси z? Сравните ваши результаты с полученными в модели оптики лучей [см. (1.10)].

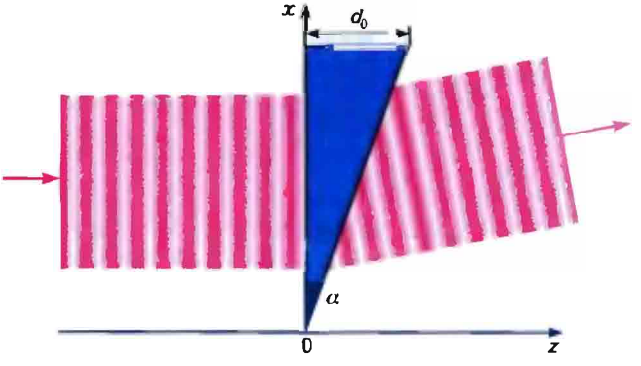


Рис. 2.14. Прохождение плоской волны через тонкую призму

Тонкая линза

Применим общее выражение (2.36) для комплексного амплитудного коэффициента пропускания тонкой прозрачной пластины переменной толщины к плосковыпуклой тонкой линзе, показанной на рис. 2.15. Поскольку линза представляет собой сегмент шара радиусом R, ее толщина в точке (x, y) есть

$$d(x,\ y) = d_0 - \overline{PQ} = d_0 - \left(R - \overline{QC}\right)$$

или

$$d(x,\ y) = d_0 - \left[R - \sqrt{R^2 - \left(x^2 + y^2\right)}\right]. \quad (2.37)$$

Это выражение можно упростить, если рассматривать только точки, в которых x и y достаточно малы по сравнению с R, так что $x^2 + y^2 = R^2$. В этом случае

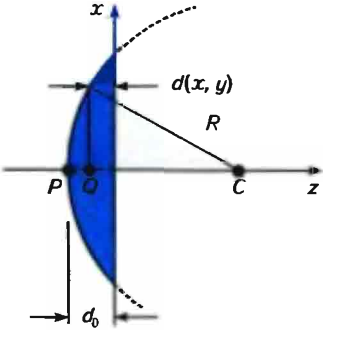


Рис. 2.15. Плосковыпуклая линза

$$\sqrt{R^2 - \left(x^2 + y^2\right)} = R\sqrt{1 - \frac{x^2 + y^2}{R^2}} \approx R\left(1 - \frac{x^2 + y^2}{2R^2}\right), \quad (2.38)$$

где использовано такое же разложение в ряд Тейлора, как в случае перехода к приближению Френеля для сферической волны (2.21). Используя приближение (2.37), получаем

$$d(x,\ y) \approx d_0 - \frac{x^2 + y^2}{2R}. \quad (2.39)$$

Подстановка в (2.36) окончательно дает

$$t(x, y) \approx h_0 \exp\left[jk_0 \frac{x^2 + y^2}{2f}\right], \tag{2.40}$$

Коэффициент пропускания
Тонкая линза

где

$$f = \frac{R}{n-1} \tag{2.41}$$

есть фокусное расстояние линзы (см. подразд. 1.2.3), а $h_0 = \exp(-jnk_0 d_0)$ — постоянный фазовый множитель, обычно не играющий какой-либо роли.

Поскольку линза вносит в фазу исходной плоской волны добавку, пропорциональную $x^2 + y^2$, она преобразует плоские волновые фронты в параболоидальные с центром, расположенным на расстоянии f от линзы, что предлагается показать в упражнении 2.4.

Упражнение 2.4

Двояковыпуклая линза

Покажите, что комплексный амплитудный коэффициент пропускания двояковыпуклой линзы (называемой также сферической линзой), показанной на рис. 2.16, определяется формулой (2.40), в которой

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \tag{2.42}$$

Вы можете доказать это либо используя общую формулу (2.36), либо рассматривая двояковыпуклую линзу как последовательность двух плосковыпуклых линз. Не забывайте, что по соглашению радиус выпуклой/вогнутой поверхности положителен/отрицателен, так что для линзы, показанной на рис. 2.16, R_1 положительно, а R_2 отрицательно. Параметр f определяет фокусное расстояние линзы [см. (1.15)].

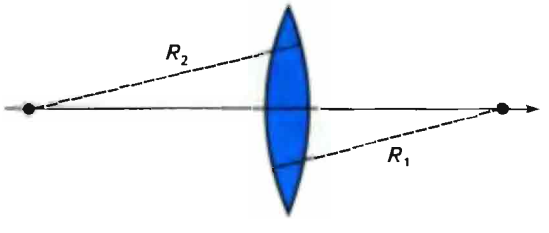


Рис. 2.16. Двояковыпуклая линза

Упражнение 2.5

Фокусировка плоской волны тонкой линзой

Покажите, что когда плоская волна проходит через тонкую линзу с фокусным расстоянием f в направлении оси линзы, она превращается в параболоидальную волну (приближение Френеля для сферической волны) с центром вблизи

точки, удаленной от линзы на расстояние f, как показано на рис. 2.17. Каково действие линзы на плоскую волну, падающую под малым углом θ к оси?

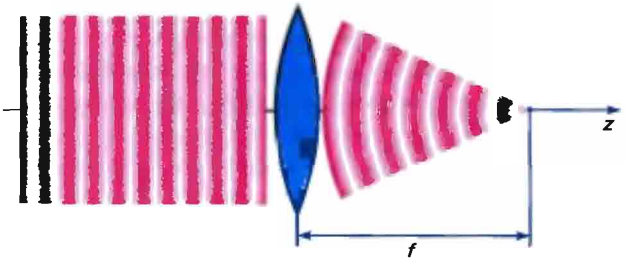


Рис. 2.17. Тонкая линза преобразует плоскую волну в параболоидальную

Упражнение 2.6

Формирование изображения линзой

Покажите, что параболоидальная волна с центром в точке P_1 (рис. 2.18) преобразуется линзой с фокусным расстоянием f в параболоидальную волну с центром около P_2, где

$$\frac{1}{z_1}+\frac{1}{z_2}=\frac{1}{f}$$

(это уравнение известно как **уравнение формирования изображения**).

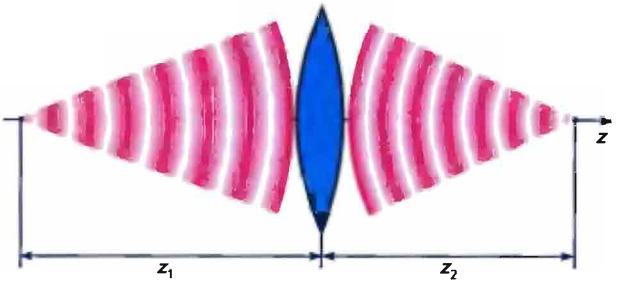


Рис. 2.18. Линза преобразует одну параболоидальную волну в другую параболоидальную волну. Центры волн находятся на расстояниях, удовлетворяющих уравнению формирования изображения

Дифракционные решетки

Дифракционная решетка — оптический элемент, служащий для периодической модуляции фазы или амплитуды падающей волны. Она может быть изготовлена в виде прозрачной пластины с периодически меняющейся толщиной или показателем преломления (см. подразд. 2.4.3). Повторяющиеся наборы элементов, на которых может происходить дифракция, таких как диафрагмы, препятствия или поглощающие элементы (см. разд. 4.3), также можно использовать для этой цели. Отражательные дифракционные решетки часто изготовляют из напыленной на стеклянную подложку тонкой пленки алюминия с периодически нанесенными штрихами.

Рассмотрим дифракционную решетку, изготовленную из тонкой прозрачной пластинки, расположенной в плоскости $z = 0$. Толщина пластинки периодически меняется в направлении x с периодом Λ (рис. 2.19). Как предлагается показать в упражнении 2.7, такая пластинка преобразует падающую плоскую волну с длиной $\lambda \ll \Lambda$, падающую под небольшим углом θ_i к оси z в несколько плоских волн, распространяющихся под малыми углами к оси z:

$$\boxed{\theta_q = \theta_i + q\frac{\lambda}{\Lambda},} \qquad (2.43)$$

Уравнение решетки

где $q = 0, \pm 1, \pm 2,$ — порядок дифракции. Соседние волны, получившиеся в результате дифракции, разделены по углу на величину $\theta = \lambda/\Lambda$, как схематически показано на рис. 2.19.

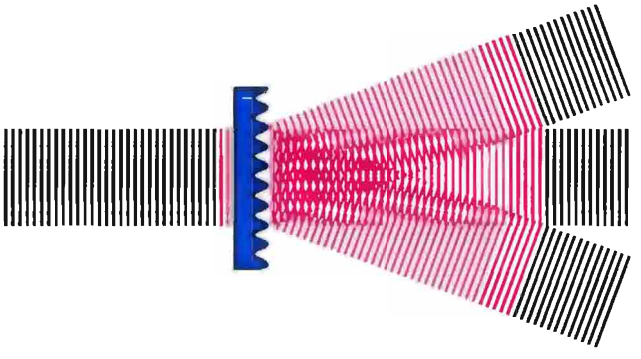


Рис. 2.19. Тонкая прозрачная пластинка с периодически меняющейся толщиной служит дифракционной решеткой. Она расщепляет падающую плоскую волну на много плоских волн, распространяющихся в различных направлениях

Упражнение 2.7

Прохождение через дифракционную решетку

а. Толщина тонкой прозрачной пластинки меняется синусоидально в направлении x:

$$d(x,\ y) = \frac{1}{2}d_0\left[1 + \cos\left(\frac{2\pi x}{\Lambda}\right)\right],$$

как показано на рис. 2.19. Покажите, что комплексный амплитудный коэффициент пропускания равен

$$t(x,\ y) = h_0 \exp\left[-j\frac{1}{2}(n-1)k_0 d_0 \cos\left(\frac{2\pi x}{\Lambda}\right)\right],$$

где

$$h_0 = \exp\left[-j\frac{1}{2}(n+1)k_0 d_0\right].$$

6. Покажите, что плоская волна, падающая под малым углом θ_i к оси z, после прохождения пластинки превращается в сумму плоских волн, распространяющихся под углами θ_q, определяемыми формулой (2.43).
Указание. Разложите периодическую функцию $t(x, y)$ в ряд Фурье.

Уравнение (2.43) справедливо только в параксиальном приближении (когда все углы малы). Это приближение применимо, когда период решетки Λ много больше длины волны λ. Более общий анализ тонких дифракционных решеток без применения параксиального приближения показывает, что падающая плоская волна превращается в несколько плоских волн, направления которых удовлетворяют уравнению[1]

$$\sin\theta_q = \sin\theta_i + q\frac{\lambda}{\Lambda}. \tag{2.44}$$

Дифракционные решетки используются как фильтры и анализаторы спектров. Поскольку углы θ_q зависят от длины волны (и, следовательно, от частоты), падающая полихроматическая волна разделяется решеткой на спектральные компоненты (рис. 2.20, см. цв. вклейку). Дифракционные решетки нашли многочисленные применения в спектроскопии.

2.4.3. Оптические элементы с градиентным показателем преломления

Действие призмы, линзы или дифракционной решетки на падающую оптическую волну состоит в придании ей фазового сдвига, который приводит к искривлению волнового фронта некоторым установленным образом. Этот фазовый сдвиг управляется изменением толщины материала по мере удаления от оптической оси (линейным, квадратичным или периодическим в случае призмы, линзы и дифракционной решетки соответственно). Вместо этого такой же фазовый сдвиг может вноситься прозрачной плоской пластинкой фиксированной толщины, но с переменным показателем преломления. Такая возможность следует из того, что толщина и показатель преломления появляются в виде произведения.

Комплексный амплитудный коэффициент пропускания тонкой прозрачной плоской пластины толщиной d_0 с градиентным показателем преломления $n(x, y)$, согласно (2.34), равен

$$\boxed{t(x, y) = \exp\left[-jn(x, y)k_0 d_0\right]}. \tag{2.45}$$

Коэффициент пропускания
Тонкая градиентная пластина

Подбирая соответствующее распределение $n(x, y)$ по x, y, можно воспроизвести действие любого тонкого оптического элемента, как показано в упражнении 2.8.

---

[1] См., например, *Hecht E., Zajac A.* Optics. Addison-Wesley, 2nd ed. 1990.

Упражнение 2.8

Линза с градиентом показателя преломления

Покажите, что тонкая пластина одинаковой толщины (рис. 2.21) с квадратичным профилем показателя преломления действует как линза с фокусным расстоянием (см. упражнение 1.8).

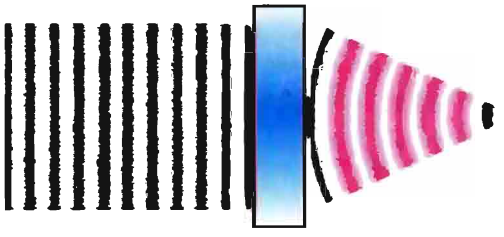


Рис. 2.21. Пластина с градиентом показателя преломления действует как линза

2.5. ИНТЕРФЕРЕНЦИЯ

Когда две или более волн одновременно присутствуют в одной и той же области пространства, полная волновая функция представляет собой сумму волновых функций отдельных волн. Фундаментальный принцип суперпозиции следует из линейности волнового уравнения. Для монохроматических волн одной и той же частоты принцип суперпозиции распространяется на их комплексные амплитуды, что следует из линейности уравнения Гельмгольца.

Принцип суперпозиции неприменим к интенсивностям волн, поскольку интенсивность суммы двух или более волн не обязательно равна сумме их интенсивностей. Это связано с интерференцией волн. Явление интерференции нельзя объяснить на основе лучевой оптики, поскольку оно зависит от соотношения между фазами наложенных друг на друга волн.

В данном разделе мы исследуем интерференцию двух и более монохроматических волн одной и той же частоты. Интерференция волн с различными частотами будет обсуждаться в разд. 2.6.

2.5.1. Интерференция двух волн

При наложении двух монохроматических волн с комплексными амплитудами $U_1(\boldsymbol{r})$ и $U_2(\boldsymbol{r})$ получается монохроматическая волна с той же частотой и комплексной амплитудой

$$U(\boldsymbol{r}) = U_1(\boldsymbol{r}) + U_2(\boldsymbol{r}). \tag{2.46}$$

В соответствии с (2.14) интенсивности составляющих волн равны $I_1 = |U_1|^2$ и $I_2 = |U_2|^2$, в то время как интенсивность полной волны есть

$$I = |U|^2 = |U_1 + U_2|^2 = |U_1|^2 + |U_2|^2 + U_1^* U_2 + U_1 U_2^*. \tag{2.47}$$

Явная зависимость от $\boldsymbol{r}$ для простоты не показана. Подставляя

$$U_1 = \sqrt{I_1}\exp(j\varphi_1); \quad U_2 = \sqrt{I_2}\exp(j\varphi_2), \tag{2.48}$$

в (2.47), где φ_1 и φ_2 — фазы волн, получаем

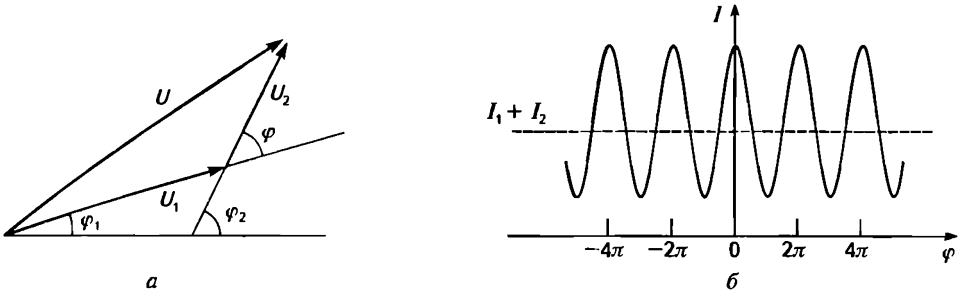
$$\boxed{I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\varphi,} \tag{2.49}$$

Уравнение интерференции

где

$$\varphi = \varphi_2 - \varphi_1. \tag{2.50}$$

Это уравнение, называемое уравнением интерференции, можно интерпретировать геометрически с помощью векторной диаграммы на комплексной плоскости, рис. 2.22, *а*, которая показывает, что длина вектора U на комплексной плоскости зависит не только от длин составляющих векторов, но и от разности их фаз φ.



а *б*

Рис. 2.22. Диаграмма векторов на комплексной плоскости для суперпозиции двух волн с интенсивностями I_1 и I_2 и разностью фаз φ (*а*); зависимость полной интенсивности I от разности фаз φ (*б*)

Ясно, что интенсивность суммы двух волн не есть сумма их интенсивностей (рис 2.22, *б*): в (2.49) присутствует дополнительный член, обязанный своим происхождением **интерференции** двух волн. Этот член может быть положительным или отрицательным, что соответствует конструктивной или деструктивной интерференции. Если, например, $I_1 = I_2 = I_0$, то (2.49) дает

$$I = 2I_0(1 + \cos\varphi) = 4I_0\cos^2\left(\frac{\varphi}{2}\right),$$

так что для $\varphi = 0$ имеем $I = 4I_0$ (т. е. полная интенсивность в четыре раза больше интенсивности каждой из интерферирующих волн). С другой стороны, при $\varphi = \pi$ налагающиеся волны гасят друг друга и полная интенсивность $I = 0$. Полное погашение волн в некоторой области пространства в общем случае невозможно, если интенсивности налагающихся волн не равны друг другу точно. Когда $\varphi = \pi/2$ или $3\pi/2$, интерференционный член исчезает и $I = 2I_0$. Для

этих особых случаев полная интенсивность равна сумме интенсивностей составляющих волн. Сильная зависимость интенсивности I от разности фаз φ позволяет измерять разности фаз по интенсивности света. Этот принцип используется в многочисленных оптических системах.

Интерференция сопровождается пространственным перераспределением оптической интенсивности без нарушения закона сохранения энергии. Например, обе волны могут иметь однородные распределения интенсивности I_1 и I_2 в некоторой плоскости, однако в результате того, что разность фаз φ зависит от положения точки наблюдения, полная интенсивность может быть меньше $I_1 + I_2$ в одних точках и больше — в других, при этом полная мощность (интеграл от интенсивности по поверхности) сохраняется.

Интерференция не наблюдается при обычных условиях освещения, поскольку случайные флуктуации фаз φ_1 и φ_2 делают их разность случайной величиной, равномерно распределенной в интервале от 0 до 2π, так что усреднение $\cos\varphi$ дает ноль и интерференционный член исчезает. Свет, обладающий такими случайными свойствами, называется *частично когерентным*, и его изучению посвящена гл. 11. Здесь мы ограничимся рассмотрением *когерентного* света.

Интерферометры

Рассмотрим суперпозицию двух плоских волн, каждая из которых имеет интенсивность I_0, распространяющихся в направлении оси z, и предположим, что одна из них отстает от другой на расстояние d, так что

$$U_1 = \sqrt{I_0}\exp(-jkz); \quad U_2 = \sqrt{I_0}\exp[-jk(z-d)].$$

Интенсивность I суммы этих двух волн определяется подстановкой $I_1 = I_2 = I_0$ и $\varphi = kd = 2\pi d/\lambda$ в уравнение интерференции (2.49)

$$I = 2I_0\left[1+\cos\left(2\pi\frac{d}{\lambda}\right)\right]. \tag{2.51}$$

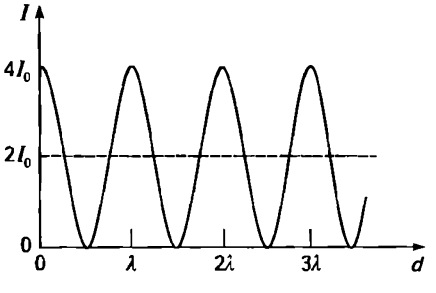


Рис. 2.23. Зависимость интенсивности I суперпозиции двух волн, каждая из которых имеет интенсивность I_0, от расстояния d. Когда d равно целому числу длин волн λ, интерференция конструктивна, когда оно равно нечетному целому числу полуволн $\lambda/2$ — деструктивна

Зависимость I от расстояния d изображена на рис. 2.23. Когда задержка равна целому кратному λ, имеет место конструктивная интерференция и полная интенсивность равна $I = 4I_0$. С другой стороны, когда d равно нечетному

целому числу $\lambda/2$, происходит полная деструктивная интерференция и $I = 0$. Средняя интенсивность есть сумма интенсивностей двух волн, т. е. $2I_0$.

Интерферометр — это оптический прибор, который расщепляет волну на две волны с помощью светоделителя, задерживает их на неодинаковые расстояния, перенаправляет их с помощью зеркал, совмещает с помощью другого (или того же самого) светоделителя и обеспечивает измерение интенсивности их суперпозиции. На рис. 2.24 приведены схемы трех важных примеров интерферометров: **интерферометра Маха—Цендера**, **интерферометра Майкельсона** и **интерферометра Саньяка**.

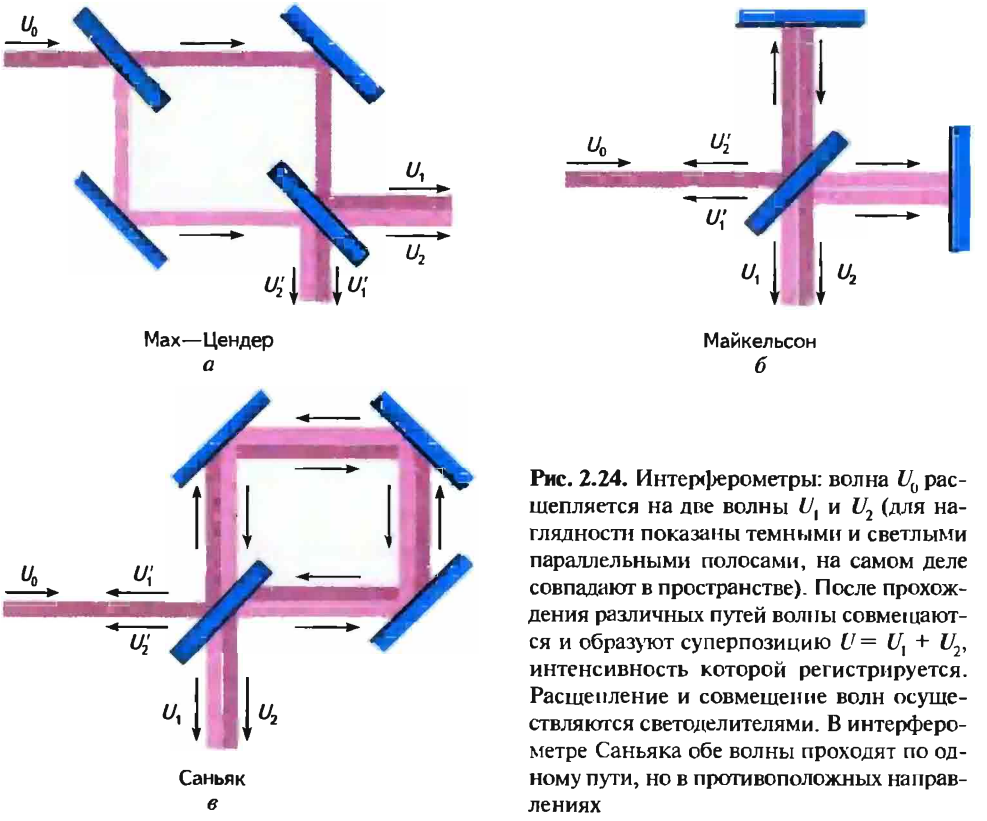


Рис. 2.24. Интерферометры: волна U_0 расщепляется на две волны U_1 и U_2 (для наглядности показаны темными и светлыми параллельными полосами, на самом деле совпадают в пространстве). После прохождения различных путей волны совмещаются и образуют суперпозицию $U = U_1 + U_2$, интенсивность которой регистрируется. Расщепление и совмещение волн осуществляются светоделителями. В интерферометре Саньяка обе волны проходят по одному пути, но в противоположных направлениях

Поскольку интенсивность I чувствительна к фазе

$$\varphi = 2\pi\frac{d}{\lambda} = 2\pi n\frac{d}{\lambda_0} = 2\pi n\nu\frac{d}{c_0},$$

где d — разность расстояний, проходимых каждой из волн, интерферометр можно использовать для измерения малых изменений расстояния d, показателя преломления n или длины волны λ_0 (или частоты ν). Например, если $d/\lambda_0 = 10^4$, изменение показателя преломления всего лишь на $\Delta n = 10^{-4}$ соответствует лег-

ко наблюдаемой разности фаз $\Delta\varphi = 2\pi$. Фаза φ меняется также на 2π при изменении d на величину длины волны λ. Приращение частоты $\Delta\nu = c/d$ дает тот же эффект.

Интерферометры имеют многочисленные применения. Сюда входит определение расстояний в метрологических приложениях, таких как измерение деформаций и профилирование поверхностей, измерение показателей преломления, а также спектрометрия для анализа полихроматического света (см. подразд. 11.2.2). В интерферометре Саньяка пути волн одинаковы, но проходятся в противоположных направлениях, поэтому вращение интерферометра приводит к появлению фазового сдвига, пропорционального угловой скорости вращения. Поэтому данная система может использоваться как гироскоп. Ввиду высокой точности оптическая интерферометрия используется также в экспериментах по обнаружению гравитационных волн.

Наконец, покажем, что закон сохранения энергии в интерферометре требует, чтобы фазы волн, отраженной от светоделителя и прошедшей через него, различались на $\pi/2$. Каждый из интерферометров, показанных на рис. 2.24, создает на выходе волну $U = U_1 + U_2$, выходящую с одной стороны светоделителя, и волну $U' = U_1' + U_2'$ с противоположной стороны. Закон сохранения энергии требует, чтобы сумма интенсивностей этих двух волн была равна интенсивности падающей волны, так что если одна из выходящих волн имеет высокую интенсивность благодаря конструктивной интерференции, то другая должна иметь низкую интенсивность из-за деструктивной интерференции. Эта дополнительная функция может быть достигнута, только если разности фаз φ и φ', соответствующие волнам U и U', различаются на π. Поскольку разности хода и число отражений от зеркал одинаковы для компонент U и U', разность фаз π необходимо приписать различию фаз, вносимых светоделителем при отражении и пропускании. Исследование трех интерферометров, показанных на рис. 2.24, приводит к выводу, что для одной выходящей волны каждая из компонент один раз проходит через светоделитель и один раз отражается от него, так что разность фаз не вносится. Для другой же выходящей волны одна из компонент дважды проходит светоделитель, а другая — дважды отражается от него, что и вносит разность фаз π. Следовательно, фазы отраженной и прошедшей через светоделитель волн различаются на $\pi/2$. Это важное свойство светоделителя более подробно описано в разд. 7.1 (см. пример 7.2).

Интерференция двух наклонных плоских волн

Рассмотрим интерференцию двух плоских волн равной интенсивности, одна из которых

$$U_1 = \sqrt{I_0}\exp(-jkz)$$

распространяется вдоль оси z, а другая,

$$U_2 = \sqrt{I_0}\exp[-j(k\cos\theta z + k\sin\theta x)]$$

под углом θ к оси z в плоскости x—z, как показано на рис. 2.25. В плоскости $z = 0$ разность фаз между волнами равна $\varphi = k \sin \theta x$, и уравнение интерференции (2.49) дает полную интенсивность

$$I = 2I_0 [1 + \cos(k \sin \theta x)]. \tag{2.52}$$

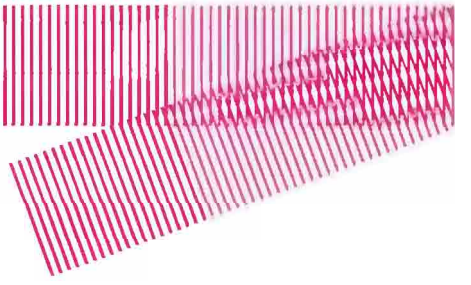


Рис. 2.25. Интерференция двух плоских волн, распространяющихся под углом θ друг к другу, порождает синусоидальное распределение интенсивности вдоль оси x с периодом $\lambda / \sin \theta$

Получается картина полос, на которой интенсивность синусоидально меняется вдоль оси x с периодом

$$\frac{2\pi}{k} \sin \theta = \lambda \sin \theta,$$

как показано на рис. 2.25. Если, например, $\theta = 30°$, этот период равен 2λ. Таким образом можно напечатать с высоким разрешением синусоидальную структуру для применения в качестве дифракционной решетки. Можно также использовать полученный результат в качестве метода для мониторинга угла наклона θ путем смешивания исследуемой волны с опорной волной и записи получающейся интерференционной картины. Обсуждение в разд. 4.5 показывает, что этот принцип лежит в основе голографии.

Упражнение 2.9

Интерференция плоской и сферической волн

Плоская волна, распространяющаяся вдоль оси z с комплексной амплитудой $A_1 \exp(-jkz)$, и сферическая волна с центром в точке $z = 0$, аппроксимирусмая параболоидальной волной с комплексной амплитудой

$$\frac{A_2}{z} \exp(-jkz) A_1 \exp\left(-jk \frac{x^2 + y^2}{2z}\right)$$

[см. (2.21)], интерферируют в плоскости $z = d$. Получите выражение для полной интенсивности $I(x, y, d)$. Предполагая, что обе волны имеют одинаковую интенсивность в плоскости $z = d$, проверьте, что точки, в которых интенсивность равна нулю, образуют систему концентрических окружностей, как показано на рис. 2.26.

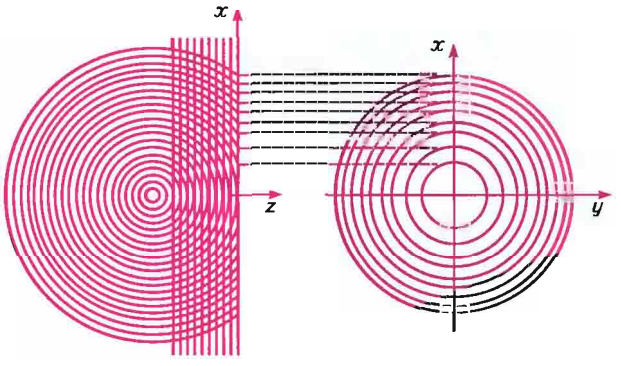


Рис. 2.26. Интерференция плоской и сферической волн создает систему концентрических колец (показаны в плоскости $z = d$)

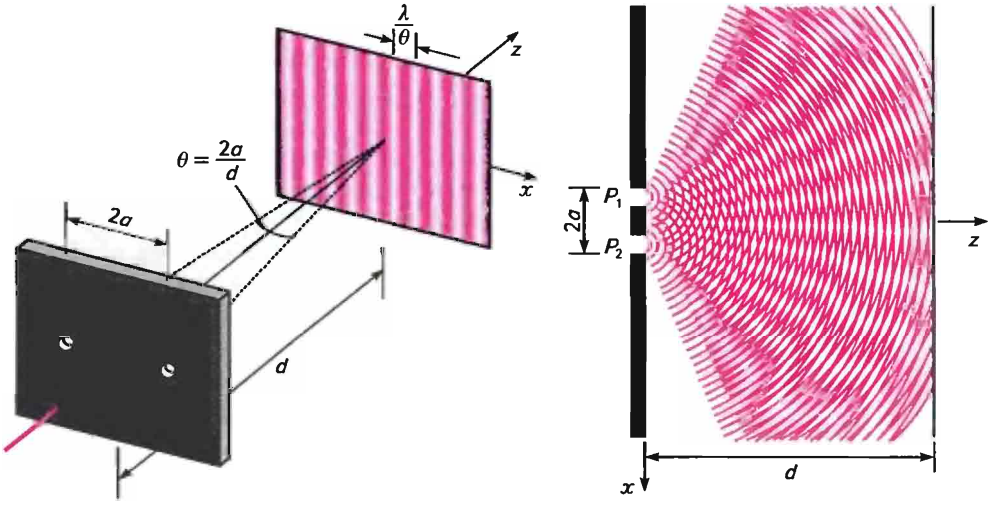


Рис. 2.27. Интерференция двух сферических волн равной интенсивности, выходящих из точек P_1 и P_2. Две такие волны можно получить, пропуская плоскую волну через два микроотверстия в непрозрачном экране. Интенсивность света в плоскости наблюдения на большом расстоянии d от отверстий принимает вид синусоидальной интерференционной картины с периодом $\approx\lambda/\theta$ вдоль направления линии, соединяющей отверстия

Упражнение 2.10

Интерференция двух сферических волн

Две сферические волны одинаковой интенсивности I_0, исходящие из точек $(-a, 0, 0)$ и $(a, 0, 0)$, интерферируют в плоскости $z = d$, как показано на рис. 2.27. Эта схема двух точечных источников аналогична схеме, использованной Томасом Юнгом в его знаменитом опыте с двумя щелями, где он впервые продемонстрировал существование интерференции. С использованием параболои-

дальной аппроксимации сферических волн покажите, что интенсивность в плоскости $z = d$ определяется выражением

$$I(x,\ y,\ d) \approx 2I_0\left(1 + \cos\frac{2\pi x\theta}{\lambda}\right), \tag{2.53}$$

где угол, под которым видны источники с плоскости наблюдения, равен $\theta \approx 2a/d$. Распределение интенсивности периодично с периодом λ/θ.

2.5.2. Многоволновая интерференция

Наложение M монохроматических волн одинаковой частоты с комплексными амплитудами $U_1, U_2, \ldots, U_M$ порождает волну той же частоты с комплексной амплитудой $U = U_1 + U_2 + \quad + U_M$. Знание интенсивностей отдельных волн $I_1, I_2, \ldots, I_M$ недостаточно для определения полной интенсивности $I = |U|^2$, поскольку требуется знать также относительные фазы волн. Роль фазы ярко иллюстрируют следующие примеры.

Интерференция M волн равной амплитуды с равными разностями фаз

Вначале рассмотрим интерференцию M волн с комплексными амплитудами

$$U_m = \sqrt{I_0}\exp[j(m-1)\varphi], \quad m = 1,\ 2,\ \ldots,\ M. \tag{2.54}$$

Эти волны имеют одинаковую интенсивность I_0 и разность фаз φ между последовательными волнами, как показано на рис. 2.28, *а*. Для вывода выражения для интенсивности суперпозиции волн удобно ввести величину $h = \exp(j\varphi)$, тогда

$$U_m = \sqrt{I_0}h^{m-1}.$$

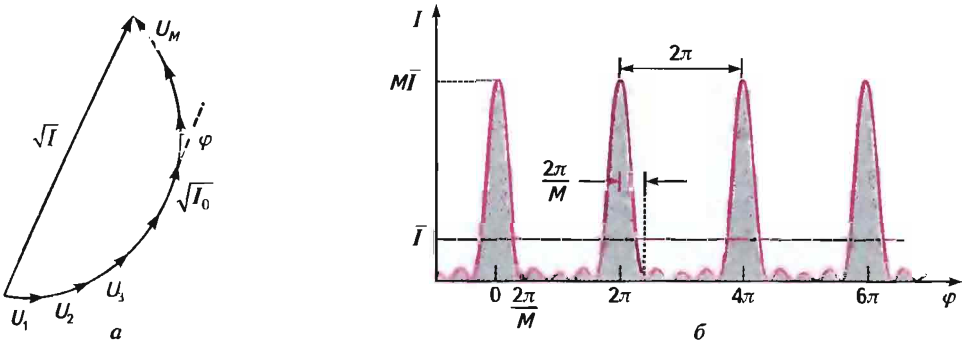


Рис. 2.28. Сумма M векторов на комплексной плоскости с равными величинами и разностями фаз (*а*); интенсивность I как функция φ (*б*). Пики интенсивности имеют место, когда векторы на комплексной плоскости параллельны, в этом случае интенсивность в M раз больше, чем средняя интенсивность MI_0. В данном случае $M = 5$

Комплексная амплитуда суперпозиции равна

$$U = \sqrt{I_0}\,(1 + h + h^2 + \ldots + h^{M-1}) = \sqrt{I_0}\,\frac{1-h^M}{1-h} = \sqrt{I_0}\,\frac{1-\exp(jM\varphi)}{1-\exp(j\varphi)}, \qquad (2.55)$$

а соответствующая интенсивность

$$I = |U|^2 = I_0 \left| \frac{\exp(-jM\varphi/2) - \exp(jM\varphi/2)}{\exp(-j\varphi/2) - \exp(j\varphi/2)} \right|^2 \qquad (2.56)$$

откуда

$$\boxed{I = I_0 \frac{\sin^2(M\varphi/2)}{\sin^2(\varphi/2)}.} \qquad (2.57)$$

Интерференция M волн

Интенсивность I, очевидно, сильно зависит от разности фаз φ, как показано на рис. 2.28, *б* при $M = 5$. Когда $\varphi = 2\pi q$, где q — целое, все векторы на комплексной плоскости выстраиваются параллельно, так что амплитуда полной волны в M раз больше амплитуды каждой из волн и интенсивность имеет максимальное значение, равное $M^2 I_0$. Интенсивность, усредненная по равномерному распределению φ,

$$\bar{I} = \frac{1}{2\pi}\int_0^{2\pi} I d\varphi = MI_0,$$

что совпадает с результатом, получаемым в отсутствие интерференции. Максимальная интенсивность, таким образом, в M раз больше средней интенсивности. Чувствительность интенсивности к разности фаз очень велика при большом M. Интенсивность в максимуме увеличена в M раз, однако она резко падает при малом отклонении φ от $2\pi q$. В частности, когда $\varphi = 2\pi/M$, интенсивность обращается в нуль. Поучительно сравнить рис. 2.28, *б*, где $M = 5$, с рис. 2.23, где $M = 2$.

Упражнение 2.11

Брэгговское отражение

Рассмотрим свет, отраженный под углом θ от M параллельных отражающих плоскостей, разделенных расстоянием Λ, как показано на рис. 2.29. Предположим, что от каждой плоскости отражается только небольшая часть света, так что амплитуды отраженных волн приблизительно равны. Покажите, что отраженные волны имеют разность фаз $\varphi = k(2\Lambda \sin\theta)$ и что угол θ, при котором полная интенсивность отраженного света максимальна, удовлетворяет соотношению

$$\boxed{\sin\theta = \frac{\lambda}{2\Lambda}.} \qquad (2.58)$$

Угол Брэгга

Оно определяет **угол Брэгга** θ. Такая ситуация встречается при отражении света от многослойных структур (см. разд. 7.1) или когда рентгеновские лучи отражаются атомными плоскостями кристаллической структуры. Такое же отражение света происходит от периодической структуры, созданной в среде акустической волной (см. гл. 19). Строгое рассмотрение брэгговского отражения содержится в подразд. 7.1.3.

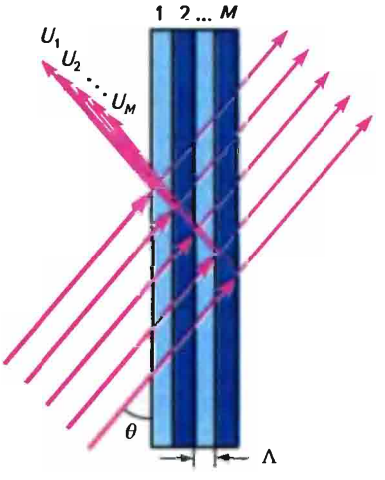


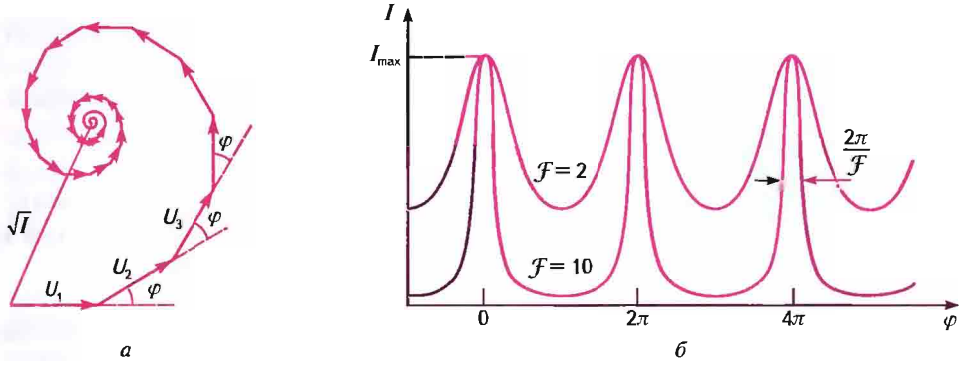
Рис. 2.29. Отражение плоской волны от M параллельных плоскостей, разделенных расстоянием Λ. Отраженные волны конструктивно интерферируют, когда угол θ равен углу Брэгга. Обратите внимание, что угол θ отсчитывается от параллельных плоскостей

Интерференция бесконечного числа волн убывающей амплитуды с равными разностями фаз

Теперь рассмотрим суперпозицию бесконечного числа волн с равной разностью фаз и амплитудами, убывающими по закону геометрической прогрессии:

$$U_1 = \sqrt{I_0}; \quad U_2 = hU_1; \quad U_3 = hU_2 = h^2U_1; \tag{2.59}$$

где $h = |h|e^{j\varphi}$; $|h| < 1$; I_0 — интенсивность первой волны. Амплитуда m-й волны меньше амплитуды $(m-1)$-й в $|h|$ раз, а фаза отличается на φ. Диаграмма векторов на комплексной плоскости показана на рис. 2.30, *а*.



а *б*

Рис. 2.30. Сумма бесконечного числа комплексных векторов, величина которых убывает по закону геометрической прогрессии, а разности фаз φ одинаковы (*а*); зависимость интенсивности I от разности фаз φ для двух значений $\mathcal{F}$ (*б*). Пиковые значения достигаются при $\varphi = 2\pi q$. Полная ширина максимума по полувысоте примерно $2\pi/\mathcal{F}$, когда $\mathcal{F} \gg 1$. Острота пиков растет с увеличением $\mathcal{F}$

Суперпозиция волн имеет комплексную амплитуду

$$U = U_1 + U_2 + U_3 + \ldots = \sqrt{I_0}\,(1 + h + h^2 + \ldots) = \frac{\sqrt{I_0}}{1-h} = \frac{\sqrt{I_0}}{1 - |h| e^{j\varphi}}. \qquad (2.60)$$

Тогда полная интенсивность

$$I = |U|^2 = \frac{I_0}{|1 - |h| e^{j\varphi}|^2} = \frac{I_0}{(1 - |h| \cos\varphi)^2 + |h|^2 \sin^2 \varphi}, \qquad (2.61)$$

откуда

$$I = \frac{I_0}{(1 - |h|)^2 + 4|h| \sin^2 \varphi/2}. \qquad (2.62)$$

Удобно переписать это соотношение в виде

$$\boxed{I = \frac{I_{\max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2(\varphi/2)};} \qquad (2.63)$$

Интенсивность суперпозиции бесконечного числа волн

$$I_{\max} = \frac{I_0}{(1 - |h|)^2},$$

где величина

$$\boxed{\mathcal{F} = \frac{\pi\sqrt{|h|}}{1 - |h|}} \qquad (2.64)$$

Параметр резкости

называется **параметром резкости.**

Интенсивность I — периодическая функция φ с периодом 2π, как показано на рис. 2.30, *б*. Она достигает максимальных значений $I_{\max}$ при $\varphi = 2\pi q$, где q — целое. Это происходит, когда векторы на комплексной плоскости выстраиваются в одном направлении. (Этот результат не отличается от показанного на рис. 2.28, *б* в случае интерференции последовательности M волн одинаковой амплитуды с постоянной разностью фаз.) Когда параметр резкости $\mathcal{F}$ велик (т. е. множитель $|h|$ близок к 1) функция $I(\varphi)$ имеет острые пики. Рассмотрим в качестве типичного примера значения φ вблизи пика $\varphi = 0$. Для $|\varphi| \ll 1$

$$\sin\frac{\varphi}{2} \approx \frac{\varphi}{2}$$

и (2.63) можно переписать как

$$I \approx \frac{I_{\max}}{1 + (\mathcal{F}/\pi)^2 \varphi^2}. \qquad (2.65)$$

Видно, что интенсивность I убывает до половины своего максимального значения при $\varphi = \pi/\mathcal{F}$, так что полная ширина на уровне полумаксимума (ПШПМ)

$$\boxed{\Delta\varphi \approx \frac{2\pi}{\mathcal{F}}.} \qquad (2.66)$$

Ширина интерференционного максимума

В режиме $\mathcal{F} \gg 1$ имеем $\Delta\varphi \ll 2\pi$, и применимо приближение $\varphi \ll 1$. Параметр резкости $\mathcal{F}$ — это отношение периода 2π к ПШПМ пиков интерференционной картины. Следовательно, он является мерой резкости интерференционной картины, т. е. чувствительности к отклонениям φ от значений $2\pi q$, соответствующих пикам интенсивности.

Полезным устройством, основанным на данном принципе, является интерферометр Фабри—Перо. Он состоит из двух параллельных зеркал, между которыми свет претерпевает многократные отражения. В течение каждого обхода свет испытывает фиксированное ослабление амплитуды $|h| = |r|$, связанное с потерями на зеркалах, и фазовый набег

$$\varphi = k2d = \frac{4\pi\nu d}{c} = \frac{2\pi\nu}{c/2d},$$

связанный с распространением волны, где d — расстояние между зеркалами. Полная интенсивность света зависит от фазового сдвига φ в соответствии с (2.63), достигая максимума, когда $\varphi/2$ кратно π. Пропорциональность фазового сдвига φ оптической частоте ν приводит к тому, что коэффициент пропускания интерферометра Фабри—Перо по интенсивности имеет пики, отстоящие друг от друга по частоте на величину $c/2d$. Ширина этих пиков равна $(c/2d)/\mathcal{F}$, где параметр резкости $\mathcal{F}$ определяется потерями [см. (2.64)]. Интерферометр Фабри—Перо, который служит также анализатором спектра, рассмотрен далее в подразд. 7.1.2. Он обычно используется в качестве резонатора в лазерах, что обсуждается в разд. 10.1 и подразд. 15.1.1.

2.6. ПОЛИХРОМАТИЧЕСКИЙ И ИМПУЛЬСНЫЙ СВЕТ

Поскольку волновая функция монохроматического света есть гармоническая функция времени на всей оси (от $-\infty$ до $+\infty$), она является идеализацией и в чистом виде не встречается в реальных ситуациях. Этот раздел посвящен волнам с произвольной зависимостью от времени, включая оптические импульсы конечной длительности во времени. Такие волны являются полихроматическими, а не монохроматическими. Более подробное введение в оптику световых импульсов дано в гл. 22.

2.6.1. Временно́е и спектральное описание

Хотя полихроматическая волна описывается волновой функцией $u(\boldsymbol{r}, t)$ с негармонической зависимостью от времени, ее можно разложить в суперпозицию гармонических функций, каждая из которых представляет монохроматическую волну. Поскольку мы знаем, как монохроматическая волна распространяется в свободном пространстве и через различные оптические элементы, можем определить действие оптической системы на полихроматический свет с помощью принципа суперпозиции.

Метод Фурье позволяет представить произвольную функцию времени $u(t)$, представляющую волновую функцию $u(\boldsymbol{r}, t)$ в фиксированной точке $\boldsymbol{r}$, в виде интегральной суперпозиции гармонических функций различной частоты, амплитуды и фазы

$$u(t) = \int_{-\infty}^{\infty} v(\nu)\exp(j2\pi\nu t)\,\mathrm{d}\nu, \tag{2.67}$$

где $v(\nu)$ определяется путем **преобразования Фурье**

$$v(\nu) = \int_{-\infty}^{\infty} u(t)\exp(-j2\pi\nu t)\,\mathrm{d}t. \tag{2.68}$$

Обзор свойств преобразования Фурье представлен в разд. А.1 приложения А. Разложение (2.67) распространяется как на положительные, так и на отрицательные частоты. Однако, поскольку $u(t)$ действительна, $v(-\nu) = v^*(\nu)$ (см. разд. А.1). Таким образом, отрицательно-частотные компоненты не являются независимыми, они являются просто комплексно сопряженными по отношению к положительно-частотным компонентам.

Комплексное представление

Удобно представить действительную функцию $u(t)$ в (2.67) комплексной функцией

$$U(t) = 2\int_{0}^{\infty} v(\nu)\exp(j2\pi\nu t)\,\mathrm{d}\nu, \tag{2.69}$$

которая включает только положительно-частные компоненты (умноженные на 2), а отрицательно-частотные компоненты подавлены. Фурье-образ $U(t)$ равен $V(\nu) = 2v(\nu)$ при $\nu \geq 0$ и 0 при $\nu < 0$.

Действительная функция $u(t)$ может быть определена по своему комплексному представлению $U(t)$ просто путем взятия действительной части:

$$u(t) = \mathrm{Re}\{U(t)\} = \frac{1}{2}\left[U(t) + U^*(t)\right]. \tag{2.70}$$

Комплексная функция $U(t)$ называется **комплексным аналитическим сигналом**. В справедливости (2.70) можно убедиться, разбивая интеграл в (2.67) на две части: в пределах от 0 до $+\infty$ и от $-\infty$ до 0. Первый интеграл равен $U(t)/2$ благодаря (2.69), а второй вычисляется как

$$\int_{-\infty}^{0} v(\nu)\exp(j2\pi\nu t)\,\mathrm{d}\nu = \int_{0}^{\infty} v(-\nu)\exp(-j2\pi\nu t)\,\mathrm{d}\nu = \\ = \int_{0}^{\infty} v^*(\nu)\exp(-j2\pi\nu t)\,\mathrm{d}\nu = \frac{1}{2}U^*(t). \tag{2.71}$$

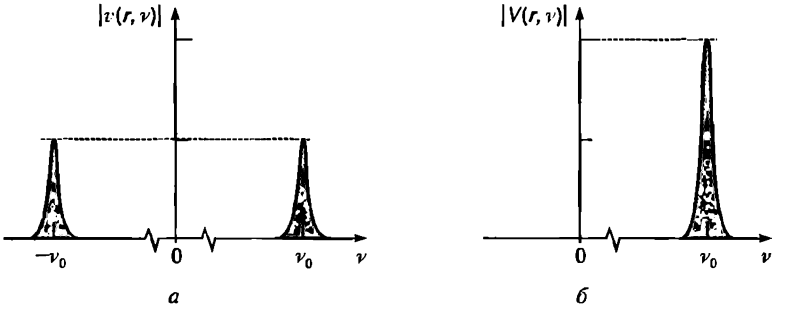
Здесь на первом шаге делается простая замена переменной ν на $-\nu$, на втором шаге используется соотношение симметрии $v(-\nu) = v^*(\nu)$. Окончательный результат состоит в том, что $u(t)$ представляется как сумма комплексной функции $U(t)/2$ и ее комплексно сопряженной, что и подтверждает (2.70).

В качестве простого примера отметим, что комплексное представление действительной гармонической функции есть комплексная гармоническая функция $U(t) = \exp(j\omega t)$. Это комплексное представление, введенное в подразд. 2.2.1 для монохроматических волн. Действительно, комплексное представление полихроматической волны, как оно описано в настоящем разделе, есть просто суперпозиция комплексных представлений каждой из монохроматических компонент Фурье.

Комплексный аналитический сигнал, соответствующий волновой функции $u(\boldsymbol{r}, t)$, называется **комплексной волновой функцией** $U(\boldsymbol{r}, t)$. Поскольку каждая из Фурье-компонент удовлетворяет волновому уравнению, удовлетворяет ему и комплексная волновая функция $U(\boldsymbol{r}, t)$

$$\boxed{\nabla^2 U - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} = 0.} \tag{2.72}$$

Волновое уравнение



а *б*

Рис. 2.31. Модуль $|v(\boldsymbol{r}, \nu)|$ Фурье-образа волновой функции $u(\boldsymbol{r}, t)$ (*а*); модуль $|V(\boldsymbol{r}, \nu)|$ Фурье-образа соответствующей комплексной волновой функции $U(\boldsymbol{r}, t)$ (*б*)

На рис. 2.31 показаны модули Фурье-образов волновой функции $u(\boldsymbol{r}, t)$ и комплексной волновой функции $U(\boldsymbol{r}, t)$. На этой иллюстрации оптическая волна **квазимонохроматическая**, т. е. имеет Фурье-компоненты, частоты которых ограничены узкой полосой ширины $\Delta\nu$ с центральной частотой ν_0, причем $\Delta\nu = \nu_0$.

Интенсивность полихроматической волны

Оптическая интенсивность связана с волновой функцией соотношением (2.3):

$$I(\boldsymbol{r},\ t) = 2\left\langle u^2(\boldsymbol{r},\ t)\right\rangle = 2\left\langle\left\{\frac{1}{2}\left[U(\boldsymbol{r},\ t)+U^*(\boldsymbol{r},\ t)\right]\right\}^2\right\rangle = \\ = \frac{1}{2}\left\langle U^2(\boldsymbol{r},\ t)\right\rangle + \frac{1}{2}\left\langle U^{*2}(\boldsymbol{r},\ t)\right\rangle + \left\langle U(\boldsymbol{r},\ t)+U^*(\boldsymbol{r},\ t)\right\rangle. \tag{2.73}$$

Для квазимонохроматической волны с центральной частотой ν_0 и спектральной шириной $\Delta\nu \ll \nu_0$ среднее $\langle\cdot\rangle$ берется по временно́му интервалу много больше оптического периода $1/\nu_0$, но много меньше, чем $1/\Delta\nu$ (см. разд. 2.1). Поскольку $U(\boldsymbol{r}, t)$ задается выражением (2.70), член U^2 в (2.73) имеет компоненты, осциллирующие на частотах $\approx 2\nu_0$. Аналогично, компоненты U^{*2} осциллируют на частотах $\approx -2\nu_0$. Эти члены, следовательно, при усреднении обращаются в нуль. Однако третий член содержит только разности частот, имеющие порядок $\Delta\nu \ll \nu_0$. Он меняется медленно, и на него не действует процедура усреднения по времени. Таким образом, третий член в (2.73) после усреднения остается, и интенсивность оказывается равной

$$\boxed{I(\boldsymbol{r},\ t) = \left|U(\boldsymbol{r},\ t)\right|^2.} \tag{2.74}$$

Оптическая интенсивность

Оптическая интенсивность квазимонохроматической волны есть квадрат модуля ее комплексной волновой функции.

Простота этого результата фактически оправдывает введение понятия комплексной волновой функции.

Импульсная плоская волна

Простейшим примером светового импульса является импульс плоской волны. Его комплексная волновая функция имеет вид

$$U(\boldsymbol{r},\ t) = \mathcal{A}\left(t-\frac{z}{c}\right)\exp\left[j2\pi\nu_0\left(t-\frac{z}{c}\right)\right], \tag{2.75}$$

где комплексная амплитуда $\mathcal{A}(t)$ есть функция времени; ν_0 — центральная оптическая частота. Монохроматическая плоская волна является частным случаем (2.75), для которого $\mathcal{A}(t)$ постоянна, т. е.

$$U(\boldsymbol{r},\ t) = \mathcal{A}\exp\left[j2\pi\nu_0\left(t-\frac{z}{c}\right)\right] = \mathcal{A}\exp(-jk_0z)\exp(-j\omega_0t),$$

где $k_0 = \omega_0/c$ и $\omega_0 = 2\pi\nu_0$.

Поскольку $U(\boldsymbol{r}, t)$ в (2.75) есть функция $(t - z/c)$, она удовлетворяет волновому уравнению (2.72), независимо от вида функции $\mathcal{A}$ (при условии, что $d^2\mathcal{A}/dt^2$ существует). Это можно проверить прямой подстановкой.

Если $\mathcal{A}(t)$ имеет конечную длительность τ, то в любой фиксированной точке z волна продолжается в течение времени τ, а в любой момент времени t имеет пространственную протяженность $c\tau$. Следовательно, это **волновой пакет** фиксированной протяженности, бегущий вдоль оси z (рис. 2.32). Например, импульс длительностью $\tau = 1$ пс занимает в свободном пространстве область длиной $c\tau = 0{,}3$ мм.

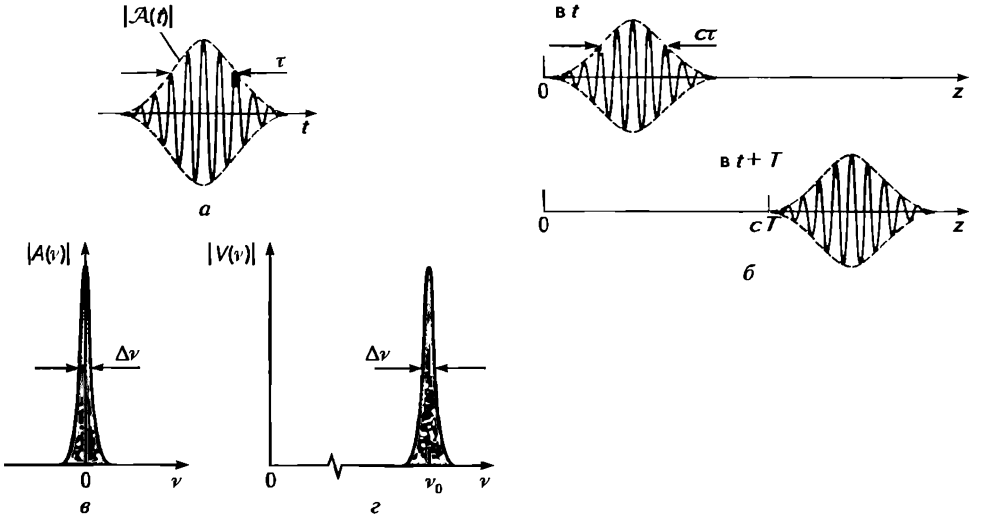


Рис. 2.32. Временны́е, пространственные и спектральные характеристики импульса плоской волны:

а — волновая функция в заданной точке имеет длительность τ; *б* — волновая функция как функция координаты в моменты времени t и $t + T$. Импульс движется со скоростью c и занимает в пространстве длину $c\tau$; *в* — модуль Фурье-образа комплексной огибающей $|A(\nu)|$; *г* — модуль $|V(\nu)|$ Фурье-образа комплексной волновой функции $|V(\nu)|$ с центром на частоте ν_0

Фурье-образ комплексной волновой функции (2.75)

$$V(\boldsymbol{r}, t) = A(\nu - \nu_0)\exp\left(-\frac{j2\pi\nu z}{c}\right), \tag{2.76}$$

где $A(\nu)$ — Фурье-образ $\mathcal{A}(t)$. Это можно показать с использованием свойства частотного сдвига преобразования Фурье (см. разд. А.1 приложения А). Комплексная огибающая $\mathcal{A}(t)$ часто является медленно меняющейся по сравнению с оптическими колебаниями, поэтому ее Фурье-образ $A(\nu)$ имеет спектральную ширину $\Delta\nu$ много меньшую, чем центральная частота ν_0. Спектральная ширина $\Delta\nu$ обратно пропорциональна временно́й длительности τ. В частности, если функция $\mathcal{A}(t)$ гауссова, то ее Фурье-образ $A(\nu)$ — тоже гауссова функция. Если

временну́ю и спектральную ширину определить как среднеквадратичные ширины по мощности, то их произведение равно $1/4\pi$ (см. разд. А.2 приложения А). Например, если $\tau = 1$ пс, то $\Delta\nu = 80$ ГГц. Если центральная частота ν_0 составляет $5 \cdot 10^{14}$ Гц (что соответствует $\lambda_0 = 0{,}6$ мкм), то $\Delta\nu/\nu_0 = 1{,}6 \cdot 10^{-4}$, так что свет квазимонохроматичен. На рис. 2.32 показаны временны́е, пространственные и спектральные характеристики импульса плоской волны.

Распространение импульса плоской волны через среду с зависящим от частоты показателем преломления (т. е. с зависящей от частоты скоростью света $c = c_0/n$) обсуждается в подразд. 5.5.2, а гл. 22 освещает другие аспекты оптики импульсов.

2.6.2. Световые биения

Зависимость интенсивности полихроматической волны от времени можно объяснить интерференцией образующих ее монохроматических компонент. Сейчас мы продемонстрируем это на двух примерах: интерференция двух монохроматических волн и интерференция конечного числа монохроматических волн.

Интерференция двух монохроматических волн с различными частотами

Оптическая волна, образованная двумя монохроматическими волнами с частотами ν_1 и ν_2 и интенсивностями I_1 и I_2, имеет в некоторой точке пространства волновую функцию

$$U(t) = \sqrt{I_1}\exp(j2\pi\nu_1 t) + \sqrt{I_2}\exp(j2\pi\nu_2 t), \tag{2.77}$$

где фазы взяты равными нулю, а зависимость от $\boldsymbol{r}$ для удобства не показана. Интенсивность полной волны определяется уравнением интерференции (2.49)

$$I(t) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\left[2\pi(\nu_2 - \nu_1)t\right]. \tag{2.78}$$

Видно, что интенсивность изменяется во времени синусоидально на разностной частоте $|\nu_1 - \nu_2|$, называемой **частотой биений**. Данное явление известно под несколькими названиями: **оптические биения, оптическое смешение, фотосмешение** и **оптическое гетеродинирование**.

Уравнение (2.78) аналогично уравнению (2.52), которое описывает *пространственную* интерференцию двух волн одинаковой частоты, распространяющихся в различных направлениях. Это можно понять на основе векторной диаграммы на комплексной плоскости (см. рис. 2.22). Два вектора U_1 и U_2 вращаются с угловыми скоростями $\omega_1 = 2\pi\nu_1$ и $\omega_2 = 2\pi\nu_2$, так что разность их углов наклона равна $\varphi = \varphi_2 - \varphi_1 = 2\pi(\nu_2 - \nu_1)t$ в соответствии с (2.78). Волны с разными частотами, распространяющиеся в разных направлениях, проявляют пространственно-временну́ю интерференцию.

В электронике говорят, что биения или смешение происходят, когда сумма двух синусоидальных сигналов регистрируется нелинейным (например, квадратичным) устройством, которое называется смесителем, вырабатывающим сигналы на разностной и суммарной частотах. Такое устройство используется в гетеродинных радиоприемниках. В оптике фотоприемники регистрируют оптическую интенсивность (см. гл. 18), которая, согласно (2.74), пропорциональна квадрату абсолютной величины комплексной волновой функции. Оптические приемники, следовательно, регистрируют только разностную частоту.

Аналогично тому, как (2.52) позволяет определить направление волны по интерференционной картине на экране, (2.78) дает возможность определить оптическую частоту по временной интерференционной картине на выходе фотоприемника. Использование оптических биений в оптических гетеродинных приемниках обсуждается в разд. 24.5. Другие виды оптического смешения позволяют нелинейным средам служить источниками излучения на суммарных и разностных частотах, как описано в гл. 21.

Упражнение 2.12

Оптический доплеровский радар

За счет **эффекта Доплера** монохроматическая волна частоты ν, отразившись от движущегося объекта, составляющая скорости которого вдоль линии наблюдения равна v, претерпевает частотный сдвиг $\Delta\nu = \pm(2v/c)\nu$, причем знак зависит от того, приближается объект к наблюдателю (+) или удаляется от него (−). Предполагая, что исходная и отраженная волны налагаются друг на друга, выведите выражение для интенсивности образующейся волны. Предложите метод измерения скорости объекта с помощью этого принципа. В случае если одно из зеркал интерферометра Майкельсона (см. рис. 2.24, *б*) движется со скоростью $\pm v$, покажите с помощью (2.51), что частота биений равна $\pm(2v/c)\nu$.

Интерференция M монохроматических волн с одинаковыми интенсивностями и равноотстоящими частотами

Интерференция большого числа монохроматических волн с одинаковыми интенсивностями, одинаковыми фазами и равноотстоящими частотами может приводить к генерации коротких импульсов света. Рассмотрим нечетное число волн $M = 2L + 1$, интенсивность каждой равна I_0, фаза равна нулю, а частоты образуют ряд

$$\nu_q = \nu_0 + q\nu_F, \quad q = -L, \ldots, 0, \ldots, L, \tag{2.79}$$

с центром на частоте ν_0 и разностью соседних частот $\nu_F \ll \nu_0$. В данной точке полная волна имеет комплексную волновую функцию

$$U(t) = \sqrt{I_0} \sum_{q=-L}^{L} \exp\left[j2\pi(\nu_0 + q\nu_F)t\right]. \tag{2.80}$$

Она представляет собой сумму M комплексных векторов равной длины и последовательностью фаз, отличающихся друг от дуга на величину $\varphi = 2\pi\nu_F t$. Интенсивность вычисляется точно таким же образом, как в подразд. 2.5.2. Используя (2.57) и рис. 2.28 и подставляя $\varphi = 2\pi t/T_F$, где $T_F = 1/\nu_F$, для полной интенсивности находим

$$I(t) = |U(t)|^2 = I_0 \frac{\sin^2(M\pi t/T_F)}{\sin^2(\pi t/T_F)}. \qquad (2.81)$$

Как показано на рис. 2.33, интенсивность $I(t)$ представляет собой периодическую последовательность импульсов с периодом T_F, пиковой интенсивностью $M^2 I_0$ и средней интенсивностью $\bar{I} = MI_0$. Таким образом, пиковая интенсивность оказывается в M раз больше средней. Длительность каждого импульса примерно T_F/M, так что при больших M импульсы становятся очень короткими. Если, например, $\nu_F = 1$ ГГц, то $T_F = 1$ нс, и для $M = 1000$ получаются импульсы длительностью в 1 пс.

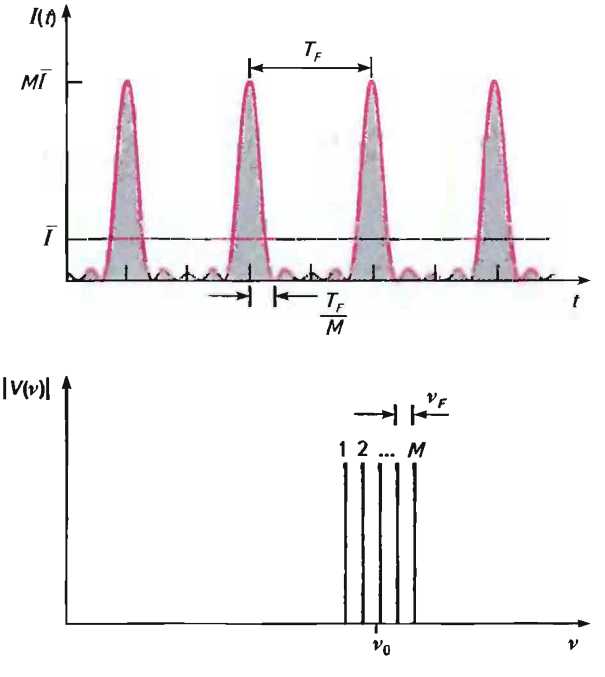


Рис. 2.33. Временна́я зависимость полной интенсивности $I(t)$ полихроматической волны, образованной сложением M монохроматических волн с одинаковыми интенсивностями, одинаковыми фазами и последовательными частотами, отстоящими друг от друга на ν_F. Интенсивность $I(t)$ представляет собой периодическую последовательность импульсов с периодом $T_F = 1/\nu_F$, пиковая интенсивность которых в M раз больше средней интенсивности $\bar{I}$. Длительность каждого импульса в M раз меньше периода следования. В данном примере $M = 5$. Сравните эти графики с изображенными на рис. 2.28. На нижнем графике показан модуль Фурье-образа $|V(\nu)|$

Этот пример убедительно показывает, как с помощью M монохроматических волн можно получить цуг очень коротких оптических импульсов. В подразд. 15.4.4 мы увидим, как моды лазера можно *синхронизировать* описанным образом для генерации последовательностей ультракоротких лазерных импульсов.

Рекомендуемая литература

КНИГИ ПО ВОЛНОВОЙ ОПТИКЕ И ИНТЕРФЕРОМЕТРИИ

См. также список литературы для чтения к гл. 1.

Pierce J.R. Almost All About Waves. MIT Press, 1974; Dover, reissued 2006.

Pain H.J. The Physics of Vibrations and Waves. Wiley, 6th ed. 2005.

Webb R.H. Elementary Wave Optics. Academic Press, 1969; Dover, 2005

Hariharan P. Optical Interferometry. Academic Press, 2nd ed. 2003.

Mansuripur M. Classical Optics and Its Applications. Cambridge University Press, 2002.

Lipson S.G., Lipson H., and *Tannhauser D.S.* Optical Physics. Cambridge University Press 3rd ed., 1998.

Akhmanov S.A., Nikitin S.Yu. Physical Optics. Oxford University Press, 1997.

Mickelson A.R. Physical Optics. Van Nostrand Reinhold, 1992.

Vaughan J.M. The Fabry-Perot Interferometer. Adam Hilger, 1989.

Young H.D. Fundamentals of Waves, Optics, and Modern Physics. McGraw-Hill, paperback 2nd ed. 1976.

Tolansky S. An Introduction to Interferometry. Wiley, 2nd ed. 1973.

Françon M., Krauzman N., Matieu J.P., May M. Experiments in Physical Optics. Gordon and Breach, 1970.

Françon M. Optical Interferometry. Academic Press, 1966.

КНИГИ ПО СПЕКТРОСКОПИИ

Hollas J.M. Modern Spectroscopy. Wiley, 4th ed. 2004.

Kauppinen J., Partanen J. Fourier Transforms in Spectroscopy. Wiley-VCH, 2001.

Christy A.A., Ozaki Y. and *Gregoriou V.G.* Modern Fourier Transform Infrared Spectroscopy. Elsevier, 2001.

Pavia D.L., Lampman G.M., Kriz G.S. Introduction to Spectroscopy. Brooks/Cole. paperback 3rd ed. 2000.

Smith B.C. Fundamentals of Fourier Transform Infrared Spectroscopy, CRC Press, 1996.

КНИГИ ПО ДИФРАКЦИОННЫМ РЕШЕТКАМ

Palmer C. Diffraction Grating Handbook. Richardson Grating Laboratory (Newport Corporation/Spectra-Physics, Irvine, CA) 4th ed. 2000.

Loewen E.G., Popov E. Diffraction Gratings and Application. Marcel Dekker, 1997.

ПОПУЛЯРНАЯ И ИСТОРИЧЕСКАЯ ЛИТЕРАТУРА

Buchwald J.Z. The Rise of the Wave Theory of Light: Optical Theory and Experiment in the Early Nineteenth Century. University of Chicago Press, paperback ed. 1989.

Kock W.E. Sound Waves and Light Waves. Doubleday/Anchor, 1965.

Huygens C. Treatise on Light. 1690, University of Chicago Press, 1945.

СТАТЬИ

Bell T.E. Waiting for Gravity. IEEE Spectrum. Vol. 43, № 7, 2006. P 40—46.

Kamerman G.W., ed. Selected Papers on Laser Radar. SPIE Optical Engineering Press (Milestone Series. Vol. 133), 1997.

Maystre D., ed. Selected Papers 011 Diffraction Gratings. SPIE Optical Engineering Press (Milestone Series. Vol. 83), 1993.

Hariharan P., ed. Selected Papers on Interferolnetry. SPIE Optical Engineering Press (Milestone Series. Vol. 28), 1991.

Задачи

К РАЗДЕЛУ 2.2

1. **Сферические волны.** Пользуясь сферической системой координат, проверьте, что комплексная амплитуда сферической волны (2.19) удовлетворяет уравнению Гельмгольца (2.11).

2. **Интенсивность сферической волны.** Получите выражение для интенсивности I сферической волны на расстоянии r от ее центра через оптическую мощность P. Какова интенсивность на расстоянии $r = 1$ м от источника мощностью $P = 100$ Вт?

3. **Цилиндрические волны.** Выведите выражения для комплексной амплитуды и интенсивности монохроматической волны, волновые фронты которой представляют собой цилиндрические поверхности с осью, направленной по координатной оси y.

4. **Параксиальное уравнение Гельмгольца.** Выведите параксиальное уравнение Гельмгольца (2.27), используя приближения в (2.25) и (2.26).

5. **Сопряженные волны.** Сравните монохроматическую волну с комплексной амплитудой $U(\boldsymbol{r})$ с монохроматической волной той же частоты, но с комплексно сопряженной амплитудой $U^*(\boldsymbol{r})$, по интенсивности, форме волновых фронтов и расположению нормалей к ним. В качестве примеров используйте плоскую волну

$$U(\boldsymbol{r}) = A \exp\left[-jk\frac{x+y}{\sqrt{2}}\right]$$

и сферическую волну $U(\boldsymbol{r}) = \exp(-jkr)$.

К РАЗДЕЛУ 2.3

1. **Волна в слое с градиентом показателя преломления.** Нарисуйте волновые фронты волны, распространяющейся в слое SELFOC с градиентом показателя преломления, описанном в примере 1.8.

К РАЗДЕЛУ 2.4

1. **Отражение сферической волны от плоского зеркала.** Сферическая волна отражается от плоского зеркала, расположенного достаточно далеко от ее источника, чтобы выполнялось приближение Френеля. Рассматривая сферическую волну локально как плоскую волну с медленно меняющимся направлением, используйте закон отражения плоской волны для определения природы отраженной волны.

2. **Оптическая длина пути.** Плоская волна распространяется в направлении, нормальном к тонкой пластине, состоящей из N тонких параллельных слоев толщины d_q с показателями преломления n_q, $q = 1, 2, \ldots, N$. Пренебрегая всеми отражениями, определите комплексный амплитудный коэффициент пропускания пластины. Если заменить пластину слоем свободного пространства, то какой должна быть его толщина d, чтобы получить то же комплексный амплитудный коэффициент пропускания? Покажите, что это расстояние есть оптическая длина пути, определенная в разд. 1.1.

3. **Дифракционная решетка.** Повторите упражнение 2.7 для тонкой прозрачной пластины, толщина которой $d(x, y)$ — квадратная (вместо синусоидальной) пери-

одическая функция x с периодом $\Lambda \gg \lambda$. Покажите, что угол θ между дифрагирующими волнами по-прежнему равен $\theta \approx \lambda/\Lambda$. Для нормального падения плоской волны на решетку определите амплитуды различных дифрагирующих плоских волн.

4. **Отражение от сферического зеркала.** Покажите, что комплексный амплитудный коэффициент отражения $r(x, y)$ (отношение комплексных амплитуд отраженной и падающей волн) тонкого сферического зеркала радиуса R дается выражением

$$r(x, y) = h_0 \exp\left[-jk_0 \frac{x^2 + y^2}{R}\right],$$

где h_0 — постоянная. Сравните это с комплексным амплитудным коэффициентом пропускания линзы с фокусным расстоянием $f = -R/2$.

К РАЗДЕЛУ 2.5

1. **Стоячие волны.** Выведите выражение для интенсивности I суперпозиции двух плоских волн с длиной волны λ, распространяющихся навстречу друг другу вдоль оси z. Постройте график I от z.

2. **Видность полос.** Видность интерференционной картины, такой как описанная (2.49) и изображенная на рис. 2.22, определяется как отношение

$$\mathcal{V} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}},$$

где I_{max} и I_{min} — максимальное и минимальное значения I. Выведите выражение для $\mathcal{V}$ как функции отношения I_1/I_2 интенсивностей интерферирующих волн и определите величину I_1/I_2, при которой видность максимальна.

3. **Интерферометр Майкельсона.** Если одно зеркало интерферометра Майкельсона (см. рис. 2.24, *б*) наклонено на малый угол $\Delta\theta$, опишите форму интерференционной картины в плоскости регистрации. Что происходит с картиной при движении другого зеркала.

К РАЗДЕЛУ 2.6

1. **Импульс сферической волны.**

а. Покажите, что импульс сферической волны имеет комплексную волновую функцию вида

$$U(r, t) = \frac{1}{r} a\left(t - \frac{r}{c}\right),$$

где $a(t)$ — произвольная функция.

б. Ультракороткий оптический импульс имеет комплексную волновую функцию с центральной частотой, соответствующей $\lambda_0 = 585$ нм, и гауссовой среднеквадратичной длительностью $\sigma_t = 6$ фс (1 фс = 10^{-15} с). Сколько оптических колебаний содержится в пределах длительности импульса? Если импульс распространяется в свободном пространстве в виде сферической волны, испущенной из начала координат в момент времени $t = 0$, опишите радиальное распределение интенсивности в пространстве в момент времени $t = 1$ пс.

ГЛАВА 3 ОПТИКА ПУЧКОВ

Гауссовы пучки носят имя великого немецкого математика **Карла Фридриха Гаусса (1777—1855)**

Лорд Рэлей (Джон Вильям Страт) (1842—1919) внес вклад во многие области оптики. Глубина резкости гауссова пучка названа в его честь

Можно ли пространственно ограничить свет и передавать его в свободном пространстве без угловой расходимости? Хотя волновая природа света препятствует такой идеализированной передаче, свет на самом деле может быть сформирован в виде ограниченного пучка, который как угодно близко приближается к пространственно локализованной и не расходящейся волне.

Два предельных случая углового и пространственного ограничения — это плоская волна и сферическая волна. Нормали к волновому фронту (лучи) плоской волны совпадают по направлению с направлением распространения волны, так что угловая расходимость отсутствует, но энергия распространяется, заполняя все пространство. Напротив, сферическая волна берет начало из одной точки, но нормали к ее волновым фронтам (лучи) расходятся во всех направлениях.

Волны, у которых нормали к волновым фронтам образуют малые углы с осью z, называются параксиальными волнами. Они должны удовлетворять параксиальному уравнению Гельмгольца, выведенному в подразд. 2.2.3. Гауссов пучок — важное решение этого уравнения, демонстрирующее характеристики

пучка, которые можно определить так. Мощность пучка главным образом сосредоточена внутри малого цилиндра, окружающего ось пучка. Распределение интенсивности по любому поперечному сечению представляет собой симметричную гауссову функцию с центром на оси пучка. Ширина этого распределения минимальна в перетяжке пучка и постепенно растет вдоль оси пучка с удалением от перетяжки в обоих направлениях. Волновые фронты, почти плоские вблизи перетяжки, постепенно искривляются по мере удаления от нее и в конце концов становятся приблизительно сферическими вдали от перетяжки. Угловая расходимость нормалей к волновым фронтам принимает наименьшее значение, разрешаемое волновым уравнением при данной ширине пучка. При этом нормали к волновым фронтам образуют тонкий пучок лучей. В идеальных условиях многие типы лазеров излучают свет в виде гауссовых пучков.

О данной главе

Выражение для комплексной амплитуды гауссова пучка предложено в разд. 3.1, при этом проводится подробное обсуждение его физических свойств (интенсивности, мощности, ширины, расходимости, глубины резкости и фазы). Формирование гауссовых пучков (фокусировка, преобразование, коллимирование и расширение) различными оптическими элементами является предметом разд. 3.2. В разд. 3.3 мы вводим более общее семейство оптических пучков, называемых пучками Гаусса—Эрмита. Наконец, в разд. 3.4 обсуждаются пучки Лагерра—Гаусса и Бесселя.

3.1. ГАУССОВ ПУЧОК

3.1.1. Комплексная амплитуда

Понятие параксиальных волн было введено в подразд. 2.2.3. Параксиальная волна — это плоская волна $\exp(-jkz)$, распространяющаяся вдоль направления z, с волновым числом $k = 2\pi/\lambda$ и длиной волны λ, модулированная комплексной огибающей $A(\boldsymbol{r})$, медленно меняющаяся в зависимости от координат (см. рис. 2.7), так что комплексная амплитуда записывается как

$$U(\boldsymbol{r}) = A(\boldsymbol{r})\exp(-jkz). \tag{3.1}$$

Огибающая берется почти постоянной в пределах длины порядка λ, так что волна остается локально плоской, а нормали к ее волновым фронтам представляют собой параксиальные лучи.

Чтобы комплексная амплитуда $U(\boldsymbol{r})$ удовлетворяла уравнению Гельмгольца $\nabla^2 U + k^2 U = 0$, комплексная огибающая $A(\boldsymbol{r})$ должна удовлетворять параксиальному уравнению Гельмгольца

$$\nabla_T^2 A - j2k\frac{\partial A}{\partial z} = 0, \tag{3.2}$$

где ∇_T^2 — поперечный оператор Лапласа,

$$\nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

Простое решение параксиального уравнения Гельмгольца дает параболическую волну (см. упражнение 2.2), для которой

$$A(\boldsymbol{r}) = \frac{A_1}{z}\exp\left(-jk\frac{\rho^2}{2z}\right); \quad \rho^2 = x^2 + y^2, \tag{3.3}$$

где A_1 — постоянная. Параболоидальная волна является параксиальным приближением сферической волны

$$U(\boldsymbol{r}) = \frac{A_1}{r}\exp(-ikr),$$

если x и y много меньше z (см. подразд. 2.2.2).

Другое решение параксиального уравнения Гельмгольца приводит к гауссову пучку. Оно получается из параболоидальной волны путем простого преобразования. Поскольку комплексная огибающая параболоидальной волны (3.3) является решением параксиального уравнения Гельмгольца (3.2), то таковым же является результат ее сдвига, т. е. замены z на $z - \xi$, где ξ — постоянная:

$$A(\boldsymbol{r}) = \frac{A_1}{q(z)}\exp\left(-jk\frac{\rho^2}{2q(z)}\right); \quad q(z) = z - \xi. \tag{3.4}$$

Это выражение представляет собой параболоидальную волну с центром в точке $z = \xi$ вместо $z = 0$. Функция (3.4) остается решением уравнения (3.2) и в случае, когда ξ — комплексное, однако, свойства этого решения резко меняются. В частности, если ξ — чисто мнимое, скажем, $\xi = -jz_0$, где z_0 — действительное, то (3.4) даст комплексную огибающую гауссова пучка

$$\boxed{A(\boldsymbol{r}) = \frac{A_1}{q(z)}\exp\left(-jk\frac{\rho^2}{2q(z)}\right); \quad q(z) = z + jz_0.} \tag{3.5}$$

Комплексная огибающая

Величина $q(z)$ называется ***q*-параметром** пучка, а параметр z_0 — **рэлеевской длиной**.

Для выделения амплитуды и фазы комплексной огибающей запишем действительную и мнимую части комплексной функции

$$\frac{1}{q(z)} = \frac{1}{z + jz_0}$$

с помощью двух новых действительных функций $R(z)$ и $W(z)$, таких что

$$\boxed{\frac{1}{q(z)} = \frac{1}{R(z)} - j\frac{\lambda}{\pi W^2(z)}.} \tag{3.6}$$

Ниже будет показано, что $W(z)$ и $R(z)$ определяют ширину пучка и радиус кривизны волнового фронта соответственно. Зависимость $W(z)$ и $R(z)$ от z и z_0 определяется выражениями (3.8) и (3.9). Подстановка (3.6) в (3.5) с использованием (3.1) приводит непосредственно к выражению для комплексной амплитуды $U(\boldsymbol{r})$ гауссова пучка

$$U(\boldsymbol{r}) = A_0 \frac{W_0}{W(z)} \exp\left[-\frac{\rho^2}{W^2(z)}\right] \exp\left[-jkz - jk\frac{\rho^2}{2R(z)} + j\zeta(z)\right]; \quad (3.7)$$

Комплексная амплитуда

$$W(z) = W_0\sqrt{1 + \left(\frac{z}{z_0}\right)^2}; \quad (3.8)$$

$$R(z) = z\left[1 + \left(\frac{z_0}{z}\right)^2\right]; \quad (3.9)$$

$$\zeta(z) = \operatorname{arctg}\frac{z}{z_0}; \quad (3.10)$$

$$W_0 = \sqrt{\frac{\lambda z_0}{\pi}}. \quad (3.11)$$

Параметры пучка

Здесь для удобства введена новая константа

$$A_0 = \frac{A_1}{jz_0}.$$

Выражение для комплексной амплитуды гауссова пучка, приведенное выше, является центральным для данной главы. Оно содержит два независимых параметра A_0 и z_0, которые определяются из начальных условий. Все остальные параметры связаны с z_0 и длиной волны λ соотношениями (3.8)—(3.11). Значение этих параметров станет ясно из дальнейшего изложения.

3.1.2. Свойства

Используем соотношения (3.8)—(3.11) для определения свойств гауссова пучка.

Интенсивность

Оптическая интенсивность $I(\boldsymbol{r}) = |U(\boldsymbol{r})|^2$ есть функция аксиальной z и радиальной $\rho = \sqrt{x^2 + y^2}$ координат точки

$$I(\rho, z) = I_0\left[\frac{W_0}{W(z)}\right]^2 \exp\left[-\frac{2\rho^2}{W^2(z)}\right], \quad (3.12)$$

где $I_0 = |A_0|^2$. При любом значении z интенсивность является гауссовой функцией радиальной координаты ρ — отсюда название «гауссов пучок». Гауссова функция имеет максимум на оси z, при $\rho = 0$, и монотонно убывает с ростом ρ. Ширина гауссова распределения интенсивности в пучке $W(z)$ растет с увеличением аксиальной координаты z, как показано на рис. 3.1.

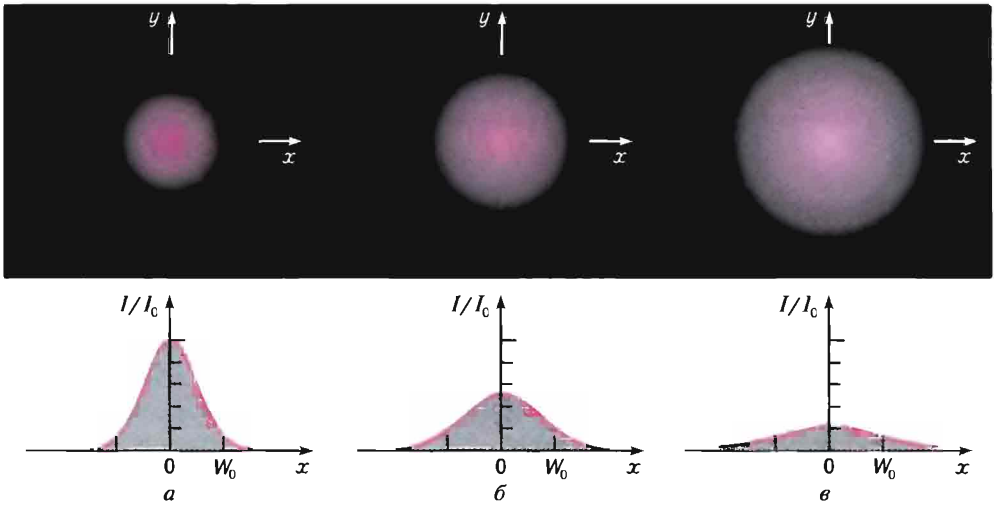


Рис. 3.1. Нормированная интенсивность пучка I/I_0 как функция радиальной координаты ρ при разных значениях аксиальной координаты:
а — $z = 0$; *б* — $z = z_0$; *в* — $z = 2z_0$

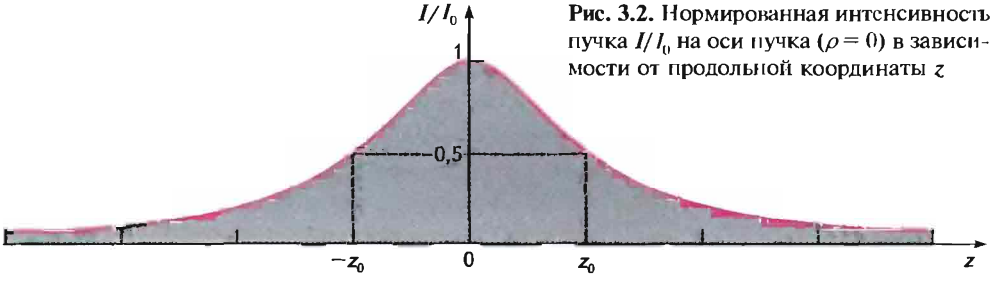


Рис. 3.2. Нормированная интенсивность пучка I/I_0 на оси пучка ($\rho = 0$) в зависимости от продольной координаты z

На оси пучка ($\rho = 0$) интенсивность в (3.12) выражается формулой

$$I(0,\ z) = I_0\left[\frac{W_0}{W(z)}\right]^2 = \frac{I_0}{1+\left(z/z_0\right)^2}. \qquad (3.13)$$

Она имеет максимальное значение I_0 при $z = 0$ и плавно убывает с ростом z, достигая половины максимального значения при $z = \pm z_0$ (рис. 3.2). Когда

$$|z| \gg z_0; \quad I(0,\ z) \approx I_0\frac{z_0^2}{z^2},$$

так что интенсивность убывает с расстоянием обратно пропорционально его квадрату, как у сферических и параболоидальных волн. Центр пучка ($z = 0$, $\rho = 0$) есть точка наибольшей интенсивности $I(0, 0) = I_0$.

Мощность

Полная оптическая мощность, переносимая пучком, есть интеграл от оптической интенсивности по любой поперечной плоскости $z = \text{const}$

$$P = \int_0^\infty I(\rho, z) 2\pi\rho \, d\rho, \tag{3.14}$$

вычисление которого дает

$$P = \frac{1}{2} I_0 (\pi W_0^2). \tag{3.15}$$

Таким образом, мощность пучка равна половине максимальной интенсивности, умноженной на площадь сечения пучка. Как и ожидалось, этот результат не зависит от z. Поскольку оптические пучки часто описываются своей мощностью P, полезно выразить через нее I_0 с помощью (3.15), после чего (3.12) можно переписать в виде

$$\boxed{I(\rho, z) = \frac{2P}{\pi W^2(z)} \exp\left[-\frac{2\rho^2}{W^2(z)}\right].} \tag{3.16}$$

Интенсивность пучка

Отношение мощности, переносимой через площадь круга радиуса ρ_0 в поперечной плоскости к полной мощности в точке z:

$$\frac{1}{P} \int_0^{\rho_0} I(\rho, z) 2\pi\rho \, d\rho = 1 - \exp\left[-\frac{2\rho_0^2}{W^2(z)}\right]. \tag{3.17}$$

Мощность, проходящая через круг радиусом $\rho_0 = W(z)$, составляет примерно 86 % полной мощности. Около 99 % мощности сосредоточено внутри круга радиусом $1{,}5W(z)$.

Ширина пучка

В любой поперечной плоскости интенсивность пучка принимает свое максимальное значение на оси пучка и спадает до $1/e^2 \approx 0{,}135$ максимального значения на расстоянии $\rho = W(z)$ от оси. Поскольку 86 % мощности переносится внутри круга радиуса $W(z)$, мы будем рассматривать $W(z)$ как радиус (или ширину) пучка. С другой стороны, среднеквадратичная ширина распределения интенсивности выражается как $\sigma = W(z)/2$ (см. различные определения ширины в приложении А, разд. А.2).

Зависимость ширины пучка от z определяется (3.8)

$$W(z) = W_0\sqrt{1+\left(\frac{z}{z_0}\right)^2}, \tag{3.18}$$

Ширина пучка (радиус пучка)

Она принимает минимальное значение W_0 в плоскости $z = 0$. Это перетяжка пучка, поэтому W_0 известно как **радиус перетяжки**. Диаметр перетяжки $2W_0$ называется также **размером пятна**. Ширина пучка монотонно растет с увеличением z и принимает значение $\sqrt{2}W_0$ при $z = \pm z_0$ (рис. 3.3).

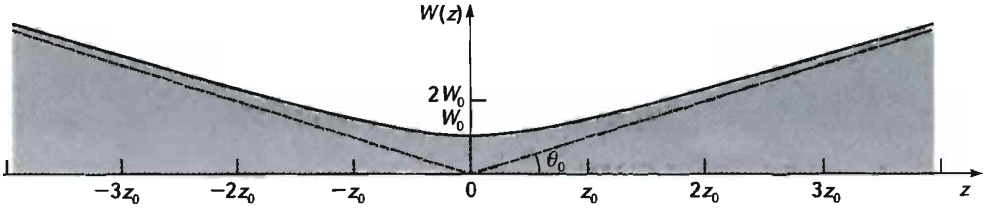


Рис. 3.3. Ширина пучка $W(z)$ принимает свое минимальное значение в перетяжке пучка ($z = 0$), достигает $\sqrt{2}W_0$ при $z = \pm z_0$ и растет пропорционально z при больших z

Расходимость пучка

При $z \gg z_0$ первым членом (3.18) можно пренебречь, что приводит к линейному соотношению

$$W(z) \approx \frac{W_0}{z_0} z = \theta_0 z. \tag{3.19}$$

Как видно на рис. 3.3, пучок расходится как конус с половиной угла при вершине

$$\theta_0 = \frac{W_0}{z_0} = \frac{\lambda}{\pi W_0}, \tag{3.20}$$

где было использовано (3.11). Примерно 86 % мощности пучка ограничено этим конусом в соответствии с (3.17).

Выражая (3.20) через размер пятна, можем записать угол расходимости как

$$2\theta_0 = \frac{4}{\pi}\frac{\lambda}{2W_0}. \tag{3.21}$$

Угол расходимости

Угол расходимости прямо пропорционален длине волны W_0 и обратно пропорционален размеру пятна $2W_0$. Сжатие пятна, таким образом, приводит к увеличению расходимости пучка. Ясно, что высоконаправленный пучок можно сформировать, используя малую длину волны и делая перетяжку достаточно широкой.

Глубина резкости

Поскольку пучок имеет минимальную ширину при $z = 0$, как видно на рис. 3.3, то в этой точке он достигает наилучшей фокусировки. В обоих направлениях пучок постепенно «выходит из фокуса». Расстояние вдоль оси, на котором ширина пучка не более чем в $\sqrt{2}$ превышает минимальное значение, так что площадь пучка удваивается по сравнению с минимальной, называется **глубиной резкости**, или **конфокальным параметром** (рис. 3.4). Из (3.18) и (3.11) ясно, что глубина резкости равна удвоенному числу Рэлея:

$$2z_0 = \frac{2\pi W_0^2}{\lambda}. \qquad (3.22)$$

Глубина резкости

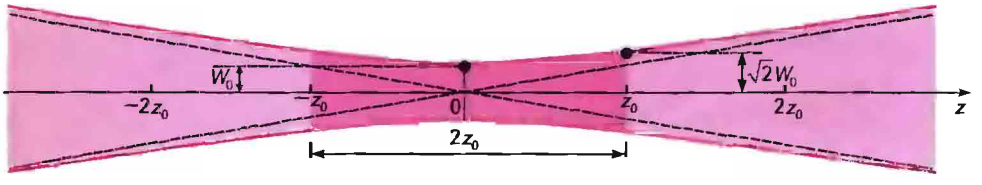


Рис. 3.4. Глубина резкости гауссова пучка

Таким образом, глубина резкости прямо пропорциональна площади сечения пучка в перетяжке πW_0^2 и обратно пропорциональна длине волны λ. Пучок, фокусируемый в малое пятно, имеет малую глубину резкости, так что установление положения фокальной плоскости требует повышенной точности. Малый размер пятна и большая глубина резкости могут быть одновременно достигнуты только при малой длине волны. Например, при $\lambda_0 = 633$ нм (обычная длина волны He—Ne-лазера) размер пятна $2W_0 = 2$ см соответствует глубине резкости $2z_0 \approx 1$ км. Намного меньший размер пятна в 20 мкм соответствует малой глубине резкости 1 мм.

Фаза

Согласно (3.7), фаза гауссова пучка есть

$$\varphi(\rho, z) = kz - \zeta(z) + \frac{k\rho^2}{2R(z)}. \qquad (3.23)$$

На оси пучка ($\rho = 0$) фаза включает два слагаемых:

$$\varphi(0, z) = kz - \zeta(z). \qquad (3.24)$$

Первое из них, kz, есть фаза плоской волны. Второе представляет собой запаздывание фазы $\zeta(z)$, выражаемое формулой (3.10), которое меняется в пределах от $-\pi/2$ при $z = -\infty$ до $+\pi/2$ при $z = \infty$, как показано на рис. 3.5. Это

запаздывание фазы соотвстствует дополнительной задержке волнового фронта по отношению к плоской или сферической волне (см. также рис. 3.8). Полное запаздывание фазы, накапливающееся при распространении гауссова пучка от $z = -\infty$ до $z = \infty$, равно π. Это явление известно как **эффект Гюи**[1].

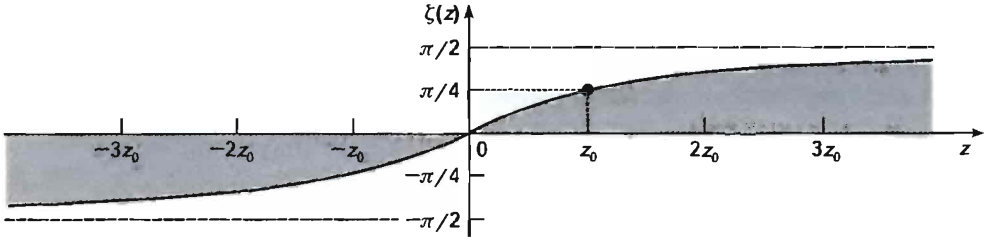


Рис. 3.5. Функция $\zeta(z)$ представляет задержку фазы гауссова пучка по отношению к однородной плоской волне в точках оси пучка

Волновые фронты

Третье слагаемое в (3.23) отвечает за искривление волнового фронта. Оно представляет собой отклонение фазы в точках, лежащих вне оси пучка в данной поперечной плоскости, от ее значения на оси пучка. Поверхности постоянной фазы удовлетворяют уравнению

$$k\left[z + \frac{\rho^2}{2R(z)}\right] - \zeta(z) = 2\pi q.$$

Так как $\zeta(z)$ и $R(z)$ — сравнительно медленно меняющиеся функции, они почти постоянны в точках, лежащих в пределах ширины пучка на каждом волновом фронте. Следовательно,

$$z + \frac{\rho^2}{2R} \approx q\lambda + \frac{\zeta\lambda}{2\pi},$$

где $R = R(z)$ и $\zeta = \zeta(z)$. Это уравнение параболоидальной поверхности с радиусом кривизны R. Таким образом, $R(z)$, график которого показан на рис. 3.6, есть радиус кривизны волнового фронта в точке z на оси пучка.

Как видно из рис. 3.6, радиус кривизны $R(z)$ бесконечен при $z = 0$, так что волновой фронт плоский и кривизны нет. По мере увеличения z радиус сначала уменьшается до минимального значения $2z_0$ при $z = z_0$, где волновой фронт имеет наибольшую кривизну (рис. 3.7). Затем радиус кривизны постепенно растет, и при $z \gg z_0$ $R(z) \approx z$. При этом волновые фронты приблизительно такие же, как у сферической волны. Картина волновых фронтов одинакова для положительных и отрицательных z, за исключением смены знака (рис. 3.8). Примем

[1] См., например, *Siegman A.E.* Lasers. University Science, 1986.

соглашение, что фронт расходящейся волны имеет положительный, а фронт сходящейся — отрицательный радиус кривизны.

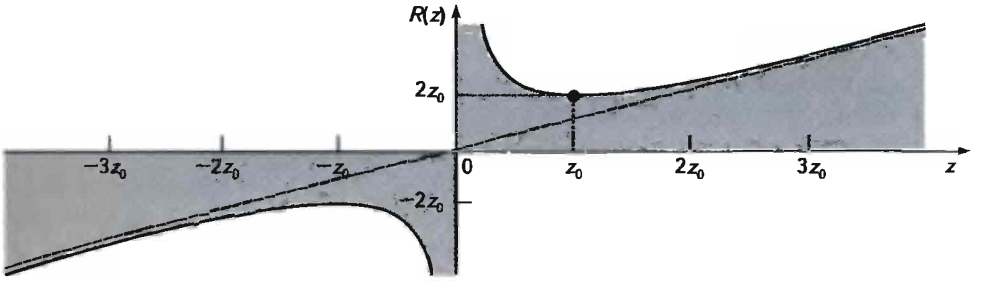


Рис. 3.6. Радиус кривизны $R(z)$ волнового фронта науссова пучка как функция положения вдоль оси пучка. Штриховая линия — радиус кривизны сферической волны

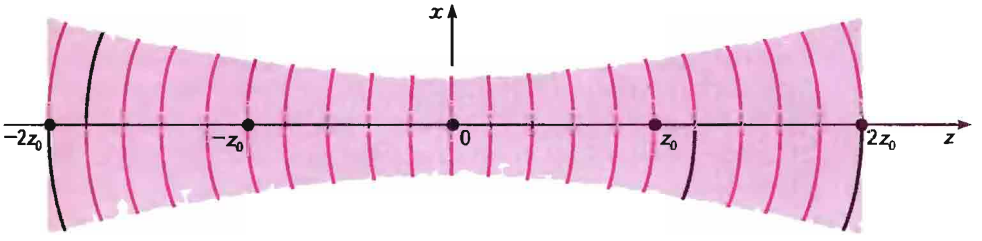


Рис. 3.7. Волновые фронты гауссова пучка

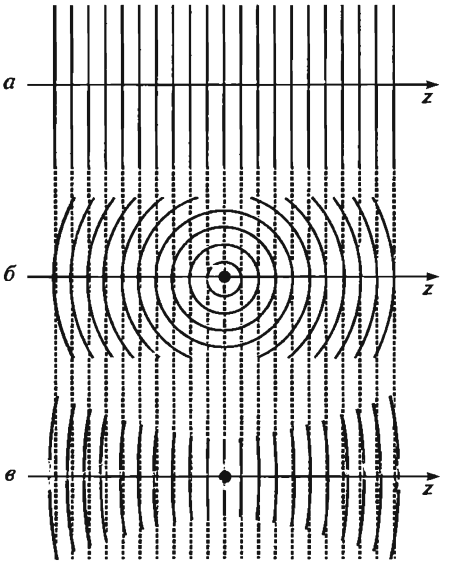


Рис. 3.8. Волновые фронты:

а — однородной плоской волны; *б* — сферической волны; *в* — гауссова пучка.

В точках, лежащих вблизи центра пучка, гауссов пучок напоминает плоскую волну. При больших z пучок ведет себя как сферическая волна, за исключением того, что его фаза задержана на $\pi/2$ (четверть расстояния между двумя соседними волновыми фронтами)

Параметры, необходимые для характеристики гауссова пучка

Предполагая, что длина волны λ известна, поставим вопрос: сколько параметров необходимо для описания плоской волны, сферической волны и гауссова пучка? Плоская волна полностью задается своей комплексной амплитудой и направлением. Сферическая волна задается своей комплексной амплитудой и положением источника. Напротив, задание гауссова пучка требует больше параметров. Для этого нужны максимальная амплитуда [определяемая параметром A_0 в (3.7)], направление (ось пучка), положение перетяжки и *еще один* дополнительный параметр, такой как радиус перетяжки W_0 *либо* рэлеевская длина z_0. Таким образом, если максимальная амплитуда и ось известны, для полного задания пучка требуется еще два дополнительных параметра.

Если известен комплексный параметр $q(z) = z + jz_0$, то расстояние z до перетяжки пучка и рэлеевская длина z_0 легко определяются как его действительная и мнимая части. Например, если $q(z)$ есть $3 + j4$ см в некоторой точке на оси пучка, можно сделать вывод, что перетяжка пучка расположена на расстоянии $z = 3$ см слева от этой точки, а глубина резкости составляет $2z_0 = 8$ см. Радиус перетяжки W_0 затем определяется по формуле (3.11). Таким образом, величины $q(z)$ достаточно для полного описания гауссова пучка с известной максимальной интенсивностью и заданной осью. Если $q(z)$ задано в одной точке, линейная зависимость от z позволяет определить его во всех точках: пусть $q(z) = q_1$ и $q(z + d) = q_2$, тогда $q_2 = q_1 + d$. Для только что приведенного примера в точке $z = 13$ см очевидно $q = 13 + 4j$.

Если ширина пучка $W(z)$ и радиус кривизны $R(z)$ известны в произвольной точке оси пучка, пучок можно полностью описать, решая (3.8), (3.9) и (3.11) для z, z_0 и W_0. Наоборот, пучок можно определить путем нахождения $q(z)$ из $W(z)$ и $R(z)$ с использованием (3.6).

Резюме
Свойства гауссова пучка в особых точках

• *При* $z = z_0$. На расстоянии z_0 от перетяжки пучка волна имеет следующие свойства:

— интенсивность на оси пучка равна половине максимальной интенсивности;

— ширина пучка в $\sqrt{2}$ раз больше ширины в перетяжке, а площадь сечения пучка больше в 2 раза;

— фаза на оси пучка отстает на $\pi/4$ по сравнению с фазой плоской волны;

— радиус кривизны волнового фронта достигает минимального значения, $R = 2z_0$, так что волновой фронт имеет наибольшую кривизну.

• *Вблизи центра пучка.* В точках, где $|z| \ll z_0$ и $\rho \ll W_0$, величина

$$\exp\left[-\frac{\rho^2}{W^2(z)}\right] \approx \exp\left(-\frac{\rho^2}{W_0^2}\right) \approx 1,$$

так что интенсивность пучка, пропорциональная квадрату этой величины, приблизительно постоянна. Также

$$R(z) \approx \frac{z_0^2}{z}; \quad \zeta(z) \approx 0,$$

так что фаза равна

$$k\left[z + \frac{\rho^2}{2R(z)}\right] \approx kz\left(1 + \frac{\rho^2}{2z_0^2}\right) \approx kz,$$

если $z_0 \gg 1$. Таким образом, вблизи центра гауссов пучок можно аппроксимировать плоской волной.

• *Вдали от перетяжки.* В пределах радиуса перетяжки по поперечной координате ($\rho < W_0$), но вдали от перетяжки ($z \gg z_0$) волна ведет себя приблизительно как сферическая. В этой области

$$W(z) \approx W_0 \frac{z}{z_0} \gg W_0; \quad \rho < W_0,$$

так что

$$\exp\left[-\frac{\rho^2}{W^2(z)}\right] \approx 1$$

и интенсивность пучка приблизительно однородна. Поскольку в этой области $R(z) \approx z$, волновые фронты приблизительно сферические. Таким образом, если не говорить о фазовой задержке Гюи $\zeta(z) \approx \pi/2$, с ростом z комплексная амплитуда гауссова пучка приближается к комплексной амплитуде параболоидальной волны, которая, в свою очередь, дает все более точное параксиальное приближение сферической волны.

Упражнение 3.1

Параметры гауссова лазерного пучка

Гелий-неоновый лазер мощностью 1 мВт генерирует гауссов пучок с длиной волны $\lambda = 633$ нм и размером пятна $2W_0 = 0{,}1$ мм.

а. Определите угловую расходимость пучка, глубину резкости и диаметр пучка при $z = 3{,}5 \cdot 10^5$ км (приблизительное расстояние до Луны).

б. Каков радиус кривизны волнового фронта при $z = 0$, $z = z_0$, $z = 2z_0$?

в. Какова оптическая интенсивность (в Вт/см^2) в центре пучка ($z = 0$, $\rho = 0$) и в точке $z = z_0$ на оси? Сравните результат с интенсивностью при $z = z_0$ сферической волны мощностью 100 Вт, создаваемой малым изотропно излучающим источником света, расположенным в точке $z = 0$.

Упражнение 3.2

Справедливость параксиального приближения для гауссова пучка

Комплексная огибающая $A(\boldsymbol{r})$ является точным решением параксиального уравнения Гельмгольца (3.2), однако соответствующая ей комплексная амплитуда $U(\boldsymbol{r}) = A(\boldsymbol{r}) \exp(-jkz)$ лишь приближенно удовлетворяет уравнению Гельмгольца (2.11). Это происходит потому, что параксиальное уравнение Гельмгольца само является приближенным. Приближение удовлетворительно, если условие (2.25) выполняется. Покажите, что если угол расходимости θ_0 гауссова пучка мал ($\theta_0 \ll 1$), то условие (2.25), необходимое для справедливости параксиального приближения, действительно выполняется.

Упражнение 3.3

Определение пучка по данной ширине и кривизне

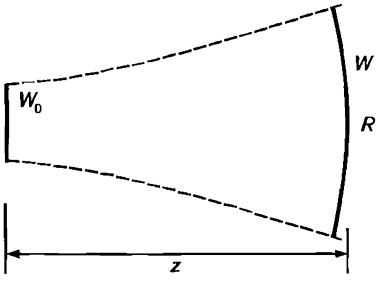


Рис. 3.9. По данным W и R определите z и W_0

Рассмотрим гауссов пучок, ширина и радиус кривизны волнового фронта известны в некоторой точке оси (рис. 3.9). Покажите, что перетяжка пучка находится слева на расстоянии

$$z = \frac{R}{1 + \left(\lambda R / \pi W^2\right)^2}, \tag{3.25}$$

а ее радиус равен

$$W_0 = \frac{W}{\sqrt{1 + \left(\pi W^2 / \lambda R\right)^2}}. \tag{3.26}$$

Упражнение 3.4

Определение ширины и кривизны в одной точке по данной ширине и кривизне в другой точке

Предположим, что ширина и радиус кривизны гауссова пучка с длиной волны $\lambda = 1$ мкм в некоторой точке на оси пучка составляют $W_1 = 1$ мм и $R_1 = 1$ м соответственно (рис. 3.10). Определите ширину пучка W_2 и радиус кривизны R_2 на расстоянии $d = 10$ см справа от данной точки.

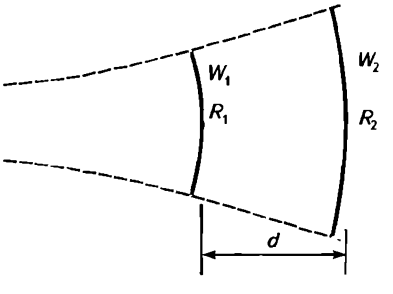


Рис. 3.10. По данным W_1, R_1 и d определите W_2 и R_2

Упражнение 3.5

Идентификация пучка по значениям кривизны в двух точках

Гауссов пучок имеет радиусы кривизны R_1 и R_2 в двух точках оси, расстояние между которыми равно d, как показано на рис. 3.11. Убедитесь, что положение центра пучка и его глубину резкости можно определить по формулам

$$z_1 = \frac{-d(R_2 - d)}{R_2 - R_1 - 2d}; \tag{3.27}$$

$$z_0^2 = \frac{-d(R_1 + d)(R_2 - d)(R_2 - R_1 - d)}{(R_2 - R_1 - 2d)^2}; \tag{3.28}$$

$$W_0 = \sqrt{\frac{\lambda z_0}{\pi}}.$$

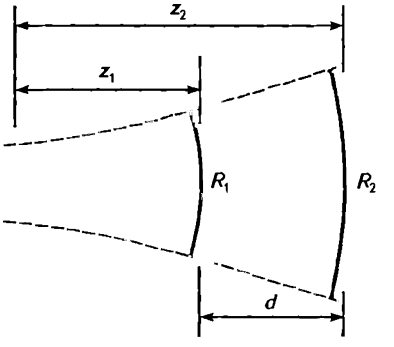


Рис. 3.11. По данным R_1 и R_2 определите z_1, z_2, z_0 и W_0

3.1.3. Качество пучка

Гауссов пучок является идеализацией, которую можно только приблизительно встретить на практике даже в хорошо сконструированных лазерных системах. Мерой качества оптического пучка является отклонение его профиля от гауссовой формы. Для пучка с диаметром перетяжки $2W_m$ и угловой расходимостью $2\theta_m$ полезной мерой качества является фактор $\mathbb{M}^2$, определяемый как отношение произведения $2W_m 2\theta_m$ фактических значений диаметра перетяжки и угла расходимости к ожидаемому значению этого произведения для гауссова пучка, которое равно $2W_0 \cdot 2\theta_0 = 4\lambda/\pi$. Таким образом,

$$\mathbb{M}^2 = \frac{2W_m 2\theta_m}{4\lambda/\pi}. \tag{3.29}$$

Если оба пучка имеют одинаковый диаметр, то фактор $\mathbb{M}^2$ есть просто отношение углов расходимости

$$\mathbb{M}^2 = \frac{\theta_m}{\theta_0}, \tag{3.30}$$

где

$$\theta_0 = \frac{\lambda}{\pi W_0} = \frac{\lambda}{\pi W_m}$$

[см. (3.21)]. Поскольку гауссов пучок имеет наименьшую возможную расходимость из всех пучков с таким же диаметром перетяжки, $\mathbb{M}^2 \geq 1$. Задание фактора $\mathbb{M}^2$ для оптического пучка означает, таким образом, тот факт, что у этого пучка расходимость в $\mathbb{M}^2$ раз больше, чем у гауссова пучка с тем же диаметром перетяжки.

Оптические пучки, создаваемые обычными гелий-неоновыми лазерами, имеют $\mathbb{M}^2 < 1{,}1$. Для ионных лазеров типичный диапазон значений $\mathbb{M}^2$ от 1,1 до 1,3. Коллимированные TEM_{00} пучки диодных лазеров обычно имеют $\mathbb{M}^2 \approx 1{,}1$—1,7, а мощные многомодовые лазеры характеризуются высокими значениями $\mathbb{M}^2$ до 3 или 4.

Для приблизительно гауссова оптического пучка фактор $\mathbb{M}^2$ можно определить с использованием ПЗС-камеры для измерения профиля интенсивности пучка в различных его сечениях. Пучок фокусируется высококачественной длиннофокусной линзой примерно до размера ПЗС-матрицы. Вначале находится центр пучка по минимальному размеру пятна, затем измеряется диаметр перетяжки $2W_m$. Расстояние вдоль оси от центра пучка до плоскости, в которой диаметр пучка увеличивается в $\sqrt{2}$ раз, дает рэлеевскую длину z_m. Оценка угловой расходимости $2\theta_m$ получается с помощью соотношения для гауссова пучка

$$\theta_m = \sqrt{\frac{\lambda}{\pi z_m}},$$

получаемого из (3.11) и (3.20). Окончательное вычисление фактора $\mathbb{M}^2$ производится с помощью соотношения (3.29).

3.2. ПРОХОЖДЕНИЕ ЧЕРЕЗ ОПТИЧЕСКИЕ ЭЛЕМЕНТЫ

Перейдем к обсуждению действия различных оптических элементов на гауссов пучок. Мы покажем, что если гауссов пучок походит через систему из оптических элементов, обладающих осевой симметрией и центрированных относительно оси пучка, то *пучок остается гауссовым* при условии, что система сохраняет параксиальную природу волны. Изменения формы пучка сводятся к изменению его перетяжки и радиуса кривизны волнового фронта. Результаты данного раздела важны для проектирования оптических приборов, рассчитанных на работу с гауссовыми пучками.

3.2.1. Прохождение через тонкую линзу

Комплексный коэффициент амплитудного пропускания тонкой линзы с фокусным расстоянием f пропорционален $\exp(jk\rho^2/2f)$ [см. (2.40)]. Когда гауссов пучок проходит через такой элемент, его комплексная амплитуда в (3.7) умножается на этот фазовый множитель. В результате ширина пучка остается неизменной ($W' = W$), а волновой фронт меняется.

Рассмотрим конкретно гауссов пучок с центром в точке $z = 0$ и радиусом перетяжки W_0, который проходит через тонкую линзу, расположенную в точке z, как показано на рис. 3.12. Фаза падающей волны в плоскости линзы есть $kz + k\rho^2/2R - \zeta$ в соответствии с (3.23), где $R = R(z)$ и $\zeta = \zeta(z)$ даются выраже-

ниями (3.9) и (3.10) соответственно. Следовательно, фаза выходящей волны получается равной

$$kz + k\frac{\rho^2}{2R} - \zeta - k\frac{\rho^2}{2f} = kz + k\frac{\rho^2}{2R'} - \zeta, \tag{3.31}$$

где

$$\frac{1}{R'} = \frac{1}{R} - \frac{1}{f}. \tag{3.32}$$

Видно, что прошедшая волна представляет собой гауссов пучок с шириной $W' = W$ и радиусом кривизны R', удовлетворяющим уравнению формирования изображения

$$\frac{1}{R} - \frac{1}{R'} = \frac{1}{f}.$$

Величина R положительна, поскольку волновой фронт падающего пучка расходящийся, а величина R', соответствующая сходящейся волне, отрицательна.

Параметры выходящего пучка определяются с помощью результатов упражнения 3.3, в котором параметры гауссова пучка определяются по его ширине и кривизне в данной точке. Уравнение (3.26) дает радиус перетяжки

$$W_0' = \frac{W}{\sqrt{1 + \left(\pi W^2/\lambda R'\right)^2}}, \tag{3.33}$$

а из (3.25) находим, что центр пучка удален от линзы на расстояние

$$-z' = \frac{R'}{1 + \left(\lambda R'/\pi W^2\right)^2}. \tag{3.34}$$

Знак «минус» в (3.34) отражает тот факт, что перетяжка пучка находится справа от линзы. Подставляя

$$W = W_0\sqrt{1 + \left(\frac{z}{z_0}\right)^2} \qquad R = z\left[1 + \left(\frac{z_0}{z}\right)^2\right]$$

из (3.8) и (3.9) в (3.32)—(3.34), получаем набор формул, связывающих параметры гауссова пучка, падающего на линзу (без штрихов) с параметрами гауссова пучка, выходящего из линзы (со штрихами), как показано на рис. 3.12:

Радиус перетяжки	$W_0' = MW_0;$	(3.35)
Положение перетяжки	$(z' - f) = M^2(z - f);$	(3.36)
Глубина резкости	$2z_0' = M^2(2z_0);$	(3.37)

Угол расходимости $2\theta_0' = \dfrac{2\theta}{M}$; (3.38)

Увеличение $M = \dfrac{M_r}{\sqrt{1+r^2}}$; (3.39)

$r = \dfrac{z_0}{z-f}, \quad M_r = \left|\dfrac{f}{z-f}\right|.$ (3.39а)

Преобразование параметров линзой

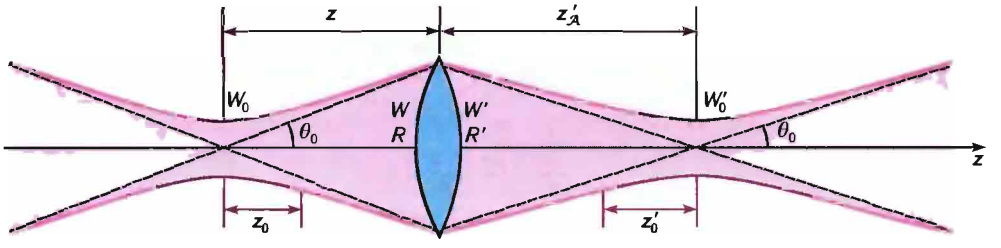


Рис. 3.12. Прохождение гауссова пучка через тонкую линзу

Важную роль, очевидно, играет фактор увеличения M. Радиус перетяжки увеличивается в M раз, глубина резкости — в M^2 раз, а угол расходимости уменьшается в M раз

Предел применимости лучевой оптики

Рассмотрим предельный случай, когда $(z - f) \gg z_0$, т. е. линза достаточно удалена от фокуса падающего пучка (рис. 3.13). Пучок можно аппроксимировать сферической волной, при этом в соответствии с (3.39) и (3.39а) $r \ll 1$, так что $M \approx M_r$. В этом случае формулы (3.35)—(3.39а) сводятся к

$$W_0' \approx MW_0; \tag{3.40}$$

$$\frac{1}{z'} + \frac{1}{z} \approx \frac{1}{f}; \tag{3.41}$$

$$M \approx M_r = \left|\frac{f}{z-f}\right|. \tag{3.42}$$

Формулы (3.40)—(3.42) в точности совпадают с соотношениями лучевой оптики для положения и размера светового пятна диаметром $2W_0$, расположенного на расстоянии z слева от тонкой линзы (см. подразд. 1.2.3). Действительно, увеличение M_r точно такое, как в оптике лучей. Поскольку (3.39) дает $M < M_r$, максимальное увеличение M для гауссова пучка есть предельное значение M_r, даваемое геометрической оптикой. По мере увеличения r^2 увеличение снижа-

ется и отклонение от геометрической оптики становится больше. Уравнения (3.40)—(3.42) согласуются также с результатами, полученными в рамках волновой оптики для фокусировки сферической волны в параксиальном приближении (см. подразд. 2.4.2).

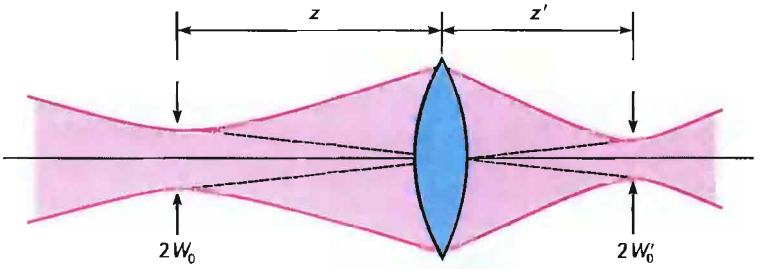


Рис. 3.13. Отображение пучка в пределе лучевой оптики

3.2.2. Формирование пучка

Линза или последовательность линз может использоваться для изменения формы гауссова пучка без нарушения его гауссовой природы. Конечно, этой же цели можно добиться с помощью элементов с градиентом показателя преломления.

Фокусировка пучка

Для линзы, помещенной в перетяжку гауссова пучка, как показано на рис. 3.14, соответствующие формулы преобразования параметров получаются простой подстановкой $z = 0$ в (3.35)—(3.39). Прошедший пучок в этом случае фокусируется в перетяжку радиусом W_0' на расстоянии z'

$$W_0' = \frac{W_0}{\sqrt{1 + (z_0/f)^2}}; \tag{3.43}$$

$$z' = \frac{f}{1 + (f/z_0)^2}. \tag{3.44}$$

В частном случае, когда глубина резкости падающего пучка $2z_0$ много больше фокусного расстояния f линзы, как показано на рис. 3.15, формула (3.43) сводится к виду

$$W_0' \approx \frac{f}{z_0} W_0.$$

Используя

$$z_0 = \frac{\pi W_0^2}{\lambda}$$

из (3.11), а также (3.20), получаем простой результат

$$W_0' \approx \frac{\lambda}{\pi W_0} f = \theta_0 f; \tag{3.45}$$

$$z' \approx f. \tag{3.46}$$

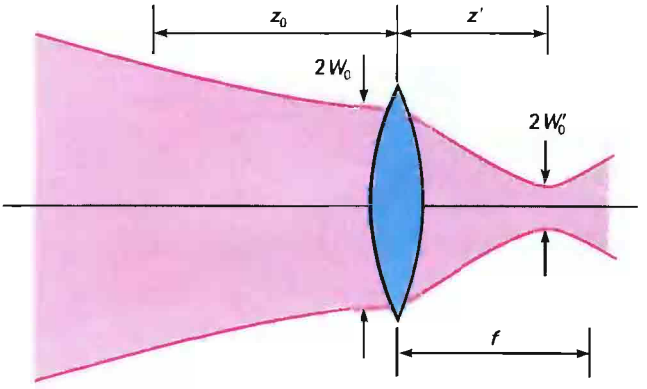


Рис. 3.14. Фокусировка гауссова пучка линзой, расположенной в перетяжке

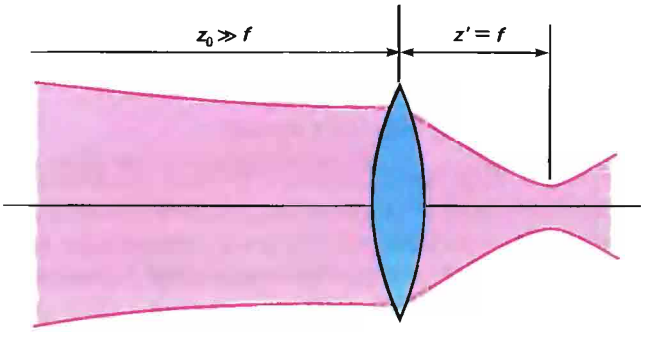


Рис. 3.15. Фокусировка коллимированного пучка

Прошедший пучок тогда фокусируется в фокальную плоскость линзы, как и следовало ожидать для коллимированного пучка параллельных лучей, падающих на линзу. Этот результат возникает в силу того, что вблизи перетяжки падающий гауссов пучок хорошо аппроксимируется плоской волной. Волновая оптика гарантирует, что радиус фокусируемого пучка в перетяжке W_0' прямо пропорционален длине волны и фокусному расстоянию и обратно пропорционален радиусу падающего пучка. Размер пятна в оптике лучей, конечно, равен нулю. Этот результат действительно получается из формул волновой оптики при $\lambda \to 0$.

Во многих приложениях, таких как лазерное сканирование, лазерная печать, прожигание компакт-дисков (CD) и лазерная сварка, желателен возможно меньший размер пятна. Из (3.45) ясно, что достичь этого можно за счет уменьшения длины волны, расширения падающего пучка и использования ко-

роткофокусной линзы. Поскольку линза должна перекрывать падающий пучок, ее диаметр D должен быть, по крайней мере, $2W_0$. Полагая $D = 2W_0$ и используя (3.45), для диаметра пятна фокусировки получим

$$\boxed{2W_0' \approx \frac{4}{\pi}\lambda F_\#;} \quad F_\# = \frac{f}{D}, \tag{3.47}$$

Размер пятна фокусировки

где $F_\#$ — относительная апертура линзы. Для получения возможно меньшего размера пятна часто используются объективы микроскопов с малыми относительными апертурами. Поскольку формулы (3.45) и (3.46) являются приближенными, их применимость всегда должна быть подтверждена перед их использованием.

Упражнение 3.6

Релейная передача пучка

Гауссов пучок радиуса W_0 с длиной волны λ многократно фокусируется последовательностью одинаковых линз с фокусным расстоянием f, разделенных расстоянием d (рис. 3.16). Радиус перетяжки преобразованного пучка равен радиусу перетяжки падающего, т. е. $W_0' = W_0$. Пользуясь (3.36), (3.39) и (3.39а), покажите, что это условие может быть выполнено, только если $d \le 4f$. Обратите внимание, что это то же самое условие ограничения, которое было выведено в примере 1.10 в лучевой оптике.

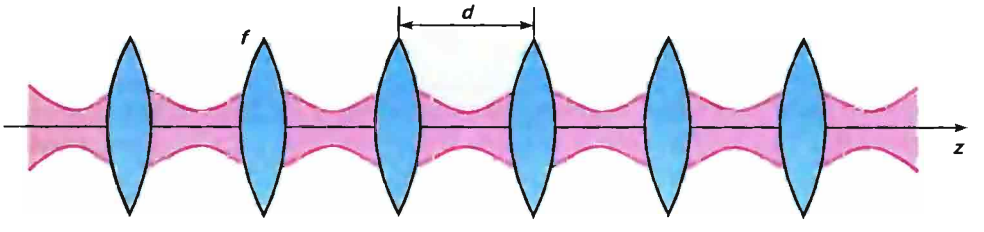


Рис. 3.16. Релейная передача пучка

Упражнение 3.7

Коллимирование пучка

Гауссов пучок пропускается через тонкую линзу с фокусным расстоянием f.

а. Покажите, что положения перетяжек падающего и прошедшего пучков z и z', соответственно, связаны соотношением

$$\frac{z'}{f} - 1 = \frac{z/f - 1}{(z/f - 1)^2 + (z_0/f)^2}. \tag{3.48}$$

Эта связь изображена графически на рис. 3.17.

б. Коллимирование пучка достигается смещением новой перетяжки z' на возможно большее расстояние от линзы. Для этого используется наименьшее возможное отношение z_0/f (малая глубина резкости в сочетании с большим фокусным расстоянием). Для данного отношения z_0/f покажите, что оптимальное значение z для коллимирования есть $z = f + z_0$.

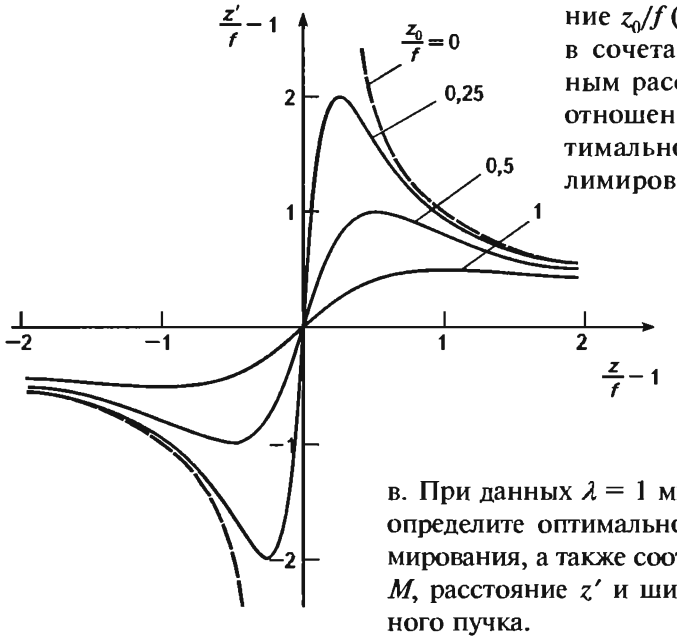


Рис. 3.17. Связь между положениями перетяжек падающего и прошедшего пучков

в. При данных $\lambda = 1$ мкм, $z_0 = 1$ см и $f = 50$ см определите оптимальное значение z для коллимирования, а также соответствующие увеличение M, расстояние z' и ширину W_0' коллимированного пучка.

Упражнение 3.8

Расширение пучка

Гауссов пучок можно расширить и коллимировать с помощью двух линз с фокусными расстояниями f_1 и f_2, как показан на рис. 3.18. Параметры исходного пучка (W_0, z_0) преобразуются первой линзой в (W_0'', z_0''), а затем второй линзой в (W_0', z_0'). Первая линза с малым фокусным расстоянием служит для

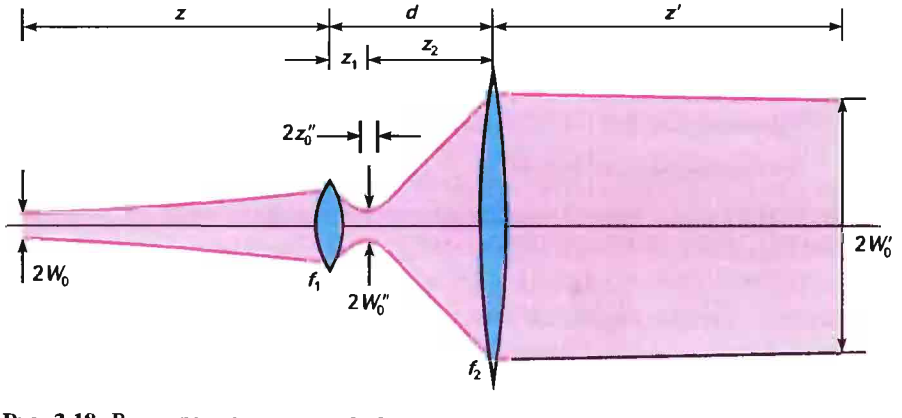


Рис. 3.18. Расширение пучка с помощью двух линз

уменьшения глубины резкости пучка $2z_0''$. Это готовит пучок к коллимированию второй линзой, которая имеет большое фокусное расстояние. Система функционирует как обращенный телескоп Кеплера.

а. Предполагая, что $f_1 \ll z$ и $z - f_1 \gg z_0$, используйте результат упражнения 3.7 для определения оптического расстояния d между линзами, обеспечивающего наибольшее возможное расстояние z' до перетяжки результирующего пучка.

б. Найдите выражение для полного увеличения $M = W_0'/W_0$ системы.

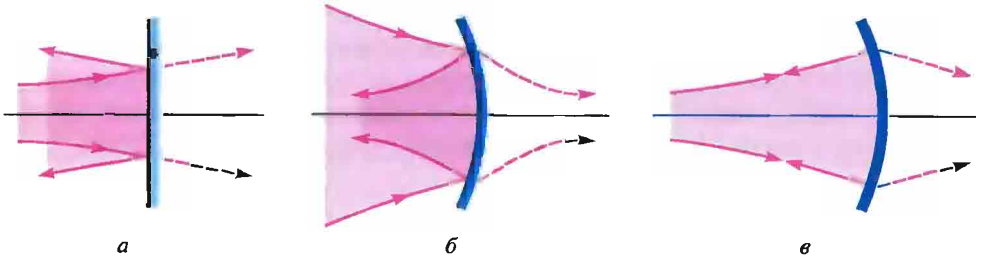
3.2.3. Отражение от сферического зеркала

Исследуем теперь отражение гауссова пучка от сферического зеркала. Комплексный амплитудный коэффициент отражения зеркала пропорционален $\exp(-jk\rho^2/R)$ (см. задачу 4 к разд. 2.4), где по соглашению $R > 0$ для выпуклого и $R < 0$ для вогнутого зеркала. Действие зеркала на гауссов пучок ширины W_1 с радиусом кривизны R_1 состоит в отражении пучка и изменении его фазы на $-k\rho^2/R$, в то время как ширина пучка остается неизменной. Отраженный пучок остается гауссовым и имеет параметры W_2 и R_2, которые определяются как

$$W_2 = W_1; \tag{3.49}$$

$$\frac{1}{R_2} = \frac{1}{R_1} + \frac{2}{R}. \tag{3.50}$$

Уравнение (3.50) совпадает с (3.32) при условии, что $f = -R/2$. Таким образом, гауссов пучок преобразуется так же, как и в линзе, за исключением обращения направления распространения.



а *б* *в*

Рис. 3.19. Отражение гауссова пучка с радиусом кривизны R_1 от зеркала радиусом R: *а* — $R = \infty$; *б* — $R_1 = \infty$; *в* — $R_1 = -R$. Штриховые линии показывают, что получится, если зеркало заменить линзой с фокусным расстоянием $f = -R/2$

Представляют интерес три частных случая, изображенные на рис. 3.19.

• Если *зеркало плоское*, т. е. $R = \infty$, то $R_2 = R_1$, так что зеркало обращает направление распространения пучка без изменения кривизны волнового фронта, как показано на рис. 3.19, *а*.

• Если $R_1 = \infty$, т. е. *если перетяжка пучка совпадает с зеркалом*, то $R_2 = R/2$. Если зеркало вогнутое ($R < 0$), то $R_2 < 0$, так что отраженный пучок приобретает отрицательную кривизну и становится сходящимся. В этом случае зеркало фокусирует пучок в пятно меньшего размера, как показано на рис. 3.19, *б*.

• Если $R_1 = -R$, т. е. падающий пучок имеет ту же кривизну, что и зеркало, то $R_2 = R$. Волновые фронты падающей и отраженной волн совпадают с зеркалом, и волна повторяет свой путь, как показано на рис 3.19, *в*. Это вполне ожидаемо, поскольку нормали к волновому фронту являются также нормалями к зеркалу, так что зеркало точно обращает волну. На рис. 3.19, *в* зеркало вогнутое ($R < 0$), падающая волна расходящаяся ($R_1 > 0$), а отраженная волна сходящаяся ($R_2 < 0$).

Упражнение 3.9

Зеркала с переменным коэффициентом отражения

Сферическое зеркало радиусом R имеет переменный коэффициент отражения по интенсивности, задаваемый формулой

$$\mathcal{R}(\rho) = |r(\rho)|^2 = \exp\left(-\frac{2\rho^2}{W_m^2}\right),$$

который является гауссовой функцией радиального расстояния ρ. Коэффициент отражения равен единице на оси и плавно спадает по мере удаления от нее, уменьшаясь в e^2 раз при удалении на $\rho = W_m$. Определите действие зеркала на гауссов пучок с радиусом кривизны R_1 и шириной W_1 на зеркале.

*3.2.4. Прохождение через произвольную оптическую систему

В параксиальном приближении лучевой оптики оптическая система полностью характеризуется 2 × 2-матрицей передачи луча, которая связывает положение и наклон луча на выходе и на входе (см. разд. 1.4). Рассмотрим, как произвольная параксиальная оптическая система, характеризуемая матрицей **M** с элементами (A, B, C, D), преобразует гауссов пучок (рис. 3.20).

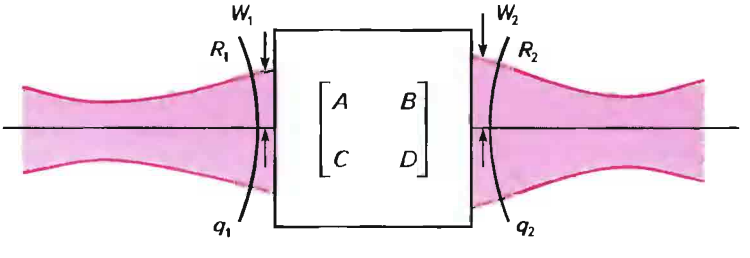


Рис. 3.20. Преобразование гауссова пучка произвольной параксиальной оптической системой, описываемой матрицей *ABCD*

Закон ABCD

Параметры q_1 и q_2 падающего и прошедшего гауссовых пучков на входной и выходной плоскостях параксиальной оптической системы, описываемой матрицей (A, B, C, D), связаны как

$$\boxed{q_2 = \frac{Aq_1 + B}{Cq_1 + D}.} \qquad (3.51)$$

Закон *ABCD*

Поскольку комплексный параметр q определяет ширину W и радиус кривизны R гауссова пучка (см. упражнение 3.3), это простое выражение, называемое **законом *ABCD***, определяет влияние произвольной параксиальной системы на гауссов пучок. Закон *ABCD* будет установлен путем проверки на частных случаях, его общность будет немедленно доказана по индукции.

Прохождение через свободное пространство

Когда оптической системой является участок свободного пространства (или произвольной однородной среды) протяженностью d, то элементы матрицы передачи луча M таковы: $A = 1$, $B = d$, $C = 0$, $D = 1$ [см. (1.43)]. Поскольку ранее было установлено, что в свободном пространстве $q = z + jz_0$, то параметр q преобразуется оптической системой по закону $q_2 = q_1 + d$. Это очевидно равно $(1 \cdot q_1 + d)/(0 \cdot q_1 + 1)$, так что закон *ABCD* выполняется.

Прохождение через тонкий оптический элемент

Произвольный тонкий оптический элемент не меняет положения луча, так что

$$y_2 = y_1, \qquad (3.52)$$

но меняет угол наклона в соответствии с

$$\theta_2 = Cy_1 + D\theta_1, \qquad (3.53)$$

как показано на рис. 3.21. Таким образом, $A = 1$, $B = 0$, но C и D произвольные. Однако у всех тонких оптических элементов, описанных в подразд. 1.4.2, $D = n_1/n_2$. По причине исчезающей малой толщины элемента ширина пучка не меняется, т. е.

$$W_2 = W_1. \qquad (3.54)$$

Далее, если пучки в плоскостях входа и выхода элемента аппроксимировать сферическими волнами с радиусами R_1 и R_2 соответственно, то в параксиальном приближении, когда θ_1 и θ_2 малы, $\theta_1 \approx y_1/R_1$ и $\theta_2 \approx y_2/R_2$. Подставляя эти выражения в (3.53), с помощью (3.52) получаем

$$\frac{1}{R_2} = C + \frac{D}{R_1}. \qquad (3.55)$$

Используя (3.6), выражающее q как функцию R и W, и замечая, что

$$D = \frac{n_1}{n_2} = \frac{\lambda_2}{\lambda_1},$$

можно скомбинировать (3.54) и (3.55) в одно уравнение

$$\frac{1}{q_2} = C + \frac{D}{q_1}, \tag{3.56}$$

из которого

$$q_2 = \frac{1q_1 + 0}{Cq_1 + D},$$

так что закон $ABCD$ снова выполняется.

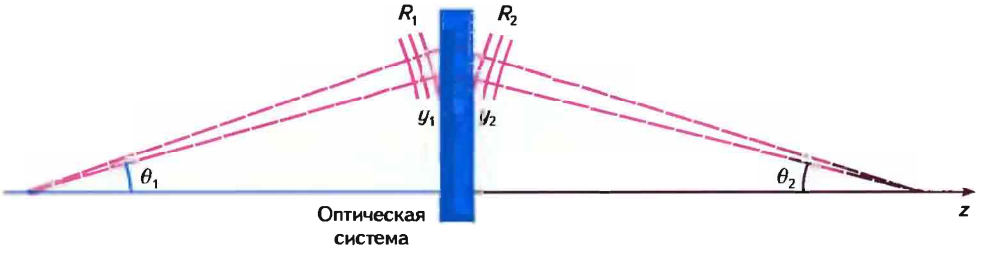


Рис. 3.21. Преобразование гауссова пучка тонкой оптической системой

Инвариантность закона ABCD по отношению к образованию каскадов

Если закон $ABCD$ выполняется для каждой из двух оптических систем с матрицами $\mathbf{M}_i = (A_i, B_i, C_i, D_i)$, $i = 1, 2$. он должен также выполняться для каскадной системы, имеющей матрицу $\mathbf{M} = \mathbf{M}_2\mathbf{M}_1$. Это можно показать прямой подстановкой.

Общность закона ABCD

Так как закон $ABCD$ применим к тонким оптическим элементам, а также к распространению в однородной среде, он применим и к любой комбинации указанных элементов. Все представляющие интерес параксиальные оптические системы являются комбинациями участков однородной среды и тонких оптических элементов, например тонких линз или зеркал. Поэтому очевидно, что закон $ABCD$ применим ко всем таким системам. Более того, поскольку неоднородная среда с плавно меняющимися параметрами может рассматриваться как каскад бесконечно тонких элементов, разделенных тонкими же слоями однородной среды, можно заключить, что закон $ABCD$ применим и к таким системам тоже при условии, что все лучи (нормали к волновым фронтам) остаются параксиальными.

Упражнение 3.10

Прохождение гауссова пучка через прозрачную пластину

Используйте закон *ABCD* для исследования прохождения гауссова пучка через находящуюся в воздухе прозрачную пластину толщины d с показателем преломления n. Ось пучка считать направленной по нормали к пластине.

3.3. ПУЧКИ ЭРМИТА—ГАУССА

Гауссовы пучки — не единственные пучковые решения параксиального уравнения Гельмгольца (3.2). Особый интерес представляют решения, которые имеют негауссово распределение интенсивности, но при этом их волновые фронты сохраняют параболоидальную форму, как у гауссовых пучков. Такие пучки имеют замечательное свойство — их волновой фронт можно согласовать по кривизне со сферическими зеркалами большого радиуса, например зеркалами оптических резонаторов, в результате чего они многократно отражаются между зеркалами без изменения своей формы. Такие самовоспроизводящиеся волны называются **модами** резонатора. Оптика резонаторов обсуждается в гл. 9.

Рассмотрим гауссов пучок с комплексной огибающей [см. (3.5)]

$$A_G(x, y, z) = \frac{A_1}{q(z)} \exp\left[-jk\frac{x^2+y^2}{2q(z)}\right], \tag{3.57}$$

где $q(z) = z + jz_0$. Ширина пучка $W(z)$ и радиус кривизны волнового фронта $R(z)$ определяются выражениями (3.8) и (3.9) соответственно. Рассмотрим теперь волну, комплексная амплитуда которой имеет дополнительную модуляцию по сравнению с гауссовым пучком

$$A(x, y, z) = X\left[\sqrt{2}\frac{x}{W(z)}\right] Y\left[\sqrt{2}\frac{y}{W(z)}\right] \exp[jZ(z)] A_G(x, y, z), \tag{3.58}$$

где $X(\ )$, $Y(\ )$, $Z(\ )$ — действительные функции. Такая волна, если она существует, обладает следующими двумя свойствами:

1. Фаза у нее такая же, как и у исходной гауссовой волны, за исключением добавки $Z(z)$, не зависящей от x и y. Если $Z(z)$ — медленно меняющаяся функция z, то обе волны имеют параболоидальные волновые фронты с одинаковым радиусом кривизны $R(z)$. Обе эти волны, следовательно, фокусируются тонкими линзами и зеркалами совершенно одинаково.

2. Величина

$$A_0 X\left[\sqrt{2}\frac{x}{W(z)}\right] Y\left[\sqrt{2}\frac{y}{W(z)}\right]\left[\frac{W_0}{W(z)}\right] \exp\left[-\frac{x^2+y^2}{W^2(z)}\right], \tag{3.59}$$

где $A_0 = A_1/jz_0$, является функцией переменных $x/W(z)$ и $y/W(z)$, поперечный масштаб которых зависит от z и определяется одним и тем же масштабным

множителем $W(z)$. С ростом z поперечное распределение интенсивности сохраняет свой вид, меняется только масштаб $W(z)$. Это распределение представляет собой гауссову функцию, промодулированную в направлениях x и y функциями $X^2(\ )$ и $Y^2(\cdot)$ соответственно.

Таким образом, модулированная волна представляет собой пучок с распределением интенсивности, отличным от гауссова, но с такими же волновыми фронтами и угловой расходимостью, как у исходной гауссовой волны.

Существование такой волны будет доказано, если удастся найти такие действительные функции $X(\cdot)$, $Y(\ )$, $Z(\cdot)$, чтобы (3.58) удовлетворяло параксиальному уравнению Гельмгольца (3.2). Подставим (3.58) в (3.2), используем тот факт, что сама A_G удовлетворяет (3.2), и определим две новые переменные

$$u = \frac{\sqrt{2}x}{W(z)}; \quad v = \frac{\sqrt{2}y}{W(z)}.$$

Тогда

$$\frac{1}{X}\left(\frac{\partial^2 X}{\partial u^2} - 2u\frac{\partial X}{\partial u}\right) + \frac{1}{Y}\left(\frac{\partial^2 Y}{\partial v^2} - 2v\frac{\partial Y}{\partial v}\right) + kW^2(z)\frac{\partial Z}{\partial z} = 0. \tag{3.60}$$

Так как левая часть этого уравнения есть сумма трех слагаемых, каждое из которых — функция одной независимой переменной u, v и z соответственно, то каждое из слагаемых должно быть постоянной величиной. Если первый член равен постоянной $-2\mu_1$, а второй $-2\mu_2$, то третий должен равняться $2(\mu_1 + \mu_2)$. Этот метод, называемый методом разделения переменных, позволяет свести дифференциальное уравнение в частных производных (3.60) к трем обыкновенным дифференциальным уравнениям для $X(u)$, $Y(v)$, $Z(z)$ соответственно:

$$-\frac{1}{2}\frac{\mathrm{d}^2 X}{\mathrm{d}u^2} + u\frac{\mathrm{d}X}{\mathrm{d}u} = \mu_1 X; \tag{3.61а}$$

$$-\frac{1}{2}\frac{\mathrm{d}^2 Y}{\mathrm{d}v^2} + v\frac{\mathrm{d}Y}{\mathrm{d}v} = \mu_2 Y; \tag{3.61б}$$

$$z_0\left[1 + \left(\frac{z}{z_0}\right)^2\right]\frac{\mathrm{d}Z}{\mathrm{d}z} = \mu_1 + \mu_2. \tag{3.61в}$$

где использовано выражение $W(z)$, данное в (3.8) и (3.11).

Уравнение (3.61а) представляет собой задачу на собственные значения, решение которой дает $\mu_1 = l$, где $l = 0, 1, 2,...$ Соответствующие собственные функции представляют собой **полиномы Эрмита** $X(u) = \mathbb{H}_l(u)$, $l = 0, 1, 2,...$ Эти полиномы определяются рекуррентным соотношением

$$\mathbb{H}_{l+1}(u) = 2u\,\mathbb{H}_l(u) - 2l\,\mathbb{H}_{l-1}(u), \tag{3.62}$$

и первыми двумя полиномами

$$\mathbb{H}_0(u) = 1; \quad \mathbb{H}_1(u) = 2u. \tag{3.63}$$

Отсюда

$$\mathbb{H}_2(u) = 4u^2 - 2; \quad \mathbb{H}_3(u) = 8u^3 - 12u; \tag{3.64}$$

Аналогично, решения (3.61б) $\mu_2 = m$ и $\mathcal{Y}(v) = \mathbb{H}_m(v)$, где $m = 0, 1, 2,$ Таким образом, получается семейство решений, нумеруемых индексами (l, m). Подставляя $\mu_1 = l$, $\mu_2 = m$ в (3.61в) и интегрируя, получаем

$$\mathcal{Z}(z) = (l + m)\zeta(z), \tag{3.65}$$

где

$$\zeta(z) = \operatorname{arctg}\left(\frac{z}{z_0}\right).$$

Таким образом, добавочная фаза медленно меняется от $-(l + m)\pi/2$ до $+(l + m)\pi/2$, по мере того как z изменяется от $-\infty$ до $+\infty$ [см. (3.10) и рис. 3.5].

Комплексная амплитуда

Окончательно подстановка в (3.58) дает выражение для комплексной огибающей пучка с индексами (l, m). Перегруппировка членов и умножение на $\exp(-jkz)$ приводит к выражению для комплексной амплитуды

$$\boxed{U_{l,m}(x, y, z) = A_{l,m}\left[\frac{W_0}{W(z)}\right]\mathbb{G}_l\left[\frac{\sqrt{2}x}{W(z)}\right]\mathbb{G}_m\left[\frac{\sqrt{2}y}{W(z)}\right]\times \exp\left[-jkz - jk\frac{x^2 + y^2}{2R(z)} + j(l + m + 1)\zeta(z)\right],} \tag{3.66}$$

Пучок Эрмита—Гаусса

где функция

$$\mathbb{G}_l(u) = \mathbb{H}_l(u)\exp\left(-\frac{u^2}{2}\right), \quad l = 0, 1, 2, \tag{3.67}$$

известна как функция Эрмита—Гаусса порядка l, а $A_{l,m}$ — постоянная.

Поскольку $\mathbb{H}_0(u) = 1$, функция Эрмита—Гаусса нулевого порядка есть просто гауссова функция. Переходя к более высоким порядкам, находим: $\mathbb{G}_1(u) = 2u\exp(-u^2/2)$ — нечетная функция; $\mathbb{G}_2(u) = (4u^2 - 2)\exp(-u^2/2)$ — четная; $\mathbb{G}_3(u) = (8u^3 - 12u)\exp(-u^2/2)$ — нечетная и т. д. Эти функции схематически показаны на рис. 3.22.

Оптическая волна с комплексной амплитудой, определяемой выражением (3.66), называется **пучком Эрмита—Гаусса** порядка (l, m). Пучок Эрмита—Гаусса порядка (0, 0) есть просто гауссов пучок.

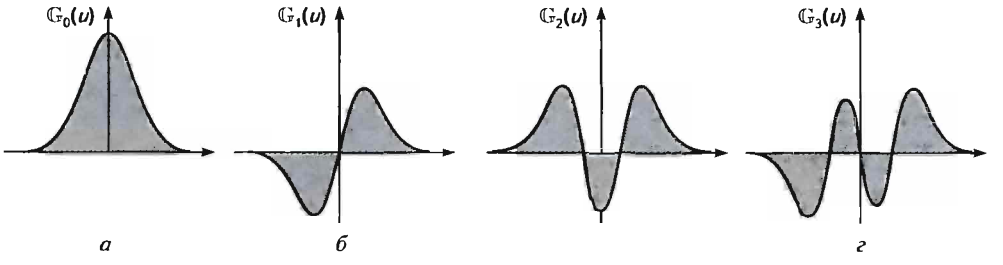


Рис. 3.22. Функции Эрмита—Гаусса низших порядков:
а — $\mathbb{G}_0(u)$; *б* — $\mathbb{G}_1(u)$; *в* — $\mathbb{G}_2(u)$; *г* — $\mathbb{G}_3(u)$

Распределение интенсивности

Оптическая интенсивность $I_{l,\,m} = |U_{l,\,m}|^2$ пучка Эрмита—Гаусса порядка $(l,\ m)$ имеет вид

$$I_{l,\,m}(x,\ y,\ z) = \left|A_{l,\,m}\right|^2 \left[\frac{W_0}{W(z)}\right]^2 \mathbb{G}_l^2\left[\frac{\sqrt{2}x}{W(z)}\right]\mathbb{G}_m^2\left[\frac{\sqrt{2}y}{W(z)}\right]. \qquad (3.68)$$

На рис. 3.23 показана зависимость интенсивности от нормированных поперечных координат

$$u = \frac{\sqrt{2}x}{W(z)}; \quad v = \frac{\sqrt{2}y}{W(z)}$$

для нескольких значений l и m. Пучки более высокого порядка имеют большую ширину, чем пучки более низкого порядка, что очевидно из рис. 3.22. Независимо от порядка, ширина пучка пропорциональна $W(z)$, так что с ростом z поперечный размер распределения интенсивности увеличивается в $W(z)/W_0$ раз, сохраняя форму профиля. Единственный аксиально-симметричный член семейства пучков Эрмита—Гаусса — это сам элементарный гауссов пучок.

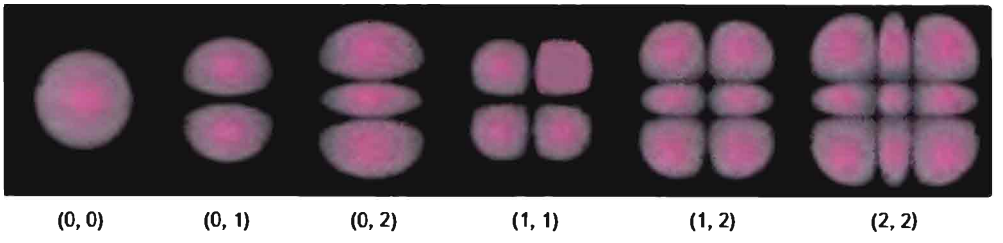


Рис. 3.23. Распределения интенсивности для нескольких пучков Эрмита—Гаусса низших порядков в поперечной плоскости. В каждом случае указан порядок $(l,\ m)$

Упражнение 3.11

Кольцевой пучок

Рассмотрим волну в виде суперпозиции двух пучком Эрмита—Гаусса порядков (1, 0) и (0, 1) одинаковой интенсивности. Пучки имеют независимые случайные фазы, так что их интенсивности складываются без интерференции. Покажите, что полная интенсивность описывается распределением в виде аксиально-симметричного кольца («бублика»). Предполагая, что $W_0 = 1$ мм, определите радиус окружности, на которой интенсивность максимальна, и радиусы окружностей, на которых она в e^2 раз меньше максимальной в перетяжке пучка.

3.4. ПУЧКИ ЛАГЕРРА—ГАУССА И БЕССЕЛЯ

Пучки Лагерра—Гаусса

Пучки Эрмита—Гаусса образуют полный набор решений параксиального уравнения Гельмгольца. Любое другое решение можно записать в виде суперпозиции этих пучков. Альтернативный полный набор решений, известный как пучки Лагерра—Гаусса, получается при решении параксиального уравнения Гельмгольца в цилиндрических координатах (ρ, ϕ, z) с помощью разделения переменных ρ и ϕ, а не x и y.

Комплексная амплитуда пучка Лагерра—Гаусса

$$U_{l,m}(\rho, \phi, z) = A_{l,m}\left[\frac{W_0}{W(z)}\right]\left(\frac{\rho}{W(z)}\right)^l \mathbb{L}_m^l\left(\frac{2\rho^2}{W^2(z)}\right)\exp\left(-\frac{\rho^2}{W^2(z)}\right)\times$$
$$\times\exp\left[-jkz - jk\frac{\rho^2}{2R(z)} - jl\phi + j(l+2m+1)\zeta(z)\right], \tag{3.69}$$

где $\mathbb{L}_m^l(\cdot)$ — обобщенный полином Лагерра[2], а величины $W(z)$, $R(z)$, $\zeta(z)$ и W_0 даются формулами (3.8)—(3.11). Пучок Лагерра—Гаусса наименьшего порядка ($l = 0$, $m = 0$) — просто гауссов пучок.

Интенсивность пучка Лагерра—Гаусса является функцией ρ и z, т. е. обладает аксиальной симметрией. Для $l \neq 0$ интенсивность в центре пучка ($\rho = 0$) равна нулю и имеет распределение интенсивности в виде концентрических колец. Фаза имеет такую же зависимость от ρ и z, как и у гауссова пучка, но содержит дополнительный член, пропорциональный азимутальному углу ϕ, а также фазу Гюи, которая в $(l + 2m + 1)$ раз больше. Из-за линейной зависимости фазы от ϕ (при $l \neq 0$) волновой фронт имеет форму винтовой поверхности,

[2] Обобщенные полиномы Лагерра определяются формулой Родригеса

$$\mathbb{L}_m^l(x) = \frac{x^{-l}e^x}{m!}\frac{d^m}{dx^m}(x^{l+m}e^{-x}).$$

Например, $\mathbb{L}_0^l(x) = 1$; $\mathbb{L}_1^0(x) = 1 - x$; $\mathbb{L}_2^0(x) = 1 - 2x + x^2/2$.

как видно из рис. 3.24. Пучки с такой винтовой фазой интересны тем, что они обладают орбитальным угловым моментом (см. разд. 5.1 и подразд. 12.1.4), и могут передавать вращательный момент облучаемой системе.

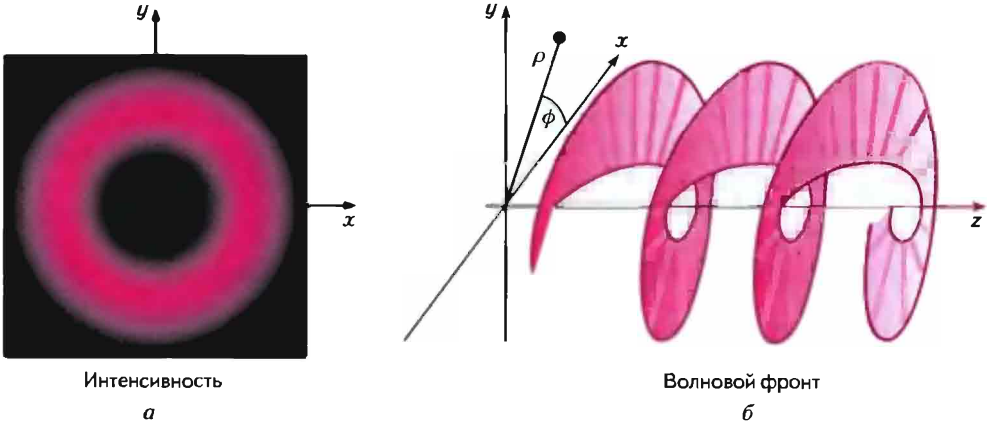


Рис. 3.24. Распределение интенсивности и волновой фронт пучка Лагерра—Гаусса с $l = 1$

Пучки Бесселя и Бесселя—Гаусса

В поисках волн, имеющих вид пучков, естественно попытаться конструировать волны, волновые фронты которых плоские, однако распределение интенсивности неоднородно в поперечной плоскости. Например, рассмотрим волну с комплексной амплитудой

$$U(\boldsymbol{r}) = A(x,\ y)e^{-j\beta z} \tag{3.70}$$

Чтобы эта волна удовлетворяла уравнению Гельмгольца (2.11), $\nabla^2 U + k^2 U = 0$, величина $A(x, y)$ должна удовлетворять

$$\nabla_T^2 A + k_T^2 A = 0, \tag{3.71}$$

где

$$k_T^2 + \beta^2 = k^2; \quad \nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

являются поперечными операторами Лапласа. Уравнение (3.71), известное как двумерное уравнение Гельмгольца, можно решить методом разделения переменных. В полярных координатах ($x = \rho\cos\phi$, $y = \rho\sin\phi$) результат оказывается следующим

$$A(x,\ y) = A_m J_m(k_T\rho)e^{jm\phi}, \quad m = 0,\ \pm 1,\ \pm 2, \tag{3.72}$$

где $J_m(\ )$ — функция Бесселя первого рода порядка m; A_m — постоянная. Решения (3.70), сингулярные в точке $\rho = 0$, не включаются в рассмотрение.

При $m = 0$ волна имеет комплексную амплитуду

$$U(\boldsymbol{r}) = A_0 J_0(k_T \rho) e^{-j\beta z}, \tag{3.73}$$

и, следовательно, имеет плоские волновые фронты. Нормали к волновым фронтам (лучи) параллельны оси z. Распределение интенсивности

$$I(\rho, \phi, z) = |A_0|^2 J_0^2(k_T \rho)$$

обладает осевой симметрией, меняется в зависимости от ρ и не зависит от z, поэтому расплывание оптической мощности не происходит (рис. 3.25). Такая волна называется **пучком Бесселя**.

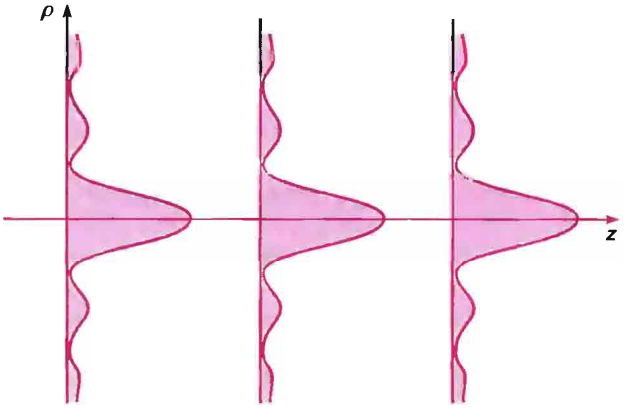


Рис. 3.25. Распределение интенсивности в пучке Бесселя в поперечном сечении не зависит от z; пучок не расходится

Полезно сравнить пучок Бесселя с гауссовым пучком. В то время как комплексная амплитуда пучка Бесселя есть *точное* решение уравнения Гельмгольца, комплексная амплитуда гауссова пучка есть лишь приближенное решение (ее комплексная огибающая есть точное решение параксиального уравнения Гельмгольца). Распределения интенсивности в этих двух пучках графически сравниваются на рис 3.26. Видно, что асимптотическое поведение этих распределений в пределе больших значений радиальной координаты существенно различается. Интенсивность гауссова пучка убывает с ростом ρ экспоненциально как $\exp[-2\rho^2/W^2(z)]$. Интенсивность пучка Бесселя, напротив, убывает как

$$J_0^2(k_T \rho) \approx \frac{2}{\pi k_T \rho} \cos^2\left(k_T \rho - \frac{\pi}{4}\right),$$

что представляет собой осциллирующую функцию, наложенную на медленное затухание обратно пропорционально ρ. Вследствие этого поперечная среднеквадратичная ширина гауссова пучка

$$\sigma = \frac{1}{2} W(z)$$

конечна, в то время как для пучка Бесселя она *бесконечна* при любом значении z (см. определение среднеквадратичной ширины в приложении А, разд. А.2), и пучок переносит бесконечную мощность. Очевидно, существует альтернатива между минимальным размером пучка и расходимостью: хотя расходимость пучка Бесселя равна нулю, его среднеквадратичная ширина бесконечна. В то время как генерация пучков Бесселя требует специальных схем[3], гауссовы пучки являются модами сферических резонаторов и, следовательно, естественно генерируются лазерами, в которых такие резонаторы используются.

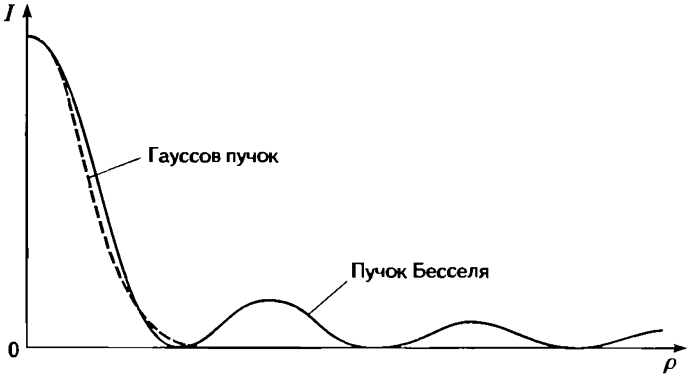


Рис. 3.26. Сравнение радиальных распределений интенсивности гауссова пучка и пучка Бесселя. Параметры выбраны так, чтобы пиковые интенсивности и ширины на уровне $1/e^2$ были одинаковы в обоих случаях

Еще один класс пучков — пучки Бесселя—Гаусса[4], которые представляют собой пучки Бесселя, промодулированные гауссовой функцией радиальной координаты ρ. Гауссова функция играет роль ограничителя, ускоряющего медленный радиальный спад пучка Бесселя.

Рекомендуемая литература

Книги

См. также книги по лазерам в списке литературы для чтения к гл. 15.

Dickey F.M., Holswade S.C., Shealy D.L. Laser Beam Shaping Applications. CRC Press, 2006.

Dickey F.M., Holswade S.C., eds. Laser Bealn Shaping: Theory and Techniques. Marcel Dekker, 2000.

Goldsmith P.F. Quasioptical Systems: Gaussian Beam Quasioptical Propagation and Applications. Wiley, 1998.

Oraevskiy A.N. Gaussian Beams and Optical Resonators. Nova Science, 1996.

Arnaud J.A. Beam and Fiber Optics, Academic Press, 1976.

---

[3] См.: *Milonni P.W., Eberly J.H.* Lasers. Wiley, 1988, sec. 14.14.

[4] См.: *Gori F., Guattari G., Padovani C.* Bessel–Gauss Beams. Optics Communications. Vol. 64, 1987. P. 491—495.

СТАТЬИ

Special issue on propagation and scattering of beam fields. Journal of the Optical Society of America A. Vol. 3, № 4, 1986.

Kogelnik H., Li T. Laser Beams and Resonators. Proceedings of the IEEE. Vol. 54, 1966. P. 1312—1329.

Boyd G.D., Gordon J.P. Confocal Multimode Resonator for Millimeter Through Optical Wavelength Masers. Bell System Technical Journal. Vol. 40, 1961. P. 489—508.

Fox A.G., Li T. Resonant Modes in a Maser Interferometer. Bell System Technical Journal. Vol. 40, 1961. P. 453—488.

Задачи

К РАЗДЕЛУ 3.1

1. **Параметры пучка.** Свет, излучаемый Nd:YAG-лазером на длине волны 1,06 мкм, представляет собой гауссов пучок с оптической мощностью 1 Вт и углом расходимости $2\theta_0 = 1$ мрад. Определите радиус перетяжки пучка, глубину резкости, максимальную интенсивность и интенсивность на оси пучка на расстоянии $z = 100$ см от перетяжки.

2. **Идентификация пучка по двум значениям ширины.** Гауссов пучок с длиной волны $\lambda_0 = 10{,}6$ мкм (излучаемый CO_2-лазером) имеет ширину $W_1 = 1{,}699$ мм и $W_2 = 3{,}380$ мм в двух точках, расстояние между которыми $d = 10$ см. Определите положение перетяжки и ее радиус.

3. **Эллиптический гауссов пучок.** Параксиальное уравнение Гельмгольца допускает решение в виде гауссова пучка с интенсивностью

$$I(x,\ y,\ 0) = |A_0|^2 \exp\left[-2\left(\frac{x^2}{W_{0x}^2} + \frac{y^2}{W_{0y}^2}\right)\right]$$

в плоскости $z = 0$ с размерами перетяжки W_{0x} и W_{0y} в направлениях x и y соответственно. Линии постоянной интенсивности у такого пучка — эллипсы, а не окружности. Напишите выражения для глубины резкости, угловой расходимости и радиусов кривизны в направлениях x и y в зависимости от W_{0x}, W_{0y} и длины волны λ. В случае $W_{0x} = 2W_{0y}$ нарисуйте форму пятна в плоскости $z = 0$ и в дальней зоне (z намного больше глубины резкости в обоих поперечных направлениях).

К РАЗДЕЛУ 3.2

1. **Фокусировка пучка.** Аргонный ионный лазер генерирует гауссов пучок с длиной волны $\lambda = 488$ нм с радиусом перетяжки $W_0 = 0{,}5$ мм. Спроектируйте однолинзовую оптическую систему для фокусировки света в пятно диаметром 100 мкм. Каково наименьшее фокусное расстояние линзы, пригодной для этого?

2. **Размер пятна.** Гауссов пучок с числом Рэлея $z_0 = 50$ см и длиной волны $\lambda = 488$ нм преобразуется в гауссов пучок с радиусом перетяжки W_0' с помощью линзы с фокусным расстоянием $f = 5$ см на расстоянии z от перетяжки, как показано на рис. 3.13. Напишите компьютерную программу для построе-

ния графика W_0' как функции z. Убедитесь, что в пределе $z - f \gg z_0$ справедливы формулы (3.40) и (3.42), а в пределе $z \ll z_0$ — формула (3.43).

3. **Рефракция пучка.** Гауссов пучок падает из воздуха ($n = 1$) в среду с плоской границей и показателем преломления $n = 1{,}5$. Ось пучка направлена по нормали в поверхности раздела, перетяжка пучка расположена на поверхности. Нарисуйте прошедший пучок. Если угол расходимости пучка в воздухе 1 мрад, то каков он в среде?

*4. **Прохождение гауссова пучка через слой с градиентным показателем преломления.** Матрица *ABCD* слоя SELFOC с квадратичной зависимостью показателя преломления (см. подразд. 1.3.2)

$$n(y) \approx n_0\left(1 - \frac{\alpha^2 y^2}{2}\right)$$

и толщиной d есть $A = \cos \alpha d$, $B = (1/\alpha) \sin \alpha d$, $C = -\alpha \sin \alpha d$, $D = \cos \alpha d$ для параксиальных лучей, направленных вдоль оси z. Гауссов пучок с длиной волны λ_0, радиусом перетяжки W_0 в свободном пространстве входит в слой вдоль оси z так, что его перетяжка совпадает с границей раздела. Используйте закон *ABCD* для нахождения ширины пучка в направлении y как функции d. Нарисуйте форму пучка, распространяющегося в среде.

К РАЗДЕЛУ 3.3

1. **Ограниченность пространственного распределения мощности в пучках Эрмита—Гаусса.** Определите отношение мощности, заключенной внутри круга радиуса $W(z)$ в поперечной плоскости к полной мощности у пучков Эрмита—Гаусса порядка (0, 0), (1, 0), (0, 1) и (1, 1). Каково отношение мощности, заключенной в круге радиуса $W(z)/10$, к полной мощности для пучков Эрмита—Гаусса порядка (0, 0) и (1, 1)?

2. **Суперпозиция двух пучков.** Нарисуйте интенсивность суперпозиции пучков (1, 0) и (0, 1) Эрмита—Гаусса, предполагая, что комплексные коэффициенты $A_{1,0}$ и $A_{0,1}$ в (3.66) равны друг другу.

3. **Фаза на оси.** Рассмотрим пучки Эрмита—Гаусса всех порядков (l, m) с числом Рэлея $z_0 = 30$ см в среде с показателем преломления $n = 1$. Определите частоты в пределах полосы $\nu = 10^{14} \pm 2 \cdot 10^9$ Гц, для которых задержка фазы на оси между плоскостями $z = -z_0$ и $z = z_0$ есть целое кратное π. Эти частоты суть моды резонатора, образованного двумя сферическими зеркалами, расположенными в плоскостях $z = \pm z_0$, как описано в подразд. 10.2.4.

ГЛАВА 4

ФУРЬЕ-ОПТИКА

Йозеф фон Фраунгофер (1787—1826) создал дифракционную решетку и внес большой вклад в наше понимание дифракции. Его эпитафия гласит: «*Approximavit sidera*» («Он приблизил звезды»)

Жан-Батист Жозеф Фурье (1768—1830) показал, что периодические функции можно конструировать в виде сумм синусоид. Гармонический анализ лежит в основе Фурье-оптики, имеющей многочисленные приложения

Денис Габор (1900—1979) изобрел голографию и внес большой вклад в ее развитие. В 1947 г. он получил первую голограмму, а в 1971 г. — Нобелевскую премию

Фурье-оптика описывает распространение световых волн на основе гармонического анализа (преобразования Фурье) в линейных системах. Метод гармонического анализа доказал свою эффективность при описании сигналов и систем во многих дисциплинах. Гармонический анализ основан на разложении произвольной функции времени $f(t)$ в суперпозицию (сумму или интеграл) гармонических функций времени с различными частотами (см. приложение А, разд. А.1). Гармоническая функция $F(\nu)\exp(j2\pi\nu t)$ с частотой ν и комплексной амплитудой $F(\nu)$ является элементарным «кирпичиком» теории. Несколько таких функций, каждая со своей амплитудой $F(\nu)$, при сложении дают функцию $f(t)$, как показано на рис. 4.1. Комплексная амплитуда $F(\nu)$, являющаяся функцией частоты, называется Фурье-образом $f(t)$. Такой подход полезен при описании линейных систем (см. приложение Б, разд. Б.1). Если

известен отклик системы на каждую из гармоник, то отклик на произвольную функцию легко определяется с помощью гармонического анализа на входе и суперпозиции на выходе.

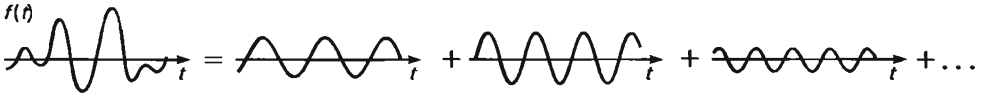


Рис. 4.1. Произвольную функцию можно представить как сумму гармонических функций с различными частотами и комплексными амплитудами

Произвольная комплексная функция $f(x, y)$ двух переменных x и y, представляющих пространственные координаты на плоскости, может быть аналогично представлена в виде суперпозиции гармонических функций от x и y, каждая из которых имеет вид

$$F(\nu_x, \nu_y) \exp[-j2\pi(\nu_x x + \nu_y y)],$$

где $F(\nu_x, \nu_y)$ — комплексная амплитуда; ν_x, ν_y — пространственные частоты (число колебаний на единицу длины, в типичном случае число колебаний на 1 мм) в направлениях x и y соответственно[1]. Гармоническая функция $F(\nu_x, \nu_y) \exp[-j2\pi(\nu_x x + \nu_y y)]$ является элементарным двумерным «строительным блоком» теории. Ее можно использовать для генерирования произвольной функции двух переменных $f(x, y)$, как показано на рис. 4.2 (см. приложение А, разд. А.3).

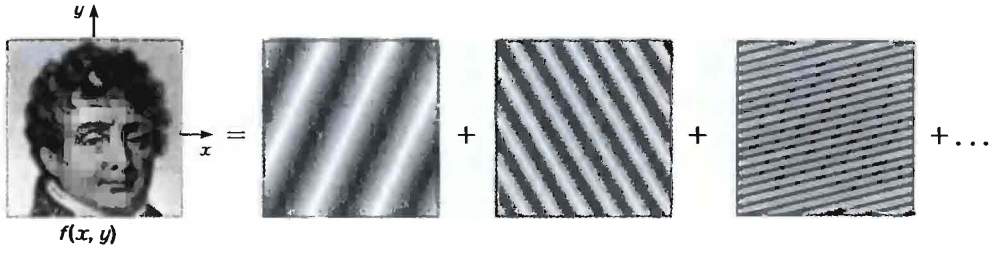


Рис. 4.2. Произвольную функцию $f(x, y)$ можно разложить на сумму гармонических функций с различными пространственными частотами и комплексными амплитудами; на рисунке эти функции схематически показаны в виде полос с плавно меняющейся яркостью

Плоская волна

$$U(x, y, z) = A \exp[-j(k_x x + k_y y + k_z z)]$$

играет важную роль в волновой оптике. Коэффициенты k_x, k_y, k_z — компоненты волнового вектора k; A — комплексная постоянная. В точках произволь-

[1] В отличие от временных пространственные гармоники определяются со знаком «минус» в показателе экспоненты (см. приложение А, разд. А.3). Такой выбор знаков соответствует виду плоской волны, распространяющейся вперед.

ной плоскости $U(x, y, z)$ является пространственной гармонической функцией. В плоскости $z = 0$, например, $U(x, y, 0)$ есть гармоническая функция

$$f(x, y) = A \exp [-j2\pi(\nu_x x + \nu_y y)],$$

где ν_x и ν_y — пространственные частоты (колеб./мм),

$$\nu_x = \frac{k_x}{2\pi}; \quad \nu_y = \frac{k_y}{2\pi}$$

k_x и k_y — пространственные угловые частоты (радиан/мм). Между плоской волной $U(x, y, z)$ и пространственной гармонической функцией $f(x, y) = U(x, y, 0)$ существует взаимно однозначное соответствие, поскольку знания k_x и k_y достаточно, чтобы определить k_z с помощью соотношения

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2} = \left(\frac{2\pi}{\lambda}\right)^2$$

Однако, как будет показано ниже, k_x и k_y не могут быть больше, чем ω/c, т. е. пространственные частоты ν_x и ν_y не могут превышать обратную длину волны $1/\lambda$. Поскольку произвольная функция $f(x, y)$ может быть представлена как суперпозиция гармонических функций, произвольная бегущая волна $U(x, y, z)$ может быть представлена как сумма плоских волн (рис. 4.3). Плоская волна является элементарным «строительным блоком» для конструирования волн произвольной сложности. Более того, если известно, как линейная оптическая система преобразует плоские волны, то с помощью принципа суперпозиции можно определить действие системы на произвольную волну.

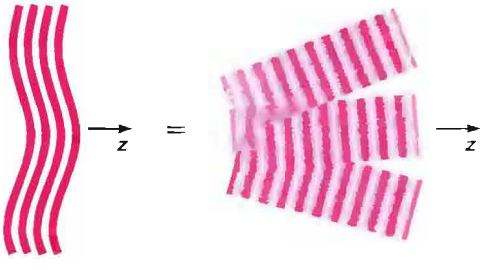


Рис. 4.3. Принцип Фурье-оптики: произвольная волна в свободном пространстве может быть представлена как суперпозиция плоских волн

Поскольку Фурье-анализ играет важную роль в описании линейных систем, полезно описать распространение света через линейные оптические элементы, включая свободное пространство, в терминах общей теории линейных систем. Комплексные амплитуды в двух плоскостях, перпендикулярных оптической оси z, будем рассматривать как вход и выход системы (рис. 4.4). Линейную систему можно охарактеризовать либо **функцией отклика на импульсное воздействие** (отклик системы на импульс или точку на входе), либо **передаточной функцией** (отклик на пространственную гармоническую функцию), как описано в приложении Б.

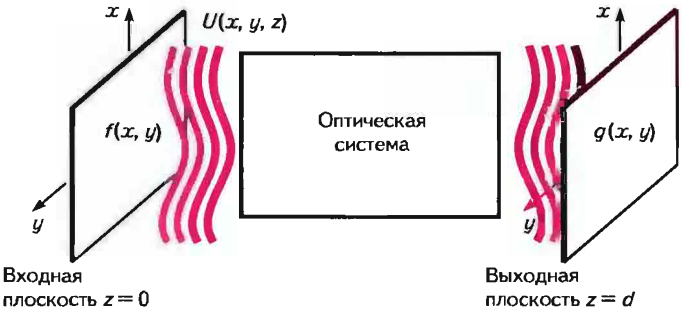


Рис. 4.4. Прохождение оптической волны $U(x, y, z)$ через оптическую систему между входной плоскостью $z = 0$ и выходной плоскостью $z = d$. Оптическая система рассматривается как линейная система, вход и выход которой характеризуются функциями $f(x, y) = U(x, y, 0)$ и $g(x, y) = U(x, y, d)$ соответственно

О данной главе

Глава начинается с Фурье-описания распространения света в свободном пространстве (см. разд. 4.1). Определяются передаточная функция и функция отклика на импульсное воздействие для свободного пространства. В разд. 4.2 показано, что линза может осуществлять пространственное преобразование Фурье. В разд. 4.3 обсуждается прохождение света через отверстия; здесь с точки зрения Фурье-оптики рассматривается явление дифракции, которое в курсах оптики для начинающих обычно описывается с помощью принципа Гюйгенса. Раздел 4.4 посвящен формированию изображений и пространственной фильтрации. Наконец, в разд. 4.5 представлено введение в голографию, запись и восстановление оптических волн. Для понимания материала данной главы необходимо знание основных свойств преобразования Фурье и линейных систем в одном и двух измерениях, обзор которых приведен в приложениях А и Б.

4.1. РАСПРОСТРАНЕНИЕ СВЕТА В СВОБОДНОМ ПРОСТРАНСТВЕ

4.4.1. Пространственные гармоники и плоские волны

Рассмотрим плоскую волну с комплексной амплитудой

$$U(x, y, z) = A \exp[-j(k_x x + k_y y + k_z z)],$$

волновым вектором $\boldsymbol{k} = (k_x, k_y, k_z)$, длиной волны λ, волновым числом

$$k = \sqrt{k_x^2 + k_y^2 + k_z^2} = \frac{2\pi}{\lambda}$$

и комплексной огибающей A. Вектор $\boldsymbol{k}$ образует с плоскостями y—z и x—z углы

$$\theta_x = \arcsin\left(\frac{k_x}{k}\right); \quad \theta_y = \arcsin\left(\frac{k_y}{k}\right),$$

соответственно, как показано на рис. 4.5. Так, если $\theta_x = 0$, компонента $\boldsymbol{k}$ в направлении x отсутствует. Комплексная амплитуда в плоскости $z = 0$, $U(x, y, 0)$ есть пространственная гармоническая функция $f(x, y) = A \exp[-j2\pi(\nu_x x + \nu_y y)]$ с пространственными частотами

$$\nu_x = \frac{k_x}{2\pi}; \quad \nu_y = \frac{k_y}{2\pi}$$

(пространственная частота выражается в колебаниях на миллиметр, а временная — в колебаниях в секунду или герцах, см. разд. 2.2). Таким образом, углы направления волнового вектора следующим образом выражаются через пространственные частоты гармонической функции:

$$\theta_x = \arcsin \lambda\nu_x; \quad \theta_y = \arcsin \lambda\nu_y. \tag{4.1}$$

Пространственные частоты и углы

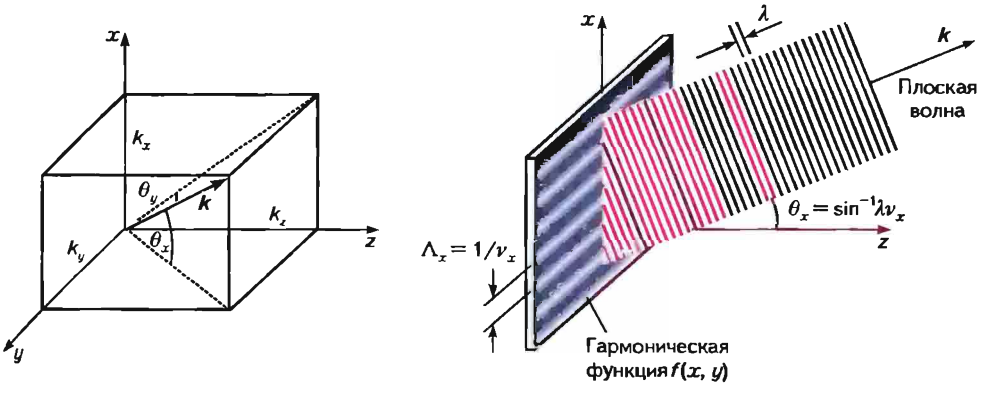


Рис. 4.5. Гармоническая функция с пространственными частотами ν_x и ν_y в плоскости $z = 0$ согласуется с плоской волной, распространяющейся под углами $\theta_x = \arcsin(\lambda\nu_x)$ и $\theta_y = \arcsin(\lambda\nu_y)$

Обозначая периоды гармонической функции (мм/колебание) в направлениях x и y через

$$\Lambda_x = \frac{1}{\nu_x}; \quad \Lambda_y = \frac{1}{\nu_y},$$

видим, что углы

$$\theta_x = \arcsin\left(\frac{\lambda}{\Lambda_x}\right); \quad \theta_y = \arcsin\left(\frac{\lambda}{\Lambda_y}\right)$$

определяются отношениями длины волны к соответствующему пространственному периоду гармонической функции в каждом направлении. Эти геометрические соотношения вытекают из согласования волновых фронтов волны с периодическим распределением гармонической функции в плоскости $z = 0$, которое иллюстрируется рис. 4.5.

Если $k_x \ll k$ и $k_y \ll k$, так что волновой вектор $\boldsymbol{k}$ является параксиальным, то углы θ_x и θ_y малы ($\sin\theta_x \approx \theta_x$, $\sin\theta_y \approx \theta_y$) и

$$\boxed{\theta_x \approx \lambda\nu_x; \quad \theta_y \approx \lambda\nu_y.} \tag{4.2}$$

Пространственные частоты и углы (параксиальное приближение)

Углы наклона волнового вектора в этом случае пропорциональны пространственным частотам соответствующей гармонической функции. Очевидно, что между плоской волной $U(x, y, z)$ и гармонической функцией $f(x, y)$ имеется взаимно однозначное соответствие. Если дана одна из них, другая легко определяется при условии, что известна длина волны λ: гармоническая функция получается, если взять значения $U(x, y, z)$ в плоскости $z = 0$, т. е. $U(x, y, 0)$. Если же, наоборот, дана гармоническая функция $f(x, y)$, то плоская волна восстанавливается по ней с помощью соотношения

$$U(x, y, z) = f(x, y)\exp(-jk_z z),$$

где

$$k_z = \pm\sqrt{k^2 - k_x^2 - k_y^2}, \quad k = \frac{2\pi}{\lambda}. \tag{4.3}$$

Условие справедливости этого соответствия:

$$k_x^2 + k_y^2 < k^2,$$

что обеспечивает действительность k_z. Это условие подразумевает, что $\lambda\nu_x < 1$ и $\lambda\nu_y < 1$, так что углы θ_x и θ_y, определяемые формулой (4.1), существуют. Знаки «+» и «–» в (4.3) представляют волны, бегущие в прямом и обратном направлениях соответственно. Далее мы будем рассматривать только волны, бегущие вперед.

Пространственный спектральный анализ

Когда плоская волна единичной амплитуды, бегущая вдоль оси z, проходит через тонкий оптический элемент с комплексным амплитудным коэффициентом пропускания

$$f(x, y) = \exp[-j2\pi(\nu_x x + \nu_y y)],$$

волна модулируется гармонической функцией, так что $U(x, y, 0) = f(x, y)$. Падающая волна преобразуется в плоскую волну с волновым вектором, направление которого характеризуется углами $\theta_x = \arcsin(\lambda\nu_x)$ и $\theta_y = \arcsin(\lambda\nu_y)$ (рис. 4.6).

Следовательно, такой элемент действует как призма, отклоняющая волну вверх на данном рисунке. Если комплексный амплитудный коэффициент пропускания задать в виде

$$f(x, y) = \exp[+j2\pi(\nu_x x + \nu_y y)],$$

то углы направления волнового вектора будут $-\theta_x$ и $-\theta_y$, так что волна будет отклоняться вниз.

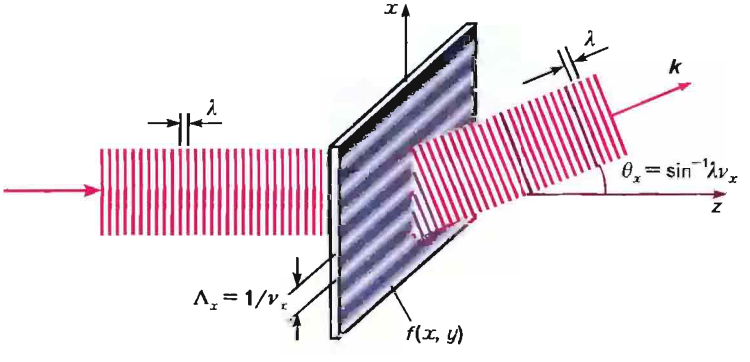


Рис. 4.6. Тонкий элемент с комплексным амплитудным коэффициентом пропускания, являющимся гармонической функцией с пространственной частотой ν_x (период $\Lambda_x = 1/\nu_x$), отклоняет плоскую волну с длиной волны λ на угол $\theta_x = \arcsin(\lambda\nu_x) = \arcsin(\lambda/\Lambda_x)$. Фиолетовым цветом подчеркивается то обстоятельство, что элемент представляет собой *фазовую* решетку, которая меняет только фазу волны

Свойство оптического элемента с гармоническим комплексным амплитудным коэффициентом пропускания изменять направление волны можно понять как интерференционное явление. В направлении, характеризуемом углом θ_x, две точки элемента, отстоящие друг от друга на период

$$\Lambda = \frac{1}{\nu_x},$$

создают волны с относительной разностью хода

$$\Lambda \sin\theta_x = \frac{1}{\nu_x}\lambda\nu_x = \lambda,$$

равной длине волны. Следовательно, все участки элемента, отстоящие друг от друга на период, вносят вклад в конструктивную интерференцию в данном направлении.

Если коэффициент пропускания оптического элемента $f(x, y)$ есть сумма нескольких гармонических функций с различными пространственными частотами, прошедшая волна также представляет собой сумму соответствующего числа плоских волн, отклоненных в различных направлениях; каждая пространственная частота отображается на соответствующее направление согласно (4.1). Ам-

плитуда каждой волны пропорциональна амплитуде соответствующей гармонической компоненты функции $f(x, y)$.

Примеры

♦ Элемент с комплексным амплитудным коэффициентом пропускания вида

$$f(x, y) = \cos(2\pi\nu_x x) = \frac{1}{2}\left[\exp(-j2\pi\nu_x x) + \exp(+j2\pi\nu_x x)\right]$$

расщепляет падающую плоскую волну на компоненты, направленные под углами $\pm\arcsin(\lambda\nu_x)$, т. е. вверх и вниз.

♦ Элемент с коэффициентом пропускания, меняющимся как $1 + \cos(2\pi\nu_y y)$, ведет себя как дифракционная решетка (см. упражнение 2.7); падающая волна расщепляется на компоненты, отклоненные вправо и влево, и часть, распространяющуюся в прямом направлении.

♦ Элемент с коэффициентом пропускания

$$f(x, y) = \mathcal{U}[\cos(2\pi\nu_x x)],$$

где $\mathcal{U}(x)$ — единичная ступенчатая функция ($\mathcal{U}(x) = 1$ при $x > 0$ и $\mathcal{U}(x) = 0$ при $x < 0$), ведет себя как периодическая система щелей, внутри которых $f(x, y) = 1$, проделанных в непрозрачном экране с $f(x, y) = 0$. Эта периодическая функция раскладывается в ряд Фурье, состоящий из гармонических функций с частотами $0, \pm\nu_x, \pm2\nu_x,$, которым соответствуют волны, распространяющиеся под углами $0, \pm\arcsin(\lambda\nu_x), \pm\arcsin(2\lambda\nu_x), \ldots$, амплитуды которых пропорциональны коэффициентам разложения Фурье. В указанных направлениях волны, прошедшие через щели, интерферируют конструктивно.

В общем случае, когда $f(x, y)$ представляет собой интегральную суперпозицию гармонических функций

$$f(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\nu_x, \nu_y)\exp\left[-j2\pi(\nu_x x + \nu_y y)\right]d\nu_x\, d\nu_y, \tag{4.4}$$

с частотами (ν_x, ν_y) и амплитудами $F(\nu_x, \nu_y)$ прошедшая волна $U(x, y, z)$ есть суперпозиция плоских волн

$$U(x, y, z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\nu_x, \nu_y)\exp\left[-j(2\pi\nu_x x + 2\pi\nu_y y)\right]\exp(-jk_z z)\,d\nu_x\, d\nu_y, \tag{4.5}$$

с комплексными огибающими $F(\nu_x, \nu_y)$, где

$$k_z = \sqrt{k^2 - k_x^2 - k_y^2} = 2\pi\sqrt{\lambda^{-2} - \nu_x^2 - \nu_y^2}.$$

Заметим, что $F(\nu_x, \nu_y)$ представляет собой Фурье-образ функции $f(x, y)$ (см. приложение А, разд. А.3).

Поскольку в результате преобразования Фурье произвольная функция может быть представлена в виде интегральной суперпозиции (4.4), свет, прошедший через тонкий оптический элемент с произвольным коэффициентом пропускания, можно описать как суперпозицию плоских волн (рис. 4.7), при условии что $\nu_x^2 + \nu_y^2 < \lambda^{-2}$.

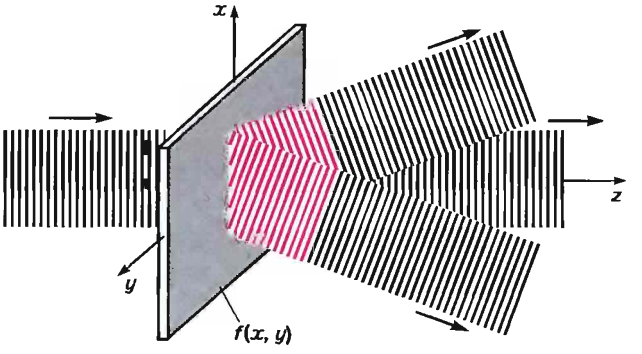


Рис. 4.7. Тонкий оптический элемент с амплитудным коэффициентом пропускания $f(x, y)$ разлагает падающую плоскую волну на множество плоских волн. Плоская волна, бегущая в направлении, задаваемом углами $\theta_x = \arcsin(\lambda\nu_x)$ и $\theta_y = \arcsin(\lambda\nu_y)$, имеет комплексную огибающую $F(\nu_x, \nu_y)$ — Фурье-образ $f(x, y)$

Описанный процесс «пространственного спектрального анализа» близок к угловой дисперсии различных частотно-временных компонент (различных длин волн), обеспечиваемой призмой. Распространение в свободном пространстве служит естественной «пространственной призмой», чувствительной к пространственной, а не временной частоте оптической волны.

Амплитудная модуляция

Рассмотрим транспарант с комплексным амплитудным коэффициентом пропускания $f_0(x, y)$. Если Фурье-образ $F_0(\nu_x, \nu_y)$ занимает область $\pm\Delta\nu_x$ и $\pm\Delta\nu_y$ в направлениях x и y, то транспарант будет отклонять падающую плоскую волну на углы θ_x и θ_y в пределах $\pm\arcsin(\lambda\Delta\nu_x)$ и $\pm\arcsin(\lambda\Delta\nu_y)$ соответственно.

Рассмотрим второй транспарант с комплексным амплитудным коэффициентом пропускания

$$f(x,\ y) = f_0(x,\ y)\exp\left[-j2\pi\left(\nu_{x_0}x + \nu_{y_0}y\right)\right],$$

где $f_0(x, y)$ — медленно меняющаяся функция по сравнению с $\exp\left[-j2\pi\left(\nu_{x_0}x + \nu_{y_0}y\right)\right]$, так что $\Delta\nu_x \ll \nu_{x_0}$ и $\Delta\nu_y \ll \nu_{y_0}$. Можно рассматривать $f(x, y)$ как амплитудно-модулированную функцию с несущими частотами ν_{x_0} и ν_{y_0} и модулирующей функцией $f_0(x, y)$. Фурье-образом $f(x, y)$ является $F_0\left(\nu_x - \nu_{x_0},\ \nu_y - \nu_{y_0}\right)$, что следует из свойства сдвига частоты преобразования Фурье (см. приложение А). Транспарант будет отклонять плоскую волну в направлениях вокруг

$$\theta_{x_0} = \arcsin\left(\lambda\nu_{x_0}\right);\quad \theta_{y_0} = \arcsin\left(\lambda\nu_{y_0}\right)$$

(рис. 4.8). Это можно легко увидеть, если рассматривать $f(x, y)$ как транспарант с коэффициентом пропускания $f_0(x, y)$, на который наложена решетка или призма с коэффициентом пропускания $\exp\left[-j2\pi\left(\nu_{x_0}x + \nu_{y_0}y\right)\right]$, вносящая отклонение на углы θ_{x_0} и θ_{y_0}.

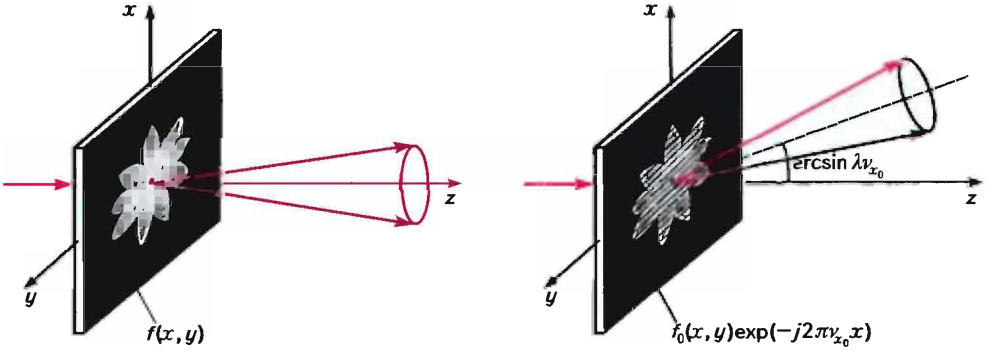


Рис. 4.8. Отклонение света транспарантами $f_0(x, y)$ и $f_0(x, y)\exp\left(-j2\pi\nu_{x_0}x\right)$. «Несущая» гармоническая функция $\exp\left(-j2\pi\nu_{x_0}x\right)$ действует как призма, отклоняющая волну на угол $\theta_{x_0} = \arcsin\left(\lambda\nu_{x_0}\right)$

Эту идею можно использовать для записи двух изображений $f_1(x, y)$ и $f_2(x, y)$ на одном транспаранте с использованием схемы пространственно-частотного мультиплексирования

$$f(x, y) = f_1(x, y)\exp\left[-j2\pi\left(\nu_{x_1}x + \nu_{y_1}x\right)\right] + f_2(x, y)\exp\left[-j2\pi\left(\nu_{x_2}x + \nu_{y_2}x\right)\right].$$

Два изображения легко разделяются при освещении транспаранта плоской волной, при этом изображения отклоняются на разные углы и таким образом пространственно разделяются. Этот принцип окажется весьма полезным в голографии (см. разд. 4.5), где часто желательно разделить два изображения, записанных на одном транспаранте.

Частотная модуляция

Теперь исследуем прохождение плоской волны через транспарант, представляющий собой «коллаж» из нескольких участков, коэффициент пропускания каждого из которых есть гармоническая функция с некоторой пространственной частотой, как показано на рис. 4.9. Если размеры каждого участка намного больше периода, каждый из них действует как дифракционная решетка или призма, отклоняющая волну в некотором направлении, так что различные порции исходного волнового фронта отклоняются в разных направлениях. Этот принцип можно использовать для создания карт оптической связи.

Транспарант может также иметь гармонический коэффициент пропускания, пространственная частота которого плавно и медленно (по сравнению с λ)

меняется от точки к точке, что в значительной степени напоминает медленные изменения частоты сигнала во времени при частотной модуляции (ЧМ). Рассмотрим, например, фазовую функцию

$$f(x, y) = \exp[-j2\pi\phi(x, y)],$$

где $\phi(x, y)$ — непрерывная медленно меняющаяся функция x и y. Вблизи точки (x_0, y_0) можно использовать разложение в ряд Тейлора

$$\phi(x, y) \approx \phi(x_0, y_0) + (x - x_0)\nu_x + (y - y_0)\nu_y,$$

где производные $\nu_x = \partial\phi/\partial x$ и $\nu_y = \partial\phi/\partial y$ вычисляются в точке (x_0, y_0). Локальное изменение $f(x, y)$, таким образом, пропорционально величине $\exp[-j2\pi(\nu_x x + \nu_y y)]$, которая является гармонической функцией с пространственными частотами

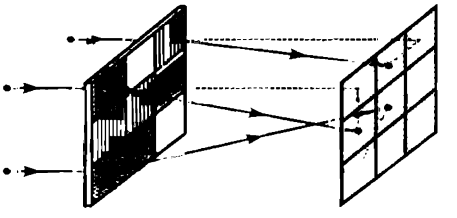


Рис. 4.9. Отклонение света транспарантом, состоящим из кусков гармонических функций (фазовых решеток) с разными пространственными частотами

$$\nu_x = \frac{\partial\phi}{\partial x}; \quad \nu_y = \frac{\partial\phi}{\partial y}.$$

Поскольку производные $\frac{\partial\phi}{\partial x}$ и $\frac{\partial\phi}{\partial y}$ меняются от точки к точке, то же самое делают и пространственные частоты. Таким образом, транспарант $f(x, y) = \exp[-j2\pi\phi(x, y)]$ отклоняет часть волны в точке (x, y) на зависящие от координат углы

$$\theta_x = \arcsin\left(\lambda\frac{\partial\phi}{\partial x}\right); \quad \theta_y = \arcsin\left(\lambda\frac{\partial\phi}{\partial y}\right).$$

Пример 4.1

Сканирование

Тонкий транспарант с комплексным амплитудным коэффициентом пропускания

$$f(x, y) = \exp\left(\frac{j\pi x^2}{\lambda f}\right)$$

вносит фазовый сдвиг $2\pi\phi(x, y)$, где

$$\phi(x, y) = -\frac{x^2}{2\lambda f},$$

так что волна в точке (x, y) отклонятся на углы

$$\theta_x = \arcsin\left(\lambda\frac{\partial\phi}{\partial x}\right) = \arcsin\left(-\frac{x}{f}\right); \quad \theta_y = 0.$$

Если $|x/f| \ll 1$, то $\theta_x \approx -x/f$, и угол отклонения прямо пропорционален поперечной координате x. Такой транспарант можно использовать для отклонения тонкого пучка света. Если двигать транспарант с постоянной скоростью, пучок будет отклоняться на линейно возрастающий угол, как показано на рис. 4.10.

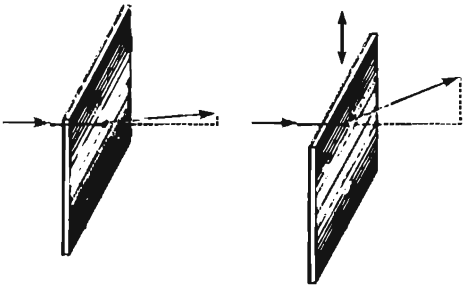


Рис. 4.10. Использование частотно-модулированного транспаранта для сканирования оптического пучка

Пример 4.2

Формирование изображения

Если транспарант в примере 4.1 освещать плоской волной, каждая ее часть будет отклоняться на свой угол, в результате волновой фронт деформируется. Локальный волновой вектор в точке x отклонится на угол $-x/f$, в результате чего все лучи встретятся на оптической оси на расстоянии f от транспаранта, как показано на рис. 4.11. Транспарант действует как цилиндрическая линза с фокусным расстоянием f. Аналогично, транспарант с коэффициентом пропускания

$$f(x,\ y) = \exp\left[\frac{j\pi\left(x^2 + y^2\right)}{\lambda f}\right]$$

действует как сферическая линза с фокусным расстоянием f. Действительно, это выражение для комплексного амплитудного коэффициента пропускания тонкой линзы [см. (2.40)].

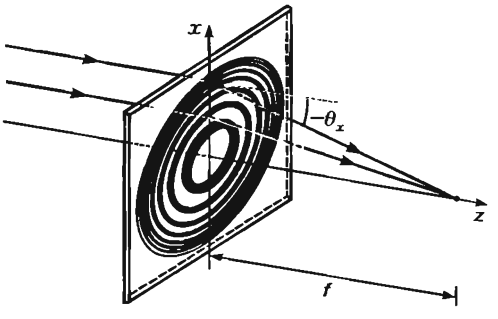


Рис. 4.11. Транспарант с коэффициентом пропускания $f(x, y) = \exp(j\pi x^2/\lambda f)$ отклоняет волну в положении x на угол $\theta_x \approx -x/f$, так что он действует как цилиндрическая линза с фокусным расстоянием f

Упражнение 4.1

Цилиндрическая линза из бинарной пластинки

Используйте гармонический анализ вблизи точки x, чтобы показать, что транспарант с комплексным амплитудным коэффициентом пропускания, равным бинарной функции

$$f(x,\ y) = \mathcal{U}\left[\cos\left(\pi\frac{x^2}{\lambda f}\right)\right], \tag{4.6}$$

где $\mathcal{U}(x)$ — единичная ступенчатая функция ($\mathcal{U}(x) = 1$ при $x > 0$ и $\mathcal{U}(x) = 0$ при $x < 0$), ведет себя как цилиндрическая линза с множественными фокусами на расстояниях ∞, $\pm f$, $\pm f/2$, (рис. 4.12).

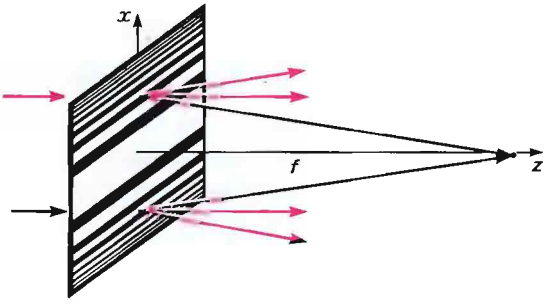


Рис. 4.12. Бинарная пластинка как цилиндрическая линза с множественными фокусами

Зонная пластинка Френеля

Двумерное обобщение бинарной пластинки из упражнения 4.1 есть аксиально-симметричный транспарант с комплексным амплитудным коэффициентом пропускания

$$f(x, y) = \mathcal{U}\left[\cos\left(\pi \frac{x^2 + y^2}{\lambda f}\right)\right], \tag{4.7}$$

известный как **зонная пластинка Френеля**. Она представляет собой систему чередующихся прозрачных и непрозрачных концентрических колец, ширина которых убывает с ростом радиуса так, что площадь остается постоянной (рис. 4.13). Луч, падающий в некоторую точку пластинки, расщепляется на множество лучей; прошедшие лучи пересекаются в множестве фокусов, расположенных на расстояниях $\pm f$, $\pm f/2$, ...; часть света проходит через пластинку без отклонения.

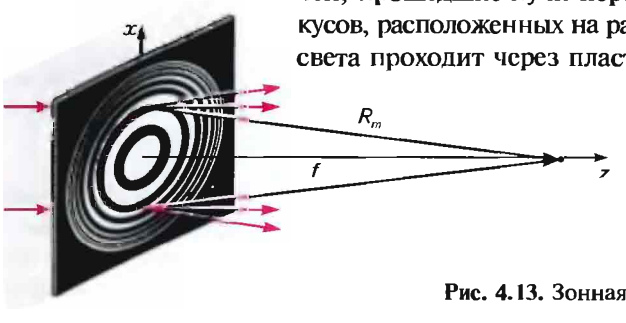


Рис. 4.13. Зонная пластинка Френеля

Действие зонной пластинки Френеля можно объяснить как интерференционный эффект (см. подразд. 2.5.2). Центр m-го кольца имеет радиус ρ_m в m-м максимуме косинуса, т. е. при

$$\frac{\pi \rho_m^2}{\lambda f} = m2\pi.$$

Расстояние R_m от фокальной точки $z = f$ до m-го кольца дается выражением

$$R_m^2 = f^2 + \rho_m^2,$$

так что

$$R_m = \sqrt{f^2 + 2m\lambda f}.$$

Если f достаточно велико, так что углы, под которыми видны кольца, малы, то $R_m \approx f + m\lambda$. Таким образом, волны, прошедшие через соседние кольца, имеют разность хода, равную длине волны, в силу чего в фокальной точке они интерферируют конструктивно. Аналогичное рассуждение справедливо для других фокальных точек.

4.1.2. Передаточная функция свободного пространства

Рассмотрим теперь распространение монохроматической световой волны с длиной волны λ и комплексной амплитудой $U(x, y, z)$ в свободном пространстве между плоскостями $z = 0$ и $z = d$, которые будем называть плоскостями входа и выхода соответственно (рис. 4.14). По данной комплексной амплитуде волны на входе $f(x, y) = U(x, y, 0)$ определим комплексную амплитуду на выходе $g(x, y) = U(x, y, d)$.

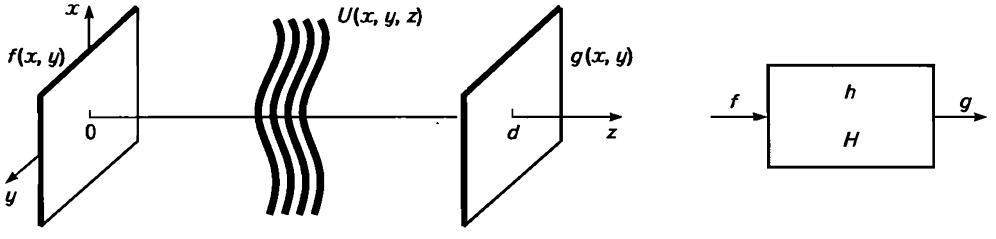


Рис. 4.14. Распространение света между двумя плоскостями рассматривается как линейная система, у которой входной и выходной сигналы — комплексная амплитуда волны в двух плоскостях

Будем рассматривать $f(x, y)$ и $g(x, y)$ как входной и выходной сигналы линейной системы. Система является линейной, поскольку уравнение Гельмгольца, которому удовлетворяет $U(x, y, z)$, линейно. Система обладает инвариантностью относительно сдвига, поскольку свободное пространство инвариантно по отношению к смещению системы координат. Как показано в приложении Б, разд. Б.2, линейная система, инвариантная относительно сдвига, может характеризоваться функцией отклика на импульсное воздействие $h(x, y)$ или передаточной функцией $H(\nu_x, \nu_y)$. Перейдем к выводу выражений для этих функций.

Передаточная функция $H(\nu_x, \nu_y)$ есть множитель, на который нужно умножить амплитуду входной пространственной гармонической функции на часто-

тах ν_x, ν_y, чтобы получить амплитуду соответствующей гармоники на выходе. Рассмотрим входную гармоническую функцию

$$f(x, y) = A \exp[-j2\pi(\nu_x x + \nu_y y)].$$

Как объяснялось ранее, она соответствует плоской волне

$$U(x, y, z) = A \exp[-j(k_x x + k_y y + k_z z)],$$

где $k_x = 2\pi\nu_x$; $k_y = 2\pi\nu_y$;

$$k_z = \sqrt{k^2 - k_x^2 - k_y^2} = 2\pi\sqrt{\lambda^{-2} - \nu_x^2 - \nu_y^2}. \tag{4.8}$$

Выходной сигнал:

$$g(x, y) = A \exp[-j(k_x x + k_y y + k_z d)],$$

так что можно записать

$$H(\nu_x, \nu_y) = \frac{g(x, y)}{f(x, y)} = \exp(-jk_z d),$$

откуда

$$\boxed{H(\nu_x, \nu_y) = \exp\left(-j2\pi d\sqrt{\lambda^{-2} - \nu_x^2 - \nu_y^2}\right).} \tag{4.9}$$

Передаточная функция свободного пространства

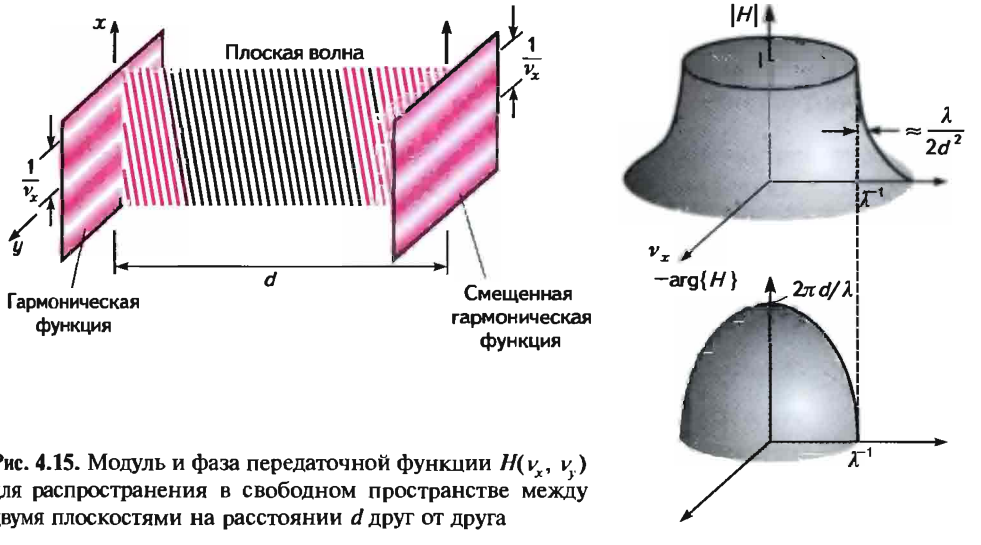


Рис. 4.15. Модуль и фаза передаточной функции $H(\nu_x, \nu_y)$ для распространения в свободном пространстве между двумя плоскостями на расстоянии d друг от друга

Таким образом, передаточная функция $H(\nu_x, \nu_y)$ — кругообразно симметричная комплексная функция пространственных частот ν_x и ν_y. Ее модуль и фаза изображены на рис. 4.15.

Вводя обозначения

$$\theta^2 = \theta_x^2 + \theta_y^2 \approx \lambda^2 \left(\nu_x^2 + \nu_y^2\right),$$

где θ — угол между волновым вектором и оптической осью, можно записать фазовый множитель в (4.9) как

$$2\pi d\sqrt{\lambda^{-2} - \nu_x^2 - \nu_y^2} = 2\pi \frac{d}{\lambda}\sqrt{1-\theta^2} = 2\pi \frac{d}{\lambda}\left(1 - \frac{\theta^2}{2} + \frac{\theta^4}{8} - \ldots\right). \qquad (4.10)$$

Пренебрегая в этом разложении малыми членами, начиная с третьего, можно приближенно записать (4.9) в виде

$$\boxed{H\left(\nu_x,\ \nu_y\right) \approx H_0 \exp\left[j\pi d\lambda\left(\nu_x^2 + \nu_y^2\right)\right],} \qquad (4.11)$$

Передаточная функция свободного пространства (приближение Френеля)

где $H_0 = \exp(-jkd)$. В этом приближении фаза является квадратичной функцией ν_x и ν_y, как показано на рис. 4.16. Данное приближение известно как **приближение Френеля.**

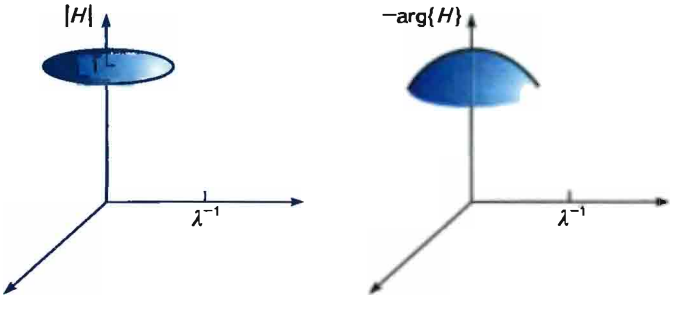
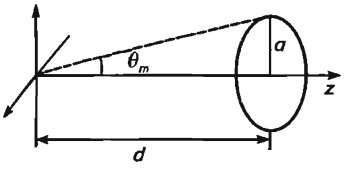


Рис. 4.16. Передаточная функция распространения в свободном пространстве для низких пространственных частот (много меньше $1/\lambda$ колебаний на мм) имеет постоянный модуль и квадратичную фазу

Условие применимости приближения Френеля состоит в том, что третий член в (4.10) мал по сравнению с π для всех θ. Это эквивалентно условию

$$\frac{\theta^4 d}{4\lambda} \ll 1. \qquad (4.12)$$

Если a — наибольшее радиальное расстояние в выходной плоскости, то наибольший угол $\theta_m \approx a/d$, и (4.12) можно записать в виде [см. (2.22)]



$$\boxed{N_F \frac{\theta_m^2}{4} \ll 1,} \qquad (4.13)$$

Приближение Френеля Условие применимости

$$\boxed{N_F = \frac{a^2}{\lambda d},} \qquad (4.14)$$

Число Френеля

где N_F — число Френеля. Например, если $a = 1$ см, $d = 100$ см и $\lambda = 0{,}5$ мкм, то $\theta_m = 10^{-2}$ радиан, $N_F = 200$ и $N_F\theta^2/4 = 5 \cdot 10^{-3}$. В этом случае приближение Френеля применимо.

Соотношение «вход—выход»

Если дана входная функция $f(x, y)$, то выходную функцию $g(x, y)$ можно определить следующим образом.

1. Найдем Фурье-образ

$$F(\nu_x, \nu_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\exp\left[j2\pi(\nu_x x + \nu_y y)\right]\mathrm{d}x\mathrm{d}y, \tag{4.15}$$

который представляет комплексные огибающие составляющих плоских волн во входной плоскости.

2. Произведение $H(\nu_x, \nu_y)F(\nu_x, \nu_y)$ дает комплексные огибающие составляющих плоских волн в выходной плоскости.

3. Комплексная амплитуда в выходной плоскости есть сумма вкладов этих плоских волн:

$$g(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(\nu_x, \nu_y)F(\nu_x, \nu_y)\exp\left[-j2\pi(\nu_x x + \nu_y y)\right]\mathrm{d}\nu_x\,\mathrm{d}\nu_y. \tag{4.16}$$

Используем для $H(\nu_x, \nu_y)$ приближение Френеля (4.11), тогда

$$\boxed{\begin{aligned} &g(x, y) = \\ &= H_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\nu_x, \nu_y)\exp\left[j\pi d\lambda(\nu_x^2 + \nu_y^2)\right]\exp\left[-j2\pi(\nu_x x + \nu_y y)\right]\mathrm{d}\nu_x\,\mathrm{d}\nu_y. \end{aligned}} \tag{4.17}$$

С помощью формул (4.17) и (4.15) можно выразить выходную функцию $g(x, y)$ через входную функцию $f(x, y)$.

4.1.3. Функция отклика на импульсное воздействие для свободного пространства

Для свободного распространения функция отклика на импульсное воздействие $h(x, y)$ системы представляет собой отклик $g(x, y)$, когда входная функция $f(x, y)$ есть точка в начале координат (0, 0). Она является результатом обратного преобразования Фурье передаточной функции $H(\nu_x, \nu_y)$. Используя результаты разд. А.3 и табл. А.2.1 приложения А, а также то, что $k = 2\pi/\lambda$, находим обратное преобразование Фурье выражения (4.11) в виде

$$\boxed{h(x, y) \approx h_0\exp\left[-jk\frac{x^2 + y^2}{2d}\right]}, \tag{4.18}$$

Функция отклика на импульсное воздействие
Свободное пространство (приближение Френеля)

где

$$h_0 = \frac{j}{\lambda d}\exp(-jkd).$$

Эта функция пропорциональна комплексной амплитуде параболоидальной волны с центром в начале координат (0, 0), взятой в плоскости $z = d$ [см. (2.21)]. Таким образом, каждая точка входной плоскости создает параболоидальную волну; все такие волны накладываются друг на друга в выходной плоскости.

Распространение в свободном пространстве как свертка

Альтернативная процедура установления связи комплексных амплитуд $f(x, y)$ и $g(x, y)$ состоит в том, что $f(x, y)$ рассматривается как суперпозиция точек (дельта-функций), каждая из которых создает параболоидальную волну. Волна, берущая начало в точке (x', y'), имеет амплитуду $f(x', y')$ и центр в точке (x', y'), так что она создает волну с амплитудой $f(x', y')h(x - x', y - y')$ в точке (x, y) выходной плоскости. Сумма таких вкладов есть двумерная свертка

$$g(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y')h(x - x', y - y')\,dx'dy', \tag{4.19}$$

которая в приближении Френеля превращается в

$$\boxed{g(x, y) = h_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y')\exp\left[-j\pi\frac{(x - x')^2 + (y - y')^2}{\lambda d}\right]dx'dy',} \tag{4.20}$$

где

$$h_0 = \frac{j}{\lambda d}\exp(-jkd).$$

Итак, в пределах применимости приближения Френеля существует два подхода к определению комплексной амплитуды $g(x, y)$ в выходной плоскости по данной комплексной амплитуде $f(x, y)$ во входной плоскости:

1) формула (4.20), основанная на пространственном представлении, в котором входная волна разлагается по элементарным параболоидальным волнам;

2) формула (4.17), основанная на частотном представлении, в котором входная волна разлагается на сумму плоских волн.

Упражнение 4.2

Гауссовы пучки с новой точки зрения

В случае если функция

$$f(x, y) = A\exp\left(-\frac{x^2 + y^2}{W_0^2}\right)$$

представляет комплексную амплитуду оптической волны $U(x, y, z)$ в плоскости $z = 0$, покажите, что $U(x, y, z)$ есть гауссов пучок (3.7), обсуждавшийся в гл. 3. Используйте как пространственное, так и частотное представление.

4.1.4. Принцип Гюйгенса—Френеля

Принцип Гюйгенса—Френеля утверждает, что каждая точка волнового фронта является источником сферической волны (рис. 4.17). Огибающая этих вторичных волн образует новый волновой фронт. Их суперпозиция образует волну в следующей плоскости. Функция отклика на импульсное воздействие системы для распространения между плоскостями $z = 0$ и $z = d$:

$$h(x, y) \propto \frac{1}{r}\exp(-jkr); \quad r = \sqrt{x^2 + y^2 + d^2}. \tag{4.21}$$

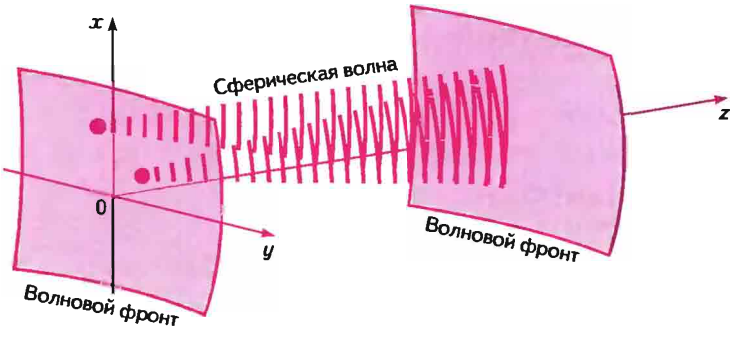


Рис. 4.17. Принцип Гюйгенса—Френеля. Каждая точка волнового фронта генерирует сферическую волну

В параксиальном приближении сферическая волна (4.21) аппроксимируется параболоидальной волной (4.18) (см. разд. 2.2.2). Наш вывод функции отклика на импульсное воздействие, таким образом, согласуется с принципом Гюйгенса—Френеля.

4.2. ОПТИЧЕСКОЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ

Как было показано в разд. 4.1, распространение света в свободном пространстве удобно описывать с помощью Фурье-анализа. Если комплексная амплитуда монохроматической волны с длиной волны λ в плоскости $z = 0$ есть функция $f(x, y)$, состоящая из гармонических компонент с различными пространственными частотами, то каждая гармоническая составляющая соответствует плоской волне. Плоская волна, распространяющаяся под углами $\theta_x = \arcsin(\lambda\nu_x)$ и $\theta_y = \arcsin(\lambda\nu_y)$, соответствует компонентам с пространственными частотами ν_x и ν_y и имеет амплитуду $F(\nu_x, \nu_y)$, являющуюся Фурье-обра-

зом $f(x, y)$. Отсюда следует, что свет может служить для осуществления преобразования Фурье двумерной функции $f(x, y)$, для чего нужно просто изготовить транспарант с амплитудным коэффициентом пропускания $f(x, y)$, через который пропускать однородную плоскую волну с единичным модулем амплитуды.

Так как каждая плоская волна бесконечно простирается в пространстве и, следовательно, перекрывается с другими плоскими волнами, необходимо найти метод разделения этих волн. Мы покажем, что на достаточно больших расстояниях только одна плоская волна вносит вклад в полную амплитуду в каждой точке выходной плоскости, так что Фурье-компоненты оказываются разделенными естественно. Более практичный подход состоит в использовании линзы для фокусировки каждой плоской волны в одну точку, как будет описано далее.

4.2.1. Преобразование Фурье в дальней зоне

Перейдем к доказательству того, что если длина распространения d достаточно велика, единственной плоской волной, дающей вклад в комплексную амплитуду в точке (x, y) выходной плоскости, является волна, направление которой образует углы

$$\theta_x \approx \frac{x}{d}; \quad \theta_y \approx \frac{y}{d}$$

с оптической осью (рис. 4.18). Это волна с компонентами волнового вектора

$$k_x \approx \frac{x}{d}k; \quad k_y \approx \frac{y}{d}k$$

и амплитудой $F(\nu_x, \nu_y)$ с

$$\nu_x = \frac{x}{\lambda d}; \quad \nu_y = \frac{y}{\lambda d}.$$

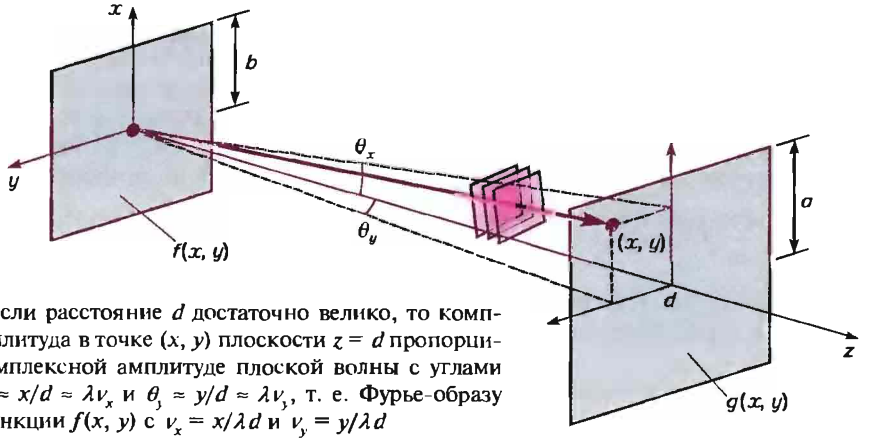


Рис. 4.18. Если расстояние d достаточно велико, то комплексная амплитуда в точке (x, y) плоскости $z = d$ пропорциональна комплексной амплитуде плоской волны с углами наклона $\theta_x \approx x/d \approx \lambda\nu_x$ и $\theta_y \approx y/d \approx \lambda\nu_y$, т. е. Фурье-образу $F(\nu_x, \nu_y)$ функции $f(x, y)$ с $\nu_x = x/\lambda d$ и $\nu_y = y/\lambda d$

Комплексные амплитуды $g(x, y)$ и $f(x, y)$ одной волны в плоскостях $z = d$ и $z = 0$ связаны соотношением

$$\boxed{g(x, y) \approx h_0 F\left(\frac{x}{\lambda d}, \frac{y}{\lambda d}\right),} \qquad (4.22)$$

Распространение в свободном пространстве как преобразование Фурье (приближение Фраунгофера)

где $F(\nu_x, \nu_y)$ — Фурье-образ $f(x, y)$;

$$h_0 = \frac{j}{\lambda d}\exp(-jkd).$$

Вклады всех других волн взаимно уничтожаются из-за деструктивной интерференции. Это приближение называется **приближением Фраунгофера.**

Как будет видно из дальнейших рассуждений, условием применимости приближения Фраунгофера является

$$\boxed{N_F \ll 1 \quad \text{и} \quad N'_F \ll 1.} \qquad (4.23)$$

Приближение Фраунгофера
Условие применимости

$$N_F = \frac{a^2}{\lambda d}; \quad N'_F = \frac{b^2}{\lambda d}$$

Таким образом, приближение Фраунгофера применимо при малых числах Френеля N_F и N'_F. Это условие более жесткое, чем условие применимости приближения Френеля, которое требует, чтобы

$$\frac{N_F \theta_m^2}{4} \ll 1$$

[см. (4.13)]. Поскольку $\theta_m \ll 1$ в параксиальном приближении, условие Френеля можно выполнить и без $N_F \ll 1$.

Доказательства свойства распространения в свободном пространстве в приближении Фраунгофера

Начнем с соотношения между $g(x, y)$ и $f(x, y)$ в (4.20). Фаза в аргументе экспоненты

$$\frac{\pi}{\lambda d}\left[(x - x')^2 + (y - y')^2\right] = \frac{\pi}{\lambda d}\left[\left(x^2 + y^2\right) + \left(x'^2 + y'^2\right) - 2(xx' + yy')\right].$$

Если $f(x, y)$ сосредоточена в малой области радиуса b, а расстояние d достаточно велико, чтобы число Френеля

$$N'_F = \frac{b^2}{\lambda d}$$

было мало, то фазовый множитель

$$\frac{\pi}{\lambda d}\left(x'^2 + y'^2\right) \le \pi \frac{b^2}{\lambda d}$$

пренебрежимо мал, и (4.20) можно аппроксимировать выражением

$$g(x,\ y) = h_0 \exp\left(-j\pi \frac{x^2 + y^2}{\lambda d}\right) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x',\ y') \exp\left(j2\pi \frac{xx' + yy'}{\lambda d}\right) \mathrm{d}x'\mathrm{d}y' \qquad (4.24)$$

Множители $x/\lambda d$ и $y/\lambda d$ можно рассматривать как частоты $\nu_x = x/\lambda d$ и $\nu_y = y/\lambda d$, так что

$$g(x,\ y) = h_0 \exp\left(-j\pi \frac{x^2 + y^2}{\lambda d}\right) F\left(\frac{x}{\lambda d},\ \frac{y}{\lambda d}\right), \qquad (4.25)$$

где $F(\nu_x,\ \nu_y)$ — Фурье-образ функции $f(x,\ y)$. Также можно пренебречь фазовым множителем $\exp|-j\pi(x^2 + y^2)/\lambda d|$ в (4.25), тогда (4.22) получается, если рассматривать только точки выходной плоскости, лежащие в круге радиусом a с центром на оси z, так что

$$\pi \frac{x^2 + y^2}{\lambda d} \le \pi \frac{a^2}{\lambda d} \ll \pi.$$

Это выполняется, если число Френеля

$$N_F = \frac{a^2}{\lambda d} \ll 1.$$

Другое доказательство основано на формуле (4.17), которая выражает комплексную амплитуду $g(x,\ y)$ в виде интеграла, содержащего плоские волны разных частот. Если d достаточно велико, чтобы фаза в интеграле была много больше 2π, то, используя метод стационарной фазы[3], можно показать, что только одно значение ν_x вносит вклад в интеграл. Это то значение, для которого производная фазы $\pi\lambda d\nu_x^2 - 2\pi\nu_x x$ по ν_x обращается в нуль, т. е. $\nu_x = x/\lambda d$. Аналогично, единственное значение ν_y, дающее вклад в интеграл, есть $\nu_y = y/\lambda d$. Это доказывает утверждение, что для дальней зоны лишь одна плоская волна вносит вклад в поле в данной точке.

Упражнение 4.3

Условия применимости приближений Френеля и Фраунгофера: сравнение

Покажите, что приближение Фраунгофера накладывает более сильные ограничения, чем приближение Френеля. Пусть длина волны света $\lambda = 0{,}5$ мкм,

[3] См., например, *Born M., Wolf E.* Cambridge University Press, 7th expanded and corrected ed., 2002, Appendix III.

точки объекта лежат в пределах круговой апертуры радиуса $b = 1$ см, а точки наблюдения — в пределах круговой апертуры радиуса $a = 2$ см. Определите диапазон расстояний d между предметной плоскостью и плоскостью наблюдения, для которых применимо каждое из этих приближений.

Резюме

В приближении Фраунгофера комплексная амплитуда $g(x, y)$ волны с длиной λ в плоскости $z = d$ пропорциональна Фурье-образу $F(\nu_x, \nu_y)$ комплексной амплитуды $f(x, y)$ в плоскости $z = 0$ при значениях пространственных частот $\nu_x = x/\lambda d$ и $\nu_y = y/\lambda d$. Приближение применимо, если функция $f(x, y)$ на входной плоскости локализована внутри круга радиусом b, удовлетворяющего условию

$$\frac{b^2}{\lambda d} \ll 1,$$

а точки наблюдения на выходной плоскости лежат внутри круга радиусом a, для которого

$$\frac{a^2}{\lambda d} \ll 1.$$

4.2.2. Преобразование Фурье с помощью линзы

Плоские волны, составляющие данную волну, можно разделить также с помощью линзы. Тонкая сферическая линза преобразует плоскую волну в параболоидальную волну, которая фокусируется в точку на фокальной плоскости линзы (см. разд. 2.4 и упражнение 2.5). Если плоская волна приходит под малыми углами θ_x и θ_y, центр параболоидальной волны лежит вблизи точки $(\theta_x f, \theta_y f)$, где f — фокусное расстояние (рис. 4.19). Таким образом, линза отображает каждое направление (θ_x, θ_y) в единственную точку $(\theta_x f, \theta_y f)$ фокальной плоскости и таким образом разделяет вклады различных плоских волн.

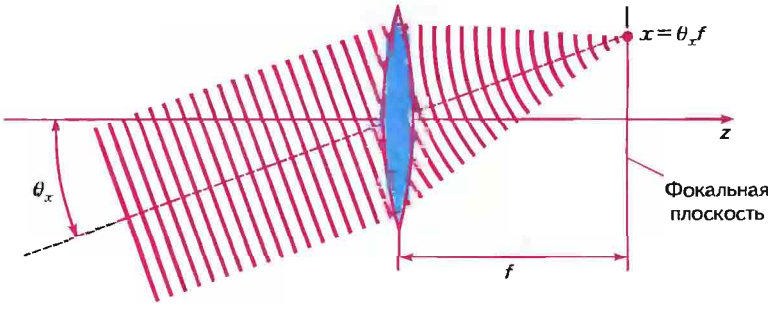


Рис. 4.19. Фокусировка плоской волны в точку. Направление (θ_x, θ_y) отображается в точку $(x, y) = (\theta_x f, \theta_y f)$ (см. упражнение 2.5)

Пусть для оптической системы, показанной на рис. 4.20, $f(x, y)$ есть комплексная амплитуда оптической волны в плоскости $z = 0$. Свет разлагается на плоские волны, причем волна, бегущая под малыми углами $\theta_x = \lambda\nu_x$ и $\theta_y = \lambda\nu_y$, имеет комплексную амплитуду, пропорциональную Фурье-образу $F(\nu_x, \nu_y)$. Эта волна фокусируется линзой в точку (x, y) фокальной плоскости, где $x = \theta_x f = \lambda f \nu_x$ и $y = \theta_y f = \lambda f \nu_y$. Следовательно, комплексная амплитуда в точке (x, y) выходной плоскости пропорциональна Фурье-образу функции $f(x, y)$, вычисленному при $\nu_x = x/\lambda f$ и $\nu_y = y/\lambda f$, так что

$$g(x, y) \propto F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right). \tag{4.26}$$

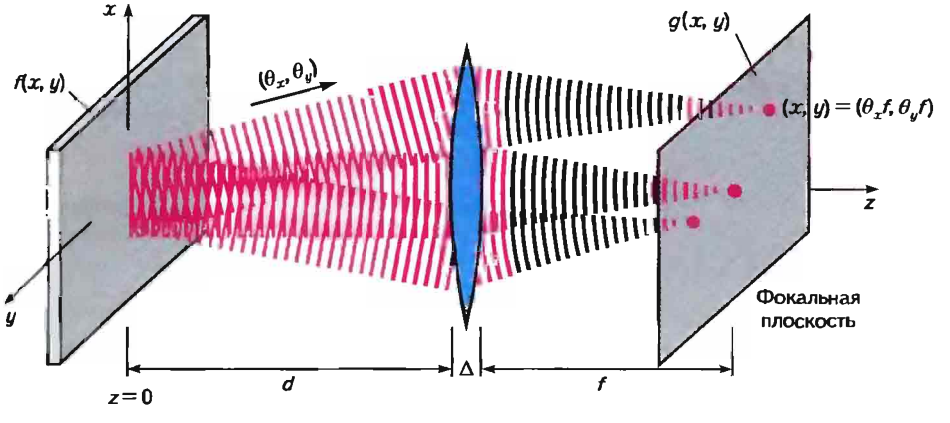


Рис. 4.20. Фокусировка плоских волн, связанных с Фурье-гармониками входной функции $f(x, y)$, в точки фокальной плоскости. Амплитуда плоской волны с направлением $(\theta_x, \theta_y) = (\lambda\nu_x, \lambda\nu_y)$ пропорциональна Фурье-образу $F(\nu_x, \nu_y)$ и фокусируется в точку $(x, y) = (\theta_x f, \theta_y f) = (\lambda f\nu_x, \lambda f\nu_y)$

Для определения коэффициента пропорциональности в (4.26) разложим входную функцию $f(x, y)$ на Фурье-компоненты и проследим за прохождением плоской волны, соответствующей каждой компоненте, через оптическую систему. Затем сложим вклады этих волн на выходной плоскости и получим $g(x, y)$. Считая волны параксиальными и применяя приближения Френеля, получим

$$\boxed{g(x, y) = h_l \exp\left(j\pi \frac{(x^2 + y^2)(d - f)}{\lambda f^2}\right) F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right),} \tag{4.27}$$

где

$$h_l = H_0 h_0 = \frac{j}{\lambda f} \exp\left[-jk(d + f)\right].$$

Таким образом, коэффициент пропорциональности в (4.26) содержит фазовый множитель, являющийся квадратичной функцией x и y.

Поскольку

$$|h_l| = \frac{1}{\lambda f},$$

из (4.27) следует, что оптическая интенсивность в выходной плоскости есть

$$I(x, y) = \frac{1}{(\lambda f)^2}\left|F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)\right|^2 \tag{4.28}$$

Интенсивность света в выходной плоскости (задней фокальной плоскости линзы) пропорциональна, таким образом, квадрату модуля Фурье-образа комплексной амплитуды волны во входной плоскости, независимо от расстояния d.

Фазовый множитель в (4.27) исчезает, если $d = f$, так что

$$\boxed{g(x, y) = h_l F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right),} \tag{4.29}$$

Фурье-преобразующее свойство линзы

где

$$h_l = \frac{j}{\lambda f}\exp(-j2kf).$$

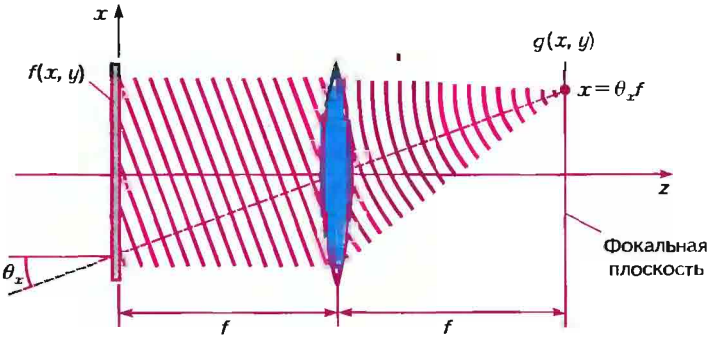


Рис. 4.21. 2*f*-система. Фурье-компонента функции $f(x, y)$ с пространственными частотами ν_x и ν_y порождает плоскую волну, направленную под углами $\theta_x = \lambda\nu_x$ и $\theta_y = \lambda\nu_y$, которая фокусируется линзой в точку $(x, y) = (\theta_x f, \theta_y f) = (\lambda f\nu_x, \lambda f\nu_y)$, так что функция $g(x, y)$ пропорциональна Фурье-образу $F(x/\lambda f, y/\lambda f)$

В системе с такой геометрией, известной как **2*f*-система** (рис. 4.21), комплексные амплитуды в передней и задней фокальных плоскостях линзы связаны преобразованием Фурье как по амплитуде, так и по фазе.

Резюме

Комплексная амплитуда света в точке (x, y) задней фокальной плоскости линзы с фокусным расстоянием f пропорциональна Фурье-образу комплексной амплитуды в передней фокальной плоскости при значениях частот $\nu_x = x/\lambda f$, $\nu_y = y/\lambda f$. Это соотношение верно, если применимо приближение Френеля. Без линзы преобразование Фурье можно получить только в приближении Фраунгофера, которое является более ограничивающим.

---

*Доказательство Фурье-преобразующего свойства линзы в приближении Френеля

Доказательство включает следующие четыре шага.

1. Плоская волна, направление которой характеризуется углами $\theta_x = \lambda\nu_x$ и $\theta_y = \lambda\nu_y$, имеет комплексную амплитуду

$$U(x, y, 0) = F(\nu_x, \nu_y)\exp\left[-j2\pi(\nu_x x + \nu_y y)\right]$$

в плоскости $z = 0$ и

$$U(x, y, d) = H(\nu_x, \nu_y)F(\nu_x, \nu_y)\exp\left[-j2\pi(\nu_x x + \nu_y y)\right]$$

в плоскости $z = d$ непосредственно перед линзой, где $H(\nu_x, \nu_y)$ — передаточная функция слоя d свободного пространства,

$$H(\nu_x, \nu_y) = H_0\exp\left[j\pi\lambda d\left(\nu_x^2 + \nu_y^2\right)\right]; \quad H_0 = \exp(-jkd).$$

2. После прохождения линзы комплексная амплитуда умножается на вносимый линзой фазовый множитель $\exp\left[j\pi(x^2 + y^2)/\lambda f\right]$ (при этом фазовый множитель $\exp(-jk\Delta)$, где Δ — толщина линзы, игнорируется). Тогда

$$U(x, y, d+\Delta) = \\ = H_0\exp\left(j\pi\frac{x^2+y^2}{\lambda f}\right)\exp\left[j\pi\lambda d\left(\nu_x^2+\nu_y^2\right)\right]F(\nu_x, \nu_y)\exp\left[-j2\pi\left(\nu_x x + \nu_y y\right)\right]. \tag{4.30}$$

Это выражение упрощается путем преобразования

$$-2\nu_x x + \frac{x^2}{\lambda f} = \frac{x^2 - 2\nu_x\lambda f x}{\lambda f} = \frac{(x-x_0)^2 - x_0^2}{\lambda f},$$

где $x_0 = \lambda\nu_x f$; аналогичное соотношение получается для y с $y_0 = \lambda\nu_y f$, тогда

$$U(x, y, d+\Delta) = A(\nu_x, \nu_y)\exp\left[j\pi\frac{(x-x_0)^2 + (y-y_0)^2}{\lambda f}\right], \tag{4.31}$$

где

$$A(\nu_x, \nu_y) = H_0\exp\left[j\pi\lambda(d-f)\left(\nu_x^2+\nu_y^2\right)\right]F(\nu_x, \nu_y). \tag{4.32}$$

Выражение (4.31) представляет собой комплексную амплитуду параболоидальной волны, сходящейся в точку (x_0, y_0) фокальной плоскости линзы $z = d + \Delta + f$.

3. Теперь рассмотрим распространение волны в свободном пространстве между линзой и выходной плоскостью, чтобы определить $U(x, y, d + \Delta + f)$. Применим к (4.10) формулу (4.20), используем соотношение

$$\int \exp\left[\frac{j2\pi(x - x_0)x'}{\lambda f}\right] dx' = \lambda f \delta(x - x_0)$$

и получим

$$U(x, y, d + \Delta + f) = h_0 (\lambda f)^2 A(\nu_x, \nu_y)\delta(x - x_0)\delta(y - y_0), \qquad (4.33)$$

где

$$h_0 = \frac{j}{\lambda f}\exp(-jkf).$$

Действительно, плоская волна фокусируется в точку с координатами $x_0 = \lambda \nu_x f$ и $y_0 = \lambda \nu_y f$.

4. Последний шаг состоит в интегрировании по всем плоским волнам, т. е. по всем ν_x и ν_y. С использованием свойств дельта-функции

$$\delta(x - x_0) = \delta(x - \lambda f \nu_x) = \frac{1}{\lambda f}\delta\left(\nu_x - \frac{x}{\lambda f}\right),$$

интегрирование дает

$$g(x, y) = h_0 A\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right).$$

Используя (4.32), окончательно получаем (4.27).

Упражнение 4.4

Обратное преобразование Фурье

В однолинзовой оптической системе, изображенной на рис. 4.21, распределение поля в передней фокальной плоскости ($z = 2f$) отличается масштабом от Фурье-образа распределения поля в задней фокальной плоскости ($z = 0$). Убедитесь, что при инверсии координат в передней фокальной плоскости $(x, y) \rightarrow (-x, -y)$ получающееся распределение поля дает обратное преобразование Фурье.

4.3. ДИФРАКЦИЯ СВЕТА

Когда оптическая волна проходит через отверстие (апертуру) в непрозрачном экране и далее распространяется в свободном пространстве на некоторое расстояние, ее распределение интенсивности называется дифракционной картиной. Если свет рассматривать как лучи, дифракционная картина пред-

ставляет собой геометрическую тень отверстия. Из-за волновой природы света дифракционная картина может слегка или существенно отличаться от тени отверстия, в зависимости от расстояния между экраном и плоскостью наблюдения, длины волны и размеров отверстия. Пример показан на рис. 4.22. Трудно точно определить, как экран изменяет падающую волну, но распространение в свободном пространстве после прохождения отверстия всегда подчиняется законам, описанным ранее в данной главе.

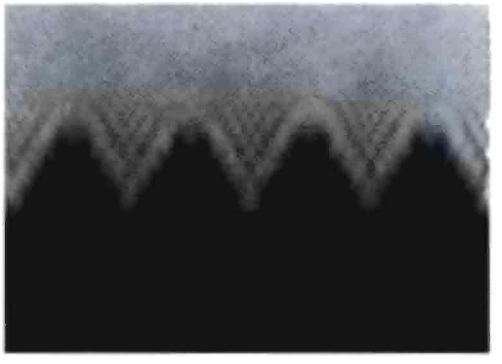


Рис. 4.22. Дифракционная картина от зубьев пилы (из *Cagnet M., Franc)on M., Thrierr J.C.* Atlas of Optical Phenomena. Springer-Verlag, 1962)

Простейшая теория дифракции основана на *предположении*, что падающая волна проходит без изменения в точках отверстия, но обращается в нуль на обратной стороне непрозрачной части экрана. Если $U(x, y)$ и $f(x, y)$ — комплексные амплитуды волны на левой и правой сторонах экрана (рис. 4.23), то в соответствии с нашим предположением

$$f(x, y) = U(x, y)\,p(x, y), \tag{4.34}$$

где

$$p(x, y) = \begin{cases} 1 & \text{внутри отверстия,} \\ 0 & \text{вне отверстия} \end{cases} \tag{4.35}$$

называется **апертурной функцией**.

При заданном $f(x, y)$ комплексная амплитуда $g(x, y)$ в плоскости наблюдения на расстоянии d от экрана можно пределить методами, описанными в разд. 4.1 и 4.2. Дифракционную картину $I(x, y) = |g(x, y)|^2$ называют **дифракцией Фраунгофера** или **дифракцией Френеля**, если распространение волны в свободном пространстве описывается приближением Фраунгофера или приближением Френеля соответственно.

Хотя такой подход в большинстве случаев дает разумно точные результаты, он не является точным. Справедливость и непротиворечивость предположения о том, что комплексная амплитуда $f(x, y)$ исчезает за пределами отверстия на задней стороне экрана, вызывает сомнения, поскольку прошедшая волна распространяется во всех направлениях и, следовательно, достигает указанных точек. Теория дифракции, основанная на строгом решении уравнения Гельмгольца с граничными условиями на отверстии, математически сложна. Лишь для нескольких геометрических структур удается получить точные решения. Однако разработано множество различных теорий дифракции, использующих те или

иные допущения и дающие различную степень точности. Точная теория дифракции не входит в круг вопросов, рассматриваемых в данной книге.

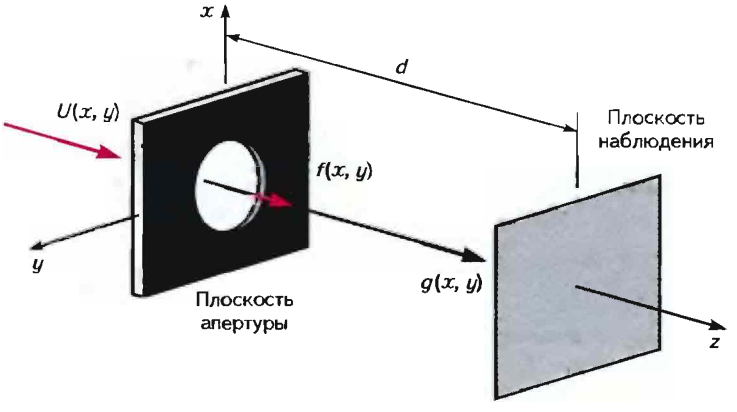


Рис. 4.23. Волна $U(x, y)$ проходит через апертуру с амплитудным пропусканием $p(x, y)$, создавая волну с комплексной амплитудой $f(x, y) = U(x, y)p(x, y)$. После прохождения расстояния d в свободном пространстве комплексная амплитуда волны $g(x, y)$, а интенсивность $I(x, y) = |g(x, y)|^2$

4.3.1. Дифракция Фраунгофера

Теория дифракции Фраунгофера основана на умножении падающей волны на апертурную функцию и использовании приближения Фраунгофера для описания последующего распространения волны в свободном пространстве. Приближение Фраунгофера справедливо, если расстояние d между экраном и плоскостью наблюдения достаточно велико, так что число Френеля

$$N_F' = \frac{b^2}{\lambda d} \ll 1,$$

где b — наибольшее радиальное расстояние в пределах отверстия.

Предполагая, что падающая волна плоская, имеет интенсивность I_i и распространяется вдоль оси z, так что

$$U(x, y) = \sqrt{I_i},$$

имеем

$$f(x, y) = \sqrt{I_i}\, p(x, y).$$

В приближении Фраунгофера [см. (4.22)]

$$g(x, y) \approx \sqrt{I_i}\, h_0 P\left(\frac{x}{\lambda d}, \frac{y}{\lambda d}\right), \tag{4.36}$$

где

$$P(\nu_x,\ \nu_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,\ y)\exp\left[j2\pi\left(\nu_x x + \nu_y y\right)\right]\mathrm{d}x\mathrm{d}y \tag{4.37}$$

есть Фурье-образ функции $p(x, y)$;

$$h_0 = \frac{j}{\lambda d}\exp(-jkd).$$

Тогда дифракционная картина

$$\boxed{I(x,\ y) = \frac{I_i}{(\lambda d)^2}\left|P\left(\frac{x}{\lambda d},\ \frac{y}{\lambda d}\right)\right|^2.} \tag{4.38}$$

Итак, интенсивность дифракционной картины Фраунгофера в точке (x, y) пропорциональна квадрату модуля Фурье-образа апертурной функции $p(x, y)$, вычисленного при пространственных частотах

$$\nu_x = \frac{x}{\lambda d};\quad \nu_y = \frac{y}{\lambda d}.$$

Упражнение 4.5

Дифракция Фраунгофера на прямоугольном отверстии

Убедитесь, что картина дифракции Фраунгофера от прямоугольного отверстия высотой D_x и шириной D_y на расстоянии d

$$I(x,\ y) = I_0\mathrm{sinc}^2\frac{D_x x}{\lambda d}\mathrm{sinc}^2\frac{D_y y}{\lambda d}, \tag{4.39}$$

где I_0 — максимальная интенсивность,

$$I_0 = \frac{D_x D_y}{\lambda d}I_i;$$

$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}.$$

Проверьте, что первые нули этой картины получаются при

$$x = \pm\frac{\lambda d}{D_x};\quad y = \pm\frac{\lambda d}{D_y},$$

так что угловая расходимость дифрагирующего света определяется формулами

$$\boxed{\theta_x = \frac{\lambda}{D_x};\quad \theta_y = \frac{\lambda}{D_y}.} \tag{4.40}$$

Если $D_y < D_x$, то дифракционная картина шире в направлении y, чем в направлении x, как показано на рис. 4.24.

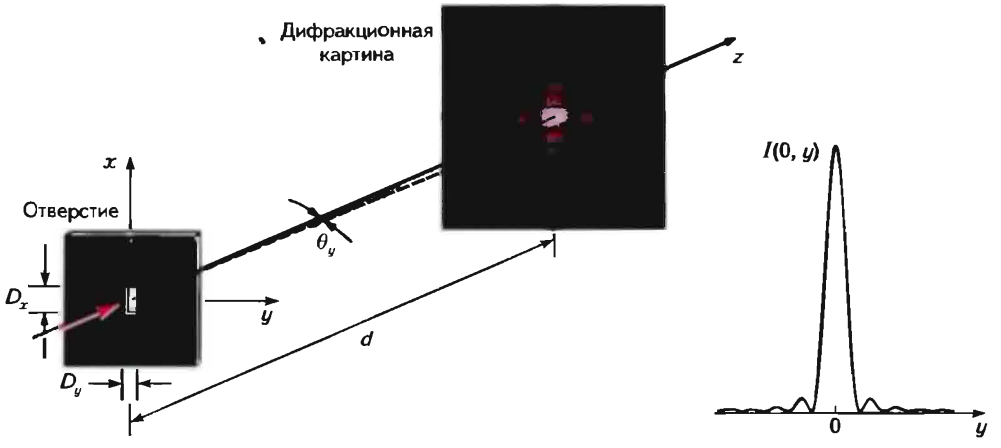


Рис. 4.24. Дифракция Фраунгофера на прямоугольном отверстии. Центральный пик картины имеет угловые полуширины $\theta_x = \lambda/D_x$, $\theta_y = \lambda/D_y$

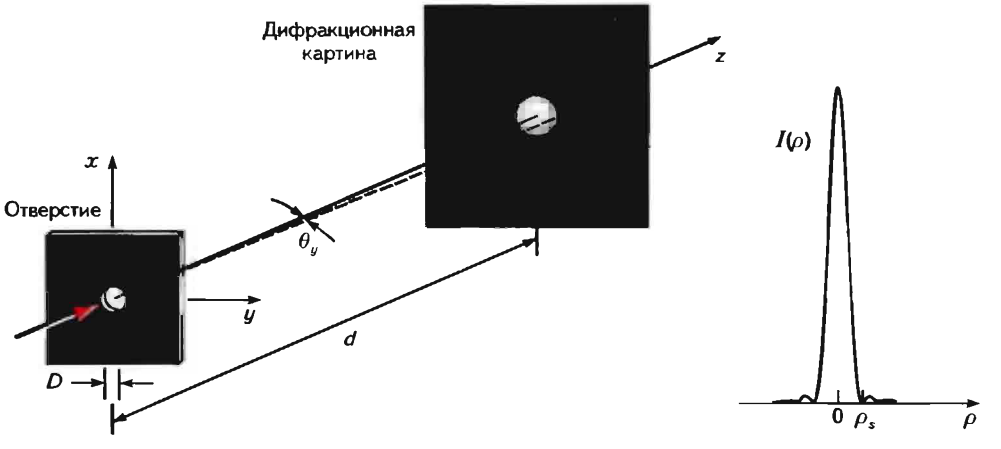


Рис. 4.25. Дифракция Фраунгофера от круглого отверстия дает картину Эйри с радиусом центрального диска, соответствующим углу $\theta = 1{,}22\lambda/D$

Упражнение 4.6

Дифракция Фраунгофера на круглом отверстии

Проверьте, что картина дифракции Фраунгофера на круглом отверстии диаметром D (рис. 4.25) имеет вид

$$I(x, y) = I_0\left[\frac{2J_1(\pi D\rho/\lambda d)}{\pi D\rho/\lambda d}\right]^2 \qquad \rho = \sqrt{x^2 + y^2}, \tag{4.41}$$

где I_0 — максимальная интенсивность,

$$I_0 = \left(\frac{\pi D^2}{4\lambda d}\right)^2 I_i;$$

$J_1(\ )$ — функция Бесселя первого порядка. Преобразование Фурье функций с круговой симметрией обсуждается в приложении А, разд. А.3. Картина (4.40) с круговой симметрией, известная как **картина Эйри**, состоит из центрального диска, окруженного кольцами. Проверьте, что радиус центрального диска, известного как **диск Эйри**, равен

$$\rho_s = 1{,}22\frac{\lambda d}{D},$$

а половина угла, под которым этот диск виден с центра отверстия, равна

$$\boxed{\theta = 1{,}22\frac{\lambda}{D}.} \qquad (4.42)$$

Половина углового размера диска Эйри

Приближение Фраунгофера обычно справедливо для предельно больших расстояний d. Это условие выполняется в приложениях, связанных с дальней оптической связью в свободном пространстве, таких как лазерный радар (лидар) и спутниковая связь. Однако, как показано в подразд. 4.2.2, если для фокусировки света, испытывающего дифракцию, использовать линзу с фокусным расстоянием f, то распределение интенсивности в фокальной плоскости линзы будет пропорционально квадрату модуля Фурье-образа апертурной функции $p(x, y)$ при значениях

$$\nu_x = \frac{x}{\lambda f}; \quad \nu_y = \frac{y}{\lambda f}.$$

Наблюдаемая картина, таким образом, идентична получаемой из (4.38) с заменой расстояния d на фокусное расстояние f.

Упражнение 4.7

Размер пятна сфокусированного оптического пучка

Пучок света фокусируется с помощью линзы с фокусным расстоянием f с круговой апертурой диаметра (рис. 4.26). Приближенно считая пучок внутри апертуры плоской волной, проверьте, что распределение интенсивности в фокусируемом пятне дается выражением

$$I(x, y) = I_0\left[\frac{2J_1(\pi D\rho/\lambda f)}{\pi D\rho/\lambda f}\right]^2 \qquad \rho = \sqrt{x^2 + y^2}, \qquad (4.43)$$

где I_0 — максимальная интенсивность. Сравните радиус пятна фокусировки

$$\rho_s = 1,22\lambda\frac{f}{D}, \tag{4.44}$$

с размером пятна при фокусировке гауссова пучка с радиусом перетяжки W_0 идеальной линзой бесконечной апертуры [см. (3.45)].

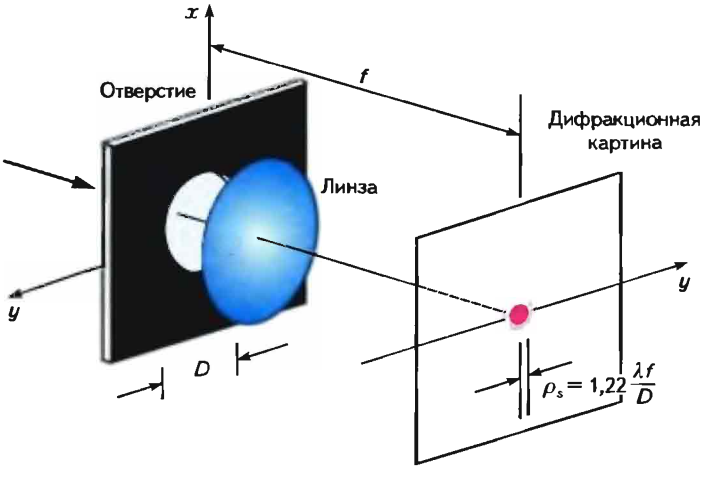


Рис. 4.26. Фокусировка плоской волны, прошедшей через круглое отверстие диаметром D

4.3.2. Дифракция Френеля

Теория дифракции Френеля основана на предположении, что падающая волна умножается на апертурную функцию $p(x, y)$ и распространяется в свободном пространстве в соответствии с приближением Френеля. Если падающая волна плоская, имеет интенсивность I_i и распространяется вдоль оси z, то комплексная амплитуда непосредственно после апертуры есть

$$f(x,\ y) = \sqrt{I_i}\,p(x,\ y).$$

Пользуясь (4.20) для дифракционной картины $I(x, y) = |g(x, y)|^2$ на расстоянии d, получаем

$$I(x,\ y) = \frac{I_i}{(\lambda d)^2}\left|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x',\ y')\exp\left[-j\pi\frac{(x-x')^2+(y-y')^2}{\lambda d}\right]dx'dy'\right|^2 \tag{4.45}$$

Удобно нормировать все расстояния, используя $\sqrt{\lambda d}$ в качестве единицы длины, так что X и X' — нормированные расстояния (аналогично для y и y'),

$$X = \frac{x}{\sqrt{\lambda d}};\quad X' = \frac{x'}{\sqrt{\lambda d}}.$$

Тогда формула (4.45) дает

$$I(X,\ Y) = \\ = I_i \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(X',\ Y') \exp\left[-j\pi (X - X')^2 + (Y - Y')^2\right] dX'\,dY' \right|^2 \tag{4.46}$$

Интеграл в (4.45) представляет собой свертку $p(X, Y)$ и $\exp[-j\pi(X^2 + Y^2)]$. Графики действительной и мнимой частей $\exp(-j\pi X^2)$, т. е. $\cos \pi X^2$ и $\sin \pi X^2$, показаны на рис. 4.27. Они осциллируют с возрастающей частотой, и их первые дифракционные максимумы лежат в интервалах $|X| < 1/\sqrt{2}$ и $|X| < 1$ соответственно. Полная площадь под графиком функции $\exp(-j\pi X^2)$ равна 1, причем основной вклад дают несколько первых дифракционных максимумов, поскольку вклады последующих максимумов компенсируют друг друга. Если a — радиус отверстия, то радиус нормированной функции $p(X, Y)$ составляет $a/\sqrt{\lambda d}$. Результат свертки, зависящий от относительного размера двух функций, определяется числом Френеля $N_F = a^2/\lambda d$.

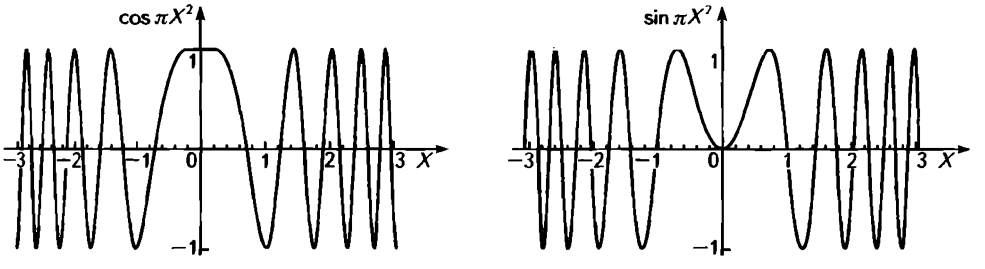


Рис. 4.27. Действительная и мнимая части $\exp(-j\pi X^2)$

Если число Френеля велико, нормированная ширина апертуры $a/\sqrt{\lambda d}$ много больше, чем ширина главного максимума, и свертка приближенно дает более широкую функцию $p(X, Y)$. При этом условии картина дифракции Френеля представляет собой тень отверстия, как следует ожидать, исходя из лучевой оптики. Заметим, что лучевая оптика применима в пределе $\lambda \to 0$, который соответствует пределу $N_F \to \infty$. В противоположном пределе, когда число Френеля мало, становится применимым приближение Фраунгофера и наблюдается картина дифракции Фраунгофера.

Пример 4.3

Дифракция Френеля на щели

Предположим, что апертура представляет собой щель шириной $D = 2a$, так что $p(x, y) = 1$ при $|x| \le a$ и 0 в противоположном случае. Нормированная координата равна $X = x/\sqrt{\lambda d}$ и

$$P(X,\ Y) = \begin{cases} 1 & |X| \le \dfrac{a}{\sqrt{\lambda d}} = \sqrt{N_F}, \\ 0 & \text{в остальных случаях,} \end{cases} \tag{4.47}$$

где N_F — число Френеля. Подставляя это в (4.45), получаем

$$I(X,\ Y) = I_i|g(X)|^2,$$

где

$$g(X) = \int_{-\sqrt{N_F}}^{\sqrt{N_F}} \exp\left[-j\pi\left(X - X'\right)^2\right] dX' = \int_{X-\sqrt{N_F}}^{X+\sqrt{N_F}} \exp\left[-j\pi X'^2\right] dX'. \qquad (4.48)$$

Этот интеграл выражается через интегралы Френеля

$$C(x) = \int_0^x \cos\frac{\pi\alpha^2}{2} d\alpha; \quad S(x) = \int_0^x \sin\frac{\pi\alpha^2}{2} d\alpha, \qquad (4.49)$$

программы для вычисления которых входят в состав стандартных компьютерных математических библиотек.

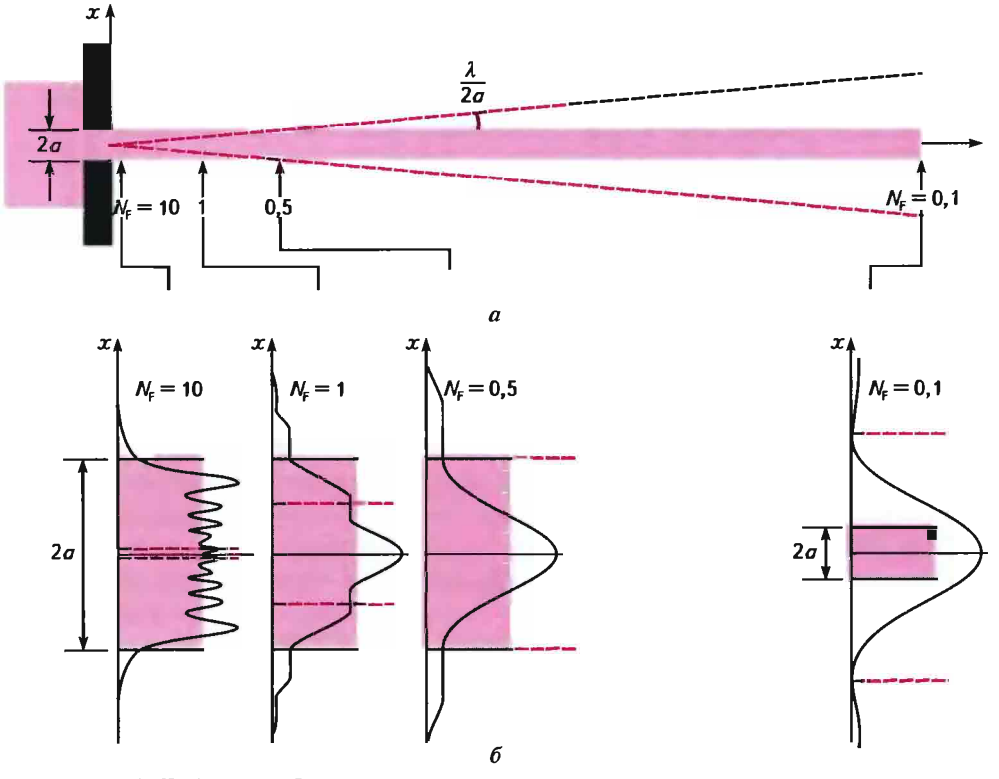


Рис. 4.28. Дифракция Френеля на шели шириной $D = 2a$:

а — затенная область — геометрическая тень апертуры. *Штриховая линия* — ширина пучка за счет дифракции Фраунгофера; *б* — дифракционная картина на четырех расстояниях от щели, отмеченных стрелками на рис. 4.28, *а* и соответствующих числам Френеля $N_F = 10$; 1; 0,5 и 0,1. *Затенные области* показывают геометрическую тень щели. *Штриховыми линиями* на уровне $|x| = (\lambda/D)d$ показана ширина дифракционной картины Фраунгофера в дальней зоне. Там, где штриховые линии совпадают с краями геометрической тени, число Френеля составляет $N_F = a^2/\lambda d = 0{,}5$

Комплексную функцию $g(X)$ можно вычислить также с помощью преобразования Фурье. Поскольку $g(X)$ представляет собой свертку прямоугольной функции с шириной $\sqrt{N_F}$ и экспоненциальной функции $\exp(-j\pi X^2)$, ее Фурье-образ:

$$G(\nu_x) \propto \operatorname{sinc}\left(\sqrt{N_F}\nu_x\right)\exp\left(-j\pi\nu_x^2\right)$$

(см. табл. А.1 в прпложении А). Таким образом, $g(X)$ можно вычислить путем нахождения обратного преобразования Фурье функции $G(\nu_x)$. Если $N_F \gg 1$, ширина функции $\operatorname{sinc}\left(\sqrt{N_F}\nu_x\right)$ намного меньше ширины первого дифракционного максимума функции $\exp\left(-j\pi\nu_x^2\right)$ (см. рис. 4.27), так что $G(\nu_x) \propto \operatorname{sinc}\left(\sqrt{N_F}\nu_x\right)$ и $g(X)$ представляет собой прямоугольную функцию, описывающую геометрическую тень апертуры.

Дифракционная картина от щели построена графически на рис. 4.28 для различных значений числа Френеля, соответствующих различным расстояниям d до апертуры. На очень малых расстояниях (очень большие N_F) дифракционная картина представляет собой идеальную тень апертуры. С ростом расстояния (уменьшением N_F) волновая природа света начинает проявляться в виде небольших колебаний интенсивности по краям тени (см. также дифракционную картину на рис. 4.22). Для очень малых N_F получается картина Фраунгофера, описываемая формулой (4.39). Это функция sinc с первым нулем под углом

$$\frac{\lambda}{D} = \frac{\lambda}{2a}.$$

Пример 4.4

Дифракция Френеля на гауссовой апертуре

В случае, когда апертурная функция $p(x, y)$ прелставляет собой функцию Гаусса

$$p(x, y) = \exp\left[-\frac{x^2 + y^2}{W_0^2}\right], \tag{4.50}$$

выражение (4.44), описывающее дифракцию Френеля, можно получить в точном аналитическом виде путем вычисления свертки функций

$$\exp\left[-\frac{x^2 + y^2}{W_0^2}\right]; \quad h_0 \exp\left[-j\pi\frac{x^2 + y^2}{\lambda d}\right]$$

с помощью, например, преобразования Фурье (см. приложение А). В результате дифракционная картина описывается формулой

$$I(x, y) = I_i\left[\frac{W_0}{W(d)}\right]^2 \exp\left[-2\frac{x^2 + y^2}{W^2(d)}\right],$$

где

$$W^2(d) = W_0^2 + \theta_0^2 d^2; \quad \theta_0 = \frac{\lambda}{\pi W_0}.$$

Дифракционная картина описывается гауссовой функцией с полушириной $W(d)$ на уровне $1/e^2$. Для малых d имеем $W(d) \approx W_0$, однако по мере роста d величина $W(d)$ приближается к $\theta_0 d$, когда d достаточно велико, чтобы можно было применять приближение Фраунгофера, так что угол, под которым видна картина дифракции Фраунгофера, равен θ_0. Эти результаты проиллюстрированы на рис. 4.29, аналогичный рис. 4.28 для дифракции на щели. Волна, испытывающая дифракцию на гауссовой апертуре, представляет собой гауссов пучок, подробно описанный в гл. 3.

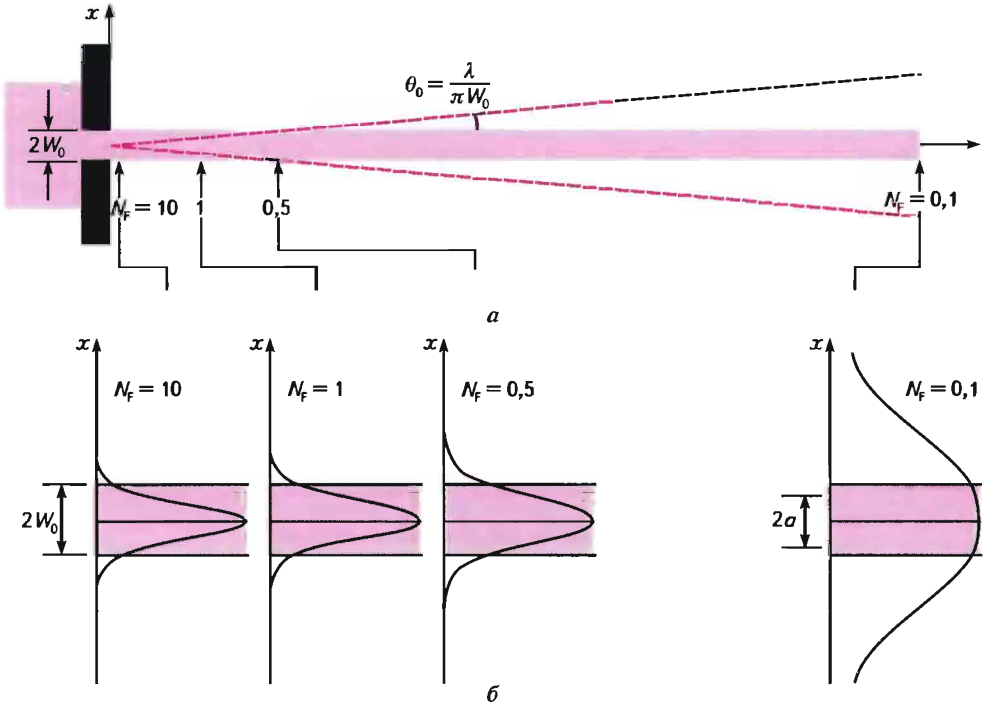


Рис. 4.29. Дифракция Френеля на гауссовой апертуре радиуса W_0 на расстояниях d таких, что параметр $\frac{\pi}{2}\frac{W_0^2}{\lambda d}$, аналогичный числу Френеля N_F на рис. 4.28, равен 10; 1; 0,5 и 0,1. Эти значения соответствуют $W(d)/W_0$ = 1,001; 1,118; 1,414 и 5,099. Дифракционная картина является гауссовой на всех расстояниях

Резюме

По мере возрастания расстояния от апертуры дифракционная картина нредставляет собой:

- тень апертуры;
- дифракционную картину Френеля: свертка нормированной апертурной функции с $\exp[-j\pi(X^2 + Y^2)]$;

• картину дифракции Фраунгофера: квадрат модуля Фурье-образа апертурной функции. В дальней зоне угол расходимости пропорционален $1/D$, где D — диаметр апертуры.

4.4. ФОРМИРОВАНИЕ ИЗОБРАЖЕНИЯ

Идеальная система, формирующая изображение, — это оптическая система, которая воспроизводит распределение света в одной плоскости — плоскости предметов — на другой плоскости — плоскости изображений. Поскольку процесс оптической передачи не бывает идеальным, изображение никогда не является точной копией предмета. Помимо изменения масштаба, связанного с увеличением системы, изображение размывается из-за неидеальной фокусировки и дифракции оптических волн. Данный раздел посвящен описанию систем для формирования изображений и их качества. Для характеристики формирования изображения используются методы описания линейных систем, такие как метод функции отклика на импульсное воздействие и метод передаточной функции (см. приложение Б). Вначале представлен простой подход на основе лучевой оптики, затем последовательно развивается описание на основе оптики волн.

4.4.1. Лучевая оптика однолинзовой изображающей системы

Рассмотрим изображающую систему с одной линзой, имеющей фокусное расстояние f и удаленной на расстояния d_1 и d_2 от плоскости предметов и плоскости изображений соответственно, показанную на рис. 4.30. Когда

$$\frac{1}{d_1}+\frac{1}{d_2}=\frac{1}{f},$$

система сфокусирована таким образом, что параксиальные лучи, исходящие из каждой точки плоскости предметов, собираются в одну точку плоскости изображений. В рамках лучевой теории света отображение «идеально» в том смысле, что каждая точка предмета дает единственную точку изображения. Функция отклика на импульсное воздействие системы представляет собой импульсную функцию.

Теперь предположим, что система не сфокусирована (рис. 4.31), причем ошибка фокусировки составляет

$$\boxed{\varepsilon=\frac{1}{d_1}+\frac{1}{d_2}-\frac{1}{f}.} \qquad (4.51)$$

Ошибка фокусировки

Точка предмета создает световое пятно на плоскости изображений, которое является тенью апертуры линзы. Распределение света в этом пятне является функцией отклика на импульсное воздействие системы. Для простоты рассмотрим точку предмета, лежащую на оптической оси, и определим распределение света $h(x, y)$, которое она создает в плоскости изображений.

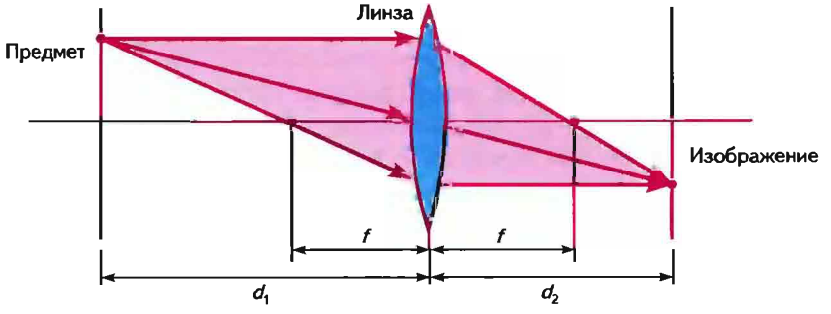


Рис. 4.30. Ход лучей в сфокусированной изображающей системе

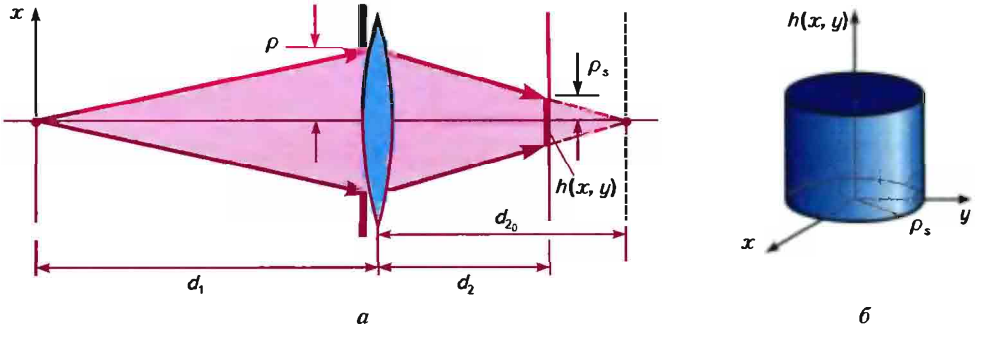


Рис. 4.31. Ход лучей в несфокусированной изображающей системе (*а*); функция отклика на импульсное воздействие изображающей системы с круговой апертурой диаметром D отлична от нуля и постоянна в круге радиусом $\rho_s = \varepsilon d_2 D/2$, где ε — ошибка фокусировки (*б*)

Предположим, что плоскость сфокусированного изображения лежит на расстоянии d_{2_0}, удовлетворяющем условию

$$\frac{1}{d_{2_0}} + \frac{1}{d_1} = \frac{1}{f}.$$

Тень точки на краю апертуры на расстоянии ρ от оси есть точка плоскости изображений на расстоянии ρ_s от оси, для которого

$$\frac{\rho_s}{\rho} = \frac{d_{2_0} - d_2}{d_{2_0}} = 1 - \frac{d_2}{d_{2_0}} = 1 - \frac{d_2}{1/f - 1/d_1} = 1 - \frac{d_2}{1/d_2 - \varepsilon} = \varepsilon d_2.$$

Если $p(x, y)$ — апертурная функция, называемая также **функцией зрачка** [$p(x, y) = 1$ для точек внутри апертуры и 0 во всех остальных точках], то $h(x, y)$ отличается от нее масштабным преобразованием с коэффициентом

$$\frac{\rho_s}{\rho} = \varepsilon d_2,$$

т. е.

$$\boxed{h(x, y) \propto p\left(\frac{x}{\varepsilon d_2}, \frac{y}{\varepsilon d_2}\right).} \qquad (4.52)$$

Функция отклика на импульсное воздействие (в лучевой оптике)

Например, круговой апертуре диаметра D соответствует функция отклика на импульсное воздействие, отличная от нуля и постоянная в круге радиусом

$$\boxed{\rho_s = \frac{1}{2}\varepsilon d_2 D,} \qquad (4.53)$$

Радиус пятна размытия

как показано на рис. 4.31. Радиус ρ_s пятна размытия является обратной мерой разрешающей силы системы и качества изображения. Низкие значения ρ_s означают способность системы разрешать мелкие детали. Поскольку ρ_s пропорционально диаметру апертуры D, качество изображения можно улучшить, используя меньший диаметр апертуры. Малая апертура соответствует пониженной чувствительности системы к ошибкам фокусировки, т. е. повышению «глубины резкости».

4.4.2. Волновая оптика формирования изображения в 4*f*-системе

Рассмотрим теперь двухлинзовую изображающую систему, показанную на рис. 4.32. Эта система, называемая 4*f*-системой, служит для получения изображений с увеличением, равном единице, как легко видеть из хода лучей.

Анализ распространения волн в такой системе становится простым, если рассматривать ее как последовательность двух подсистем, осуществляющих преобразование Фурье. Первая подсистема (между плоскостью предметов и Фурье-плоскостью) производит прямое преобразование Фурье, а вторая (между Фурье-плоскостью и плоскостью изображений) — обратное преобразование Фурье, поскольку система координат в плоскости изображений инвертирована (см. упражнение 4.4). В результате в отсутствие апертуры изображение является точной копией предмета.

Пусть $f(x, y)$ — комплексный амплитудный коэффициент пропускания транспаранта, помещенного в плоскости предметов и освещаемого плоской волной $\exp(-jkz)$, распространяющейся вдоль оси z, как показано на рис. 4.33, а $g(x, y)$ —

комплексная амплитуда волны в плоскости изображений. Первая линза осуществляет пространственное преобразование Фурье функции $f(x, y)$ и разделяет ее Фурье-компоненты таким образом, что каждая точка Фурье-плоскости соответствует единственной пространственной частоте. Затем эти компоненты воссоединяются с помощью второй линзы, и исходное распределение полностью восстанавливается.

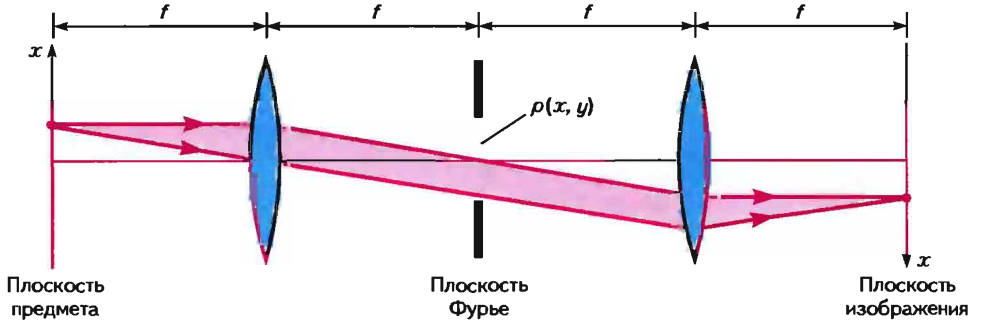


Рис. 4.32. Изображение в 4*f*-системе. При использовании инвертированной системы координат в плоскости изображений увеличение равно единице

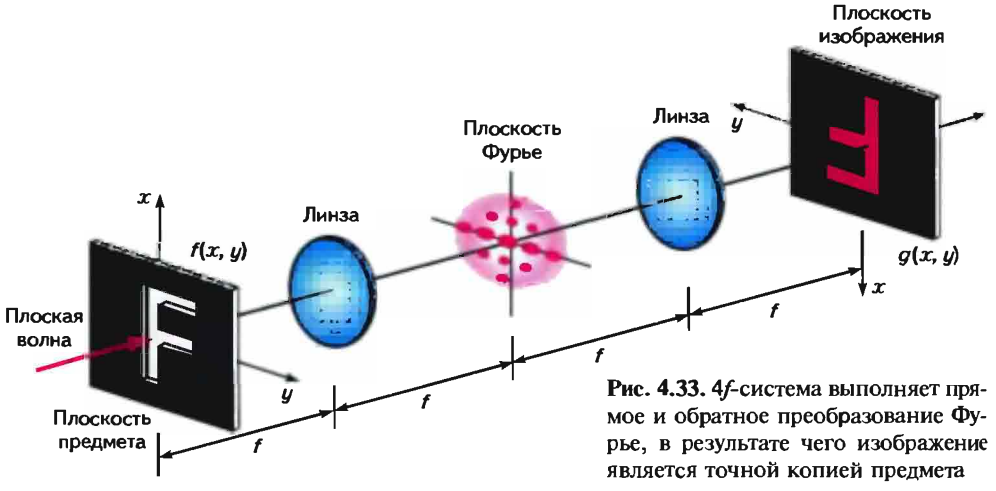


Рис. 4.33. 4*f*-система выполняет прямое и обратное преобразование Фурье, в результате чего изображение является точной копией предмета

4*f*-систему можно использовать как пространственный фильтр, в котором изображение $g(x, y)$ есть фильтрованная копия предмета $f(x, y)$. Поскольку Фурье-компоненты функции $f(x, y)$ доступны в виде точек Фурье-плоскости, для их селективного отбора можно использовать маску, блокирующую одни компоненты и пропускающую другие, как показано на рис. 4.34. Фурье-компонента $f(x, y)$ с пространственной частотой (ν_x, ν_y) локализуется на Фурье-плоскости в точках $x = \lambda f \nu_x$, $y = \lambda f \nu_y$. Для реализации фильтра с передаточной функцией $H(\nu_x, \nu_y)$ комплексный амплитудный коэффициент пропускания маски

$p(x, y)$ должен быть пропорционален $H(x/\lambda f, y/\lambda f)$. Таким образом, передаточная функция фильтра, использующего маску с пропусканием $p(x, y)$,

$$H(\nu_x, \nu_y) = p(\lambda f \nu_x, \lambda f \nu_y), \tag{4.54}$$

Передаточная функция 4*f*-системы

где мы пренебрегли фазовым множителем $j\exp(-j2kf)$, связанным с каждым преобразованием Фурье [аргумент h_l в (4.29)]. Фурье-образы $G(\nu_x, \nu_y)$ и $F(\nu_x, \nu_y)$ функций $g(x, y)$ и $f(x, y)$ связаны между собой соотношением $G(\nu_x, \nu_y) = H(\nu_x, \nu_y)F(\nu_x, \nu_y)$.

Это весьма простой результат: *передаточная функция имеет ту же форму, что и функция зрачка.* Соответствующая функция отклика на импульсное воздействие $h(\nu_x, \nu_y)$ получается из $H(\nu_x, \nu_y)$ обратным Фурье-преобразованием

$$h(x, y) = \frac{1}{(\lambda f)^2} P\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right), \tag{4.55}$$

Функция отклика на импульсное воздействие 4*f*-системы

где $P(\nu_x, \nu_y)$ — Фурье-образ $p(x, y)$.

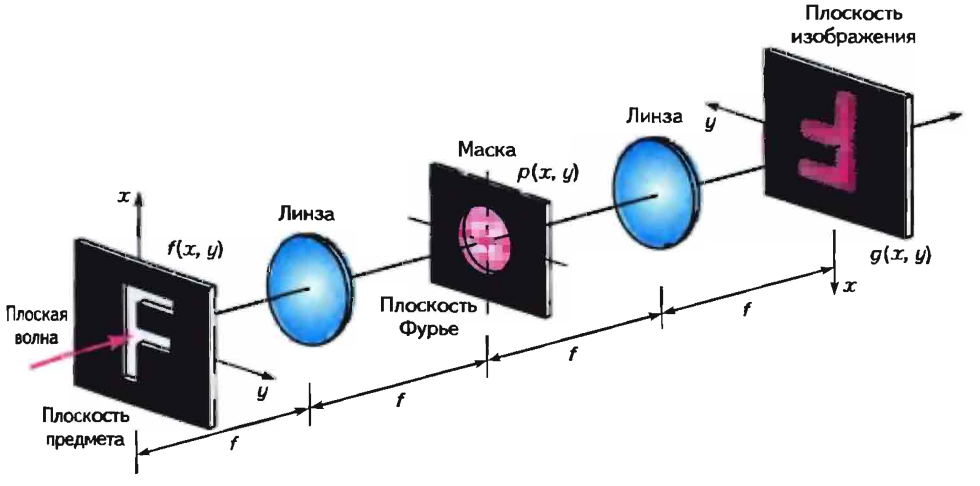


Рис. 4.34. Пространственная фильтрация. Транспаранты, расположенные в предметной и фурье-плоскости, имеют комплексные амплитудные коэффициенты пропускания $f(x, y)$ и $p(x, y)$. Плоская волна, распространяющаяся в направлении оси z, модулируется предметным транспарантом, подвергается Фурье-преобразованию первой линзой, умножается на коэффициент пропускания маски, расположенной в Фурье-плоскости, после чего вторая линза производит обратное преобразование Фурье. В результате комплексная амплитуда $g(x, y)$ в плоскости изображений представляет собой фильтрованную версию $f(x, y)$. Система имеет передаточную функцию $H(\nu_x, \nu_y) = p(\lambda f \nu_x, \lambda f \nu_y)$

Примеры пространственных фильтров

♦ Идеальный кругообразно-симметричный *фильтр нижних частот* имеет передаточную функцию $H(\nu_x, \nu_y) = 1$ при $\nu_x^2 + \nu_y^2 < \nu_s^2$ и $H(\nu_x, \nu_y) = 0$ в остальных случаях. Он пропускает пространственные частоты, меньшие частоты отсечки ν_s, и блокирует более высокие частоты. Этот фильтр представляет собой маску в виде кругового отверстия диаметром $D = 2\lambda f \nu_s$. Например, если $D = 2$ см, $\lambda = 1$ мкм и $f = 100$ см, частота отсечки (пространственная ширина полосы) $\nu_s = D/2\lambda f = 10$ линий/мм. Этот фильтр подавляет пространственные частоты больше, чем 10 линий/мм, так что наименьший размер различимых деталей фильтрованного изображения составляет примерно 0,1 мм.

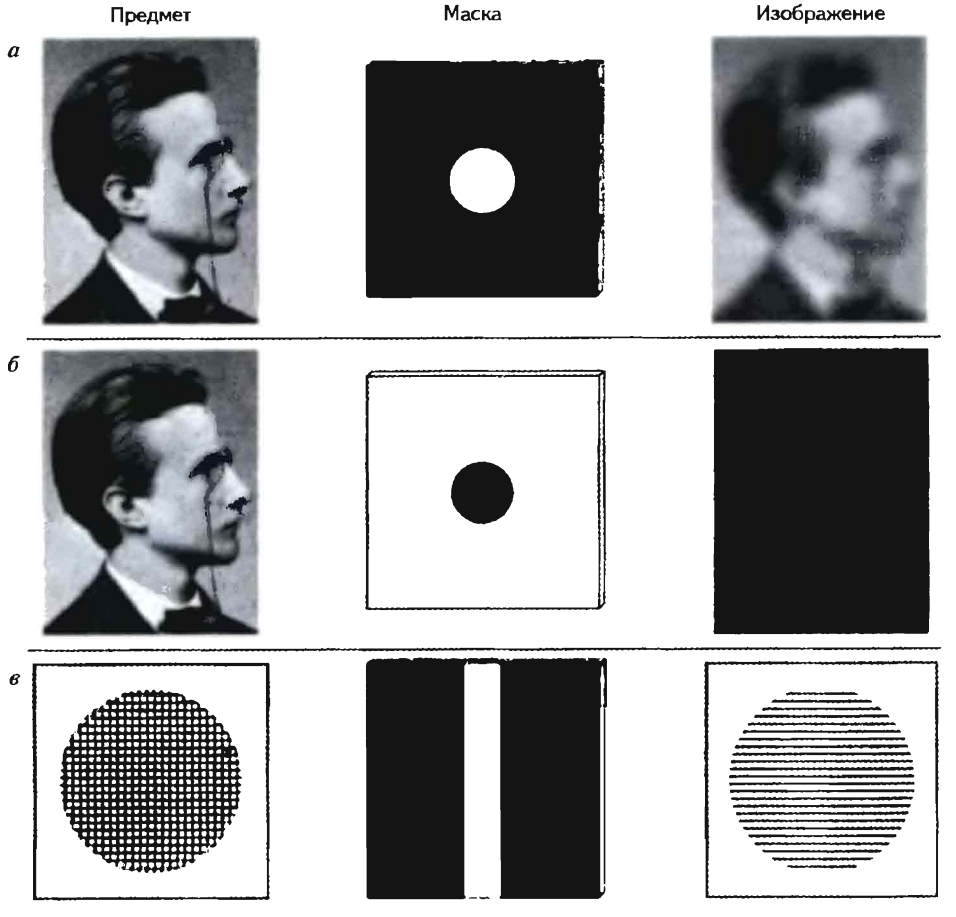


Рис. 4.35. Примеры объектов, масок и фильтрованных изображений для трех пространственных фильтров:

а — фильтр нижних частот; *б* — фильтр высоких частот; *в* — фильтр вертикальных компонент. Черный цвет означает, что коэффициент пропускания равен нулю, а белый — единице

♦ *Фильтр высоких частот* является дополнением фильтра низких частот. Он блокирует низкие частоты и пропускает высокие. Маска представляет собой прозрачный транспарант с непрозрачным кругом в центре. Выходной сигнал фильтра велик в областях, где интенсивность быстро меняется от точки к точке предмета, и мал в областях плавного изменения. Таким образом, фильтр полезен для усиления краев при обработке изображений.

♦ *Фильтр вертикальных компонент* блокирует частоты горизонтальных компонент и пропускает частоты вертикальных. Передаются только изменения в направлении x. Если маска представляет собой вертикальную щель шириной D, то наибольшая пропускаемая частота:

$$\nu_y = \frac{D/2}{\lambda f}.$$

Примеры этих фильтров и их действие на изображения иллюстрирует рис. 4.35.

4.4.3. Волновая оптика однолинзовой изображающей системы

Рассмотрим формирование изображения однолинзовой системой, показанной на рис. 4.36, с точки зрения волновой оптики. Вначале определим функцию отклика на импульсное воздействие, а затем получим передаточную функцию. Эти функции определяются ошибкой фокусировки ε, которая дается формулой (4.50), и функцией зрачка $p(x, y)$ (коэффициентом пропускания апертуры, расположенной в плоскости линзы). Функция зрачка в такой однолинзовой системе играет ту же роль, которую играет функция маски в 4f-системе, описанной в предыдущем разделе.

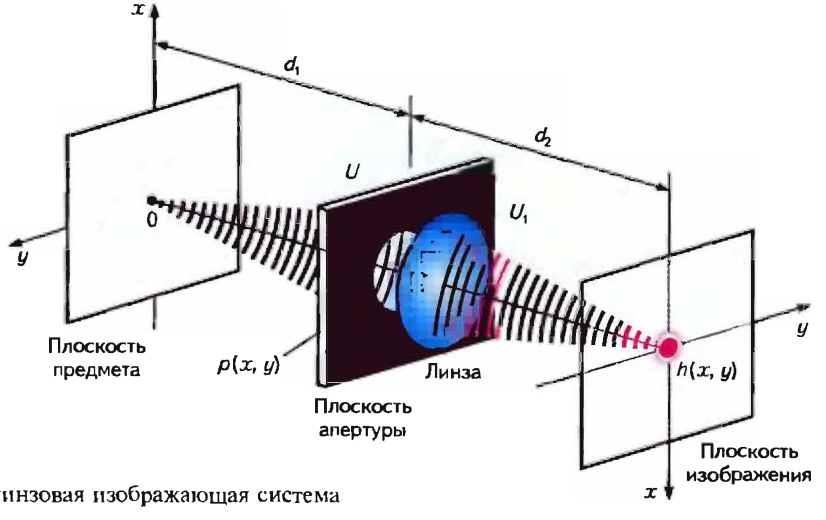


Рис. 4.36. Однолинзовая изображающая система

Функция отклика на импульсное воздействие

Для определения функции отклика на импульсное воздействие рассмотрим предмет, состоящий из одной точки (импульс) на оптической оси с координатами (0, 0), и проследим за распространением испущенной оптической волны от источника до плоскости изображений. Результирующая комплексная амплитуда и есть функция отклика на импульсное воздействие $h(x, y)$.

Точка (импульс) на плоскости предметов создает в плоскости апертуры сферическую волну, приближенно описываемую формулой [см. (4.18)]:

$$U(x, y) \approx h_1 \exp\left[-jk\frac{x^2 + y^2}{2d_1}\right], \tag{4.56}$$

где

$$h_1 = \frac{j}{\lambda d_1}\exp(-jkd_1).$$

При прохождении апертуры и линзы $U(x, y)$ умножается на функцию зрачка $p(x, y)$ и вносимый линзой квадратичный фазовый множитель $\exp[jk(x^2 + y^2)/2f]$, превращаясь в

$$U_1(x, y) = U(x, y)\exp\left(jk\frac{x^2 + y^2}{2f}\right)p(x, y). \tag{4.57}$$

Образовавшееся поле $U_1(x, y)$ затем распространяется в свободном пространстве на расстояние d_2. В соответствии с (4.20) оно дает функцию отклика на импульсное воздействие вида

$$h(x, y) = h_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}U_1(x', y')\exp\left[-j\pi\frac{(x - x')^2 + (y - y')^2}{\lambda d_2}\right]dx'dy', \tag{4.58}$$

где

$$h_2 = \frac{j}{\lambda d_2}\exp(-jkd_2).$$

Подставляя (4.55) и (4.56) в (4.57) и рассматривая интегралы как преобразования Фурье, получаем

$$h(x, y) = h_1h_2\exp\left(-j\pi\frac{x^2 + y^2}{\lambda d_2}\right)P_1\left(\frac{x}{\lambda d_2}, \frac{y}{\lambda d_2}\right), \tag{4.59}$$

где $P(\nu_x, \nu_y)$ — Фурье-образ функции

$$\boxed{p_1(x, y) = p(x, y)\exp\left(-j\pi\varepsilon\frac{x^2 + y^2}{\lambda}\right),} \tag{4.60}$$

Обобщенная функция зрачка

известной как **обобщенная функция зрачка**. Множитель ε — ошибка фокусировки, определяемая соотношением (4.51).

Для изображающих систем высокого качества функция отклика на импульсное воздействие имеет вид узкого пика, т. е. отлична от нуля только в малом интервале значений x и y. Если добавка $\pi(x^2 + y^2)/\lambda d_2$ к фазе в (4.58) много меньше единицы для всех x и y, лежащих в этом интервале, ею можно пренебречь, так что

$$\boxed{h(x,\ y) = h_0 P_1\left(\frac{x}{\lambda d_2},\ \frac{y}{\lambda d_2}\right),} \qquad (4.61)$$

Функция отклика на импульсное воздействие

где $h_0 = h_1 h_2$ — постоянная величины $(1/\lambda d_1)(1/\lambda d_2)$. Следовательно, функция отклика системы на импульсное воздействие пропорциональна Фурье-образу обобщенной функции зрачка $p_1(x, y)$, взятому при

$$\nu_x = \frac{x}{\lambda d_2}; \quad \nu_y = \frac{y}{\lambda d_2}.$$

Если система сфокусирована ($\varepsilon = 0$), то $p_1(x, y) = p(x, y)$ и

$$h(x,\ y) \approx h_0 P\left(\frac{x}{\lambda d_2},\ \frac{y}{\lambda d_2}\right), \qquad (4.62)$$

где $P(\nu_x, \nu_y)$ — Фурье-образ $p(x, y)$. Этот результат аналогичен соответствующему результату (4.54) для $4f$-системы.

Пример 4.5

Функция отклика на импульсное воздействие для сфокусированной изображающей системы с круговой апертурой

Если апертура представляет собой круглое отверстие диаметром D, так что $p(x, y) = 1$ при

$$\rho = \sqrt{x^2 + y^2} \le \frac{D}{2} \qquad (4.63)$$

и 0 в остальных случаях, то функция отклика на импульсное воздействие равна

$$h(x,\ y) = h(0,\ 0)\frac{2J_1(\pi D\rho/\lambda d_2)}{\pi D\rho/\lambda d_2}, \quad \rho = \sqrt{x^2 + y^2};$$

$$|h(0,\ 0)| = \frac{\pi D^2}{4\lambda^2 d_1 d_2}.$$

Это функция с круговой симметрией, сечение которой показано на рис. 4.37. Она спадает к нулю при радиусе

$$\rho_s = 1{,}22\lambda\frac{d_2}{D} \qquad (4.64)$$

и слегка колеблется, прежде чем обратиться в нуль. Следовательно, радиус ρ_s является мерой диаметра круга размытости. Если система сфокусирована на бесконечность, то $d_1 = \infty$, а $d_2 = f$ и

$$\boxed{\rho_s = 1,22\lambda F_\#,} \qquad (4.65)$$

Радиус пятна

где $F_\# = f/D$ — F-число линзы. Таким образом, системы с меньшим $F_\#$ (большей апертурой) дают изображение более высокого качества. При этом, конечно, предполагается, что линза большего диаметра не вносит дополнительных геометрических аберраций.

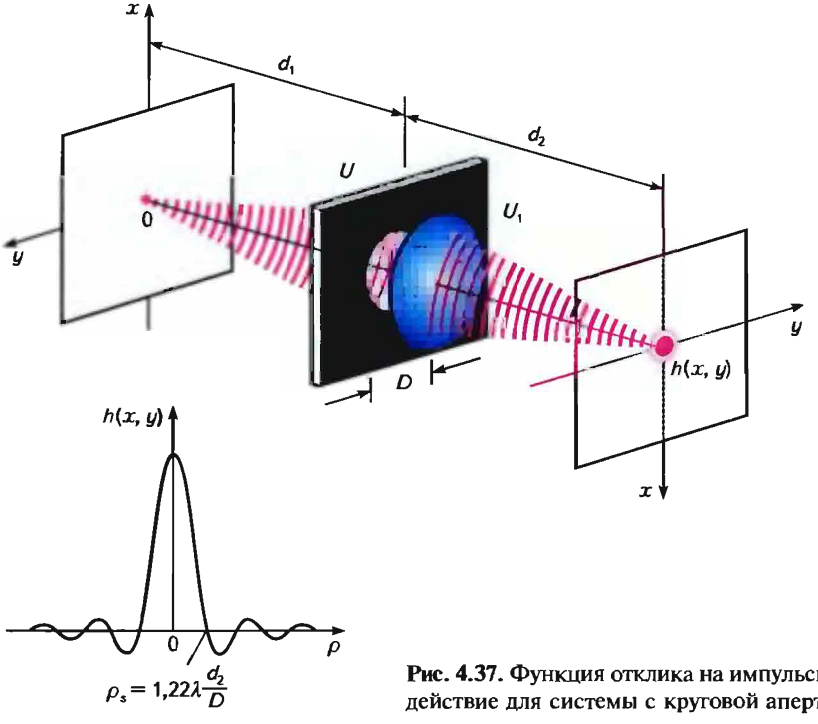


Рис. 4.37. Функция отклика на импульсное воздействие для системы с круговой апертурой

Передаточная функция

Передаточная функция линейной системы может быть определена, только если система инвариантна относительно сдвига (см. приложение Б). Очевидно, однолинзовая изображающая система не обладает инвариантностью относительно сдвига, поскольку сдвиг Δ точки в плоскости предметов сопровождается *другим* сдвигом $M\Delta$ точки в плоскости изображений, где $M = -d_2/d_1$ — увеличение.

Изображение отличается от предмета по двум признакам. Первое, изображение есть увеличенная копия предмета, т. е. точка (x, y) предмета отображает-

ся в новую точку (Mx, My) изображения. Второе, каждая точка размывается в пятно в результате неточной фокусировки или дифракции. Таким образом, формирование изображения можно приписать каскаду двух систем — одна обеспечивает идеальное увеличение, а другая размывает изображение, как показано на рис. 4.38. По своей природе система увеличения неинвариантна относительно сдвига. Для точек, лежащих вблизи оптической оси, размывающая система приближенно инвариантна относительно сдвига и, следовательно, может быть описана передаточной функцией.

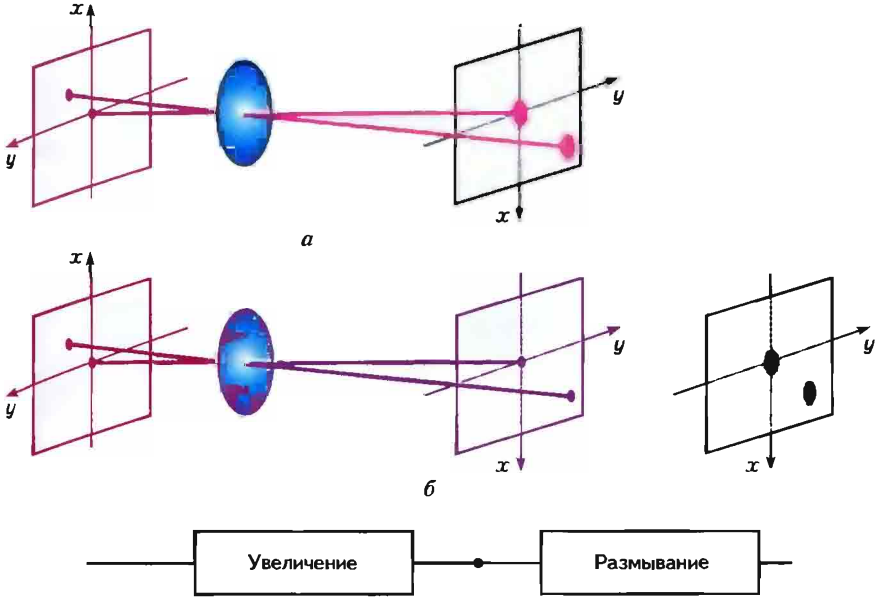


Рис. 4.38. Изображающая система (*а*) рассматривается как комбинация идеальной изображающей системы (*б*), обеспечивающей только увеличение, за которой следует обладающая сдвиговой инвариантностью система, размывающая каждую точку изображения в пятно с распределением, равным функции отклика на импульсное воздействие

Передаточная функция $H(\nu_x, \nu_y)$ размывающей системы определяется путем получения Фурье-образа функции отклика на импульсное воздействие $h(x, y)$ в (4.61). В результате имеем

$$\boxed{H(\nu_x, \nu_y) \approx p_1(\lambda d_2 \nu_x, \lambda d_2 \nu_y),} \qquad (4.66)$$

Передаточная функция

где $p_1(x, y)$ — обощенная функция зрачка, поэтому можно пренебречь постоянным фазовым фактором $\exp(-jkd_1)\exp(-jkd_2)$. Если система сфокусирована, то

$$H(\nu_x, \nu_y) \approx p(\lambda d_2 \nu_x, \lambda d_2 \nu_y), \qquad (4.67)$$

где $p(x, y)$ — функция зрачка. Этот результат идентичен полученному для 4f-системы [см. (4.53)]. Если, например, апертура представляет собой круг диаметром D, то передаточная функция постоянна в пределах круга радиусом ν_s, где

$$\nu_s = \frac{D}{2\lambda d_2}, \tag{4.68}$$

и равна нулю во всех остальных точках, как показано на рис. 4.39.

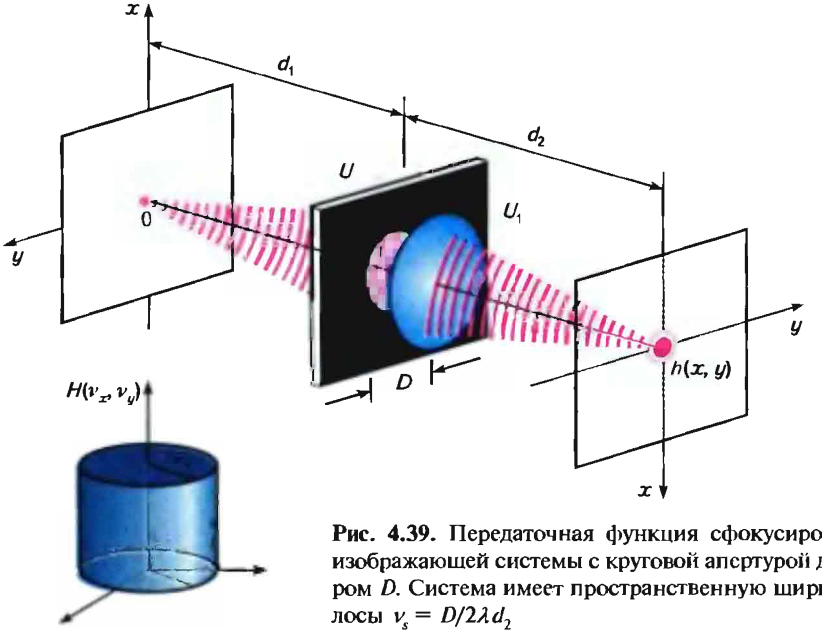


Рис. 4.39. Передаточная функция сфокусированной изображающей системы с круговой апертурой диаметром D. Система имеет пространственную ширину полосы $\nu_s = D/2\lambda d_2$

Если линза сфокусирована на бесконечность, т. е. $d_2 = f$, то

$$\boxed{\nu_s = \frac{1}{2\lambda F_\#}}, \tag{4.69}$$

Пространственная ширина полосы

где $F_\# = f/D$ — F-число линзы. Например, для F-2 линзы с $F_\# = f/D = 2$ при $\lambda = 0{,}5$ мкм $\nu_s = 500$ линий/мм. Частота ν_s представляет собой пространственную ширину полосы, т. е. наибольшую пространственную частоту, передаваемую изображающей системой.

4.4.4. Формирование изображения в ближнем поле

В разд. 4.1 было показано, что пространственная ширина полосы света, распространяющегося в свободном пространстве, при длине волны λ равна λ^{-1} периодов/мм. Фурье-компоненты предмета с пространственными ча-

стотами больше λ^{-1} порождают нераспространяющиеся волны, которые быстро затухают и исчезают на расстояниях порядка длины волны от него, так что детали предмета меньше длины волны невозможно передать. Более того, в подразд. 4.4.3 было показано, что изображающая система, состоящая из линзы с заданным $F_{\#}$, имеет функцию отклика на импульсное воздействие, радиус которой равен $1{,}22F_{\#}$, так что точки, отстоящие друг от друга на расстояние меньше $1{,}22F_{\#}$, невозможно различить (рис. 4.40, *а*). Другой способ получения изображений, использующий для сканирования объекта лазерный пучок, сфокусированный линзой (рис. 4.40, *б*), имеет те же ограничения. Разрешение этой системы диктуется размером пятна фокусировки, которое имеет радиус $1{,}22F_{\#}$, как было показано в примере 4.5. В обоих этих случаях детали предмета, размеры которых меньше длины волны, не передаются в его изображении. Этот фундаментальный предел разрешающей способности изображающих систем часто называют **дифракционным пределом**.

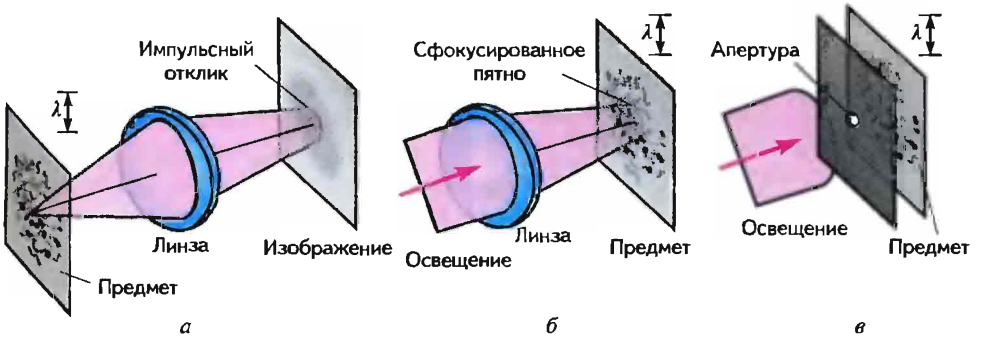


Рис. 4.40. В однолинзовой изображающей системе пространственные детали предмета меньше длины волны размываются при формировании изображений одной линзой (*а*) или системой со сканированием сфокусированного лазерного пучка (*б*). Сканирующая изображающая система (*в*), в которой освещение предмета производится светом, пропущенным через отверстие диаметром меньше длины волны, сохраняет детали предмета размером меньше длины волны при условии, что плоскость предмета удалена от плоскости апертуры на расстояние, также меньшее длины волны

Однако дифракционный предел можно преодолеть. Свет *можно* локализовать в пятно размером много меньше длины волны *в одной плоскости*. Трудность состоит в том, что нераспространяющиеся волны полностью затухают на малых расстояниях от этой плоскости, сразу после чего пятно расплывается и приобретает размеры больше длины волны. На еще больших расстояниях волна в конце концов становится сферической. Следовательно, дифракционный предел можно преодолеть, если предмет расположить в непосредственной близости к осветителю субволновых размеров. Это можно реализовать в сканирующей конфигурации путем пропускания пучка через апертуру с диаметром, много меньшим длины волны, как показано на рис. 4.40, *в*.

Предмет располагается на расстоянии меньше длины волны от отверстия (обычно это менее половины радиуса отверстия), так что освещению подвергается участок, много меньший длины волны. После прохождения сквозь предмет распространяющиеся компоненты волны образуют сферическую волну, амплитуда которой пропорциональна локальному коэффициенту пропускания в месте освещения. Разрешение такой системы порядка размеров отверстия, которое намного меньше длины волны. Изображение строится путем растрового сканирования поверхности отверстием предмета и регистрации оптического отклика с помощью обычной изображающей системы дальнего поля. Эта методика известна как оптическая ближнеполевая визуализация, или **сканирующая оптическая микроскопия ближнего поля (СОМБП)**. Она относится к области нанофотоники, поскольку формирование изображения характеризуется субволновым (нанометровым) пространственным масштабом.

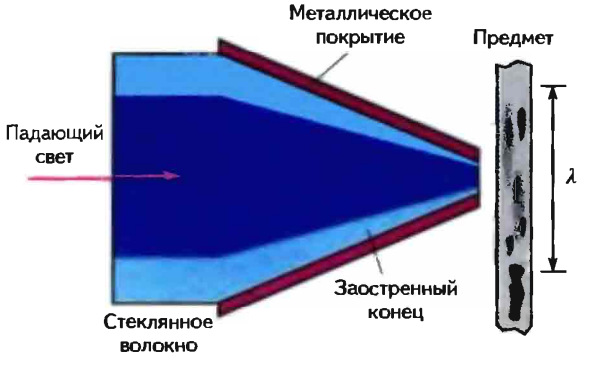


Рис. 4.41. Оптическое волокно с заостренным концом, имеющим металлическое покрытие, для ближнеполевой визуализации

Обычно для реализации СОМБП свет подводится по оптическому волокну, конец которого заострен и покрыт алюминием, как показано на рис. 4.41. Свет распространяется внутри волокна за счет полного внутреннего отражения. Когда диаметр волокна начинает уменьшаться, свет удерживается в нем за счет отражения от металлической поверхности, которая действует как коническое зеркало. Когда вблизи острия диаметр волокна становится еще меньше, волна перестает быть направляемой (см. разд. 8.1) и становится нераспространяющейся. Распределение освещающей волны в области острия сложно и требует численного описания. Диаметр апертуры и пространственное разрешение составляют десятки нанометров и достигаются СОМБП в видимом диапазоне. Поскольку острие световода при сканировании предмета должно находиться на расстоянии всего в несколько нанометров от его поверхности, требуется весьма сложная система обратной связи, чтобы поддерживать это расстояние постоянным для предметов с произвольной топографией. Приложения СОМБП включают неразрушающее определение характеристик неорганических, органических, композитных и биологических материалов и наноструктур.

4.5. ГОЛОГРАФИЯ

Голография включает запись и восстановление оптических волн. **Голограмма** представляет собой транспарант, содержащий зашифрованную запись оптической волны, включая ее амплитудные и фазовые свойства. Рассмотрим монохроматическую оптическую волну, комплексная амплитуда которой в некоторой плоскости, скажем $z = 0$, равна $U_0(x, y)$. Если каким-либо образом можно было бы изготовить тонкий оптический элемент (транспарант) с комплексным амплитудным коэффициентом пропускания $t(x, y)$, равным $U_0(x, y)$, то он обеспечил бы полную запись волны. Затем можно было бы восстановить волну просто путем освещения транспаранта однородной плоской волной, распространяющейся вдоль оси z и имеющей единичную амплитуду. Прошедшая волна в этом случае имела бы в плоскости $z = 0$ комплексную амплитуду $U(x, y) = 1 \quad t(x, y) = U_0(x, y)$. Исходная волна, такими образом, была бы воспроизведена во всех точках плоскости $z = 0$ и, следовательно, восстановлена во всех точках полупространства $z > 0$.

В качестве примера рассмотрим известный случай: однородная плоская волна, распространяющаяся под углом θ к оси z в плоскости x—z, имеет комплексную амплитуду $U_0(x, y) = \exp[-jkx \sin \theta]$. Записью этой волны будет транспарант с комплексным амплитудным коэффициентом пропускания $t(x, y) = \exp[-jkx \sin \theta]$. Такой транспарант действует как призма, отклоняющая падающую волну $\exp[-jkz]$ на угол θ (см. подразд. 2.4.2), воспроизводя, таким образом, исходную волну.

Вопрос состоит в том, как изготовить транспарант $t(x, y)$ по исходной волне $U_0(x, y)$. Ключевая трудность состоит в том, что оптические детекторы, включая фотоэмульсии, применяемые для изготовления транспарантов, реагируют на интенсивность света $|U_0(x, y)|^2$ и, следовательно, нечувствительны к фазе $\arg\{U_0(x, y)\}$. Однако фазовая информация существенна и не может быть упущена. Например, если фаза наклонной волны $U_0(x, y) = \exp[-jkx \sin \theta]$ не была бы записана, то не было бы записано и направление распространения волны. Для записи фазы $U_0(x, y)$ необходимо найти способ ее кодирования путем преобразования фазы в интенсивность. Тогда для восстановления волны записанная информация могла бы быть декодирована оптическим путем.

Голографический код

Голографический код основан на смешении исходной волны (здесь и далее будем называть ее **предметной волной**) U_0 с известной **опорной волной** U_r и записи картины их интерференции в плоскости $z = 0$. Интенсивность суммы двух волн регистрируется фотографическим способом. Получается транспарант с комплексным амплитудным коэффициентом пропускания t, пропорциональ-

ным интенсивности (рис. 4.42, *а*). Коэффициент пропускания определяется формулой

$$t \propto |U_0 + U_r|^2 = |U_0|^2 + |U_r|^2 + U_r^* U_0 + U_r U_0^* = \\ = I_r + I_0 + U_r^* U_0 + U_r U_0^* = I_r + I_0 + 2\sqrt{I_r I_0}\cos\left[\arg\{U_r\} - \arg\{U_0\}\right], \tag{4.70}$$

где I_r и I_0 — интенсивности опорной и предметной волн, соответственно, в плоскости $z = 0$.

Транспарант, называемый голограммой, очевидно, несет закодированную информацию как об амплитуде, как и о фазе волны U_0. Действительно, будучи записью интерференционной картины, транспарант обладает коэффициентом пропускания, весьма чувствительным к разности фаз обеих волн, как было показано в разд. 2.5 (временной аналог голографии — это гетеродинирование, обсуждавшееся в разд. 2.6). Как отмечалось ранее, обычная фотография чувствительна только к интенсивности падающей волны и не записывает фазовую информацию.

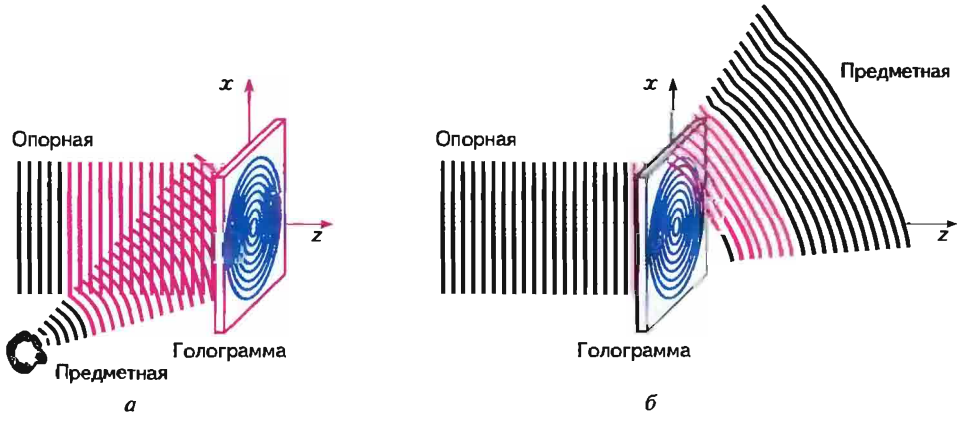


Рис. 4.42. Голограмма — транспарант с записью картины интерференции исходной (предметной) и опорной волн (*а*); исходная волна восстанавливается путем освещения голограммы опорной волной (*б*):

а — запись; *б* — восстановление

Для декодирования содержащейся в голограмме информации и восстановления предметной волны голограмма вновь освещается опорной волной U_r (рис. 4.42, *б*). В результате получается волна с комплексной амплитудой

$$U = tU_r \propto U_r I_r + U_r I_0 + I_r U_0 + U_r^2 U_0^* \tag{4.71}$$

в плоскости голограммы $z = 0$. Третий член в правой части — исходная волна, умноженная на интенсивность I_r опорной волны. Если I_r распределена однородно (не зависит от x и y), то этот член представляет собой искомую восстановленную волну. Однако его нужно выделить на фоне остальных трех членов.

Четвертый член представляет собой комплексно сопряженную исходную волну, умноженную на U_r^2. Первые два члена описывают опорную волну, модулированную суммой интенсивностей обеих волн.

Если в качестве опорной выбрать однородную плоскую волну, распространяющуюся вдоль направления оси z, $\sqrt{I_r}\exp(-jkz)$, то в плоскости $z = 0$ $U_r(x, y) = \sqrt{I_r}$ постоянна и не зависит от x и y. Деля (4.71) на $U_r = \sqrt{I_r}$, получаем

$$U(x, y) \propto I_r + I_0(x, y) + \sqrt{I_r}U_0(x, y) + \sqrt{I_r}U_0^*(x, y). \qquad (4.72)$$

Восстановленная волна в плоскости голограммы

Значение различных членов в (4.71) и методы выделения исходной волны (третьего члена) поясняются на ряде примеров.

Пример 4.6

Голограмма наклонной плоской волны

Если предметная волна плоская и падает под углом θ (рис. 4.43, *а*),

$$U_0(x, y) = \sqrt{I_0}\exp(-jkx\sin\theta),$$

то (4.71) дает

$$U(x, y) \propto I_r + I_0 + \sqrt{I_r I_0}\exp(-jkx\sin\theta) + \sqrt{I_r I_0}\exp(+jkx\sin\theta).$$

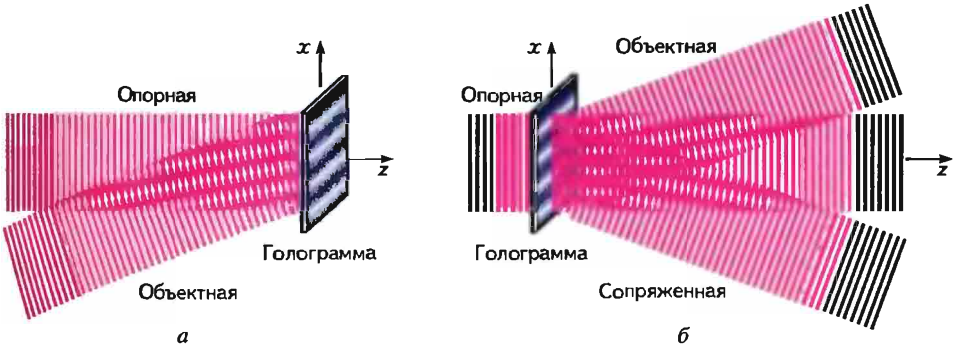


Рис. 4.43. Голограмма наклонной плоской волны — синусоидальная дифракционная решетка:

а — запись; *б* — восстановление

Первые два члена постоянны и соответствуют волне, распространяющейся в направлении оси z (продолжение опорной волны). Третий член соответствует исходной предметной волне, а четвертый представляет собой **сопря-**

женную волну — плоскую волну, распространяющуюся под углом $-\theta$. Таким образом, предметную волну можно выделить среди остальных. Фактически данная голограмма — не что иное, как запись интерференционной картины от двух плоских волн, распространяющихся под углом θ друг к другу. Она служит синусоидальной дифракционной решеткой, которая расщепляет падающую опорную волну на три волны под углами 0, θ и $-\theta$ (см. рис. 4.43, *б* и разд. 2.4.2).

Пример 4.7

Голограмма точечного источника

В этом случае предметная волна сферическая с началом в точке $\boldsymbol{r}_0 = (0, 0, -d)$, как показано на рис. 4.44, так что

$$U_0(x, y) \propto \frac{\exp\left(-jk|\boldsymbol{r} - \boldsymbol{r}_0|\right)}{|\boldsymbol{r} - \boldsymbol{r}_0|},$$

где $\boldsymbol{r} = (x, y, 0)$. Первый член в (4.71) соответствует плоской волне, распространяющейся в направлении оси z, а третий член пропорционален амплитуде исходной сферической волны с началом в точке $(0, 0, -d)$. Четвертый член пропорционален амплитуде сопряженной волны

$$U_0^*(x, y) \propto \frac{\exp\left(jk|\boldsymbol{r} - \boldsymbol{r}_0|\right)}{|\boldsymbol{r} - \boldsymbol{r}_0|},$$

которая представляет собой сферическую волну, сходящуюся в точку $(0, 0, -d)$. Второй член пропорционален $1/|\boldsymbol{r} - \boldsymbol{r}_0|^2$, соответствующая ему волна распространяется вдоль оси z с очень малой угловой расходимостью, поскольку ее интенсивность медленно меняется в поперечной плоскости.

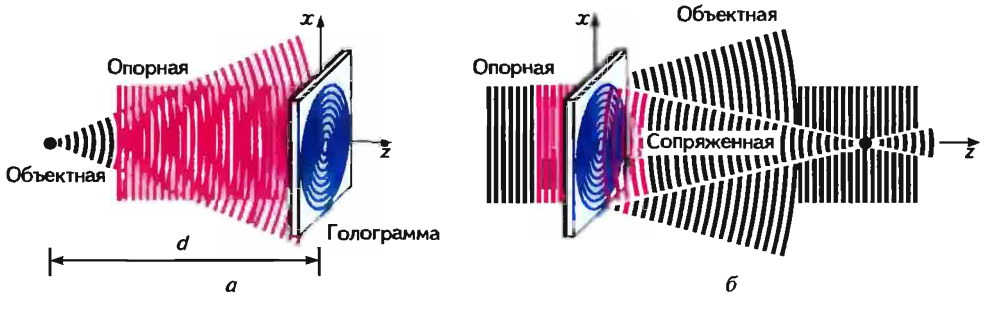


Рис. 4.44. Голограмма сферической волны от точечного источника. Сопряженная волна формирует действительное изображение точки:

а — запись; *б* — восстановление

Внеосевая голография

Один из способов разделения четырех компонент восстановленной волны основан на том, что они меняются на далеко отстоящих друг от друга пространственных частотах и поэтому имеют сильно отличающиеся направления. Этот вид мультиплексирования (разделения) пространственных частот (см. подразд. 4.1.1) обеспечивается при условии, что направления прихода предметной и опорной волн различаются достаточно сильно.

Рассмотрим случай, когда предметная волна имеет комплексную амплитуду $U_0(x, y) = f(x, y) \exp[-jkx \sin\theta]$. Это волна с комплексной огибающей $f(x, y)$ и таким же фазовым множителем, как тот, который вносит призма с углом отклонения θ. Предполагается, что функция $f(x, y)$ меняется медленно, так что ее максимальная пространственная частота ν_s соответствуст углу $\theta_s = \arcsin \lambda\nu_s = \theta$. Тогда направления предметной волны распределены в небольшой окрестности угла θ, как показано на рис. 4.45. Формула (4.71) дает

$$U(x, y) \propto I_r + |f(x, y)|^2 + \sqrt{I_r} f(x, y) \exp(-jkx \sin\theta) + \\ + \sqrt{I_r} f^*(x, y) \exp(+jkx \sin\theta). \tag{4.73}$$

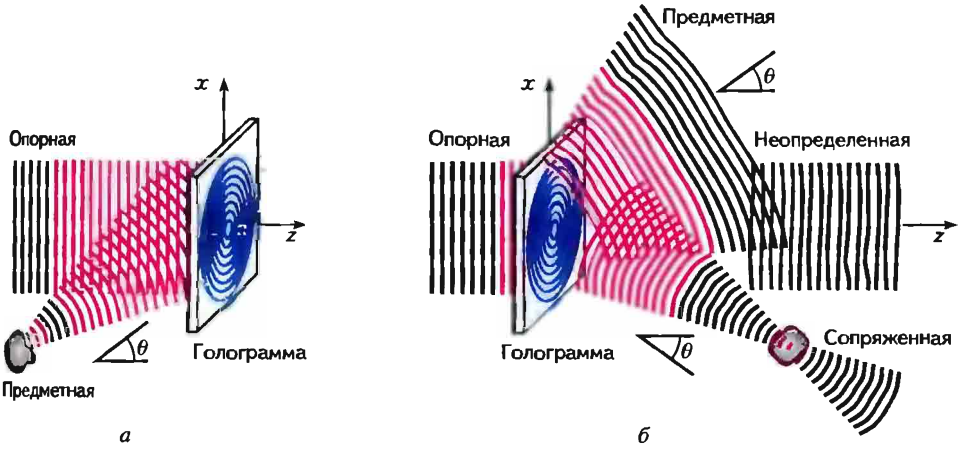


Рис. 4.45. Голограмма внеосевой объектной волны. Объектная волна отделяется от опорной и сопряженной волн:

а — запись; *б* — восстановление

Третий член, очевидно, представляет собой копию предметной волны, падающей под углом θ. Присутствие фазового множителя $\exp[+jkx \sin\theta]$ в четвертом члене отражает тот факт, что соответствующая волна отклоняется в направлении $-\theta$. Первый член соответствует плоской волне, бегущей вдоль оси z. Второй член, известный как **неопределенная волна**, соответствует неоднородной квазиплоской волне, направления которой лежат внутри конуса с малым углом $2\theta_s$ вокруг оси z. Разница направлений предметной и опорной волн

приводит к естественному угловому отделению предметной и сопряженной волн друг от друга и от других двух волн, если $\theta > 3\theta_s$, что позволяет однозначно восстановить исходную волну. Альтернативный способ уменьшить влияние неопределенности — сделать интенсивность опорной волны много большей, чем у предметной. Тогда волна, вносящая неопределенность [второй член в (4.71)], становится намного меньше других, поскольку в соответствующее слагаемое входит только амплитуда предметной волны, в силу чего им можно пренебречь.

Голография Фурье-образа

Фурье-образ $F(\nu_x, \nu_y)$ функции $f(x, y)$ можно получить оптическим путем с помощью линзы (см. разд. 4.2). Если $f(x, y)$ — комплексная амплитуда волны в одной фокальной плоскости линзы, то $F(x/\lambda f, y/\lambda f)$ — комплексная амплитуда в другой фокальной плоскости, где f — фокусное расстояние линзы, а λ — длина волны. Поскольку Фурье-образ — как правило, комплексная функция, его невозможно записать непосредственно.

Фурье-образ $F(x/\lambda f, y/\lambda f)$ можно записать с помощью голографии, рассматривая его в качестве предметной волны

$$U_0(x, y) = F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right),$$

смешивая с опорной волной $U_f(x, y)$ и записывая суперпозицию в виде голограммы (рис. 4.46, *а*). Как обычно, восстановление осуществляется путем освещения голограммы опорной волной. Затем можно произвести обратное преобразование Фурье с помощью той же линзы и восстановить исходную функцию $f(x, y)$ (рис. 4.46, *б*).

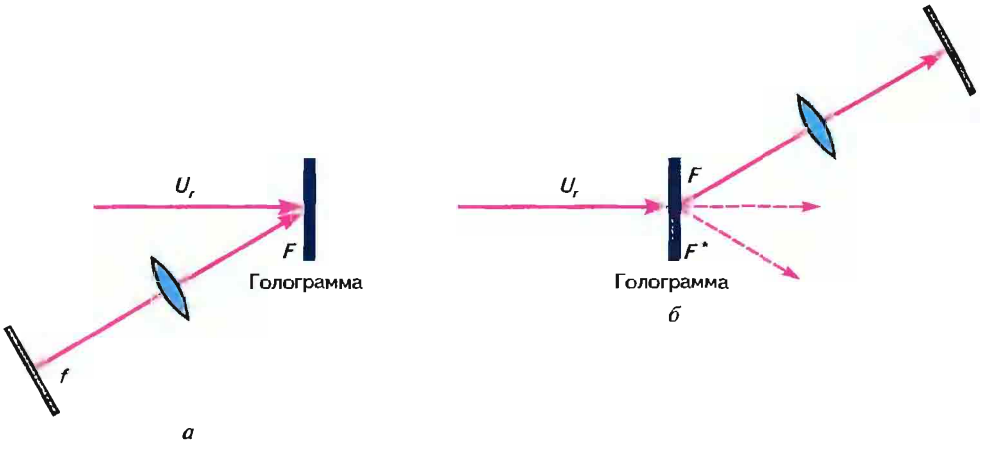


Рис. 4.46. Голограмма волны, комплексная амплитуда которой представляет собой Фурье-образ функции $f(x, y)$ (*а*); восстановление $f(x, y)$ с помощью линзы, осуществляющей преобразование Фурье (*б*):

а — запись; *б* — восстановление

Голографические пространственные фильтры

Пространственный фильтр с передаточной функцией $H(\nu_x, \nu_y)$ можно реализовать с помощью оптической $4f$-системы с маской, имеющей комплексный амплитудный коэффициент пропускания

$$p(x, y) = H\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)$$

и помещенной в Фурье-плоскости (см. подразд. 4.4.2). Поскольку передаточная функция $H(\nu_x, \nu_y)$ обычно принимает комплексные значения, коэффициент пропускания маски $p(x, y)$ должен иметь фазовый множитель. Такую маску трудно изготовить с помощью обычной печатной техники. Однако, если функция отклика фильтра на импульсное воздействие $h(x, y)$ действительная, можно создать голограмму ее Фурье-образа. Для этого производится голографическая запись Фурье-образа

$$U_0(x, y) = H\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right).$$

При использовании Фурье-образа входной функции $f(x, y)$ в качестве опорной волны

$$U_r(x, y) = F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right),$$

голограмма создает волну

$$U_r(x, y)U_0(x, y) = F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)H\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right). \qquad (4.74)$$

Обратное преобразование Фурье восстановленной предметной волны с помощью линзы с фокусным расстоянием f (рис 4.47, *б*) дает комплексную амплитуду $g(x, y)$, Фурье-образ которой есть $G(\nu_x, \nu_y) = H(\nu_x, \nu_y)F(\nu_x, \nu_y)$. Таким образом, $g(x, y)$ есть свертка $f(x, y)$ с $h(x, y)$. Система в целом, называемая **фильтром Ван дер Люгта**, осуществляет операцию свертки, лежащую в основе пространственной фильтрации.

Если вместо $U_0(x, y)$ взять для обратного преобразования Фурье сопряженную волну

$$U_r(x, y)U_0^*(x, y) = F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)H^*\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right),$$

то вместо свертки функций $f(x, y)$ и $h(x, y)$ получим их корреляцию. Операция корреляции полезна в приложениях, связанных с обработкой изображений, включая распознавание образов.

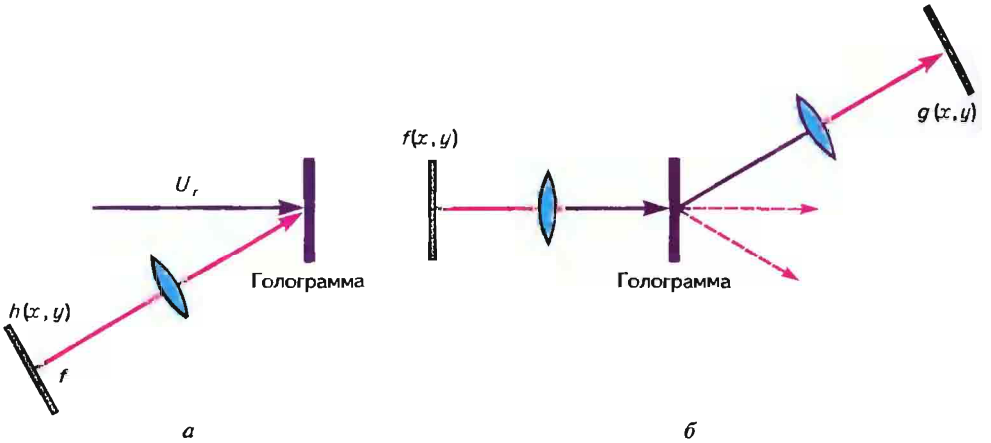


Рис. 4.47. Голографический фильтр Ван дер Люгта:

а — записывается голограмма Фурье-образа функции $h(x, y)$; *б* — Фурье-образ функции $f(x, y)$ пропускается через голограмму и подвергается обратному преобразованию Фурье с помощью линзы. В результате получается функция $g(x, y)$, пропорциональная свертке функций $f(x, y)$ и $h(x, y)$. В итоге получается пространственный фильтр с функцией отклика на импульсное воздействие $h(x, y)$

Голографическая аппаратура

Важным условием успешного изготовления голограммы является наличие монохроматического источника света с минимальными фазовыми флуктуациями. Присутствие фазовых флуктуаций приводит к случайному сдвигу интерференционной картины и размыванию голограммы. По этой причине когерентный источник света (обычно лазер) является необходимой частью оборудования. Требования к когерентности для интерференции световых волн обсуждаются в гл. 10.

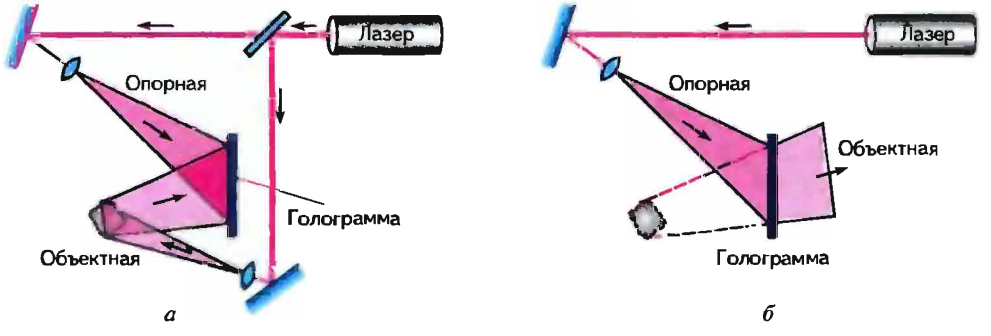


Рис. 4.48. Запись и восстановление голограммы:

а — запись; *б* — восстановление

На рис. 4.48 показана схема типичной экспериментальной установки для записи голограмм и восстановления оптических волн, рассеянных поверхнос-

тью физического объекта. С помощью светоделителя лазерный пучок расщепляется на две части, одна из которых используется как опорная волна, а другая рассеивается объектом и образует предметную волну. Оптическая разность хода между этими волнами должна быть как можно меньше, чтобы гарантировать сохранение неслучайной разности фаз [член $\arg\{U_r\} - \arg\{U_0\}$ в (4.69)].

Поскольку интерференционная картина, образующая голограмму, состоит из тонких линий, расстояние между которыми порядка $\lambda/\sin\theta$, где θ — угол между направлениями опорной и предметной волн, фотографическая пленка должна иметь высокое разрешение, а система не должна испытывать вибраций во время экспозиции. Чем больше угол θ, тем меньше расстояние между линиями голограммы и тем более строгими являются указанные требования. Предметная волна восстанавливается при освещении голограммы опорной волной, так что наблюдатель видит предмет, как если бы он в самом деле находился в нужном месте, при этом сохраняется его трехмерный характер.

Объемная голография

До сих пор предполагалось, что голограмма представляет собой тонкий плоский транспарант, на котором записана картина интерференции предметной и опорной волн. Рассмотрим теперь запись голограммы в сравнительно толстом слое среды и покажем возникающие при этом преимущества. Рассмотрим простой случай, когда предметная и опорная волны плоские с волновыми векторами $\boldsymbol{k}_r$ и $\boldsymbol{k}_0$. Записывающая среда находится между плоскостями $z = 0$ и $z = \Delta$, как показано на рис. 4.49. Теперь интерференционная картина описывается функцией трех координат x, y и z:

$$\begin{aligned} I(x, y, z) &= \left|\sqrt{I_r}\exp(-j\boldsymbol{k}_r\boldsymbol{r}) + \sqrt{I_0}\exp(-j\boldsymbol{k}_0\boldsymbol{r})\right|^2 = \\ &= I_r + I_0 + 2\sqrt{I_r I_0}\cos(\boldsymbol{k}_0\boldsymbol{r} - \boldsymbol{k}_r\boldsymbol{r}) = I_r + I_0 + 2\sqrt{I_r I_0}\cos(\boldsymbol{k}_g\boldsymbol{r}), \end{aligned} \tag{4.75}$$

где $\boldsymbol{k}_g = \boldsymbol{k}_0 - \boldsymbol{k}_r$. Это синусоидальное распределение с периодом

$$\Lambda = \frac{2\pi}{|\boldsymbol{k}_g|}$$

и поверхностями одинаковой интенсивности, перпендикулярными вектору $\boldsymbol{k}_g$.

Например, если опорная волна распространяется вдоль оси z, а предметная волна образует с ней угол θ, то

$$|\boldsymbol{k}_g| = 2k\sin\left(\frac{\theta}{2}\right)$$

и период равен

$$\Lambda = \frac{\lambda}{2\sin(\theta/2)}, \tag{4.76}$$

как видно из рис. 4.49.

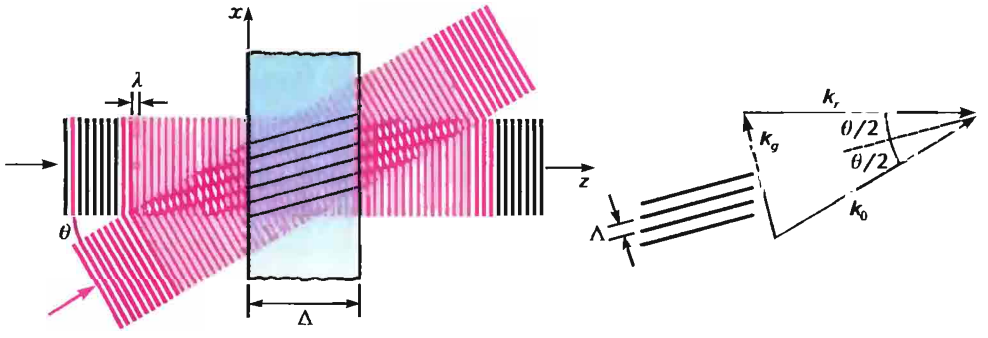


Рис. 4.49. Интерференционная картина в случае, когда предметная и опорная волны плоские. Поскольку $|\boldsymbol{k}_r| = |\boldsymbol{k}_0| = 2\pi/\lambda$ и $|\boldsymbol{k}_g| = 2\pi/\Lambda$, из векторной диаграммы следует, что $2\pi/\Lambda = 2(2\pi/\lambda)\sin(\theta/2)$, так что $\Lambda = \lambda/2\sin(\theta/2)$

Записанная в фотоэмульсии, эта картина представляет собой толстую дифракционную решетку — **объемную голограмму**. Вектор $\boldsymbol{k}_g$ называется **вектором решетки**. При освещении опорной волной, как показано на рис. 4.50, параллельные плоскости решетки отражают волну только при выполнении условия Брэгга

$$\sin\phi = \frac{\lambda}{2\Lambda},$$

где ϕ — угол между плоскостями решетки и падающей опорной волной (см. упражнение 2.11). В нашем случае $\phi = \theta/2$, так что

$$\sin\left(\frac{\theta}{2}\right) = \frac{\lambda}{2\Lambda}.$$

Ввиду (4.75) условие Брэгга действительно выполняется, так что опорная волна действительно отражается. Как очевидно из геометрии, отраженная волна является продолжением предметной волны, так что восстановление успешно осуществлено.

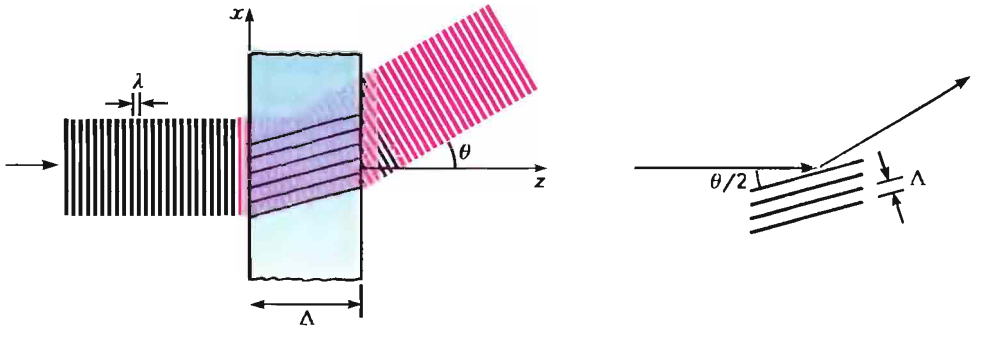


Рис. 4.50. Опорная волна отражается по условию Брэгга от объемной голограммы, предметная волна восстанавливается

Теперь предположим, что голограмма освещается опорной волной с другой длиной волны λ'. Очевидно, условие Брэгга

$$\sin\left(\frac{\theta}{2}\right) = \frac{\lambda'}{2\Lambda}$$

не будет выполняться и волна не отразится. Это значит, что предметная волна восстанавливается, только если длина волны источника, используемого при восстановлении, точно совпадает с источником, использованным при записи. Если для восстановления используется свет с широким спектром (белый), то лишь свет с «правильной» длиной волны будет отражаться и обеспечивать успешное восстановление.

Итак, хотя для процесса записи нужен монохроматический свет, восстановление можно осуществить с белым светом. Это является очевидным преимуществом во многих приложениях голографии. Другие схемы записи и восстановления объемных голограмм показаны на рис. 4.51.

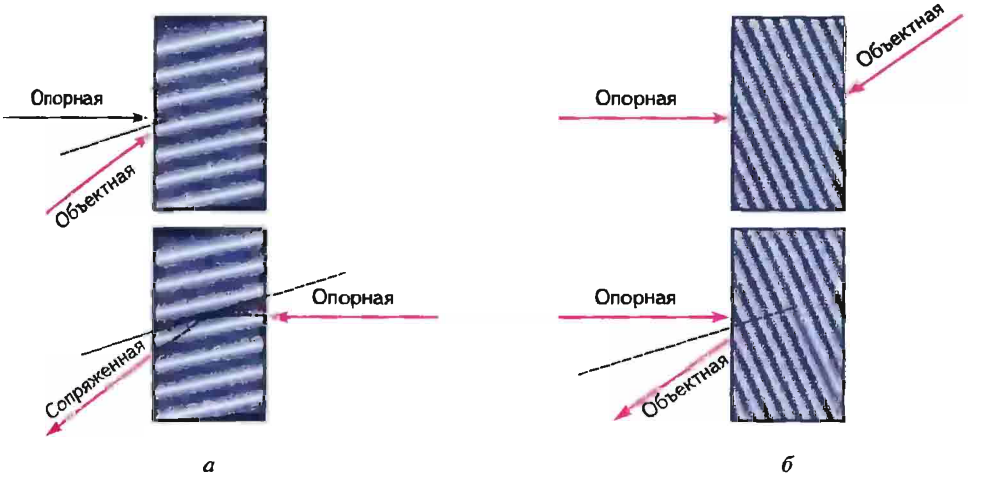


Рис. 4.51. Схемы записи и восстановления объемных голограмм:

а — эта голограмма записывается с опорной и предметной волнами, падающими с одной стороны, и восстанавливается с использованием обращенной опорной волны; восстановленная волна представляет собой сопряженную волну, распространяющуюся в направлении, противоположном по отношению к исходной предметной волне; *б* — отражающая голограмма записывается с опорной и предметной волнами, приходящими с разных сторон; предметная волна восстанавливается за счет отражения от решетки

Еще одним типом голограммы, которую можно рассматривать в белом свете, является **радужная голограмма**. Эта голограмма записывается через узкую щель, так что восстановленное изображение, естественно, также выглядит как рассматриваемое через щель. Однако, если длина волны при восстановлении отличается от использованной при записи, то восстановленная волна выглядит как приходящая от смещенной щели из-за вносимого увеличения. При восста-

новлении в белом свете восстановленная волна выглядит так, как будто предмет виден через множество смещенных друг относительно друга щелей, каждая с различной длиной волны (цветом). В результате получаем радугу изображений, видимых через параллельные щели. Каждая щель отображает предмет с эффектом параллакса в направлении щели, но не в ортогональном направлении. Помимо научных, радужные голограммы имеют многочисленные практические применения в рекламе, полиграфии, защите документов от подделки и т. п.

Рекомендуемая литература

ФУРЬЕ-ОПТИКА И ОПТИЧЕСКАЯ ОБРАБОТКА СИГНАЛОВ

Goodman J.W. Introduction to Fourier Optics. Roberts, 3rd ed. 2005.

Steward E.G. Fourier Optics: An Introduction. Halsted Press, 2nd ed. 1987; Dover, reissued 2004.

Lauterbom W., Kurz T. Coherent Optics: Fundamentals and Applications. Springer-Verlag, 2nd ed. 2003.

O'Neill E.L. Introduction to Statistical Optics. Addison-Wesley, 1963; Dover, reissued 2003.

Abushagur M.A., Caulfield H., eds. Selected Papers on Fourier Optics. SPIE Optical Engineering Press (Milestone Series. Vol. 105), 1995.

Hooijmans P.W. Coherent Optical System Design. Wiley, 1994.

Yu F.T., Yin S., eds. Selected Papers on Coherent Optical Processing. SPIE Optical Engineering Press (Milestone Series. Vol. 52), 1992.

Reynolds G., DeVelis J.B., Parrent G.B., Thompson B.J. The New Physical Optics Notebook: Tutorials in Fourier Optics. SPIE Optical Engineering Press, 1989.

Homer J.L., ed. Optical Signal Processing. Academic Press, 1987.

Papoulis A. Systems and Transforms with Applications in Optics. McGraw-Hill, 1968; Krieger, reissued, 1986.

Yu F.T.S. White-Light Optical Signal Processing. Wiley, 1985.

Duffieux P.M. Fourier Transform and Its Applications to Optics. Wiley, 2nd ed. 1983.

Stark H., ed. Applications of Optical Fourier Transforms. Academic Press, 1982.

Gaskill J.D. Linear Systems, Fourier Transforms and Optics. Wiley, 1978.

Carlson F.P. Introduction to Applied Optics for Engineers. Academic Press, 1977.

Harbum G., Taylor C.A., Welberry T.R. Atlas of Optical Transforms. Cornell University Press, 1975.

Cathey W.T. Optical Information Processing and Holography. Wiley, 1974.

Lipson H.S., ed., Optical Transforms. Academic Press, 1972.

Cagnet M., Françon M., Mallick S. Atlas of Optical Phenomena. Springer-Verlag, reprinted with supplement 1971.

Mertz L. Transformations in Optics. Wiley, 1965.

Cagnet M., Françon M., Thrierr J.C. Atlas of Optical Phenomena. Springer-Verlag, 1962.

ДИФРАКЦИЯ

Ersoy O.K. Diffraction, Fourier Optics, and Imaging. Wiley, 2007.

Nieto-Vesperinas M. Scattering and Diffraction in Physical Optics. World Scientific, 2nd ed. 2006.

Sommerfeld A. Mathematical Theory of Diffraction. Mathematische Annalen, 1896; Birkhauser, 2004.

O'Shea D.C., Suleski T.J., Kathman A.D. and *Prather D.W.* Diffractive Optics: Design, Fabrication and Test. SPIE Optical Engineering Press, 2003.

Cowley J.M. Diffraction Physics. Elsevier, 3rd revised ed. 1995.

Nussenzveig H.M. Diffraction Effects in Semiclassical Scattering. Cambridge University Press, 1992.

Oughstun K.E., ed. Selected Papers on Scalar Wave Diffraction. SPIE Optical Engineering Press (Milestone Series Volume 51), 1992.

Solimeno S., Crosignani B., Di Porto P. Guiding, Diffraction, and Confinement of Optical Radiation. Academic Press, 1986.

Françon M. Diffraction: Coherence in Optics. Pergamon, 1966.

ФОРМИРОВАНИЕ ИЗОБРАЖЕНИЙ

Novotny L., Hecht B. Principles of Nano-Optics. Cambridge University Press, 2006.

Barrett H., Myers K. Foundations of Image Science. Wiley, 2003.

Courjon D. Near-Field Microscopy and Near-Field Optics. Imperial College Press, 2003.

Williams C.S., Becklund O.A. Introduction to the Optical Transfer Function. Wiley, 1989; SPIE Optical Engineering Press, 2002.

Kawata S., ed. Near-Field Optics and Surface Plasmon Polaritons. Springer, 2001.

de Fomel F. Evanescent Waves: From Newtonian Optics to Atomic Optics. Springer-Verlag, 2001.

Gu M. Advanced Optical Imaging Theory. Springer-Verlag, 1999.

Herzig H.P., ed. Micro-Optics: Elements, Systems and Applications. Taylor & Francis, 1997.

Kufner M., Kufner S. Micro-Optics and Lithography. VUB Press, 1997.

Fillard J. Near Field Optics and Nanoscopy. World Scientific, 1996.

Françon M. Optical Image Formation and Processing. Academic Press, 1979.

Dainty J.C., Shaw R. Image Science: Principles, Analysis and Evaluation Of Photographic-Type Imaging Processes. Academic Press, 1974.

Barnes K.R. The Optical Transfer Function. Elsevier, 1971.

ГОЛОГРАФИЯ

Saxby G. Practical Holography. Institute of Physics, 3rd ed. 2004.

Schnars U., Jueptner W. Digital Holography: Digital Hologram Recording, Numerical Reconstruction and Related Techniques. Springer-Verlag, 2004.

Yaroslavsky L. Digital Holography and Digital Image Processing: Principles, Methods, Algorithms. Kluwer, 2004.

Hariharan P. Basics of Holography. Cambridge University Press, 2002.

Bjelkhagen H.I., Caulfield H.J., eds. Selected Papers on Fundamental Techniques in Holography. SPIE Optical Engineering Press (Milestone Series. Vol. 171), 2001.

Kasper J.E., Feller S.A. Complete Book of Holograms: How They Work and How to Make Them. Wiley, 1987; Dover, reissued 2001.

Sirohi R.S., Hinsch K.D., eds. Selected Papers on Holographic Interferometry Principles and Techniques. SPIE Optical Engineering Press (Milestone Series. Vol. 144), 1998.

Hariharan P. Optical Holography: Principles, Techniques and Applications. Cambridge University Press, 2nd ed. 1996.

Soifer V.A., Golub M.V. Laser Beam Mode Selection by Computer Generated Holograms. CRC Press, 1994.

Smith H.M. Principles of Holography. Wiley, 2nd ed. 1975, reprinted 1988.

Schumann W., Zurcher J.-P., Cuche D. Holography and Deformation Analysis. Springer-Verlag, 1985.

Abramson N. The Making and Evaluation of Holograms. Academic Press, 1981.

Ostrovsky Yu.I., Butusov M.M. and *Ostrovskaya G.V.* Interferometry by Holography. Springer-Verlag, 1980.

Soroko L.M. Holography and Coherent Optics. Plenum, 1980.

Caulfield H.J., ed. Handbook of Optical Holography. Academic Press, 1979.

Schumann W., Dubas M. Holographic Interferometry. Springer-Verlag, 1979.

Vest C.M. Holographic Interferometry. Springer-Verlag, 1979.

Collier R.J., Burckhardt C.B., Lin L.H. Optical Holography. Academic Press, paperback ed., 1977.

Françon M. Holography. Academic Press, 1974.

Caulfield H.J., Sun L. The Applications of Holography. Wiley, 1970.

Задачи

К РАЗДЕЛУ 4.1

1. **Соответствие между гармоническими функциями и плоскими волнами.** Комплексные амплитуды монохроматических волн с длиной волны λ в плоскостях $z = 0$ и $z = d$ равны $f(x, y)$ и $g(x, y)$ соответственно. Предполагая, что $d = 10^4\lambda$, с помощью гармонического анализа определите $g(x, y)$ в следующих случаях:

а) $f(x, y) = 1$;

б) $f(x, y) = \exp\left[-\frac{j\pi}{\lambda}(x + y)\right]$;

в) $f(x, y) = \cos\left(\frac{\pi x}{2\lambda}\right)$;

г) $f(x, y) = \cos^2\left(\frac{\pi y}{2\lambda}\right)$;

д) $f(x, y) = \sum_m \operatorname{rect}\left(\frac{x}{10\lambda} - 2m\right)$, $m = 0, \pm 1, \pm 2, \ldots$, где $\operatorname{rect}(x) = 1$, если $|x| \le 1/2$, и 0 в остальных случаях.

Опишите физическую природу волны в каждом случае.

2. В задаче 1 при условии, что функция $f(x, y)$ обладает круговой симметрией и максимальной пространственной частотой 200 линий/мм, определите угол при вершине конуса, внутри которого находятся направления всех волн. Предположите, что $\lambda = 633$ нм.

3. **Логарифмическая карта связи.** Транспарант с амплитудным коэффициентом пропускания $t(x, y) = \exp[-j2\pi\phi(x)]$ освещается однородной плоской волной с длиной волны $\lambda = 1$ мкм. Прошедший свет фокусируется плотно прижатой линзой с фокусным расстоянием $f = 100$ см. Какой должна быть функция $\phi(x)$, чтобы луч, падающий на транспарант в точке x, отклонялся и фокусировался в точку $x' = \ln(x)$ для всех $x > 0$? (Замечание: x и x' измеряются в миллиметрах.) Если убрать линзу, как нужно изменить $\phi(x)$, чтобы система выполняла ту же функцию? Такая система может использоваться для реализации логарифмического преобразования координат, что обсуждается в гл. 21.

К РАЗДЕЛУ 4.2

1. **Доказательство Фурье-преобразования линзой.**

а. Покажите, что свертку $f(x)$ и $\exp(-j\pi x^2/\lambda d)$ можно получить в три этапа: умножить $f(x)$ на $\exp(-j\pi x^2/\lambda d)$, вычислить Фурье-образ произведения на частоте $\nu_x = x/\lambda d$ и умножить результат на $\exp(-j\pi x^2/\lambda d)$.

б. Система, реализующая преобразование Фурье, показанная на рис. 4.20, представляет собой каскад из трех элементов — распространение в свободном пространстве на расстояние f, прохождение через линзу с фокусным расстоянием f, распространение в свободном пространстве на расстояние f. Замечая, что прохождение участка свободного пространства длиной d эквивалентно свертке с $\exp(-j\pi x^2/\lambda d)$ [см. (4.20)], и используя результат пункта (а), выведите уравнение Фурье-преобразования линзой (4.29). Для простоты пренебрегите зависимостью от y.

2. **Фурье-преобразование функций, описывающих линии.** Транспарант с амплитудным коэффициентом пропускания $t(x, y)$ освещается плоской волной с длиной волны $\lambda = 1$ мкм, которая затем фокусируется линзой с фокусным расстоянием $f = 100$ см. Нарисуйте распределение интенсивности в плоскости транспаранта и в фокальной плоскости линзы в следующих случаях (все расстояния измерены в мм):

а) $t(x, y) = \delta(x - y)$;

б) $t(x, y) = \delta(x + a) + \delta(x - a)$, $a = 1$ мм;

в) $t(x, y) = \delta(x + a) + j\delta(x - a)$, $a = 1$ мм,

где $\delta(\ )$ — дельта-функция (см. приложение А, разд. А.1).

3. **Проектирование оптической системы для преобразования Фурье.** Линза используется для изображения Фурье-образа функции двух переменных с пространственными частотами от 20 до 200 линий/мм. Если длина волны света $\lambda = 488$ нм, каким должно быть фокусное расстояние линзы, чтобы в Фурье-плоскости наибольшая и наименьшая пространственные частоты были разделены расстоянием 9 см?

К РАЗДЕЛУ 4.3

1. **Дифракция Фраунгофера на дифракционной решетке.** Получите формулу для картины дифракции Фраунгофера от апертуры, состоящей из $M = 2L + 1$ параллельных щелей бесконечно малой ширины, если расстояние между соседними щелями $a = 10\lambda$:

$$p(x, y) = \sum_{m=-L}^{L} \delta(x - ma). \tag{4.77}$$

Нарисуйте картину распределения интенсивности в зависимости от угла наблюдения $\theta = x/d$, где d — дистанция наблюдения.

2. **Дифракция Фраунгофера при наклонном падении исходной волны.** Дифракционная картина от апертуры с апертурной функцией $p(x, y)$ пропорциональна $|P(x/\lambda d, y/\lambda d)|^2$, где $P(\nu_x, \nu_y)$ — Фурье-образ функции $p(x, y)$, а d — расстояние

между плоскостью апертуры и плоскостью наблюдения. Какой будет дифракционная картина, если направление распространения падающей волны образует малый угол $\theta_x \ll 1$ с осью z в плоскости x—z?

*3. **Дифракция Френеля от двух точечных отверстий.** Покажите, что картина дифракции Френеля от двух точечных отверстий, расстояние между которыми $2a$, что соответствует апертурной функции $p(x, y) = |\delta(x + a) + \delta(x - a)|\delta(y)$, на расстоянии d представляет собой периодическую структуру

$$I(x, y) = \left(\frac{2}{\lambda d}\right)^2 \cos^2\left(\frac{2\pi a x}{\lambda d}\right).$$

*4. **Связь между дифракцией Френеля и дифракцией Фраунгофера.** Покажите, что картина дифракции Френеля для апертурной функции $p(x, y)$ совпадает с картиной дифракции Фраунгофера для апертурной функции

$$p(x, y)\exp\left[-\frac{j\pi(x^2 + y^2)}{\lambda d}\right].$$

К РАЗДЕЛУ 4.4

1. **Размывание синусоидальной решетки.** Объект

$$f(x, y) = \cos^2\left(\frac{2\pi x}{a}\right)$$

отображается расфокусированной однолинзовой изображающей системой с функцией отклик на импульсное воздействие $h(x, y) = 1$ внутри квадрата со стороной D и 0 во всех других точках. Выведите выражение для распределения изображения $g(x, 0)$ в направлении x. Получите выражение для контраста изображения, выраженного через D/a. Контраст определяется как отношение (max − min)/(max + min), где max и min — наибольшее и наименьшее значения $g(x, 0)$.

2. **Изображение фазового объекта.** Изображающая система имеет функцию отклика на импульсное воздействие $\mathrm{rect}(x)\delta(y)$. Если волна на входе есть

$$f(x, y) = \begin{cases} \exp\left(j\frac{\pi}{2}\right) & \text{при } x > 0; \\ \exp\left(-j\frac{\pi}{2}\right) & \text{при } x \le 0, \end{cases} \tag{4.78}$$

определите и изобразите графически интенсивность $|g(x, y)|^2$ выходной волны $g(x, y)$. Проверьте, что даже если интенсивность входной волны $|f(x, y)|^2 = 1$, интенсивность выходной волны неоднородна.

3. **Оптическая пространственная фильтрация.** Рассмотрим систему пространственной фильтрации, показанную на рис. 4.34, с $f = 1000$ мм. Система осве-

щается однородной плоской волной с единичной амплитудой и длиной волны $\lambda = 10^{-3}$ мм. Входной транспарант имеет амплитудный коэффициент пропускания $f(x, y)$, а маска имеет амплитудный коэффициент пропускания $p(x, y)$. Напишите выражение, связывающее комплексную амплитуду света $g(x, y)$ в плоскости изображений с $f(x, y)$ и $p(x, y)$. Предполагая, что все расстояния измеряются в миллиметрах, нарисуйте график $g(x, 0)$ в следующих случаях:

а) $f(x, y) = \delta(x - 5)$, $p(x, y) = \text{rect}\,(x)$;

б) $f(x, y) = \text{rect}\,(x)$, $p(x, y) = \text{sinc}\,(x)$.

Найдите такую функцию $p(x, y)$, что

$$g(x, y) = \nabla_T^2 f(x, y),$$

где ∇_T^2 — поперечный оператор Лапласа,

$$\nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

4. **Оптическая кросскорреляция**. Покажите, что пространственный фильтр можно использовать для выполнения операции кросскорреляции (ее определение см. в приложении А) между двумя изображениями, описываемыми действительными функциями $f_1(x, y)$ и $f_2(x, y)$. При каких условиях комплексные амплитудные коэффициенты пропускания используемых масок и транспарантов были бы действительными?

*5. **Функция отклика на импульсное воздействие сильно расфокусированной системы.** С использованием волновой оптики покажите, что функция отклика на импульсное воздействие у сильно расфокусированной системы (т. е. такой, у которой ошибка фокусировки ε очень велика) приближенно равна

$$h(x, y) = p\left(\frac{x}{\varepsilon d_2}, \frac{y}{\varepsilon d_2}\right),$$

где $p(x, y)$ — функция зрачка.

Указание. Используйте метод стационарной фазы, описанный на с. 172 (второе доказательство) для оценки интеграла, получающегося при использовании (4.60) и (4.59). Обратите внимание, что этот результат совпадает с предсказанием лучевой теории света [см. (4.51)].

6. **Разрешение двух точек**.

а. Рассмотрим однолинзовую изображающую систему, обсуждавшуюся в подразд. 4.4.3. В случае квадратной апертуры со стороной D, единичного увеличения и точной фокусировки запишите выражение для функции отклика на импульсное воздействие $h(x, y)$.

б. Определите отклик системы на объект, состоящий из двух точек на расстоянии b друг от друга, т. е.

$$f(x, y) = \delta(x)\delta(y) + \delta(x - b)\delta(y). \tag{4.79}$$

в. Считая, что $\lambda d_2/D = 0{,}1$ мм, нарисуйте график модуля изображения $g(x, 0)$ как функции x для случаев, когда расстояние между точками $b = 0{,}5$; 1 и 2 мм. Каково минимальное расстояние между двумя точками, при котором изображение остается состоящим из двух пятен, а не одного пятна, т. е. имеет два пика?

7. **Кольцевая апертура.**

а. Сфокусированная однолинзовая изображающая система с увеличением $M = 1$ и фокусным расстоянием $f = 100$ см имеет апертуру в виде кольца

$$p(x, y) = \begin{cases} 1, & a \le \sqrt{x^2 + y^2} \le b; \\ 0, & \text{в остальных случаях,} \end{cases} \tag{4.80}$$

где $a = 5$ мм и $b = 6$ мм. Определите передаточную функцию $H(\nu_x, \nu_y)$ системы и изобразите ее поперечное сечение $H(\nu_x, 0)$. Длина волны $\lambda = 1$ мкм.

б. Теперь сдвинем плоскость изображений ближе к линзе так, чтобы расстояние до нее составило $d_2 = 25$ см, сохранив расстояние от линзы до предметной плоскости таким же, как и в случае (а). В приближении лучевой оптики определите функцию отклика системы на импульсное воздействие $h(x, y)$ и постройте график $h(x, 0)$.

К РАЗДЕЛУ 4.5

1. **Голография с использованием сферической опорной волны.** Выбор однородной плоской волны в качестве опорной не является существенным для голографии, и можно использовать другие волны. Предполагая, что опорная волна является сферической волной с центром в точке $(0, 0, -d)$, определите вид голографической картины и исследуйте восстановленную волну, когда:

а) предметная волна является плоской волной, распространяющейся под углом θ_x;

б) предметная волна — сферическая с центром в точке $(-x_0, 0, -d_1)$.

Аппроксимируйте сферические волны параболоидальными.

2. **Оптическая корреляция.** Фурье-образ транспаранта с амплитудным коэффициентом пропускания $f(x, y) = f_1(x - a, y) + f_2(x + a, y)$ получается с помощью линзы, и его интенсивность записывается в виде транспаранта (голограммы). Затем голограмма освещается опорной волной, и восстановленная волна подвергается преобразованию Фурье с помощью линзы, что дает функцию $g(x, y)$. Выведите выражение, связывающее $g(x, y)$ с $f_1(x, y)$ и $f_2(x, y)$. Покажите, как можно определить корреляцию двух функций $f_1(x, y)$ и $f_2(x, y)$ с помощью этой системы.

ГЛАВА 5 ЭЛЕКТРОМАГНИТНАЯ ОПТИКА

Джеймс Клерк Максвелл (1831—1879) выдвинул теорию, согласно которой свет представляет собой электромагнитное явление

Из результатов, представленных в гл. 2—4, ясно, что волновая оптика охватывает намного более широкий круг явлений, чем оптика лучей. Замечательно, что оба подхода дают одинаковые результаты для многих простых оптических явлений с участием параксиальных волн, таких как фокусировка света линзой и поведение света в средах с плавно меняющимся показателем преломления и периодических системах. Ясно, однако, что волновая оптика дает нечто большее, чем может дать оптика лучей: возможность описывать такие явления, как дифракция и интерференция, в которых важная роль принадлежит фазе и которые поэтому далеко выходят за рамки такой простой теории, как оптика лучей. Несмотря на свои многочисленные достижения, волновая оптика, как и оптика лучей, не способна количественно описать некоторые простые явления, наблюдаемые в оптических экспериментах, например расщепление пучка светоделителем. Доля отраженного и прошедшего света оказывается зависящей от поляризации падающего света, а это означает необходимость рассматривать свет в рамках векторной, а не скалярной теории. Вот здесь правильное описание и обеспечивает электромагнитная оптика. Вместе с рентгеновскими лучами и радиоволнами, свет представляет собой электромагнитное явление, описываемое *векторной* волновой теорией. Как показано на рис. 5.1 (см. цветную вклейку) электромагнитное излучение распространяется

в виде двух взаимно связанных векторных волн — электрической и магнитной. С этой точки зрения волновая оптика, описанная в гл. 2 и развитая в гл. 3 и 4, представляет собой просто *скалярное* приближение по отношению к более полной электромагнитной теории.

Таким образом, электромагнитная оптика включает в себя волновую оптику, которая, в свою очередь, сводится к лучевой оптике в коротковолновом пределе, как было показано в гл. 2. Эта иерархия показана на рис. 5.2.

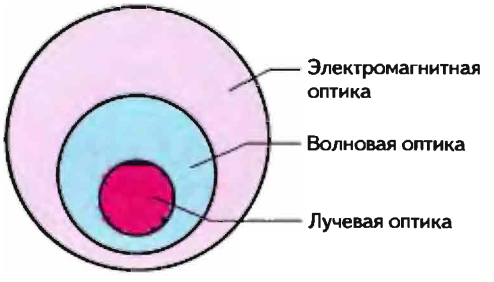


Рис. 5.2. Электромагнитная оптика — векторная теория, описывающая электрическое и магнитное поля, меняющиеся во времени и пространстве. Волновая оптика — приближение электромагнитной оптики, в котором используется скалярная волновая функция координат и времени. Лучевая оптика — предел волновой оптики при очень малых длинах волны

Оптические частоты занимают полосу электромагнитного спектра, включающую инфракрасный, видимый и ультрафиолетовый диапазоны, как показано на рис. 5.1. Область длин волн, обычно относимая к оптическому диапазону, простирается от 10 нм до 300 мкм (как более подробно показано на рис. 2.1). Поскольку эти длины волн существенно короче, чем в радио или микроволновом диапазоне, аппаратура и методы для их генерации, передачи и регистрации традиционно сильно различались. В последние годы движение в направлении миниатюризации привело к заметному размыванию этих различий, так, например, лазеры и оптические волноводы с размерами порядка длины волны, а также миниатюрные фотоприемники стали обычными в практике.

О данной главе

В данной главе приводится краткий обзор тех аспектов электромагнитной теории, которые имеют важнейшее значение в оптике. Фундаментальная теоретическая основа — уравнения Максвелла — представлена в разд. 5.1. Поведение оптических электромагнитных волн в диэлектрических средах изучается в разд. 5.2. Вместе эти разделы составляют основу электромагнитной оптики и обеспечивают набор правил, определяющих содержание оставшихся разделов главы. Эти правила существенно упрощаются в частном случае монохроматического света, что обсуждается в разд. 5.3. Элементарные электромагнитные волны (плоские, сферические волны и гауссовы пучки), введенные в разд. 5.4, представляют собой важные примеры волн, часто встречающихся на практике. Наконец, разд. 5.5 посвящен исследованию распространения света в средах с дисперсией, у которых поглощение и преломление зависят от длины волны, что всегда присуще реальным средам. Мы вернемся к этой теме далее в гл. 22.

На основе представленной здесь электромагнитной теории оптических явлений в гл. 6 непосредственно рассматривается поляризация света и взаимодействие поляризованного света с диэлектрическими и анизотропными средами, в частности, жидкими кристаллами. Материал данной главы образует основу также для гл. 8—11, где рассмотрена оптика слоистых и периодических структур, оптика волноводов, волоконная оптика, оптика резонаторов и статистическая оптика соответственно. Глава 21 посвящена электромагнитной оптике нелинейных сред.

5.1. ЭЛЕКТРОМАГНИТНАЯ ТЕОРИЯ СВЕТА

Электромагнитная волна описывается двумя связанными векторными полями, зависящими от координат и времени: **электрическим полем** $\mathcal{E}(r, t)$ и **магнитным полем** $\mathcal{H}(r, t)$. В общем случае имеется шесть скалярных функций координат и времени, необходимых для описания света в свободном пространстве. К счастью, эти шесть функций связаны друг с другом, поскольку они должны удовлетворять знаменитой системе связанных дифференциальных уравнений в частных производных, известной как система уравнений Максвелла.

Уравнения Максвелла в свободном пространстве

Векторы электрического и магнитного полей в свободном пространстве удовлетворяют **уравнениям Максвелла**

$$\nabla \times \mathcal{H} = \varepsilon_0 \frac{\partial \mathcal{E}}{\partial t}; \tag{5.1}$$

$$\nabla \times \mathcal{E} = -\mu_0 \frac{\partial \mathcal{H}}{\partial t}; \tag{5.2}$$

$$\nabla \cdot \mathcal{E} = 0; \tag{5.3}$$

$$\nabla \cdot \mathcal{H} = 0, \tag{5.4}$$

Уравнения Максвелла (свободное пространство)

где постоянные $\varepsilon_0 \approx (1/36\pi) \cdot 10^9$ Ф/м и $\mu_0 = 4\pi \cdot 10^{-7}$ Гн/м (система МКС) — **диэлектрическая** и **магнитная проницаемости** свободного пространства соответственно. Векторные операторы $\nabla\cdot$ и $\nabla\times$ обозначают дивергенцию и ротор соответственно[1].

[1] В декартовой системе координат $\nabla \cdot \mathcal{E} = \frac{\partial \mathcal{E}_x}{\partial x} + \frac{\partial \mathcal{E}_y}{\partial y} + \frac{\partial \mathcal{E}_z}{\partial z}$, а $\nabla \times \mathcal{E}$ — вектор с декартовыми компонентами $\left(\frac{\partial \mathcal{E}_z}{\partial y} - \frac{\partial \mathcal{E}_y}{\partial z}\right)$, $\left(\frac{\partial \mathcal{E}_x}{\partial z} - \frac{\partial \mathcal{E}_z}{\partial x}\right)$ и $\left(\frac{\partial \mathcal{E}_y}{\partial x} - \frac{\partial \mathcal{E}_x}{\partial y}\right)$.

Волновое уравнение

Необходимое условие того, чтобы $\mathcal{E}$ и $\mathcal{H}$ удовлетворяли уравнениям Максвелла, состоит в том, чтобы их компоненты удовлетворяли волновому уравнению

$$\boxed{\nabla^2 u - \frac{1}{c_0^2}\frac{\partial^2 u}{\partial t^2} = 0.} \qquad (5.5)$$

Волновое уравнение (свободное пространство)

Здесь

$$\boxed{c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}} \approx 3 \cdot 10^8 \text{ м/с} \qquad (5.6)$$

Скорость света (свободное пространство)

есть скорость света в вакууме, а скалярная функция $u(\boldsymbol{r}, t)$ представляет собой любую из трех компонент ($\mathcal{E}_x$, $\mathcal{E}_y$, $\mathcal{E}_z$) вектора $\mathcal{E}$ или трех компонент ($\mathcal{H}_x$, $\mathcal{H}_y$, $\mathcal{H}_z$) вектора $\mathcal{H}$.

Волновое уравнение для компонент вектора $\mathcal{E}$ можно вывести из уравнений Максвелла, если подействовать оператором ротора $\nabla\times$ на обе части (5.2), применить векторное тождество $\nabla \times (\nabla \times \mathcal{E}) = \nabla \cdot (\nabla \cdot \mathcal{E}) - \nabla^2 \mathcal{E}$ и затем использовать (5.1) и (5.3). Аналогичная процедура может быть проделана для $\mathcal{H}$. Поскольку уравнения Максвелла и волновое уравнение линейны, к ним применим принцип суперпозиции: если два набора компонент электрического и магнитного полей в отдельности являются решениями этих уравнений, то их сумма также является решением.

Теперь становится очевидной связь между электромагнитной и волновой оптикой. Волновое уравнение (2.2), составляющее основу волновой оптики, укладывается в структуру электромагнитной теории. Скорость света выражается через электромагнитные постоянные ε_0 и μ_0 соотношением (5.6), а скалярная волновая функция $u(\boldsymbol{r}, t)$ из гл. 2 представляет любую из шести компонент векторов электрического и магнитного полей. Электромагнитная оптика сводится к волновой оптике в задачах, где векторная природа электромагнитного поля несущественна. Как мы увидим далее в этой и последующих главах, векторный характер света лежит в основе явлений поляризации и определяет количество света, отраженного от границы раздела между различными средами или прошедшего через нее, что, в свою очередь, определяет характеристики распространения света в волноводах, слоистых средах и оптических резонаторах.

Уравнения Максвелла в среде

Для описания поля в среде, не содержащей свободных электрических зарядов и токов, требуется ввести два дополнительных вектора — **вектор электрической индукции** $\mathcal{D}(\boldsymbol{r}, t)$ (также называемый вектором электрического

смещения) и вектор магнитной индукции $\mathcal{B}(\boldsymbol{r}, t)$. Четыре векторных поля $\mathcal{E}$, $\mathcal{H}$, $\mathcal{D}$ и $\mathcal{B}$ связаны уравнениями Максвелла в среде, свободной от источников:

$$\nabla \times \mathcal{H} = \varepsilon_0 \frac{\partial \mathcal{D}}{\partial t}; \tag{5.7}$$

$$\nabla \times \mathcal{E} = -\frac{\partial \mathcal{B}}{\partial t}; \tag{5.8}$$

$$\nabla \cdot \mathcal{D} = 0; \tag{5.9}$$

$$\nabla \cdot \mathcal{B} = 0. \tag{5.10}$$

Уравнения Максвелла
(среда, свободная от источников)

Связь между вектором электрической индукции $\mathcal{D}$ и электрическим полем $\mathcal{E}$ зависит от электрических свойств среды, которые характеризуются вектором **плотности поляризации** $\mathcal{P}$. В диэлектрической среде плотность поляризации есть макроскопическая сумма электрических дипольных моментов, наведенных электрическим полем. Аналогично, связь между вектором магнитной индукции $\mathcal{B}$ и магнитным полем $\mathcal{H}$ зависит от магнитных свойств среды, описываемых вектором **плотности магнитного момента** $\mathcal{M}$, который определяется аналогично вектору плотности поляризации. Уравнения, связывающие векторы индукции с полями, имеют вид

$$\mathcal{D} = \varepsilon_0 \mathcal{E} + \mathcal{P}; \tag{5.11}$$

$$\mathcal{B} = \mu_0 \mathcal{H} + \mu_0 \mathcal{M}. \tag{5.12}$$

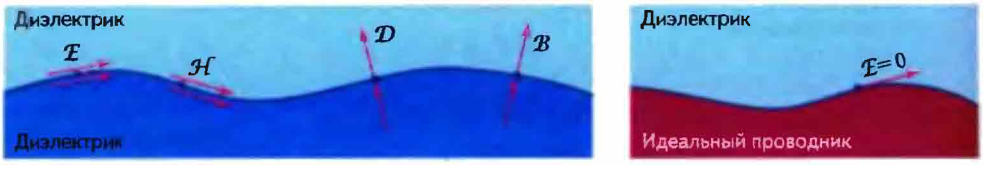
Векторы $\mathcal{P}$ и $\mathcal{M}$, в свою очередь, связаны с приложенными извне электрическим и магнитным полями посредством соотношений, вид которых зависит от электрических и магнитных свойств среды, как будет показано в разд. 5.2. Уравнения, связывающие $\mathcal{P}$ с $\mathcal{E}$ и $\mathcal{M}$ с $\mathcal{H}$, устанавливаются при определении среды (по этой причине эти уравнения в литературе обычно называются **материальными уравнениями** — *примеч. перев.*). При подстановке материальных уравнений в уравнения Максвелла в отсутствие источников из них исчезают векторы индукции.

В свободном пространстве $\mathcal{P} = \mathcal{M} = 0$, так что $\mathcal{D} = \varepsilon_0 \mathcal{E}$ и $\mathcal{B} = \mu_0 \mathcal{H}$, и уравнения (5.7)—(5.10) сводятся к уравнениям Максвелла для свободного пространства (5.1)—(5.4).

Граничные условия

В однородной среде все компоненты полей $\mathcal{E}$, $\mathcal{H}$, $\mathcal{D}$ и $\mathcal{B}$ являются непрерывными функциями координат. На границе между двумя диэлектрическими средами в отсутствие свободных электрических зарядов и токов тангенциальные компоненты электрического и магнитного полей $\mathcal{E}$ и $\mathcal{H}$ и нормальные компоненты векторов индукции $\mathcal{D}$ и $\mathcal{B}$ должны быть непрерывны (рис. 5.3).

На границе между диэлектриком и идеально проводящей средой тангенциальные компоненты вектора электрического поля должны исчезать. Поскольку идеальное зеркало изготовляется из идеально проводящего материала (металла), компонента электрического поля, параллельная поверхности зеркала, должна равняться нулю. Отсюда при нормальном падении электрические поля отраженной и падающей волн должны иметь равные величины и сдвиг по фазе на π так, чтобы их сумма равнялась нулю.



а *б*

Рис. 5.3. Граничные условия на границе раздела двух диэлектрических сред (*а*); на границе раздела между идеальным проводником и диэлектрическим материалом (*б*)

Эти граничные условия являются неотъемлемой составной частью уравнений Максвелла. Они используются для определения коэффициентов отражения и пропускания волн на различных границах (см. разд. 6.2) и для описания распространения волн в периодических структурах (см. разд. 8.2).

Интенсивность, мощность и энергия

Поток электромагнитной энергии определяется вектором плотности потока S

$$S = \mathcal{E} \times \mathcal{H}, \tag{5.13}$$

который называется **вектором Пойнтинга**. Направление потока в каждой точке совпадает с направлением вектора Пойнтинга, т. е. ортогонально $\mathcal{E}$ и $\mathcal{H}$. **Оптическая интенсивность** $I(\boldsymbol{r}, t)$ (поток энергии через единичную площадку, нормальную по отношению к вектору S)[2] — модуль вектора Пойнтинга, усредненного по времени $\langle S \rangle$. Усреднение проводится по промежутку времени, который много больше оптического периода, но много меньше других времен, представляющих интерес. Эквивалентное определение интенсивности в волновой оптике дается выражением (2.3).

С помощью векторного тождества $\nabla \cdot (\mathcal{E} \times \mathcal{H}) = (\nabla \times \mathcal{E}) \cdot \mathcal{H} - (\nabla \times \mathcal{H}) \cdot \mathcal{E}$ вместе с уравнениями Максвелла (5.7)—(5.8) и (5.11)—(5.12) получаем

$$\nabla \cdot S = -\frac{\partial}{\partial t}\left(\frac{1}{2}\varepsilon_0 \mathcal{E}^2 + \frac{1}{2}\mu_0 \mathcal{H}^2\right) + \mathcal{E} \cdot \frac{\partial \mathcal{P}}{\partial t} + \mu_0 \mathcal{H} \cdot \frac{\partial \mathcal{M}}{\partial t}. \tag{5.14}$$

Первый и второй члены в скобках в (5.14) представляют объемные плотности энергии, запасенной в электрическом и магнитном полях соответственно. Тре-

[2] Обсуждение этой интерпретации см. в книге *Борн М., Вольф Е.* Основы оптики. М., Наука, 1973.

тий и четвертый члены представляют плотности энергии, сообщенной электрическим и магнитным диполям материальной среды. Уравнение (5.14), известное как **теорема Пойнтинга**, таким образом, выражает закон сохранения энергии: поток энергии, выходящий из поверхности бесконечно малого элемента объема, равен скорости изменения энергии, запасенной внутри этого объема.

Импульс

Электромагнитное поле обладает импульсом, что проявляется в радиационном давлении на предметы со стороны отражаемой или рассеиваемой ими волны. В свободном пространстве объемная плотность импульса есть вектор

$$\varepsilon_0 \mathcal{E} \times \mathcal{B} = \frac{1}{c^2} S, \qquad (5.15)$$

Плотность импульса

пропорциональный вектору Пойнтинга S. Средний импульс поля в цилиндре длиной c с единичной площадью основания:

$$\frac{\langle S \rangle}{c^2} \cdot c = \frac{\langle S \rangle}{c}.$$

Этот импульс пересекает единичную площадку за единицу времени, так что средняя плотность потока импульса через единичную площадку, ориентированную перпендикулярно направлению S, равна $\langle S \rangle / c$.

Электромагнитная волна может также переносить момент импульса (угловой момент) и, следовательно, оказывать на предметы вращающее воздействие. Средняя плотность потока углового момента электромагнитным полем: $\boldsymbol{r} \times \langle S \rangle / c$. Например, пучки Лагерра—Гаусса, введенные в разд. 3.4, имеют винтовые волновые фронты, при этом вектор Пойнтинга имеет азимутальную компоненту, что приводит к возникновению орбитального углового момента.

5.2. ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ В ДИЭЛЕКТРИЧЕСКИХ СРЕДАХ

Характер среды заложен в соотношении между плотностями поляризации и магнитного момента, $\mathcal{P}$ и $\mathcal{M}$, с одной стороны, и электрическим и магнитным полями, $\mathcal{E}$ и $\mathcal{H}$, с другой стороны. Это соотношение, как уже говорилось ранее, называется **материальным уравнением**. В большинстве сред материальное уравнение распадается на два соотношения, одно из которых связывает $\mathcal{P}$ и $\mathcal{E}$, а другое — $\mathcal{M}$ и $\mathcal{H}$. Первое описывает диэлектрические свойства среды, а второе — магнитные. За исключением важных случаев, рассматриваемых в разд. 5.7 и 6.4, основной акцент в данной книге делается на диэлектрических свойствах. Поэтому сосредоточим наше внимание на соотношении между $\mathcal{P}$ и $\mathcal{E}$; соотношение между $\mathcal{M}$ и $\mathcal{H}$ подчиняется тем же закономерностям при аналогичных условиях.

Материальное соотношение между $\mathcal{P}$ и $\mathcal{E}$ удобно рассматривать как характеристику некоторой системы, для которой $\mathcal{E}$ представляет собой входной сигнал, а $\mathcal{P}$ — выходной сигнал, или отклик (рис. 5.4). Напомним, что и $\mathcal{E} = \mathcal{E}(\boldsymbol{r}, t)$, и $\mathcal{P} = \mathcal{P}(\boldsymbol{r}, t)$ являются функциями координат и времени.

Рис. 5.4. В ответ на приложенное электрическое поле $\mathcal{E}$ в диэлектрической среде создается плотность поляризации $\mathcal{P}$

Определения

• Диэлектрическая среда называется **линейной**, если векторное поле $\mathcal{P}(\boldsymbol{r}, t)$ линейно связано с векторным полем $\mathcal{E}(\boldsymbol{r}, t)$. В этом случае можно применять принцип суперпозиции.

• Среда называется **недиспергирующей**, если ее отклик мгновенен, т. е. если $\mathcal{P}$ в момент t определяется полем $\mathcal{E}$ в тот же момент времени t, а не предшествующими значениями $\mathcal{E}$. Отсутствие дисперсии, очевидно, является идеализацией, поскольку все физические системы, как бы ни быстро они откликались на внешнее воздействие, требуют для этого конечного времени.

• Среда называется **однородной**, если соотношение между $\mathcal{P}$ и $\mathcal{E}$ не зависит от положения точки $\boldsymbol{r}$.

• Среда называется **изотропной**, если соотношение между векторами $\mathcal{P}$ и $\mathcal{E}$ не зависит от направления вектора $\mathcal{E}$, так что свойства среды одинаковы по всем направлениям. В этом случае векторы $\mathcal{P}$ и $\mathcal{E}$ параллельны.

• Среда называется **пространственно недиспергирующей**, если связь между $\mathcal{P}$ и $\mathcal{E}$ является локальной, т. е. если $\mathcal{P}$ в каждой точке $\boldsymbol{r}$ определяется только значением поля $\mathcal{E}$ в той же точке $\boldsymbol{r}$. Во всей данной главе среда будет предполагаться пространственно недиспергирующей (оптически активные среды, рассматриваемые в подразд. 6.4.1, являются пространственно диспергирующими).

5.2.1. Линейные, недиспергирующие, однородные и изотропные среды

Рассмотрим вначале простейший случай *линейной*, *недиспергирующей*, *однородной* и *изотропной* диэлектрической среды. Векторы $\mathcal{P}$ и $\mathcal{E}$ в каждой точке и в каждый момент времени в этом случае параллельны и пропорциональны друг другу, т. е.

$$\boxed{\mathcal{P} = \varepsilon_0 \chi \mathcal{E},} \tag{5.16}$$

где скалярная постоянная χ называется **электрической восприимчивостью** (рис. 5.5).

Подстановка (5.16) в (5.11) показывает, что $\mathcal{D}$ и $\mathcal{E}$ также параллельны и пропорциональны друг другу

$$\mathcal{D} = \varepsilon \mathcal{E}, \tag{5.17}$$

где скалярная величина

$$\varepsilon = \varepsilon(1 + \chi) \tag{5.18}$$

определяется как **диэлектрическая проницаемость** среды. **Относительная проницаемость**

$$\frac{\varepsilon}{\varepsilon_0} = 1 + \chi$$

называется также **диэлектрической постоянной** среды.

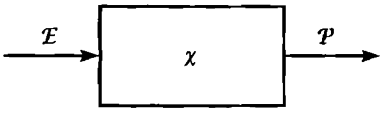


Рис. 5.5. Линейная недиспергирующая однородная и изотропная среда полностью характеризуется одной постоянной — электрической восприимчивостью χ

При тех же условиях соотношение для магнитного поля можно записать в виде

$$\mathcal{B} = \mu \mathcal{H}, \tag{5.19}$$

где μ — **магнитная восприимчивость** среды.

С учетом соотношений (5.17) и (5.19) уравнения Максвелла связывают только два векторных поля $\mathcal{E}(\boldsymbol{r}, t)$ и $\mathcal{H}(\boldsymbol{r}, t)$ и упрощаются к виду

$$\nabla \times \mathcal{H} = \varepsilon \frac{\partial \mathcal{E}}{\partial t}; \tag{5.20}$$

$$\nabla \times \mathcal{E} = -\mu \frac{\partial \mathcal{H}}{\partial t}; \tag{5.21}$$

$$\nabla \cdot \mathcal{E} = 0; \tag{5.22}$$

$$\nabla \cdot \mathcal{H} = 0, \tag{5.23}$$

Уравнения Максвелла (линейная, недиспергирующая, однородная и изотропная среда без источников)

Очевидно, что уравнения (5.20)—(5.23) идентичны по форме уравнениям Максвелла для свободного пространства (5.1)—(5.4) с заменой ε_0 на ε и μ_0 на μ. Каждая компонента $\mathcal{E}$ и $\mathcal{H}$ в этом случае удовлетворяет волновому уравнению

$$\nabla^2 u - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = 0, \tag{5.24}$$

Волновое уравнение (в среде)

где скорость света в среде обозначена через c:

$$\boxed{c = \frac{1}{\sqrt{\varepsilon\mu}}.} \qquad (5.25)$$

Скорость света (в среде)

Отношение скорости света в свободном пространстве к скорости света в среде, c_0/c, называется показателем преломления n:

$$\boxed{n = \frac{c_0}{c} = \sqrt{\frac{\varepsilon}{\varepsilon_0}\frac{\mu}{\mu_0}},} \qquad (5.26)$$

Показатель преломления (немагнитная среда)

где из (5.6) следует

$$c_0 = \frac{1}{\sqrt{\varepsilon_0\mu_0}}. \qquad (5.27)$$

Для немагнитного материала $\mu = \mu_0$ и

$$\boxed{n = \sqrt{\frac{\varepsilon}{\varepsilon_0}} = \sqrt{1+\chi},} \qquad (5.28)$$

Показатель преломления (немагнитная среда)

так что показатель преломления есть корень квадратный из диэлектрической постоянной. Эти соотношения обеспечивают еще один пункт связи со скалярной волновой оптикой (см. разд. 2.1), как обсуждается далее в подразд. 5.4.2.

Наконец, теорема Пойнтинга (5.14), основанная на уравнениях Максвелла (5.20) и (5.21), принимает вид уравнения непрерывности

$$\nabla \cdot S = -\frac{\partial \mathcal{W}}{\partial t}, \qquad (5.29)$$

где

$$\mathcal{W} = \frac{1}{2}\varepsilon\mathcal{E}^2 + \frac{1}{2}\mu\mathcal{H}^2 \qquad (5.30)$$

есть плотность энергии, запасенной в среде.

5.2.2. Нелинейные, диспергирующие, неоднородные, или неизотропные, среды

Рассмотрим теперь немагнитную диэлектрическую среду, для которой одно или более из свойств (линейность, бездисперсность, однородность, изотропия) не имеет места.

Неоднородные среды

Сначала рассмотрим неоднородный диэлектрик (такой, как среды с плавно меняющимся показателем преломления), который при этом линеен, изотропен и не имеет дисперсии. Простые соотношения пропорциональности $\mathcal{P} = \varepsilon_0 \chi \mathcal{E}$ и $\mathcal{D} = \varepsilon \mathcal{E}$ остаются в силе, но коэффициенты χ и ε становятся функциями координат, $\chi = \chi(\boldsymbol{r})$, $\varepsilon = \varepsilon(\boldsymbol{r})$ (рис. 5.6). Тогда показатель преломления также становится функцией координат, $n = n(\boldsymbol{r})$.

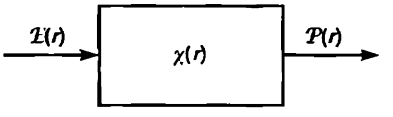


Рис. 5.6. Неоднородная (однако, линейная, недиспергирующая и изотропная) среда характеризуется восприимчивостью, зависящей от координат

Начиная с уравнений Максвелла (5.7)—(5.10) и замечая, что $\varepsilon = \varepsilon(\boldsymbol{r})$ — функция координат, применим операцию ротора $\nabla\times$ к обеим частям уравнения (5.8). Тогда, используя (5.7), можем записать

$$\boxed{\frac{\varepsilon_0}{\varepsilon}\nabla\times(\nabla\times\mathcal{E}) = -\frac{1}{c_0^2}\frac{\partial^2\mathcal{E}}{\partial t^2}.} \qquad (5.31)$$

Волновое уравнение (неоднородная среда)

Магнитное поле удовлетворяет другому уравнению

$$\boxed{\nabla\times\left(\frac{\varepsilon_0}{\varepsilon}\nabla\times\mathcal{H}\right) = -\frac{1}{c_0}\frac{\partial^2\mathcal{H}}{\partial t^2}.} \qquad (5.32)$$

Волновое уравнение (неоднородная среда)

Уравнение (5.31) также можно записать в виде

$$\nabla^2\mathcal{E} + \nabla\left(\frac{1}{\varepsilon}\nabla\varepsilon\cdot\mathcal{E}\right) - \mu_0\varepsilon\frac{\partial^2\mathcal{E}}{\partial t^2} = 0. \qquad (5.33)$$

Справедливость (5.33) можно продемонстрировать следующим образом. Используем тождество

$$\nabla\times(\nabla\times\mathcal{E}) = \nabla(\nabla\cdot\mathcal{E}) - \nabla^2\mathcal{E},$$

верное в прямоугольной системе координат. Привлечем (5.9), что дает $\nabla\cdot\varepsilon\mathcal{E} = 0$, а также тождество

$$\nabla\cdot\varepsilon\mathcal{E} = \varepsilon\nabla\ \mathcal{E} + \nabla\varepsilon\cdot\mathcal{E},$$

после чего

$$\nabla\ \mathcal{E} = -\frac{1}{\varepsilon}\nabla\varepsilon\cdot\mathcal{E}.$$

Наконец, подставим это в (5.31) и получим (5.33).

Для сред с плавно меняющимися диэлектрическими свойствами, т. е. когда величина $\varepsilon(\boldsymbol{r})$ меняется достаточно медленно, чтобы ее можно было считать постоянной на расстояниях порядка длины волны, вторым членом в левой части (5.33) можно пренебречь по сравнению с первым, так что

$$\nabla^2 \mathcal{E} - \frac{1}{c^2(\boldsymbol{r})} \frac{\partial^2 \mathcal{E}}{\partial t^2} \approx 0, \tag{5.34}$$

где

$$c(\boldsymbol{r}) = \frac{1}{\sqrt{\mu_0 \varepsilon}} = \frac{c_0}{n(\boldsymbol{r})}$$

зависит от координат; $n(\boldsymbol{r})$ — показатель преломления в точке $\boldsymbol{r}$,

$$n(\boldsymbol{r}) = \sqrt{\frac{\varepsilon(\boldsymbol{r})}{\varepsilon_0}}.$$

Это соотношение было использовано без доказательства в гл. 2, однако теперь совершенно ясно, что оно является приближенным следствием уравнений Максвелла.

Для однородной диэлектрической среды с показателем преломления n, возмущенным малой и медленно меняющейся добавкой Δn, часто бывает полезно записать (5.34) в виде

$$\nabla^2 \mathcal{E} - \frac{1}{c^2} \frac{\partial^2 \mathcal{E}}{\partial t^2} \approx -S; \quad S = -\mu_0 \frac{\partial^2 \Delta \mathcal{P}}{\partial t^2}; \quad \Delta \mathcal{P} = 2\varepsilon_0 n \Delta n \mathcal{E}, \tag{5.35}$$

где $c = c_0/n$ — скорость света в однородной среде. Таким образом, $\mathcal{E}$ удовлетворяет волновому уравнению с радиационным источником S, создаваемым возмущением плотности поляризации $\Delta \mathcal{P}$, которое, в свою очередь, пропорционально самим Δn и $\mathcal{E}$. В справедливости этих уравнений можно убедиться посредством разложения члена $1/c^2(\boldsymbol{r})$ в (5.34) как

$$\frac{(n + \Delta n)^2}{c_0^2} \approx \frac{n^2 + 2n\Delta n}{c_0^2}$$

и переноса члена, содержащего возмущение, в правую часть уравнения. Член $\Delta \mathcal{P}$ является возмущенной частью $\mathcal{P}$, в чем нетрудно убедиться, замечая что

$$\mathcal{P} = \varepsilon_0 \chi \mathcal{E} = \varepsilon_0 \left(\frac{\varepsilon}{\varepsilon_0} - 1 \right) \mathcal{E} = \varepsilon_0 (n^2 - 1) \mathcal{E},$$

так что

$$\Delta \mathcal{P} = \varepsilon_0 \Delta (n^2 - 1) \mathcal{E} = 2\varepsilon_0 n \Delta n \mathcal{E}.$$

Анизотропные среды

Связь между векторами $\mathcal{P}$ и $\mathcal{E}$ в анизотропной среде зависит от направления вектора $\mathcal{E}$; требование параллельности этих векторов в этом случае не выполняется. Если среда линейная, недиспергирующая и однородная, то каждая компонента $\mathcal{P}$ представляет собой линейную комбинацию трех компонент $\mathcal{E}$:

$$\mathcal{P}_i = \sum_j \varepsilon_0 \chi_{ij} \mathcal{E}_j, \tag{5.36}$$

где индексы $i, j = 1, 2, 3$ нумеруют компоненты x, y и z соответственно.

Диэлектрические свойства среды теперь описываются набором из 3×3 постоянных $\{\chi_{ij}\}$, являющихся **элементами тензора электрической восприимчивости** χ (рис. 5.7). Аналогичный вид приобретает соотношение между $\mathcal{D}$ и $\mathcal{E}$

$$\mathcal{D}_i = \sum_j \varepsilon_{ij} \mathcal{E}_j, \tag{5.37}$$

где $\{\varepsilon_{ij}\}$ — элементы тензора диэлектрической проницаемости.

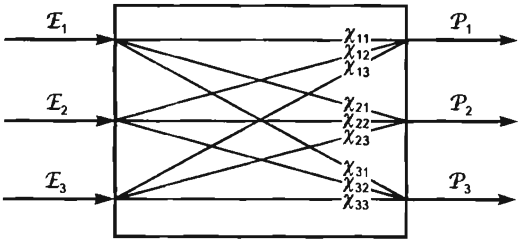


Рис. 5.7. Анизотропная (но линейная, однородная и недиспергирующая) среда характеризуется девятью постоянными — элементами тензора восприимчивости χ_{ij}. Каждая компонента $\mathcal{P}$ — взвешенная суперпозиция трех компонент $\mathcal{E}$

Оптические свойства анизотропных сред рассматриваются в гл. 6. Связь между $\mathcal{B}(t)$ и $\mathcal{H}(t)$ для магнитных анизотропных сред принимает вид, аналогичный (5.37) при тех же допущениях.

Диспергирующие среды

Связь между векторами $\mathcal{P}$ и $\mathcal{E}$ в диспергирующей диэлектрической среде динамическая, а не мгновенная. Вектор $\mathcal{E}(t)$ можно представлять как входной сигнал, возбуждающий колебания связанных электронов в атомах среды. Эти колебания коллективно порождают вектор плотности поляризации $\mathcal{P}(t)$ — выходной сигнал. Наличие временной задержки между выходным и входным сигналами отражает тот факт, что система обладает памятью. Лишь когда указанное время мало в сравнении с другими временами, можно считать отклик системы мгновенным, и тогда среда приближенно оказывается недиспергирующей.

Для диспергирующей среды, которая является линейной, однородной и изотропной, динамическая связь между $\mathcal{P}(t)$ и $\mathcal{E}(t)$ может быть описана, на-

пример, линейным дифференциальным уравнением, описывающим модель вынужденных колебаний гармонического осциллятора:

$$a_1 \frac{d^2\mathcal{P}}{dt^2} + a_2 \frac{d\mathcal{P}}{dt} + a_3 \mathcal{P} = \mathcal{E},$$

где a_1, a_2 и a_3 — постоянные. Простой анализ в этом направлении дает физическое объяснение возникновения дисперсии (и поглошения).

Более общий подход, представленный в приложении Б, можно использовать для исследования произвольной линейной системы, характеризуемой своим откликом на импульс (функцией отклика на импульсное воздействие). Импульс электрического поля $\delta(t)$, приложенный в момент времени $t = 0$, индуцирует протяженную во времени плотность поляризации $\varepsilon_0 x(t)$, где $x(t)$ — скалярная функция времени с конечной продолжительностью и началом при $t = 0$. Поскольку среда линейная, произвольное электрическое поле $\mathcal{E}(t)$ тогда индуцирует плотность поляризации, которая является суперпозицией откликов на $\mathcal{E}(t')$ для всех $t' \le t$ и может быть выражена в виде свертки, определенной в приложении А:

$$\mathcal{P}(t) = \varepsilon_0 \int_{-\infty}^{\infty} x(t - t')\mathcal{E}(t')\,dt' \tag{5.38}$$

Динамическая линейная система может быть писана также своей передаточной функцией, которая описывает отклик на гармоническое воздействие. Передаточная функция представляет собой Фурье-образ функции отклика на импульсное воздействие (см. приложение Б). В рассматриваемом примере передаточная функция на частоте ν есть $\varepsilon_0 \chi(\nu)$, где $\chi(\nu)$ — Фурье-образ функции $x(t)$, являющийся зависящей от частоты восприимчивостью (рис. 5.8). Это понятие обсуждается далее в разд. 5.3 и 5.5.

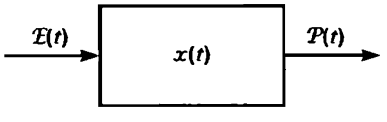


Рис. 5.8. В диспергирующей (но линейной, однородной и изотропной) среде связь между $\mathcal{P}(t)$ и $\mathcal{E}(t)$ задается динамической линейной системой, описываемой функцией отклика на импульсное воздействие $\varepsilon_0 x(t)$, которая соответствует зависящей от частоты восприимчивости $\chi(\nu)$

Для магнитных материалов при тех же допущениях связь между $\mathcal{M}(t)$ и $\mathcal{H}(t)$ аналогична (5.38).

Нелинейные среды

Нелинейная диэлектрическая среда определяется как характеризуемая нелинейной связью $\mathcal{P}$ и $\mathcal{E}$. В этом случае волновое уравнение в виде (5.24) неприменимо. Необходимо использовать уравнения Максвелла для вывода нелинейного волнового уравнения, которому подчиняются электромагнитные волны в такой среде.

Сначала выведем общее волновое уравнение, справедливое для однородной и изотропной немагнитной среды. Действуя на обе части уравнения Максвелла (5.8) оператором ротора $\nabla\times$ и используя соотношение $\mathcal{B} = \mu_0\mathcal{H}$ из (5.19) совместно с (5.7), мы получим

$$\nabla\times(\nabla\times\mathcal{E}) = -\frac{\mu_0\varepsilon^2\mathcal{D}}{\varepsilon t^2}.$$

Используя векторное тождество $\nabla\times(\nabla\times\mathcal{E}) = \nabla(\nabla\ \mathcal{E}) - \nabla^2\mathcal{E}$ и соотношение $\mathcal{D} = \varepsilon_0\mathcal{E} + \mathcal{P}$ из (5.11), находим

$$\nabla(\nabla\cdot\mathcal{E}) - \nabla^2\mathcal{E} = -\varepsilon_0\mu_0\frac{\partial^2\mathcal{E}}{\partial t^2} - \mu_0\frac{\partial^2\mathcal{P}}{\partial t^2}. \tag{5.39}$$

Для однородных и изотропных сред $\mathcal{D} = \varepsilon\mathcal{E}$, таким образом, $\nabla\ \mathcal{D} = 0$ из (5.9) эквивалентно $\nabla\ \mathcal{E} = 0$. Подставляя это вместе с

$$\varepsilon_0\mu_0 = \frac{1}{c_0^2}$$

из (5.6) в (5.39), находим

$$\boxed{\nabla^2\mathcal{E} - \frac{1}{c_0^2}\frac{\partial^2\mathcal{E}}{\partial t^2} = \mu_0\frac{\partial^2\mathcal{P}}{\partial t^2}.} \tag{5.40}$$

Волновое уравнение (однородная и изотропная среда)

Уравнение (5.40) применимо для всех однородных и изотропных диэлектрических сред, как линейных, так и нелинейных, диспергирующих или недиспергирующих.

Если среда нелинейная, недиспергирующая и немагнитная, то плотность поляризации $\mathcal{P}$ можно записать как нелинейную функцию $\mathcal{E}$ без памяти, скажем, $\mathcal{P} = \Psi(\mathcal{E})$, для любой точки и момента времени (простейший пример такой функции $\mathcal{P} = a_1\mathcal{E} + a_2\mathcal{E}^2$, где a_1, a_2 — постоянные). При этих условиях (5.40) становится нелинейным дифференциальным уравнением в частных производных относительно вектора электрического поля $\mathcal{E}(\boldsymbol{r}, t)$:

$$\nabla^2\mathcal{E} - \frac{1}{c_0^2}\frac{\partial^2\mathcal{E}}{\partial t^2} = \mu_0\frac{\partial^2\Psi(\mathcal{E})}{\partial t^2}. \tag{5.41}$$

Принцип суперпозиции больше применять нельзя из-за нелинейной природы этого волнового уравнения. Нелинейные магнитные материалы описываются аналогичным образом.

Большинство диэлектрических сред приближенно линейны, если только оптическая интенсивность не слишком велика, как, например, в сфокусированных лазерных пучках. Нелинейная оптика обсуждается в гл. 21.

5.3. МОНОХРОМАТИЧЕСКИЕ ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ

В частном случае монохроматических электромагнитных волн в оптической среде все компоненты электрического и магнитного полей являются гармоническими функциями одной частоты ν и соответствующей угловой частоты $\omega = 2\pi\nu$. Принимая правила комплексного представления, использованные в подразд. 2.2.1, эти шесть действительных компонент можно выразить как

$$\begin{aligned} \mathcal{E}(\boldsymbol{r}, t) &= \mathrm{Re}\{\boldsymbol{E}(\boldsymbol{r})\exp(j\omega t)\}; \\ \mathcal{H}(\boldsymbol{r}, t) &= \mathrm{Re}\{\boldsymbol{H}(\boldsymbol{r})\exp(j\omega t)\}, \end{aligned} \tag{5.42}$$

где $\boldsymbol{E}(\boldsymbol{r})$ и $\boldsymbol{H}(\boldsymbol{r})$ — комплексные векторные амплитуды электрического и магнитного полей соответственно. Аналогично вводятся комплексные амплитуды $\boldsymbol{P}$, $\boldsymbol{D}$, $\boldsymbol{M}$, $\boldsymbol{B}$ для всех действительных векторов $\mathcal{P}$, $\mathcal{D}$, $\mathcal{M}$ и $\mathcal{B}$ соответственно.

Уравнения Максвелла в среде

Подставляя (5.42) в уравнения Максвелла (5.7)—(5.10) и используя соотношение

$$\frac{\partial}{\partial t} e^{j\omega t} = j\omega e^{j\omega t}$$

для монохроматических волн с угловой частотой ω, получаем систему уравнений для векторов комплексных амплитуд

$$\boxed{\begin{aligned} \nabla \times \boldsymbol{H} &= j\omega \boldsymbol{D}; && (5.43) \\ \nabla \times \boldsymbol{E} &= -j\omega \boldsymbol{B}; && (5.44) \\ \nabla \cdot \boldsymbol{D} &= 0; && (5.45) \\ \nabla \quad \boldsymbol{B} &= 0. && (5.46) \end{aligned}}$$

Уравнения Максвелла
(среда без источников, монохроматические поля)

Аналогично, (5.11) и (5.12) порождают соотношения

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P}; \tag{5.47}$$

$$\boldsymbol{B} = \mu_0 \boldsymbol{H} + \mu_0 \boldsymbol{M}. \tag{5.48}$$

Интенсивность и мощность

Как отмечалось в разд. 5.1, плотность потока электромагнитной энергии определяется усредненным по времени вектором Пойнтинга $\boldsymbol{S} = \mathcal{E} \times \mathcal{H}$. Для комплексных амплитуд это выражение дает

$$\begin{aligned} \boldsymbol{S} &= \mathrm{Re}\{\boldsymbol{E}e^{j\omega t}\} \times \mathrm{Re}\{\boldsymbol{H}e^{j\omega t}\} = \frac{1}{2}(\boldsymbol{E}e^{j\omega t} + \boldsymbol{E}^* e^{-j\omega t}) \times \frac{1}{2}(\boldsymbol{H}e^{j\omega t} + \boldsymbol{H} \\ &= \frac{1}{4}(\boldsymbol{E} \times \boldsymbol{H}^* + \boldsymbol{E}^* \times \boldsymbol{H} + e^{2j\omega t}\boldsymbol{E} \times \boldsymbol{H} + e^{-2j\omega t}\boldsymbol{E}^* \times \boldsymbol{H}^*). \end{aligned} \tag{5.49}$$

Члены, содержащие множители $e^{2j\omega t}$ и $e^{-2j\omega t}$, осциллируют на оптической частоте и поэтому обращаются в нуль в процессе усреднения по времени, большому по сравнению с оптическим периодом. Таким образом,

$$\langle S\rangle = \frac{1}{4}(\boldsymbol{E}\times\boldsymbol{H}^* + \boldsymbol{E}^*\times\boldsymbol{H}) = \frac{1}{2}(\boldsymbol{S}+\boldsymbol{S}^*) = \mathrm{Re}\{\boldsymbol{S}\}, \tag{5.50}$$

где вектор

$$\boxed{\boldsymbol{S} = \frac{1}{2}(\boldsymbol{E}\times\boldsymbol{H}^*)} \tag{5.51}$$

Комплексный вектор Пойнтинга

можно рассматривать как комплексный вектор Пойнтинга. Оптическая интенсивность есть модуль вектора $\mathrm{Re}\{\boldsymbol{S}\}$.

Линейные, недиспергирующие, однородные и изотропные среды

Для монохроматических волн соотношения (5.17) и (5.19) превращаются в следующие **материальные уравнения**:

$$\boldsymbol{D} = \varepsilon\boldsymbol{E} \quad \text{и} \quad \boldsymbol{B} = \mu\boldsymbol{H}, \tag{5.52}$$

так что уравнения Максвелла (5.43)—(5.46) содержат только связанные между собой векторы комплексных амплитуд $\boldsymbol{E}$ и $\boldsymbol{H}$:

$$\boxed{\nabla\times\boldsymbol{H} = j\omega\varepsilon\boldsymbol{E};} \tag{5.53}$$

$$\boxed{\nabla\times\boldsymbol{E} = -j\omega\mu\boldsymbol{H};} \tag{5.54}$$

$$\boxed{\nabla\cdot\boldsymbol{E} = 0;} \tag{5.55}$$

$$\boxed{\nabla\ \boldsymbol{H} = 0.} \tag{5.56}$$

Уравнения Максвелла (линейная недиспергирующая однородная изотропная среда без источников, монохроматический свет)

Подстановка выражения (5.42) для электрического и магнитного полей $\mathcal{E}$ и $\mathcal{H}$ в волновое уравнение (5.24) приводит к уравнению Гельмгольца

$$\boxed{\nabla^2 U + k^2 U = 0}; \quad k = nk_0 = \omega\sqrt{\varepsilon\mu}, \tag{5.57}$$

Уравнение Гельмгольца

где скалярная функция $U = U(\boldsymbol{r})$ представляет собой комплексную амплитуду любой из трех компонент (E_x, E_y, E_z) вектора $\boldsymbol{E}$ или трех компонент (H_x, H_y, H_z) вектора $\boldsymbol{H}$ и

$$n = \sqrt{\frac{\varepsilon}{\varepsilon_0}\frac{\mu}{\mu_0}}; \quad k_0 = \frac{\omega}{c_0}; \quad c = \frac{c_0}{n}.$$

В волновой оптике уравнение Гельмгольца (2.22) было записано для комплексной амплитуды $U(\boldsymbol{r})$ действительной волновой функции $u(\boldsymbol{r}, t)$.

Неоднородные среды

В неоднородной немагнитной среде уравнения Максвелла (5.53)—(5.56) остаются справедливыми, но диэлектрическая проницаемость среды становится зависящей от координат, $\varepsilon = \varepsilon(\boldsymbol{r})$. Для локально однородных сред, в которых $\varepsilon(\boldsymbol{r})$ меняется медленно по сравнению с длиной волны, уравнение Гельмгольца (5.57) остается приближенно верным при подстановках

$$k = n(\boldsymbol{r})k_0; \quad n(\boldsymbol{r}) = \sqrt{\frac{\varepsilon(\boldsymbol{r})}{\varepsilon_0}}.$$

Диспергирующие среды

В диспергирующей диэлектрической среде $\mathcal{P}(t)$ и $\mathcal{E}(t)$ связаны динамическим соотношением (5.38). Для определения соответствующей связи между векторами комплексных амплитуд $\boldsymbol{P}$ и $\boldsymbol{E}$ подставим (5.42) в (5.38), что дает

$$\boxed{\boldsymbol{P} = \varepsilon_0 \chi(\nu)\boldsymbol{E},} \tag{5.58}$$

где

$$\chi(\nu) = \int_{-\infty}^{\infty} x(t)\exp(-j2\pi\nu t)\,dt \tag{5.59}$$

есть Фурье-образ функции $x(t)$. Уравнение (5.58) может быть также непосредственно получено из (5.38) с привлечением теоремы свертки: свертка во временной области соответствует умножению в частотной области (см. разд. А.1 и Б.1 приложений А и Б соответственно). При этом нужно заметить, что $\boldsymbol{P}$ и $\boldsymbol{E}$ являются фурье-компонентами $\mathcal{P}$ и $\mathcal{E}$ на частоте ν. Функция $\varepsilon_0\chi(\nu)$ может, следовательно, рассматриваться как передаточная функция линейной системы, связывающей $\mathcal{P}(t)$ и $\mathcal{E}(t)$.

Соотношение между $\mathcal{D}$ и $\mathcal{E}$ аналогично,

$$\boldsymbol{D} = \varepsilon(\nu)\boldsymbol{E}, \tag{5.60}$$

где

$$\varepsilon(\nu) = \varepsilon_0[1 + \chi(\nu)]. \tag{5.61}$$

Следовательно, в диспергирующих средах восприимчивость χ и диэлектрическая проницаемость ε зависят от частоты и в общем случае являются комплексными величинами. Таким образом, уравнение Гельмгольца (5.57) легко адаптируется для применения к диспергирующим немагнитным средам, если положить

$$k = \omega\sqrt{\varepsilon(\nu)\mu_0}. \tag{5.62}$$

Когда $\chi(\nu)$ и $\varepsilon(\nu)$ приблизительно постоянны в исследуемом диапазоне частот, среда может рассматриваться как почти не диспергирующая. Следствия комплексной природы χ и k в диспергирующих средах обсуждаются далее в разд. 5.5.

5.4. ЭЛЕМЕНТАРНЫЕ ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ

5.4.1. Плоские, сферические и гауссовы электромагнитные волны

Теперь изучим простейшие решения уравнений Максвелла, имеющие большое значение в оптике: плоские и сферические волны, которые обсуждались в подразд. 2.2.2, а также гауссовы пучки, исследованные в гл. 3, причем в обоих случаях с использованием формализма волновой оптики. Среда предполагается линейной, однородной, недиспергирующей и изотропной, а волны — монохроматическими.

Поперечная электромагнитная (TEM) плоская волна

Рассмотрим монохроматическую электромагнитную волну, векторы комплексных амплитуд магнитного и электрического полей которой — плоские волны с волновым вектором $\boldsymbol{k}$ (см. подразд. 2.2.2), так что

$$\boldsymbol{H}(\boldsymbol{r}) = \boldsymbol{H}_0 \exp(-j\boldsymbol{k}\boldsymbol{r}); \tag{5.63}$$

$$\boldsymbol{E}(\boldsymbol{r}) = \boldsymbol{E}_0 \exp(-j\boldsymbol{k}\boldsymbol{r}), \tag{5.64}$$

где комплексные огибающие $\boldsymbol{H}_0$ и $\boldsymbol{E}_0$ — постоянные векторы. Все компоненты $\boldsymbol{H}(\boldsymbol{r})$ и $\boldsymbol{E}(\boldsymbol{r})$ удовлетворяют уравнению Гельмгольца при условии, что модуль вектора $\boldsymbol{k}$ равен $k = nk_0$, где n — показатель преломления среды.

Исследуем условия, которым должны подчиняться $\boldsymbol{H}_0$ и $\boldsymbol{E}_0$, чтобы удовлетворялись уравнения Максвелла. Подстановка (5.63) и (5.64), соответственно, в уравнения Максвелла (5.53) и (5.54) приводит к уравнениям

$$\boldsymbol{k} \times \boldsymbol{H}_0 = -\omega\varepsilon\boldsymbol{E}_0; \tag{5.65}$$

$$\boldsymbol{k} \times \boldsymbol{E}_0 = \omega\mu\boldsymbol{H}_0. \tag{5.66}$$

Два других уравнения Максвелла (5.55) и (5.56) удовлетворяются тождественно, поскольку дивергенция однородной плоской волны равна нулю.

Из (5.65) следует, что $\boldsymbol{E}$ должно быть перпендикулярно как $\boldsymbol{k}$, так и $\boldsymbol{H}$, а из (5.66) — что $\boldsymbol{H}$ должно быть перпендикулярно как $\boldsymbol{k}$, так и $\boldsymbol{E}$. Таким образом, $\boldsymbol{E}$, $\boldsymbol{H}$ и $\boldsymbol{k}$ взаимно ортогональны, как показано на рис. 5.9. Поскольку $\boldsymbol{E}$ и $\boldsymbol{H}$ лежат в плоскости, перпендикулярной направлению распространения $\boldsymbol{k}$, волна называется **поперечной электромагнитной (TEM)** волной.

В соответствии с (5.65) величины H_0 и E_0 связаны соотношением

$$H_0 = \frac{\omega\varepsilon}{k} E_0.$$

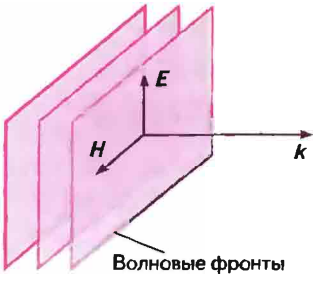


Рис. 5.9. Плоская ТЕМ-волна. Векторы $\boldsymbol{E}$, $\boldsymbol{H}$ и $\boldsymbol{k}$ взаимно ортогональны. Волновые фронты (поверхности постоянной фазы) нормальны $\boldsymbol{k}$

Аналогично, (5.66) дает

$$H_0 = \frac{k}{\omega\mu} E_0.$$

Чтобы эти два уравнения были совместны, необходимо соблюдение соотношения

$$\frac{\omega\varepsilon}{k} = \frac{k}{\omega\mu}$$

или

$$k = \omega\sqrt{\varepsilon\mu} = \frac{\omega}{c} = \frac{n\omega}{c_0} = nk_0.$$

Фактически это то же самое условие, которое требуется для того, чтобы волна удовлетворяла уравнению Гельмгольца.

Отношение амплитуд электрического и магнитного полей

$$\frac{E_0}{H_0} = \frac{\omega\mu}{k} = c\mu = \sqrt{\frac{\mu}{\varepsilon}}.$$

Эта величина называется импедансом среды,

$$\boxed{\eta = \frac{E_0}{H_0} = \sqrt{\frac{\mu}{\varepsilon}}.} \qquad (5.67)$$

Импеданс

Для немагнитных сред $\mu = \mu_0$, в результате чего

$$\eta = \sqrt{\frac{\mu_0}{\varepsilon}}$$

и можно определить через импеданс свободного пространства посредством соотношения

$$\boxed{\eta = \frac{\eta_0}{n},} \qquad (5.68)$$

Импеданс (немагнитные среды)

где

$$\eta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \approx 120\pi \approx 377 \text{ Ом}. \qquad (5.69)$$

Комплексный вектор Пойнтинга $\boldsymbol{S} = (1/2)\boldsymbol{E} \times \boldsymbol{H}^*$ [см. (5.51)] параллелен волновому вектору $\boldsymbol{k}$, так что поток энергии направлен по нормали к волновому фронту. Его величина равна

$$\frac{1}{2}E_0H_0 = \frac{|E_0|^2}{2\eta},$$

следовательно, интенсивность дается формулой

$$\boxed{I = \frac{|E_0|^2}{2\eta}.} \qquad (5.70)$$

Интенсивность

Видно, что интенсивность ТЕМ-волны пропорциональна квадрату модуля комплексной огибающей электрического поля. Например, интенсивность в 10 Вт/см^2 в свободном пространстве соответствует электрическому полю около 87 В/см. Обратите внимание на совпадение (5.70) и формулы $I = |U|^2$, согласно которой определялась интенсивность для скалярных волн в подразд. 2.2.1.

Из формулы (5.30) следует, что усредненная по времени плотность энергии $W = \langle \mathcal{W} \rangle$ для плоской волны

$$W = \frac{1}{2}\varepsilon|E_0|^2 \qquad (5.71)$$

поскольку электрический и магнитный вклады равны, т. е.

$$\frac{1}{2}\varepsilon\frac{|E_0|^2}{2} = \frac{1}{2}\mu\frac{|H_0|^2}{2}.$$

Следовательно, интенсивность (5.70) и усредненная по времени плотность энергии (5.71) связаны соотношением

$$I = cW, \qquad (5.72)$$

которое означает, что усредненная по времени плотность потока энергии I является результатом переноса усредненной по времени плотности энергии со скоростью c. Это легко увидеть, рассматривая цилиндр с площадью основания A и длиной c, ось которого параллельна направлению распространения. Энергия, запасенная в цилиндре, cAW, пройдет через его основание за одну секунду, что согласуется с формулой (5.72).

Плотность импульса, переносимого плоской волной, равна

$$\frac{1}{c^2}S = \frac{1}{c^2}I\hat{k} = \frac{W}{c}\hat{k}.$$

Сферическая волна

Осциллирующий электрический диполь излучает волну со свойствами, напоминающими скалярную сферическую волну, обсуждавшуюся в подразд. 2.2.2. Эта электромагнитная волна легко строится из вспомогательного векторного поля $\boldsymbol{A}(\boldsymbol{r})$, известного как векторный потенциал, которое часто используется для облегчения решения уравнений Максвелла в электродинамике. Для рассматриваемого случая

$$\boldsymbol{A}(\boldsymbol{r}) = A_0 U(\boldsymbol{r})\hat{\boldsymbol{x}}, \tag{5.73}$$

где A_0 — постоянная; $\hat{\boldsymbol{x}}$ — единичный вектор в направлении оси x. Величина $U(\boldsymbol{r})$ представляет скалярную сферическую волну с началом в точке $r = 0$:

$$U(\boldsymbol{r}) = \frac{1}{r}\exp(-jkr). \tag{5.74}$$

Так как $U(\boldsymbol{r})$ удовлетворяет уравнению Гельмгольца, как было установлено в подразд. 2.2.2, то $\boldsymbol{A}(\boldsymbol{r})$ тоже будет удовлетворять уравнению Гельмгольца $\nabla^2\boldsymbol{A} + k^2\boldsymbol{A} = 0$.

Выразим теперь магнитное поле как ротор этого вектора

$$\boldsymbol{H} = \frac{1}{\mu}\nabla \times \boldsymbol{A} \tag{5.75}$$

и определим соответствующее электрическое поле из уравнения Максвелла (5.53)

$$\boldsymbol{E} = \frac{1}{j\omega\varepsilon}\nabla \times \boldsymbol{H}. \tag{5.76}$$

Формулы (5.75) и (5.76) гарантируют выполнение равенств $\nabla \cdot \boldsymbol{E} = 0$ и $\nabla \cdot \boldsymbol{H} = 0$, поскольку дивергенция ротора любого вектора равна нулю. Поскольку $\boldsymbol{A}(\boldsymbol{r})$ удовлетворяет уравнению Гельмгольца, легко показать, что остающееся уравнение Максвелла, $\nabla \times \boldsymbol{E} = -j\omega\mu\boldsymbol{H}$, также удовлетворяется. Таким образом, формулы (5.73)—(5.76) полностью определяют правильную электромагнитную волну, удовлетворяющую уравнениям Максвелла.

Для получения явных выражений $\boldsymbol{E}$ и $\boldsymbol{H}$ нужно вычислить ротор в (5.75) и (5.76). Это удобно делать в сферических координатах (r, θ, ϕ), определение которых дано на рис. 5.10, *а*. Для точек, удаленных от начала координат на расстояние, много большее длины волны ($r \gg 1$ или $kr \gg 2\pi$), векторы комплексных амплитуд можно аппроксимировать выражениями

$$\boldsymbol{E}(\boldsymbol{r}) \approx E_0 \sin\theta\, U(\boldsymbol{r})\hat{\theta}; \tag{5.77}$$

$$\boldsymbol{H}(\boldsymbol{r}) \approx H_0 \sin\theta\, U(\boldsymbol{r})\hat{\phi}, \tag{5.78}$$

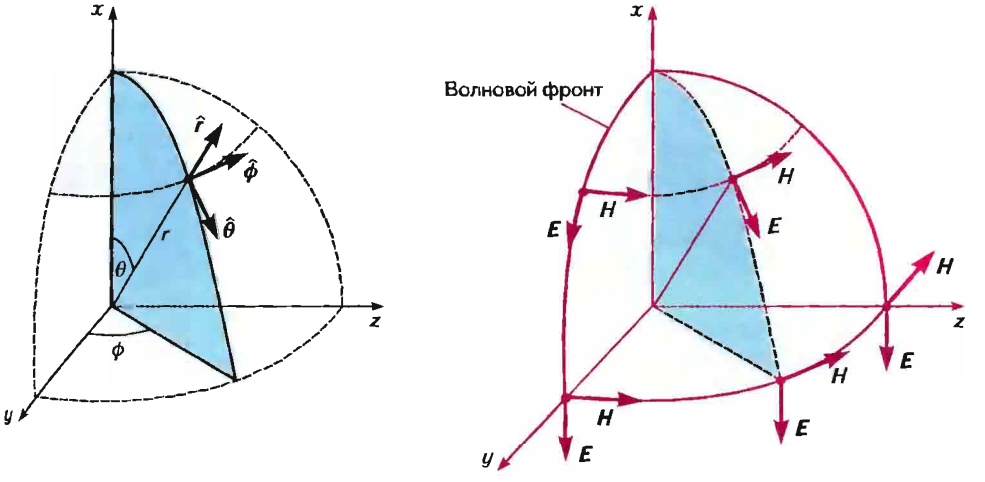
где

$$H_0 = \frac{jk}{\mu}A_0; \quad E_0 = \eta H_0; \quad \theta = \arccos\left(\frac{x}{r}\right);$$

$\hat{\theta}$ и $\hat{\phi}$ — единичные векторы в сферических координатах. Следовательно, волновые фронты представляют собой сферы, а электрическое и магнитное поля ортогональны друг другу и радиальному направлению $\hat{r}$, как показано на рис. 5.10, *б*. В отличие от скалярной сферической волны, амплитуда этой векторной волны меняется как $\sin\theta$. В точках, близких к оси z, и удаленных от начала координат,

$$\theta \approx \frac{\pi}{2}; \quad \phi \approx \frac{\pi}{2},$$

так что нормали к волновому фронту почти параллельны оси z, что соответствует параксиальным лучам, а $\sin\theta \approx 1$.



а *б*

Рис. 5.10. Сферическая система координат (*а*). Векторы электрического и магнитного полей и волновые фронты на расстояниях $r \gg 1$ для электромагнитного поля, излучаемого осциллирующим электрическим диполем (*б*)

В декартовой системе координат

$$\hat{\theta} = -\sin\theta\hat{x} + \cos\theta\cos\phi\hat{y} + \cos\theta\sin\phi\hat{z} \approx -\hat{x} + \frac{x}{z}\frac{y}{z}\hat{y} + \frac{x}{z}\hat{z} \approx$$

так что

$$\boldsymbol{E}(\boldsymbol{r}) \approx E_0\left(-\hat{x} + \frac{x}{z}\hat{z}\right)U(\boldsymbol{r}), \tag{5.79}$$

где $U(\boldsymbol{r})$ — параксиальное приближение для сферической волны, а именно, параболоидальная волна, рассмотренная в подразд. 2.2.2. Для достаточно больших значений z членом (x/z) в (5.79) также можно пренебречь, в результате чего имеем

$$\boldsymbol{E}(\boldsymbol{r}) \approx -E_0U(\boldsymbol{r})\hat{x}; \tag{5.80}$$

$$\boldsymbol{H}(\boldsymbol{r}) \approx H_0U(\boldsymbol{r})\hat{y}. \tag{5.81}$$

В этом приближении

$$U(\boldsymbol{r}) \to \frac{1}{z} e^{-jkz},$$

так что в конце концов получается плоская ТЕМ-волна.

Гауссов пучок

В разд. 3.1 было показано, что скалярный гауссов пучок легко получается из параболоидальной волны (параксиального приближения для сферической волны) посредством замены координаты z на $z + jz_0$, где z_0 — действительная постоянная.

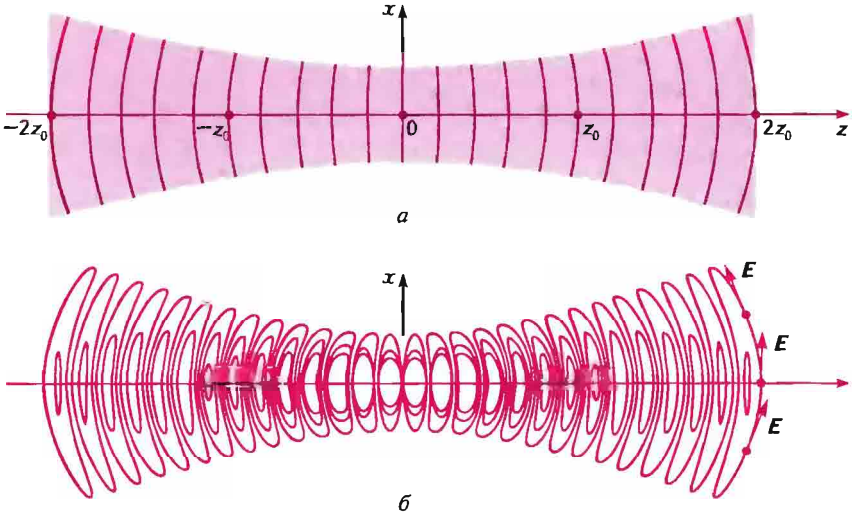


Рис. 5.11. Волновые фронты скалярного гауссова пучка $U(\boldsymbol{r})$ в плоскости x—z (*а*). Линии электрического поля электромагнитного гауссова пучка в плоскости x—z (*б*). (Из *Haus H.A.* Waves and Fields in Optoelectronics. Prentice Hall, 1984, рис. 5.3, *а*)

Такое же преобразование соответствующей электромагнитной волны приводит к электромагнитному гауссову пучку. Замена z на $z + jz_0$ в (5.79) приводит к выражению

$$\boldsymbol{E}(\boldsymbol{r}) = E_0 \left(-\hat{x} + \frac{x}{z + jz_0} \hat{z} \right) U(\boldsymbol{r}), \tag{5.82}$$

где $U(\boldsymbol{r})$ теперь представляет собой скалярную комплексную амплитуду гауссова пучка (3.7). Волновые фронты гауссова пучка показаны на рис. 5.11, *а* (они также показаны на рис. 3.7), а линии электрического поля $\boldsymbol{E}$, заданного формулой (5.82), показаны на рис. 5.11, *б*. В этом случае пространственное распределение направлений поля $\boldsymbol{E}$ неоднородно.

5.4.2. Связь между электромагнитной и скалярной волновой оптикой

Параксиальная скалярная волна, определенная в подразд. 2.2.3, имеет нормали к волновому фронту, образующие малые углы с продольной координатной осью z. Волновые фронты локально ведут себя как плоские волны, а комплексная огибающая и направление распространения медленно меняются вдоль z.

Эти представления применимы и к электромагнитным волнам в линейных изотропных средах. Параксиальная электромагнитная волна локально аппроксимируется плоской ТЕМ-волной. В каждой точке векторы $\boldsymbol{E}$ и $\boldsymbol{H}$ лежат в плоскости, касательной по отношению к волновому фронту и нормальной к волновому вектору $\boldsymbol{k}$ (рис 5.12). Оптическая энергия течет в направлении $\boldsymbol{E} \times \boldsymbol{H}$, параллельном $\boldsymbol{k}$ и приблизительно параллельном оси z.

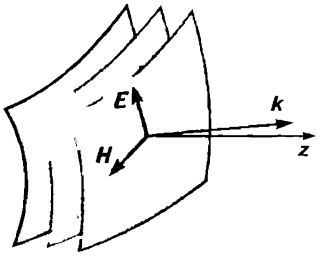


Рис. 5.12. Параксиальная электромагнитная волна

Параксиальную скалярную волну интенсивности $I = |U|^2$ [см. (2.14)] можно связать с параксиальной электромагнитной волной той же интенсивности $I = |E|^2/2\eta$ [см. (5.70)], полагая комплексную амплитуду равной

$$U = \frac{E}{\sqrt{2\eta}}$$

и совмещая волновые фронты. Как говорилось в гл. 2—4, скалярная теория света обеспечивает очень хорошее приближение для решения множества задач, включающих интерференцию, дифракцию, распространение света и формирование изображений в параксиальных волнах. Удачным примером является гауссов пучок с малой расходимостью, рассмотренный в гл. 3. Большинство свойств таких пучков, как интенсивность, фокусировка линзой, отражение от зеркала и интерференция, можно удовлетворительно исследовать в рамках скалярной волновой оптики. Конечно, когда в игру вступает поляризация света, скалярная волновая оптика молчит и *необходимо* обращаться к электромагнитной оптике.

Интересно заметить, что U (как определено выше) и E не удовлетворяют одним и тем же граничным условиям. Например, для электрического поля, касательного по отношению к границе между двумя диэлектрическими средами, E непрерывно, однако, $U = E/\sqrt{2\eta}$ терпит разрыв, поскольку значение η меняется на границе. Таким образом, задачи, включающие отражение и преломление на границах, нельзя полностью решить на основе скалярной волновой теории, хотя процедура обеспечения фазового синхронизма, приводящая к закону отражения и закону Снелла, адекватно выполняется в ее рамках (см. разд. 2.4). В самом деле, расчет коэффициентов отражения и пропускания на границе зависит от состояния поляризации света и поэтому требует применения элект-

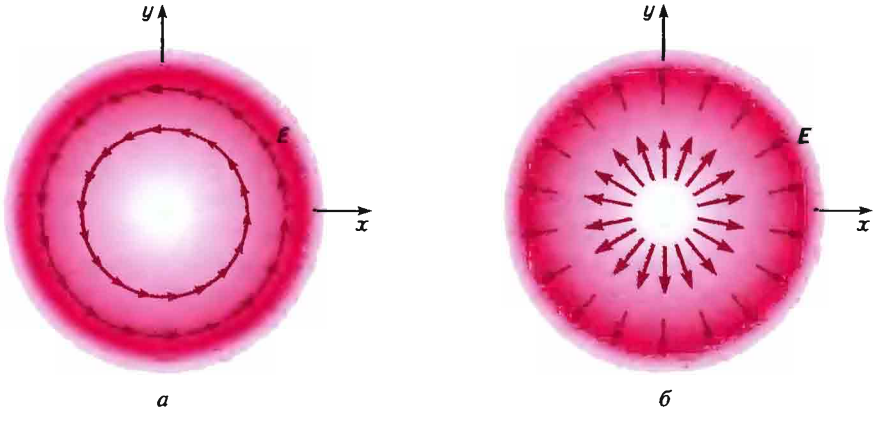
ромагнитной оптики (см. разд. 6.2). Аналогично, задачи распространения света в диэлектрических волноводах требуют анализа, основанного на электромагнитной теории, что обсуждается в гл. 8 и 9.

5.4.3. Векторные пучки

Уравнения Максвелла в параксиальном приближении допускают другие решения в виде цилиндрически симметричных пучков, для которых распределение направлений вектора электрического поля пространственно неоднородно. Один из примеров — пучок, для которого электрическое поле имеет азимутальную ориентацию по отношению к оси пучка (рис. 5.13, *а*), т. е.

$$\boldsymbol{E}(\boldsymbol{r}) = U(\rho, z)\exp(-jkz)\hat{\phi}. \tag{5.83}$$

Скалярная функция $U(\rho, z)$ оказывается решением Бесселя—Гаусса уравнения Гельмгольца, как обсуждалось в разд. 3.4. Интенсивность пучка исчезает на оси ($\rho = 0$) и имеет кольцевое пространственное распределение. Пучок расходится в продольном направлении и размер его пятна увеличивается, сильно напоминая гауссов пучок[3].



а *б*

Рис. 5.13. Векторные пучки с цилиндрической симметрией:

а — вектор электрического поля ориентирован в азимутальном направлении; *б* — вектор электрического поля ориентирован в радиальном направлении. Темным показано пространственное распределение оптической интенсивности в поперечной плоскости

Еще один цилиндрически симметричный пучок имеет азимутально-ориентированный вектор магнитного поля, при этом электрическое поле направлено радиально, как схематически показано на рис. 5.13, *б*. Он также имеет пространственное распределение с нулем на оси. Распределение векторного поля

[3] *Hall D.G.* Vector-Beam Solutions of Maxwell's Wave Equation. Optics Letters. Vol. 21, 1996. P. 9—11.

для этого пучка аналогично электромагнитному полю, излучаемому диполем, ориентированным вдоль оси пучка.

Показано, что векторный пучок с радиальным электрическим полем можно сфокусировать линзой большой числовой апертуры в пятно значительно меньших размеров, чем для обычного скалярного гауссова пучка[4]. Очевидно, что такие пучки могут найти полезные приложения в микроскопии высокого разрешения.

5.5. ПОГЛОЩЕНИЕ И ДИСПЕРСИЯ

В данном разделе мы рассмотрим поглощение и дисперсию в немагнитных средах.

5.5.1. Поглощение

Рассмотренные до сих пор диэлектрические среды предполагались полностью прозрачными, т. е. не поглощающими свет. Стекло является материалом, поглощающим свет в ультрафиолетовом и инфракрасном диапазонах видимого спектра. Оптические элементы, пропускающие свет в этих диапазонах, изготовляются из других материалов, примерами которых являются кварц и фтористый магний в ультрафиолетовом диапазоне, германий и фтористый барий — в инфракрасном. На рис. 5.14 показаны спектральные окна, внутри которых некоторые часто встречающиеся оптические материалы прозрачны (для дальнейшего обсуждения см. подразд. 13.1.3).

В данном разделе мы будем исходить из феноменологического подхода к поглощению света в линейных средах. Рассмотрим комплексную электрическую восприимчивость

$$\chi = \chi' + j\chi'', \tag{5.84}$$

соответствующую комплексной диэлектрической проницаемости $\varepsilon = \varepsilon_0(1 + \chi)$ и комплексной диэлектрической постоянной $\varepsilon/\varepsilon_0 = (1 + \chi)$. Для монохроматического света уравнение Гельмгольца (5.57) для комплексной амплитуды $U(\boldsymbol{r})$ остается справедливым, $\nabla^2 U + k^2 U = 0$, однако само волновое число k становится комплексным:

$$k = \omega\sqrt{\varepsilon\mu_0} = k_0\sqrt{1 + \chi} = k_0\sqrt{1 + \chi' + j\chi''}, \tag{5.85}$$

где $k_0 = \omega/c_0$ — волновое число в свободном пространстве.

Расписывая действительную и мнимую части k, $k = \beta - j\alpha/2$, можно связать β и α с действительной и мнимой частями восприимчивости χ' и χ'':

$$\beta - j\frac{\alpha}{2} = k_0\sqrt{1 + \chi' + j\chi''} \tag{5.86}$$

---

[4] *Dorn R., Quabis S., Leuchs G.* Sharper Focus for a Radially Polarized Light Beam. Physical Review Letters. Vol. 91, 2003. P. 233901.

В результате появления мнимой части у k плоская волна с комплексной амплитудой $U = A\exp(-jkz)$, распространяющаяся в такой среде в направлении z, кроме обычного изменения фазы испытывает изменение амплитуды. Подставляя $k = \beta - j\alpha/2$ в показатель экспоненты этой плоской волны, получаем

$$U = A\exp\left(-\frac{1}{2}\alpha z\right)\exp(-j\beta z).$$

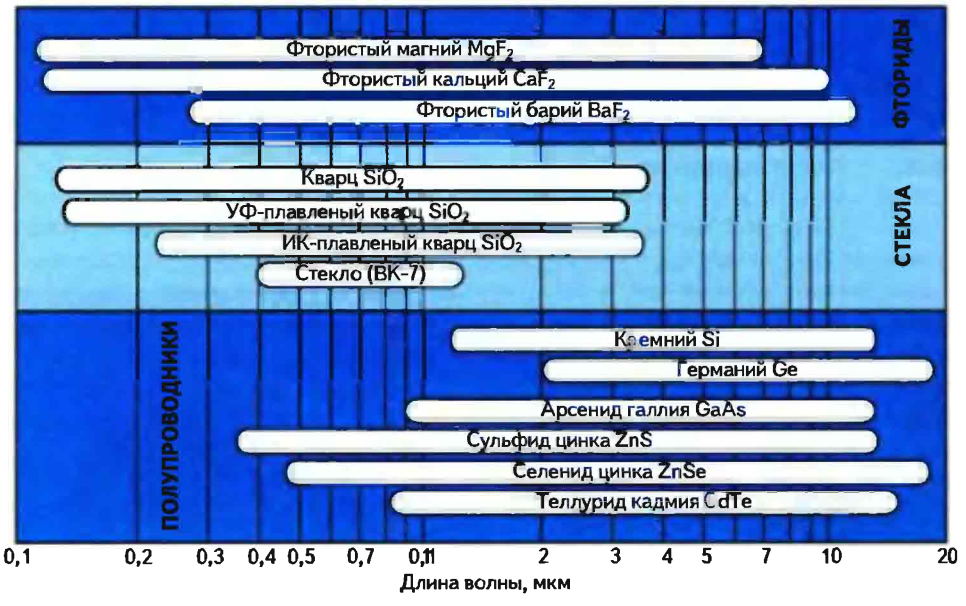


Рис. 5.14. Спектральные области, в которых некоторые оптические материалы пропускают свет

При $\alpha > 0$, что соответствует поглощению в среде, огибающая A исходной плоской волны ослабляется множителем $\exp(-\alpha z/2)$, так что интенсивность, пропорциональная $|U|^2$, убывает как

$$\left|\exp\left(-\frac{1}{2}\alpha z\right)\right|^2 = \exp(-\alpha z).$$

Коэффициент α тогда можно интерпретировать как **коэффициент поглощения** (называемый также **коэффициентом ослабления** или **коэффициентом экстинкции**) среды. Этот простой экспоненциальный закон ослабления интенсивности дает основания для представления мнимой части k в виде $-\alpha/2$. В подразд. 14.1.1 будет показано, что в некоторых средах, таких как активные среды лазеров, возможно $\alpha < 0$, в этом случае величина $\gamma \equiv -\alpha$ называется коэффициентом усиления и среда не поглощает, а усиливает свет.

Поскольку параметр β есть скорость изменения фазы вдоль z, он представляет собой постоянную распространения волны. Следовательно, среда имеет эффективный показатель преломления n, определяемый как

$$\beta = nk_0, \tag{5.87}$$

и волна распространяется с фазовой скоростью $c = c_0/n$.

Поставляя (5.87) в (5.86), получаем связь между показателем преломления n и коэффициентом поглощения α, с одной стороны, и действительной и мнимой частями восприимчивости χ' и χ'' — с другой:

$$\boxed{n - j\frac{1}{2}\frac{\alpha}{k_0} = \sqrt{\frac{\varepsilon}{\varepsilon_0}} = \sqrt{1 + \chi' + \chi''}.} \tag{5.88}$$

Коэффициент поглощения и показатель преломления

Заметим, что квадратный корень в правой части (5.88) дает два комплексных числа противоположных знаков (с разностью фаз π). Знак выбирается таким, чтобы если χ'' отрицательно, т. е. среда поглощающая, то α было положительным, т. е. волна затухала. Если $1 + \chi'$ положительно, то комплексное число $1 + \chi' + j\chi''$ лежит в четвертом квадранте, а квадратный корень из него — в четвертом или втором квадранте. Выбирая корень в четвертом квадранте, мы гарантируем положительность α, и тогда n тоже положительно. Аналогично, если $1 + \chi'$ отрицательно, то $1 + \chi' + j\chi''$ лежит в третьем квадранте, а корень из него выбирается в четвертом квадранте, чтобы и α, и n были положительными. Импеданс, соответствующий комплексной восприимчивости χ, также комплексный и дается выражением

$$\boxed{\eta = \sqrt{\frac{\mu_0}{\varepsilon}} = \frac{\eta_0}{\sqrt{1 + \chi}}} \tag{5.89}$$

Импеданс

В общем случае χ, k, ε и η — комплексные, а α, β и n — действительные.

Слабо поглощающие среды

В слабо поглощающей среде $\chi'' \ll 1 + \chi'$, так что

$$\sqrt{1 + \chi' + \chi''} = \sqrt{1 + \chi'}\sqrt{1 + j\delta} \approx \sqrt{1 + \chi'}\left(1 + \frac{j\delta}{2}\right),$$

где

$$\delta = \frac{\chi''}{1 + \chi'}.$$

Из (5.88) следует, что

$$\boxed{n \approx \sqrt{1+\chi'};} \tag{5.90}$$

$$\boxed{\alpha \approx -\frac{k_0}{n}\chi''.} \tag{5.91}$$

Слабо поглощающая среда

При этих условиях показатель преломления определяется действительной частью восприимчивости, а коэффициент поглощения — мнимой. В поглощающей среде χ'' отрицательна, так что коэффициент α положителен, а в усиливающей среде — наоборот, χ'' положительна, так что коэффициент α отрицателен.

Упражнение 5.1

Разбавленная поглощающая среда

Прозрачная среда с показателем преломления n_0 служит базовым веществом для разбавленной суспензии примеси, характеризуемой восприимчивостью $\chi = \chi' + \chi''$, где $\chi' \ll 1$; $\chi'' \ll 1$. Определите полную восприимчивость среды и покажите, что показатель преломления и коэффициент поглощения приближенно выражаются как

$$n \approx n_0 + \frac{\chi'}{2n_0}; \tag{5.92}$$

$$\alpha \approx -\frac{k_0\chi''}{n_0}. \tag{5.93}$$

Сильно поглощающие среды

В сильно поглощающей среде $|\chi''| \gg |1+\chi'|$, что по (5.88) дает

$$n - \frac{j\alpha}{2k_0} \approx \sqrt{j\chi''} = \sqrt{-j}\sqrt{(-\chi'')} = \pm\frac{1}{\sqrt{2}}(1-j)\sqrt{(-\chi'')},$$

откуда

$$\boxed{n \approx \sqrt{\frac{(-\chi'')}{2}};} \tag{5.94}$$

$$\boxed{\alpha \approx -2k_0\sqrt{\frac{(-\chi'')}{2}}.} \tag{5.95}$$

Сильно поглощающая среда

Поскольку χ'' отрицательна для поглощающей среды, знак «плюс» у квадратного корня был выбран для гарантии того, чтобы α и n были положительны.

5.5.2. Дисперсия

Диспергирующие среды характеризуются зависящими от частоты (и, следовательно, длины волны) восприимчивостью $\chi(\nu)$, диэлектрической проницаемостью $\varepsilon(\nu)$, показателем преломления $n(\nu)$ и скоростью света $c_0/n(\nu)$. Поскольку угол преломления в законе Снелла зависит от показателя преломления, который зависит от длины волны, оптические элементы, изготовленные из диспергирующих материалов, такие как призмы и линзы, отклоняют лучи с различной длиной волны на разные углы. Это является причиной способности преломляющих поверхностей разлагать свет по длинам волн, а также зависимости фокусного расстояния линз от длины волны (и возникающих вследствие этого хроматических аберраций в системах формирования изображений). Полихроматический свет преломляется в целый диапазон направлений. Эти эффекты проиллюстрированы на рис. 5.15 (см. цв. вклейку).

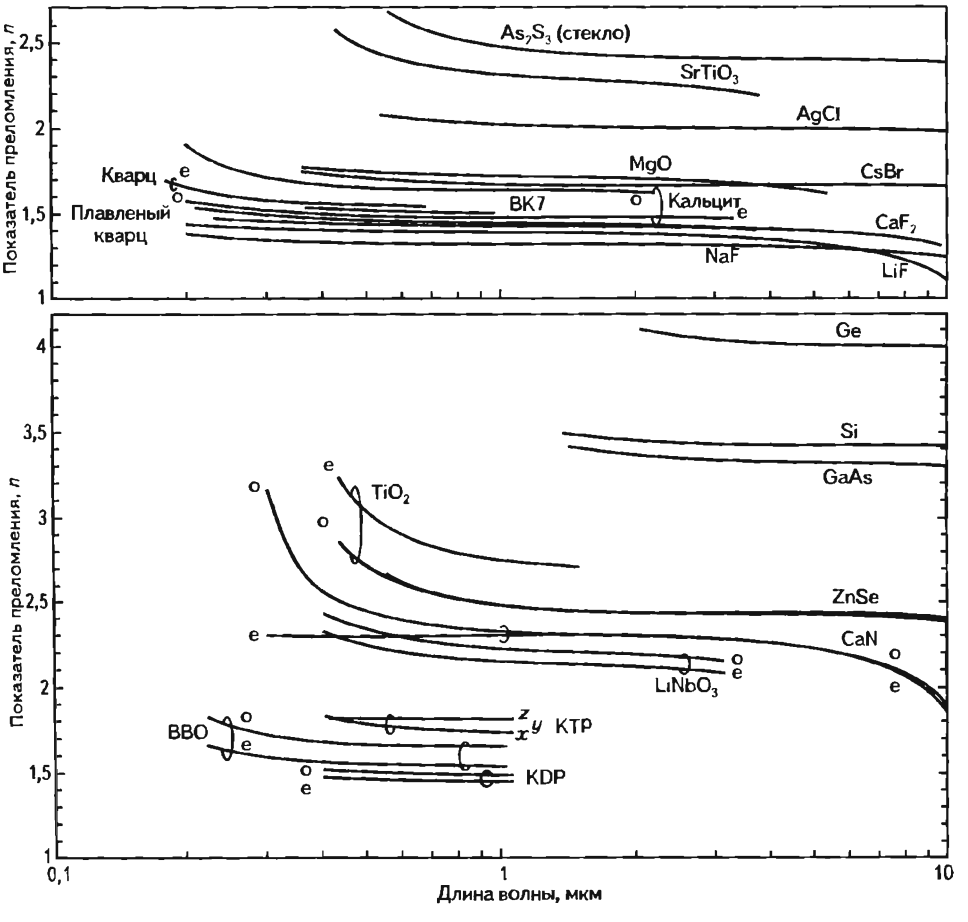


Рис. 5.17. Зависимость показателя преломления от длины волны некоторых оптических материалов, включая стекла, кристаллы и полупроводники

Кроме того, из-за частотной зависимости скорости света в диспергирующей среде каждая частотная компонента короткого импульса претерпевает различную временную задержку. Если длина распространения импульса в среде значительна, как, например, часто бывает в случае оптических волокон, короткий импульс на входе существенно расплывается по времени, так что его длительность на выходе увеличивается, как показано на рис. 5.16 (см. цв. вклейку).

Зависимость показателя преломления от длины волны для некоторых распространенных оптических материалов показана на рис. 5.17.

Меры дисперсии

Дисперсия материалов может быть количественно оценена множеством различных способов. Для стеклянных оптических элементов и широкополосного света, покрывающего видимый диапазон (белого света), обычно используется число Аббе

$$\mathbb{V} = \frac{n_d - 1}{n_F - n_C},$$

где n_F, n_d и n_C — показатели преломления стекла на трех стандартных длинах волны: 486,1 нм (синий), 587,6 нм (желтый) и 656,3 нм (красный) соответственно. Для стекла сорта флинт $\mathbb{V} \approx 38$, а для плавленого кварца $\mathbb{V} \approx 68$.

С другой стороны, если интерес представляет дисперсия в окрестности некоторой длины волны λ_0, в качестве меры часто используется величина производной $\mathrm{d}n/\mathrm{d}\lambda_0$ в точке λ_0. Такая мера подходит, например, для призм, у которых угол отклонения луча θ_d является функцией n [см. (1.9)]. В этом случае угловая дисперсия

$$\frac{\mathrm{d}\theta_d}{\mathrm{d}\lambda_0} = \frac{\mathrm{d}\theta_d}{\mathrm{d}n}\frac{\mathrm{d}n}{\mathrm{d}\lambda_0}$$

является произведением материальной дисперсии $\mathrm{d}n/\mathrm{d}\lambda_0$ на множитель $\mathrm{d}\theta_d/\mathrm{d}n$, зависящий от геометрии призмы и показателя преломления материала, из которого она сделана.

Влияние дисперсии материала на распространение коротких импульсов света определяется не только показателем преломления n и его первой производной $\mathrm{d}n/\mathrm{d}\lambda_0$, но и второй производной $\mathrm{d}^2n/\mathrm{d}\lambda_0^2$, как будет выяснено в разд. 5.6 и 22.3.

Поглощение и дисперсия: соотношения Крамерса—Кронига

Поглощение и дисперсия неразрывно связаны друг с другом. В самом деле, диспергирующий материал, т. е. материал, показатель преломления которого зависит от длины волны, *должен* быть поглощающим, причем его коэффициент поглощения также должен зависеть от длины волны. Связь между коэффициентом поглощения и показателем преломления вытекает из со-

отношений Крамерса—Кронига, связывающих действительную и мнимую части восприимчивости среды $\chi'(\nu)$ и $\chi''(\nu)$:

$$\chi'(\nu) = \frac{2}{\pi}\int_0^\infty \frac{s\chi''(s)}{s^2 - \nu^2}\,ds; \tag{5.96}$$

$$\chi''(\nu) = \frac{2}{\pi}\int_0^\infty \frac{\nu\chi'(s)}{\nu^2 - s^2}\,ds. \tag{5.97}$$

Соотношения Крамерса—Кронига

Если для всех ν известна действительная или мнимая часть восприимчивости $\chi(\nu)$, то с помощью этих мощных формул можно определить ее мнимую или действительную часть, соответственно, для всех ν. Соотношения Крамерса—Кронига для $\chi'(\nu)$ и $\chi''(\nu)$ превращаются в связь между коэффициентом поглощения $\alpha(\nu)$ и показателем преломления $n(\nu)$ с помощью формул (5.88), выражающих α и n через χ' и χ''

Как следует из теории линейных систем, соотношения Крамерса—Кронига представляют собой специальный случай пары преобразований Гильберта [см. разд. Б.1 приложения Б]. Они применимы для всех линейных инвариантных относительно сдвига причинных систем с действительной функцией отклика на импульсное воздействие. В рассматриваемой линейной системе плотность поляризации $\mathcal{P}(t)$ является откликом на электрическое поле $\mathcal{E}(t)$, согласно (5.38). Поскольку $\mathcal{P}(t)$ и $\mathcal{E}(t)$ действительны, действительной является и функция отклика на импульсное воздействие $\varepsilon_0 x(t)$. Вследствие этого ее Фурье-образ — передаточная функция $\varepsilon_0\chi(\nu)$ — обладает эрмитовой симметрией $\chi(-\nu) = \chi^*(\nu)$ [см. разд. А.1 приложения А]. Таким образом, данная система обладает всеми свойствами, необходимыми для применимости соотношений Крамерса—Кронига. Следовательно, действительная и мнимая части передаточной функции $\varepsilon_0\chi(\nu)$ связаны соотношениями (Б.6) и (Б.7) и, в частности, (5.96) и (5.97).

5.5.3. Резонансная среда

Построим простую классическую микроскопическую теорию, приводящую к комплексной восприимчивости среды и дающую понимание того, почему оптическая среда обладает коэффициентом поглощения и показателем преломления, зависящими от частоты света. Этот подход известен как **модель осциллятора Лоренца**. Более глубокое обсуждение взаимодействия излучения с веществом будет проведено в гл. 13.

Представим диэлектрическую среду как ансамблю атомов, в котором динамическая связь между плотностью поляризации среды $\mathcal{P}(t)$ и электрическим полем $\mathcal{E}(t)$, которое предполагается линейно поляризованным, описы-

вается линейным обыкновенным дифференциальным уравнением второго порядка

$$\boxed{\frac{d^2\mathcal{P}}{dt^2} + \sigma\frac{d\mathcal{P}}{dt} + \omega_0^2\mathcal{P} = \omega_0^2\varepsilon_0\chi_0\mathcal{E},} \qquad (5.98)$$

Резонансная диэлектрическая среда

где σ, ω_0 и χ_0 — постоянные.

Уравнение такого вида возникает при феноменологическом описании движения связанного заряда в атоме с помощью модели классического гармонического осциллятора, у которого смещение $x(t)$ и приложенная извне сила $\mathcal{F}(t)$ связаны уравнением

$$\frac{d^2x}{dt^2} + \sigma\frac{dx}{dt} + \omega_0^2 x = \frac{\mathcal{F}}{m}. \qquad (5.99)$$

Здесь m — масса связанной заряженной частицы; $\omega_0 = \sqrt{k/m}$ — резонансная угловая частота; k — постоянная упругости возвращающей силы; σ — коэффициент затухания.

Если дипольный момент, связанный с каждым отдельным атомом, есть $p = -ex$, то плотность поляризации среды в целом связана со смещением зарядов формулой $\mathcal{P} = Np = -Nex$, где $-e$ — заряд электрона; N — число атомов в единице объема среды. Электрическое поле и сила связаны соотношением $\mathcal{E} = \mathcal{F}/(-e)$. Таким образом, величины $\mathcal{P}$ и $\mathcal{E}$ пропорциональны x и $\mathcal{F}$, соответственно, и сравнение (5.98) с (5.99) дает

$$\chi_0 = \frac{Ne^2}{\varepsilon_0 m\omega_0^2}. \qquad (5.100)$$

Таким образом, можно считать, что приложенное электрическое поле индуцирует электрический дипольный момент у каждого атома, как изображено на рис. 5.18, и, следовательно, зависящую от времени плотность поляризации среды в целом.

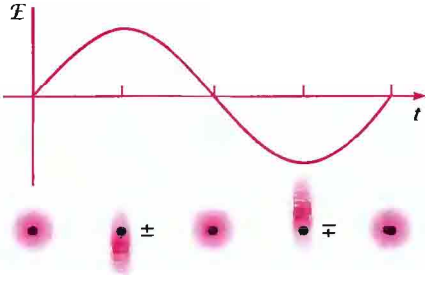


Рис. 5.18. Зависящее от времени электрическое поле $\mathcal{E}$, приложенное к атому, рассматриваемому как осциллятор Лоренца, индуцирует зависящий от времени дипольный момент p, дающий вклад в полную плотность поляризации $\mathcal{P}$

Среда полностью характеризуется функцией отклика на импульсное воздействие $\varepsilon_0 x(t)$, которая представляет собой экспоненциально затухающие гармо-

нические колебания, или, что равносильно, передаточной функцией $\varepsilon_0\chi(\nu)$, для получения которой нужно решить (5.98) для каждой частотной гармоники поля следующим образом. Подставим $\mathcal{E}(t) = \mathrm{Re}\{E\exp(j\omega t)\}$ и $\mathcal{P}(t) = \mathrm{Re}\{P\exp(j\omega t)\}$ в (5.98), тогда

$$\left(-\omega^2 + j\sigma\omega + \omega_0^2\right)P = \omega_0^2\varepsilon_0\chi_0 E, \tag{5.101}$$

откуда

$$P = \varepsilon_0 \frac{\chi_0\omega_0^2}{\omega_0^2 - \omega^2 + j\sigma\omega} E.$$

Придавая этому соотношению вид $P = \varepsilon_0\chi(\nu)E$ и подставляя $\omega = 2\pi\nu$, получаем выражение для восприимчивости, зависящей от частоты:

$$\boxed{\chi(\nu) = \chi_0 \frac{\nu_0^2}{\nu_0^2 - \nu^2 + j\nu\Delta\nu},} \tag{5.102}$$

Восприимчивость (резонансная среда)

где $\nu_0 = \omega_0/2\pi$ — резонансная частота, а $\Delta\nu = \sigma/2\pi$.

Действительная и мнимая части $\chi(\nu)$, обозначаемые $\chi'(\nu)$ и $\chi''(\nu)$, имеют тогда вид

$$\chi'(\nu) = \chi_0 \frac{\nu_0^2\left(\nu_0^2 - \nu^2\right)}{\left(\nu_0^2 - \nu^2\right)^2 + \left(\nu\Delta\nu\right)^2}; \tag{5.103}$$

$$\chi''(\nu) = -\chi_0 \frac{\nu_0^2\nu\Delta\nu}{\left(\nu_0^2 - \nu^2\right)^2 + \left(\nu\Delta\nu\right)^2}. \tag{5.104}$$

График этих выражений приведен на рис. 5.19.

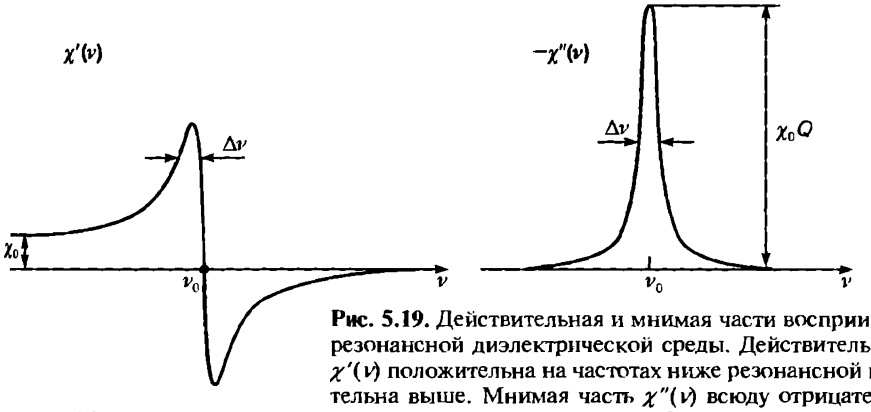


Рис. 5.19. Действительная и мнимая части восприимчивости резонансной диэлектрической среды. Действительная часть $\chi'(\nu)$ положительна на частотах ниже резонансной и отрицательна выше. Мнимая часть $\chi''(\nu)$ всюду отрицательна, так что $-\chi''(\nu)$ всюду положительна и имеет пиковое значение $\chi_0 Q$ при $\nu = \nu_0$, где $Q = \nu_0/\Delta\nu$. График построен для $Q = 10$

На частотах намного ниже резонансной ($\nu \ll \nu_0$) $\chi'(\nu) \approx \chi_0$ и $\chi''(\nu) \approx 0$, так что низкочастотная восприимчивость — это просто χ_0. На очень высоких частотах ($\nu \gg \nu_0$) $\chi'(\nu) \approx \chi''(\nu) \approx 0$, так что среда ведет себя как свободное пространство. При точном резонансе ($\nu = \nu_0$) $\chi'(\nu) = 0$, а $-\chi''(\nu_0)$ достигает своего максимального значения $\chi_0 Q$, где $Q = \nu_0/\Delta\nu$. Резонансная частота ν_0 обычно намного больше, чем $\Delta\nu$, так что $Q \gg 1$. Поэтому величина резонансного пика $-\chi''(\nu)$ при $\chi_0 Q$ намного больше низкочастотного значения восприимчивости $\chi'(\nu)$ при χ_0. Максимальное и минимальное значения

$$\chi'(\nu) = \pm\frac{\chi_0 Q}{2 \pm 1/Q}$$

достигаются на частотах

$$\nu_0\sqrt{1 \mp \frac{1}{Q}}$$

соответственно. Для больших Q перепад между максимумом и минимумом $\chi'(\nu)$ составляет примерно $\chi_0 Q/2$, т. е. половину пикового значения χ''. Знаки χ' и χ'' определяют фазу χ, которая задает угол между векторами P и E на комплексной плоскости.

Поведение $\chi(\nu)$ в окрестности резонанса ($\nu \sim \nu_0$) часто представляет особый интерес. В этой области можно использовать приближение

$$\nu_0^2 - \nu^2 = (\nu_0 - \nu)(\nu_0 + \nu) \approx 2\nu_0(\nu_0 - \nu)$$

в действительной части знаменателя (5.102) и заменить ν на ν_0 в его мнимой части, что дает

$$\chi(\nu \sim \nu_0) \approx \chi_0 \frac{\nu_0/2}{\nu_0 - \nu + j\Delta\nu/2}, \tag{5.105}$$

откуда

$$\chi''(\nu) \approx -\chi_0 \frac{\nu_0 \Delta\nu}{4} \frac{1}{(\nu_0 - \nu)^2 + (\Delta\nu/2)^2}; \tag{5.106}$$

$$\chi'(\nu) \approx 2\frac{\nu_0 - \nu}{\Delta\nu}\chi''(\nu). \tag{5.107}$$

Восприимчивость (вблизи резонанса)

Функция $\chi''(\nu)$ в (5.106), известная как функция Лоренца, спадает до половины максимума, когда $|\nu_0 - \nu| = \Delta\nu/2$. Таким образом, параметр $\Delta\nu$ представляет собой полную ширину пика $\chi''(\nu)$ на уровне половины максимума.

Поведение $\chi(\nu)$ вдали от резонанса также представляет интерес. В предельном случае $|\nu_0 - \nu| \gg \Delta\nu$ восприимчивость (5.102) приближенно действительна

$$\chi(\nu) \approx \chi_0 \frac{\nu_0^2}{\nu_0^2 - \nu^2}, \tag{5.108}$$

Восприимчивость (вдали от резонанса)

так что среда имеет пренебрежимо малое поглощение.

Коэффициент поглощения и показатель преломления резонансной среды можно получить, подставляя выражения для $\chi'(\nu)$ и $\chi''(\nu)$, например, (5.106) и (5.107) в (5.85). Каждый из этих параметров в общем случае зависит и от $\chi'(\nu)$, и от $\chi''(\nu)$. Однако, в том случае, когда резонансные атомы погружены в недиспергирующее базовое вещество с показателем преломления n_0 и их концентрация достаточно мала, так что $\chi'(\nu)$ и $\chi''(\nu)$ много меньше единицы, указанная зависимость сильно упрощается: показатель преломления зависит только от $\chi'(\nu)$, а коэффициент поглощения — только от $\chi''(\nu)$. Используя результаты упражнения 5.1, можно показать, что

$$\alpha(\nu) \approx -\left(\frac{2\pi\nu}{n_0 c_0}\right)\chi''(\nu); \tag{5.109}$$

$$n(\nu) \approx n_0 + \frac{\chi'(\nu)}{2n_0}. \tag{5.110}$$

Зависимость этих параметров от ν показана на рис. 5.20.

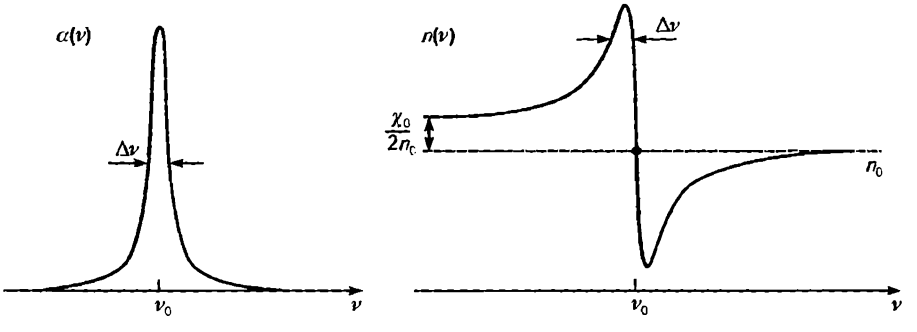


Рис. 5.20. Коэффициент поглощения $\alpha(\nu)$ и показатель преломления $n(\nu)$ диэлектрической среды с показателем преломления n_0, содержащей разбавленную до малой концентрации примесь атомов с резонансной частотой ν_0

Среды, имеющие много резонансов

Типичные диэлектрические среды имеют много резонансов, соответствующих различным колебаниям решетки и электронов. Полная восприимчивость является суперпозицией вкладов этих резонансов. В то время как

мнимая часть восприимчивости отлична от нуля в небольшой окрестности каждого резонанса, действительная часть вносит вклад на всех частотах выше и *ниже* резонанса, как показано на рис. 5.19. Это проявляется в частотной зависимости коэффициента поглощения и показателя преломления, показанной на рис. 5.21. Поглощение и дисперсия наиболее сильны вблизи резонансных частот. Вдали от резонансов показатель преломления постоянен и среда приближенно является недиспергирующей и непоглощающей. Каждый резонанс, однако, вносит постоянный вклад в показатель преломления на всех частотах ниже резонансной частоты.

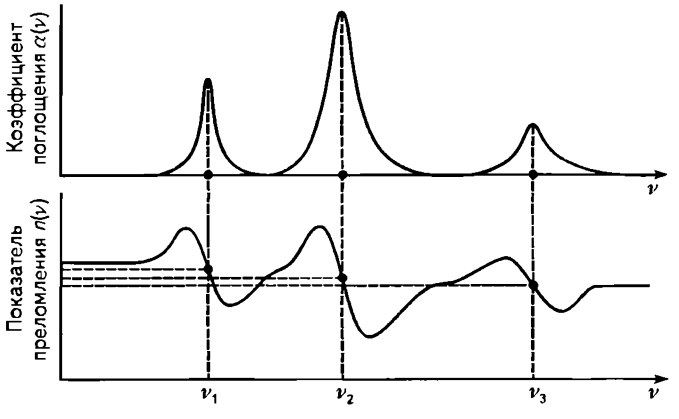


Рис. 5.21. Частотная зависимость коэффициента поглощения $\alpha(\nu)$ и показателя преломления $n(\nu)$ для среды с тремя резонансами

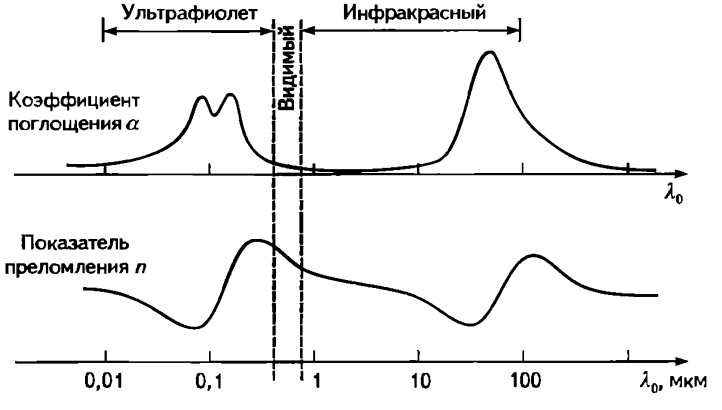


Рис. 5.22. Типичная зависимость коэффициента поглощения и показателя преломления от длины волны для диэлектрической среды с резонансным поглощением в ультрафиолетовой и инфракрасной областях, сопутствующая низкому поглощению в видимой области. Обратите внимание на то, что по оси абсцисс отложена длина волны, а не частота

Другие сложные процессы также могут вносить свой вклад в коэффициент поглощения и показатель преломления материала, так что могут возникать различные виды их частотной зависимости. На рис. 5.22 продемонстрирована за-

висимость коэффициента поглощения и показателя преломления от длины волны для диэлектрического материала, существенно прозрачного на длинах волн видимого света. Иллюстрация показывает убывание показателя преломления с ростом длины волны в видимом диапазоне из-за близкого резонанса в ультрафиолетовой области. За счет этого материал обладает более сильной дисперсией в коротковолновой части видимого диапазона, где скорость убывания показателя преломления наивысшая. Эта зависимость согласуется с показанной на рис. 5.14 и 5.17 для различных реальных диэлектрических материалов.

Таблица 5.1. **Формулы Зельмейера, описывающие зависимость показателя преломления некоторых материалов от длины волны при комнатной температуре. Через n_o и $n_н$ обозначены обыкновенный и необыкновенный показатели преломления, соответственно, для анизотропных сред (см. разд. 6.3). В правом столбце показаны области длин волн, в которых справедливы приведенные результаты**

Материал	Формула Зельмейера (длина волны λ, мкм)	Область длин волн, мкм
Плавленый кварц	$n^2 = 1 + \frac{0,6962\lambda^2}{\lambda^2 - (0,06840)^2} + \frac{0,4079\lambda^2}{\lambda^2 - (0,1162)^2} + \frac{0,8975\lambda^2}{\lambda^2 - (9,8962)^2}$	0,21—3,71
Si	$n^2 = 1 + \frac{10,6684\lambda^2}{\lambda^2 - (0,3015)^2} + \frac{0,0030\lambda^2}{\lambda^2 - (1,1347)^2} + \frac{1,5413\lambda^2}{\lambda^2 - (1104,0)^2}$	1,36 — 11,00
GaAs	$n^2 = 3,5 + \frac{7,4969\lambda^2}{\lambda^2 - (0,4082)^2} + \frac{1,9347\lambda^2}{\lambda^2 - (37,17)^2}$	1,40 — 11,00
BBO	$n_o^2 = 2,7359 + \frac{0,01878}{\lambda^2 - 0,01822} - 0,01354\lambda^2$ $n_н^2 = 2,3753 + \frac{0,01224}{\lambda^2 - 0,01667} - 0,01516\lambda^2$	0,22 — 1,06
KDP	$n_o^2 = 1 + \frac{1,2566\lambda^2}{\lambda^2 - (0,09191)^2} + \frac{33,8991\lambda^2}{\lambda^2 - (33,3752)^2}$ $n_н^2 = 1 + \frac{1,1311\lambda^2}{\lambda^2 - (0,09026)^2} + \frac{5,7568\lambda^2}{\lambda^2 - (28,4913)^2}$	0,40 — 1,06
$LiNbO_3$	$n_o^2 = 2,3920 + \frac{2,5112\lambda^2}{\lambda^2 - (0,217)^2} + \frac{7,1333\lambda^2}{\lambda^2 - (16,502)^2}$ $n_н^2 = 2,3247 + \frac{2,2565\lambda^2}{\lambda^2 - (0,210)^2} + \frac{14,503\lambda^2}{\lambda^2 - (25,915)^2}$	0,40 — 3,10

Формула Зельмейера

В среде с множеством резонансов, нумеруемых индексом $i = 1, 2, ...$, восприимчивость вдали от всех резонансных частот приближенно определяется суммой членов вида (5.108). С использованием связи (5.28) между показателем преломления и действительной восприимчивостью, $n^2 = 1 + \chi$, зависимость n от частоты и длины волны приобретает вид **формулы Зельмейера**

$$\boxed{n^2 \approx 1 + \sum_i \chi_{0i} \frac{\nu_i^2}{\nu_i^2 - \nu^2} = 1 + \sum_i \chi_{0i} \frac{\lambda^2}{\lambda^2 - \lambda_i^2}.} \qquad (5.111)$$

Формула Зельмейера

В табл. 5.1 приведены результаты подгонки формулы Зельмейера для некоторых материалов к экспериментальным данным по методу наименьших квадратов. Формула Зельмейера обеспечивает хорошее описание показателя преломления для большинства оптически прозрачных материалов.

На длинах волн, для которых $\lambda \ll \lambda_i$, i-й член становится приблизительно пропорциональным λ^2, а для $\lambda \gg \lambda_i$ — почти постоянным. Например, дисперсия плавленого кварца, рассмотренная в примере 5.2, хорошо описывается тремя резонансами. Для некоторых материалов формулу Зельмейера удобно аппроксимировать степенным рядом.

5.5.4. Оптика проводящих сред

Проводящие материалы, такие как металлы, полупроводники, примесные диэлектрики и ионизированные газы, содержат свободные электрические заряды и имеют связанную с ними плотность электрического тока $\boldsymbol{J} = \sigma\eta_0 d_p$. В таких средах первое из уравнений Максвелла (5.7) должно быть модифицировано путем добавления плотности тока проводимости к плотности тока смещения $\partial\mathcal{D}/\partial t$:

$$\nabla \times \mathcal{H} = \frac{\partial \mathcal{D}}{\partial t} + \boldsymbol{J}. \qquad (5.112)$$

Другие три уравнения Максвелла остаются без изменения. Для монохроматической волны с угловой частотой ω это уравнение приобретает вид

$$\nabla \times \boldsymbol{H} = j\omega \boldsymbol{D} + \boldsymbol{J}, \qquad (5.113)$$

который представляет собой модифицированную версию уравнения (5.43).

Для среды с линейными диэлектрическим свойствами $\boldsymbol{D} = \varepsilon \boldsymbol{E} = \varepsilon_0(1 + \chi)\boldsymbol{E}$. Для среды с линейными проводящими свойствами и **удельной электропроводностью** σ плотность электрического тока пропорциональна электрическому полю,

$$\boldsymbol{J} = \sigma \boldsymbol{E}, \qquad (5.114)$$

Это соотношение является одной из форм закона Ома [см. (18.13) и (18.14)]. Тогда правая часть (5.113) приобретает вид

$$(j\omega\varepsilon + \sigma)\boldsymbol{E} = j\omega\left(\varepsilon + \frac{\sigma}{j\omega}\right)\boldsymbol{E},$$

так что

$$\nabla \times \boldsymbol{H} = j\omega\varepsilon_e \boldsymbol{E}, \tag{5.115}$$

где эффективная диэлектрическая проницаемость ε_e определяется как

$$\boxed{\varepsilon_e = \varepsilon + \frac{\sigma}{j\omega}.} \tag{5.116}$$

Эффективная диэлектрическая проницаемость

Эффективная диэлектрическая проницаемость ε_e — комплексный частотно-зависимый параметр, описывающий комбинацию диэлектрических и проводящих свойств среды. Поскольку второе слагаемое в (5.116) обратно пропорционально частоте, вклад проводимости уменьшается с ростом частоты.

Поскольку (5.115) имеет такой же вид, как и аналогичное уравнение для диэлектрической среды, законы распространения, полученные в разд. 5.3—5.5, остаются справедливыми и при наличии проводимости. Так, волновое число в (5.85) и (5.86) становится равным

$$k = \beta - j\frac{1}{2}\alpha = \omega\sqrt{\varepsilon_e \mu_0},$$

а импеданс в (5.89)

$$\eta = \sqrt{\frac{\mu_0}{\varepsilon_e}}.$$

Показатель преломления n и коэффициент поглощения α в (5.88) определяются из комплексного уравнения

$$n - \frac{j\alpha}{2k_0} = \sqrt{\frac{\varepsilon_e}{\varepsilon_0}}.$$

Когда $\sigma/\omega \gg \varepsilon$, эффекты проводимости преобладают и $\varepsilon_e \approx \sigma/j\omega$. Тогда

$$n - \frac{j\alpha}{2k_0} \approx \sqrt{\frac{\sigma}{j\omega\varepsilon_0}}; \quad \eta \approx \sqrt{\frac{j\omega\mu_0}{\sigma}},$$

откуда

$$\boxed{n \approx \sqrt{\frac{\sigma}{2\omega\varepsilon_0}};} \tag{5.117}$$

$$\boxed{\alpha \approx \sqrt{2\omega\mu_0\sigma};} \tag{5.118}$$

$$\boxed{\eta \approx (1+j)\sqrt{\frac{\omega\mu_0}{2\sigma}},} \tag{5.119}$$

Проводящая среда

где использовано соотношение

$$k_0 = \frac{\omega}{c_0} = \omega\sqrt{\varepsilon_0\mu_0}.$$

Оптическая интенсивность ослабляется в e^{-1} раз на глубине

$$d_p = \frac{1}{\alpha} = \frac{1}{\sqrt{2\omega\mu_0\sigma}},$$

которая называется **глубиной проникновения**, или **толщиной скин-слоя**[5]. И d_p и n меняются как $1/\sqrt{\omega}$.

У металлов σ очень велико, следовательно, велико и α. Это означает, что оптические волны значительно ослабляются как только они пересекают поверхность материала. Однако импеданс η очень мал, поэтому эти материалы обладают высокой рефракцией (см. упражнение 6.6)

Пример 5.1

Глубина проникновения и показатель преломления меди. Медь имеет удельную электропроводность $\sigma = 0{,}58 \cdot 10^8$ (Ом · м)$^{-1}$, так что глубина проникновения ограничена $d_p = 1{,}9$ нм на длине волны $\lambda_0 = 1$ мкм. В соответствии с (5.117) и (5.118) показатель преломления связан с глубиной проникновения как $n = \sigma\eta_0 d_p$, что в рассматриваемом случае дает $n = 41{,}6$.

Модель Друде

Поскольку соотношение между J и $\mathcal{E}$ динамическое, удельная электропроводность σ должна зависеть от частоты поля и иметь конечную спектральную ширину. В модели Друде электроны проводимости рассматриваются как независимые частицы идеального газа, движущиеся свободно между актами столкновения, что приводит к выражению для удельной электропроводности, зависящей от частоты

$$\sigma = \frac{\sigma_0}{1 + j\omega\tau}, \tag{5.120}$$

где σ_0 — низкочастотная удельная электропроводность; τ — время релаксации. Тогда из (5.116) следует, что

$$\varepsilon_e = \varepsilon + \frac{\sigma_0}{j\omega(1 + j\omega\tau)}. \tag{5.121}$$

Для $\omega \gg 1/\tau$ выражение (5.121) имеет вид

$$\varepsilon_e \approx \varepsilon - \frac{\sigma_0}{\omega^2 t}.$$

[5] Глубину проникновения иногда определяют как расстояние, на котором поле (а не интенсивность) спадает в e^{-1} раз.

Очевидно, что в этом случае проводимость снижает действительную часть диэлектрической проницаемости, давая отрицательный вклад в диэлектрическую постоянную, обратно пропорциональный квадрату частоты. В частности, если среда имеет диэлектрические свойства свободного пространства с $\varepsilon = \varepsilon_0$, то эффективную диэлектрическую проницаемость можно записать в виде

$$\varepsilon_e = \varepsilon_0\left(1 - \frac{\omega_p^2}{\omega^2}\right), \tag{5.122}$$

где ω_p называется **плазменной частотой,**

$$\omega_p = \sqrt{\frac{\sigma_0}{\varepsilon_0 \tau}}.$$

Простая классическая микроскопическая теория дает обоснование результатов модели Друде. Рассмотрение аналогично модели Лоренца, за исключением того, что представляющие интерес электроны в проводящей среде свободные, а не связанные. В силу этого как возвращающая сила, так и затухание отсутствуют ($k = 0$ и $\sigma = 0$). При этих условиях уравнение движения Лоренца (5.99) принимает вид

$$m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = -e\mathcal{E},$$

так что соответствующая плотность поляризации $\mathcal{P} = -Nex$ удовлетворяет простому уравнению

$$\frac{\mathrm{d}^2 \mathcal{P}}{\mathrm{d}t^2} = \frac{Ne^2}{m}\mathcal{E},$$

где N — плотность числа электронов в среде. Для поля, осциллирующего с угловой частотой ω, это приводит к уравнению

$$-\omega^2 P = \frac{Ne^2}{m}\mathcal{E},$$

эквивалентному уменьшению диэлектрической постоянной из-за проводимости на величину

$$\frac{P}{\varepsilon_0 E} = -\frac{Ne^2/\varepsilon_0 m}{\omega^2}.$$

Это согласуется с (5.122) при условии, что плазменная частота равна

$$\omega_p = \sqrt{\frac{Ne^2}{\varepsilon_0 m}}. \tag{5.123}$$

Комбинируя (5.123) с результатом Друде

$$\omega_p = \sqrt{\frac{\sigma_0}{\varepsilon_0 \tau}},$$

получаем

$$\sigma_0 = \frac{Ne^2\tau}{m},$$

что согласуется с (18.13).

Из (5.122) очевидно, что распространение волны в среде, описываемой моделью Друде, подразумевает четкое различие в поведении ниже и выше плазменной частоты, как показано на рис. 5.23.

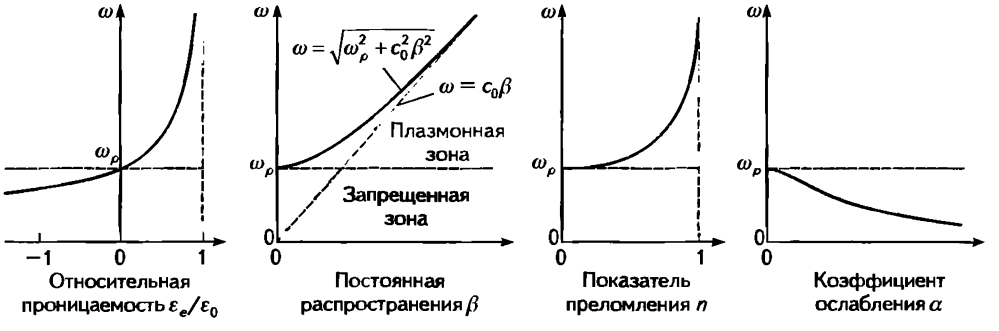


Рис. 5.23. Частотная зависимость относительной проницаемости $\varepsilon_e/\varepsilon_0$, постоянной распространения β, показателя преломления n и коэффициента ослабления α среды, описываемой моделью Друде

• На частотах ниже плазменной ($\omega < \omega_p$) эффективная проницаемость отрицательна, так что $k = \omega\sqrt{\varepsilon_e \mu_0}$ мнимое, что соответствует затуханию без распространения. Эта спектральная область, следовательно, может быть названа **запрещенной зоной**. Коэффициент ослабления

$$\alpha = 2k_0\left(\frac{\omega_p^2}{\omega^2} - 1\right)^{1/2}$$

монотонно убывает с ростом частоты и обращается в нуль на плазменной частоте. Отрицательная проницаемость соответствует мнимому импедансу. Следовательно, на границе между обычной средой с действительным импедансом и проводящей средой с мнимым импедансом свет полностью отражается (см. разд. 6.2), так что граница служит идеальным зеркалом.

• На частотах выше плазменной ($\omega > \omega_p$) эффективная проницаемость положительна и действительна, так что проводящая среда ведет себя как диэлект-

рик без потерь, хотя и с уникальными дисперсионными характеристиками. Постоянная распространения становится равной

$$\beta = \frac{\left(\omega^2 - \omega_p^2\right)^{1/2}}{c_0},$$

а показатель преломления

$$n = \left(1 - \frac{\omega_p^2}{\omega^2}\right)^{1/2}$$

и принимает очень малые значения вблизи плазменной частоты. Эта спектральная область называется **плазмонной зоной**.

• На плазменной частоте ($\omega = \omega_p$) постоянная распространения обращается в нуль, поэтому волна не распространяется в проводящей среде. Однако плотность электрического тока осциллирует и свободные электроны испытывают продольные колебания; квантовые частицы, сопоставляемые этим колебаниям, называются **плазмонами** (аналогично фотонам, сопоставляемым колебаниям оптического поля, см. гл. 12).

В большинстве металлов плазменная частота лежит в ультрафиолетовом диапазоне, поэтому в видимой области они отражают свет и имеют блестящую поверхность. Однако у некоторых металлов, например меди, плазменная частота находится в видимом диапазоне, так что отражается только часть белого света, и металл обладает выраженным цветом. У легированных полупроводников плазменная частота обычно находится в инфракрасной области.

5.6. РАСПРОСТРАНЕНИЕ ИМПУЛЬСОВ В СРЕДАХ С ДИСПЕРСИЕЙ

Распространение импульсов света в диспергирующих средах важно для многих приложений, включающих волоконно-оптические системы связи, что будет подробно рассмотрено в гл. 9 и 24. Как отмечалось ранее, диспергирующая среда характеризуется показателем преломления, зависящим от частоты, так что монохроматические волны разных частот распространяются в такой среде с различными скоростями и различным затуханием. Поскольку импульс света представляет собой суперпозицию монохроматических волн, каждая из которых изменяется по-разному, при прохождении среды он задерживается и уширяется; в общем случае меняется и его форма. В данном разделе мы проведем упрощенный анализ этих эффектов, подробное их описание читатель найдет в гл. 22.

Групповая скорость

Рассмотрим импульс в виде плоской волны, распространяющийся вдоль оси z в диспергирующей среде без потерь с показателем преломления

$n(\omega)$. Следуя примеру, приведенному в разд. 2.6, предположим, что начальная комплексная волновая функция при $z = 0$

$$U(0, t) = \mathcal{A}(t) \exp(j\omega_0 t),$$

где ω_0 — центральная угловая частота; $\mathcal{A}(t)$ — комплексная огибающая волны. Далее будет показано, что если дисперсия слабая, т. е. n незначительно меняется в пределах спектральной ширины импульса, то комплексная волновая функция на расстоянии z приближенно равна

$$U(z, t) = \mathcal{A}\left(t - \frac{z}{v}\right)\exp\left[j\omega_0\left(t - \frac{z}{c}\right)\right],$$

где c — скорость света в среде на центральной частоте,

$$c = \frac{c_0}{n(\omega_0)};$$

v — скорость, с которой распространяется огибающая (рис. 5.24). Параметр v, называемый **групповой скоростью**, дается выражением

$$\boxed{\frac{1}{v} = \beta' = \frac{d\beta}{d\omega},} \qquad (5.124)$$

Групповая скорость

где β — частотно-зависимая постоянная распространения,

$$\beta = \frac{\omega n(\omega)}{c_0};$$

производная в (5.124), которая часто обозначается β', берется на центральной частоте ω_0. Групповая скорость является характеристикой диспергирующей среды и в общем случае меняется в зависимости от центральной частоты. Соответствующая временна́я задержка $\tau_d = z/v$ называется **групповой задержкой**.

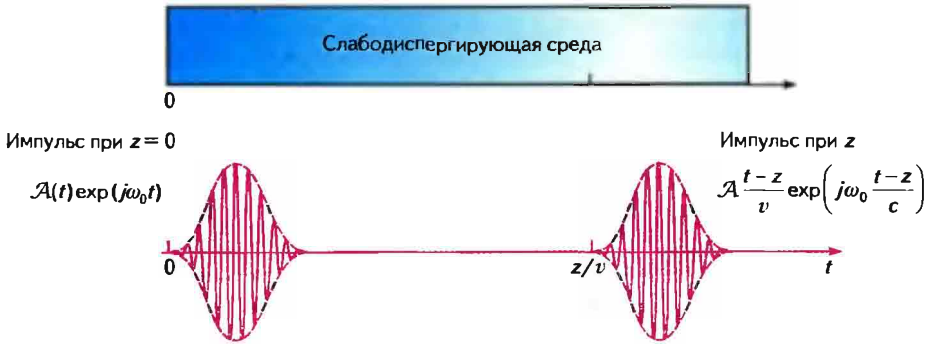


Рис. 5.24. Оптический импульс, распространяющийся в среде с достаточно слабой дисперсией, чтобы групповая скорость не зависела от частоты. Огибающая распространяется с групповой скоростью v, а несущая волна — с фазовой скоростью c

Поскольку фазовый множитель $\exp[j\omega_0(t - z/c)]$ — функция $(t - z/c)$, скорость света c, определяемая как

$$\frac{1}{c} = \frac{\beta(\omega_0)}{\omega_0},$$

часто называется **фазовой скоростью**. В идеальной (недиспергирующей) среде $\beta(\omega) = \omega/c$ и, следовательно, $v = c$, т. е. фазовая и групповая скорости совпадают.

Вывод формулы для групповой скорости

Доказательство (5.124) опирается на разложение Фурье огибающей $\mathcal{A}(t)$. Составляющая на частоте Ω с амплитудой $A(\Omega)$ соответствует монохроматической волне с частотой $\omega = \omega_0 + \Omega$, постоянная распространения которой $\beta(\omega_0 + \Omega)$. Эта составляющая импульса в виде плоской волны распространяется как $A(\Omega)\exp\{j[\beta(\omega_0 + \Omega)]z\}\exp[j(\omega_0 + \Omega)t]$. Если функция $\beta(\omega)$ медленно меняется вблизи центральной частоты ω_0, ее можно разложить в ряд Тейлора с точностью до линейного члена

$$\beta(\omega_0 + \Omega) \approx \beta(\omega_0) + \Omega\frac{d\beta}{d\omega} = \frac{\omega_0}{c} + \frac{\Omega}{v}.$$

Тогда составляющую комплексной волновой функции на частоте Ω можно приближенно представить как $A(\Omega)\exp[j\Omega(t - z/v)]\exp[j\omega_0(t - z/c)]$. Отсюда следует, что при распространении на расстояние z огибающая Фурье-компоненты $A(\Omega)\exp(j\Omega t)$ превращается в $A(\Omega)\exp[j\Omega(t - z/v)]$ для любого значения частоты Ω, следовательно, огибающая импульса $\mathcal{A}(t)$ превращается в $\mathcal{A}(t - z/v)$. Значит, импульс распространяется с групповой скоростью

$$v = \frac{1}{d\beta/d\omega},$$

что совпадает с (5.124).

Поскольку показатель преломления большинства материалов, как правило, измеряется и табулируется как функция длины волны, а не частоты, удобно выразить групповую скорость v через $n(\lambda)$. Соотношения

$$\beta = \frac{\omega n(\lambda)}{c_0} = \frac{2\pi n(\lambda_0)}{\lambda_0} \quad \text{и} \quad \lambda_0 = \frac{2\pi c_0}{\omega},$$

а также правило дифференцирования сложной функции

$$\frac{d\beta}{d\omega} = \frac{d\beta}{d\lambda}\frac{d\lambda}{d\omega}$$

дают

$$\boxed{v = \frac{c_0}{N}}; \quad \boxed{N = n - \lambda_0 \frac{\mathrm{d}n}{\mathrm{d}\lambda_0}}. \tag{5.125}$$

Групповая скорость и групповой показатель преломления

Параметр N часто называют **групповым показателем преломления.**

Дисперсия групповой скорости (ДГС)

Поскольку групповая скорость

$$v = \frac{1}{\mathrm{d}\beta/\mathrm{d}\omega}$$

сама, как правило, зависит от частоты, различные частотные компоненты импульса испытывают различную задержку $\tau_d = z/v$. В результате импульс расплывается во времени. Это явление известно как **дисперсия групповой скорости** (ДГС). Для оценки расплывания из-за ДГС заметим, что при распространении на расстояние z два одинаковых импульса с центральными частотами ν и $\nu + \delta\nu$ имеют разность групповых задержек

$$\delta\tau = \frac{\mathrm{d}\tau_d}{\mathrm{d}\nu}\delta\nu = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{z}{v}\right)\delta\nu = D_\nu z \delta\nu, \tag{5.126}$$

где величина

$$\boxed{D_\nu = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{1}{v}\right) = 2\pi\beta''} \tag{5.127}$$

Коэффициент дисперсии

называется **коэффициентом дисперсии**, а

$$\beta'' = \left.\frac{d^2\beta}{d\omega^2}\right|_{\omega_0}$$

На самом деле этот эффект связан с членами высших порядков в разложении Тейлора $\beta(\omega)$, которыми мы пренебрегли в приведенном выше выводе формулы для групповой скорости; более полное исследование будет проведено в гл. 22.

Если импульс имеет начальную спектральную ширину σ_ν (Гц), то согласно (5.126) хорошую оценку временного уширения дает выражение

$$\boxed{\sigma_\tau = |D_\nu|\sigma_\nu z}. \tag{5.128}$$

Уширение импульса

Коэффициент дисперсии D_ν является мерой увеличения длительности импульса при распространении на единицу длины, приходящегося на единицу

спектральной ширины [с/(м · Гц)]. Этот процесс временного уширения схематически проиллюстрирован на рис. 5.25. Если показатель преломления задан как функция длины волны, $n(\lambda_0)$, то (5.125) и (5.127) дают

$$\boxed{D_\nu = \frac{\lambda_0^3}{c_0^2}\frac{\mathrm{d}^2 n}{\mathrm{d}\lambda_0^2}.} \qquad (5.129)$$

Коэффициент дисперсии [с/(м · Гц)]

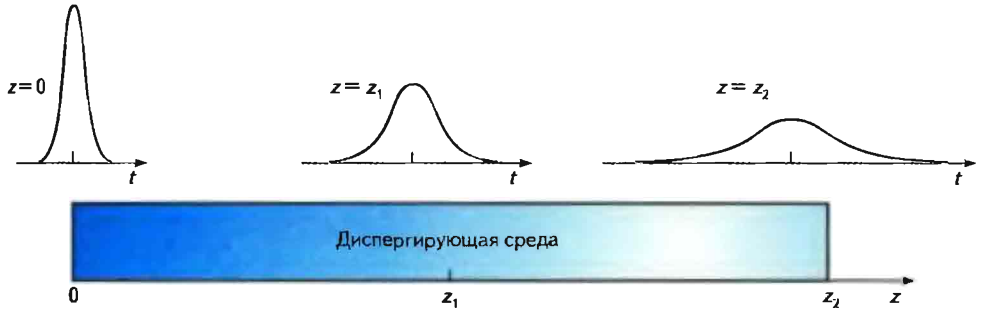


Рис. 5.25. Оптический импульс, распространяющийся в диспергирующей среде, увеличивает свою длительность со скоростью, пропорциональной произведению коэффициента дисперсии D_ν, спектральной ширины σ_ν и пройденного расстояния z

Часто используют коэффициент дисперсии D_λ, выраженный через длину волны[6], а не частоту. С помощью $D_\lambda \mathrm{d}\lambda = D_\nu \mathrm{d}\nu$ находим

$$D_\lambda = D_\nu \frac{\mathrm{d}\nu}{\mathrm{d}\lambda_0} = D_\nu \left(-\frac{c_0}{\lambda_0^2}\right),$$

откуда окончательно

$$\boxed{D_\lambda = -\frac{\lambda_0}{c_0}\frac{\mathrm{d}^2 n}{\mathrm{d}\lambda_0^2}.} \qquad (5.130)$$

Коэффициент дисперсии [с/(м · нм)]

По аналогии с (5.128) для источника со спектральной шириной σ_λ временное уширение светового импульса

$$\boxed{\sigma_\tau = |D_\lambda| \sigma_\lambda z.} \qquad (5.131)$$

Расплывание импульса

Как обсуждается в разд. 9.3, 24.1 и 22.3, в волоконно-оптических приложениях D_λ обычно выражается в единицах пс/(км · нм): уширение импульса измеряется в пикосекундах, пройденный путь в километрах и спектральная ширина источника в нанометрах.

[6] Альтернативное определение коэффициента дисперсии $M = -D_\lambda$ также широко распространено в литературе.

Нормальная и аномальная дисперсия

Хотя знак коэффициента дисперсии D_ν (или D_λ) не влияет на скорость уширения импульса, он влияет на фазу комплексной огибающей оптического импульса. Как таковой, знак может играть важную роль в распространении импульсов через среды, состоящие из каскадов материалов с разными дисперсионными свойствами, что рассматривается в гл. 22. Если $D_\nu > 0$ ($D_\lambda < 0$), говорят, что среда обладает **нормальной дисперсией**. В этом случае время распространения для высокочастотных компонент больше, чем для низкочастотных, так что более коротковолновые компоненты импульса приходят позже, чем более длинноволновые, как это схематически показано на рис. 5.26. Если, напротив, $D_\nu < 0$ ($D_\lambda > 0$), дисперсия называется **аномальной**, и в этом случае более коротковолновые компоненты распространяются быстрее и приходят раньше. Большинство стекол имеет нормальную дисперсию в видимом диапазоне, однако при больших длинах волн дисперсия становится аномальной.

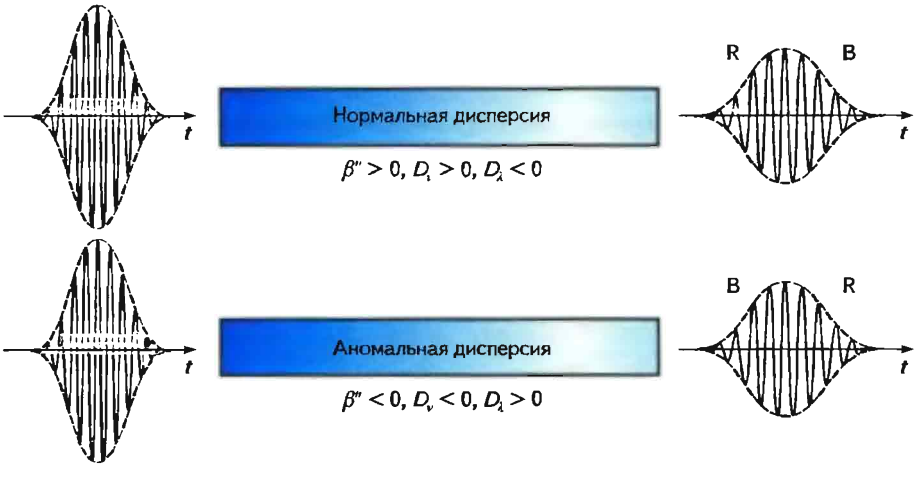


Рис. 5.26. Распространение оптического импульса в среде с нормальной и аномальной дисперсией. В среде с нормальной дисперсией коротковолновые компоненты импульса (B) приходят позже длинноволновых (R). Среда с аномальной дисперсией демонстрирует противоположное поведение. Говорят, что импульс приобретает чирп, поскольку мгновенная частота меняется со временем

Среда с одним резонансом

Групповая скорость и коэффициент дисперсии для оптического импульса, распространяющегося в среде с одним резонансом, определяются путем подстановки (5.103) и (5.104) в (5.88) с использованием (5.125) и (5.130). Для иллюстрации поведения импульса в такой среде на рис. 5.27 приведены графики зависимости показателя преломления n, группового показателя преломления N и коэффициента дисперсии D_λ от нормированной длины волны λ/λ_0 для среды с параметрами $\chi_0 = 0{,}05$ и $\Delta\nu/\nu_0 = 0{,}1$.

В окрестности резонанса (затененная область рисунка) n меняется с длиной волны достаточно быстро, чтобы N и D_λ, определяемые на основе разложения Тейлора с небольшим числом удерживаемых членов, имели смысл. Вдали от резонанса с обеих сторон от него показатель преломления монотонно убывает с ростом длины волны и имеет точки перегиба (отмечены темными точками на графике). Первая производная показателя преломления (и, следовательно, групповой показатель преломления N) в этих точках достигает локальных максимумов. Вторая производная в этих точках обращается в нуль, так что коэффициент дисперсии меняет знак. При приближении длины волны к резонансу снизу характер дисперсии меняется с нормального на аномальный; то же самое происходит при приближении к резонансу сверху, как хорошо видно из рис. 5.27.

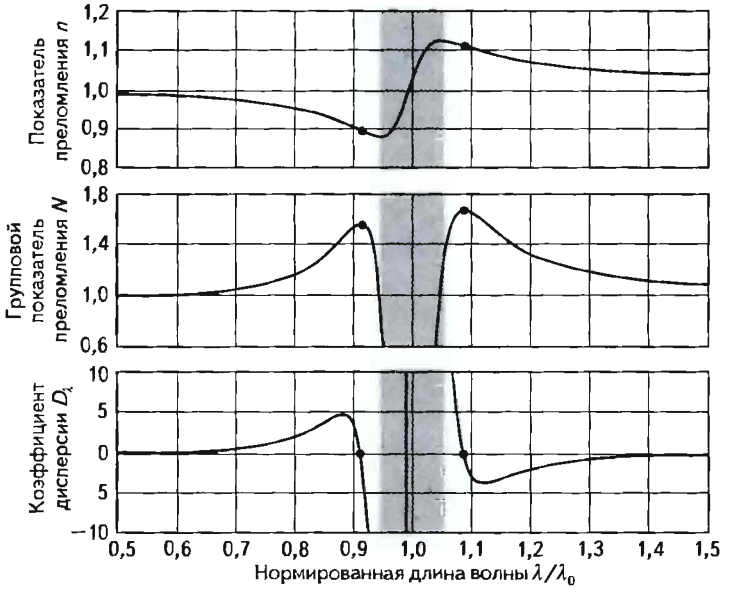


Рис. 5.27. Зависимость оптических параметров среды с одним резонансом от длины волны, нормированной на ее резонансное значение λ/λ_0: показатель преломления $n = c_0/c$ (точками отмечены перегибы), групповой показатель преломления $N = c_0/v$ (точками отмечены максимумы), коэффициент дисперсии D_λ (точками отмечены нули). Параметры N и D_λ не имеют смысла вблизи резонанса (затененная область)

Медленный и быстрый свет в резонансных средах

Как видно из рис. 5.27, в резонансной среде показатель преломления n и групповой показатель преломления N испытывают быстрые изменения вблизи резонансной частоты и могут быть существенно больше или существенно меньше единицы. Следовательно, фазовая скорость $c = c_0/n$ и групповая скорость $v = c_0/N$ могут быть существенно больше или существенно меньше, чем скорость света в вакууме c_0. Групповой показатель преломления и, следо-

вательно, групповая скорость могут быть даже отрицательными. Возникает вопрос о совместимости этих фактов с принципом причинности и специальной теорией относительности, согласно которой скорость передачи информации не может превышать c_0. Оказывается, что противоречия здесь нет, поскольку ни фазовая, ни групповая скорости не являются **информационной скоростью**, т. е. скоростью, с которой осуществляется передача информации между двумя точками. Информационная скорость может быть определена по скорости, с которой движется точка неаналитичности импульса, например фронт прямоугольного импульса. Она не может превосходить c_0.

Понятия фазовой и групповой скорости рассматривались ранее для оптического импульса в слабодиспергирующей среде, т. е. среде, у которой постоянная распространения $\beta(\omega)$ меняется приблизительно по линейному закону вблизи резонансной частоты ω_0. После прохождения расстояния z импульс задерживается на время z/v и приобретает фазовый множитель $\exp(-j\omega_0 z/c)$. Эта фаза, распространяющаяся со скоростью c, не переносит информации. Время «прихода» импульса определяется групповой скоростью v. Поскольку в данном приближении форма огибающей импульса сохраняется в процессе его распространения (см. рис. 5.24), групповая скорость является хорошим приближением для информационной скорости. В резонансных средах это происходит вдали от резонанса, где групповой показатель преломления больше единицы и, соответственно, групповая скорость меньше c_0.

На частотах, близких к резонансу, существенными становятся вклады в дисперсию высших порядков. Если присутствует дисперсия второго порядка (ДГС), но вкладами более высоких порядков можно пренебречь, то, например, гауссов импульс остается гауссовым в процессе распространения, увеличивая лишь свою длительность; его пик движется со скоростью v. Однако, поскольку гауссов импульс имеет непрерывный профиль, простирающийся до бесконечности (т. е. существует все время), скорость, с которой движется его пик, не обязательно является информационной скоростью. В самом деле, она может превышать скорость света в вакууме.

В непосредственной близости к резонансу, где групповая скорость может быть значительно больше c_0 или даже быть отрицательной (см. рис. 5.27), необходимо рассматривать вклады в дисперсию высших порядков. При этом форма импульса может существенно меняться и групповая скорость уже не может рассматриваться как возможная информационная скорость. На достаточно коротких расстояниях, однако, форму импульса можно считать неизменной, и это происходит при групповой скорости много больше c_0. Импульс может также распространяться с отрицательной групповой скоростью, что означает появление точки, идентифицируемой как пик, на выходе из среды до того, как соответствующая точка вошла в среду! В противоположном предельном случае «медленного света» в некоторых специально подобранных средах групповую скорость можно сделать предельно малой, так что световой импульс можно сильно замедлить или даже остановить. Следует, однако, подчеркнуть, что ни в одной из этих ситуаций информационная скорость не превышает c_0.

Поскольку быстрый и медленный свет могут наблюдаться только вблизи резонанса, где коэффициент поглощения велик и зависит от частоты, для усиления этих эффектов требуется механизм усиления; для этой же цели часто используются нелинейные эффекты.

Пример 5.2

Дисперсия в среде с многими резонансами: плавленый кварц

В области между 0,21 и 3,71 мкм зависимость показателя преломления n от длины волны для плавленого кварца при комнатной температуре хорошо описывается формулой Зельмейера (5.111). Если выражать длины волн в мкм, то для этого достаточно использовать три резонансные длины волны $\lambda_1 = 0{,}06840$ мкм, $\lambda_2 = 0{,}1162$ мкм и $\lambda_3 = 9{,}8962$ мкм с весами $\chi_{01} = 0{,}6962$, $\chi_{02} = 0{,}4079$ и $\chi_{03} = 0{,}8975$ соответственно. Выражения для группового показателя преломления N и коэффициента дисперсии D_λ легко выводятся из этой формулы с помощью (5.125) и (5.130). Результаты этого расчета в диапазоне длин волн 600—1600 нм представлены на рис. 5.28.

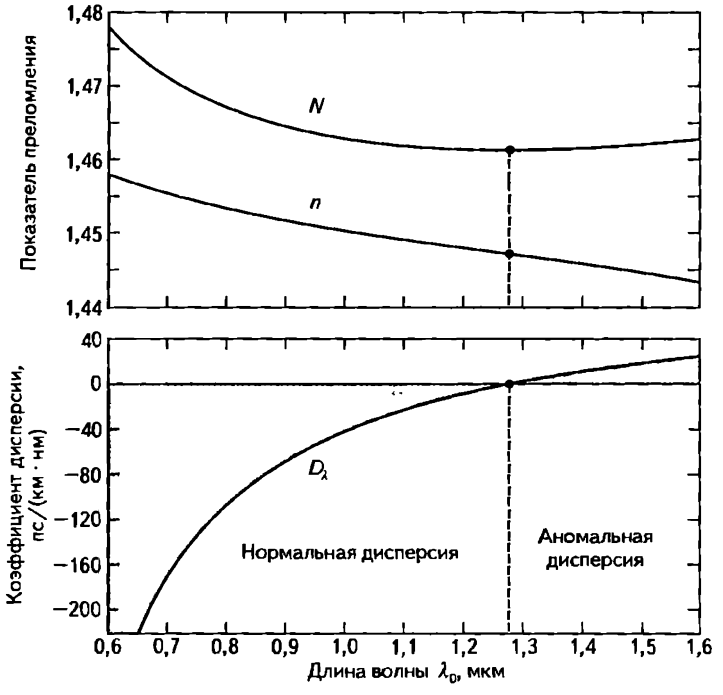


Рис. 5.28. Зависимость оптических параметров от длины волны для плавленого кварца, рассчитанная на основе формулы Зельмейера (5.111): показатель преломления $n = c_0/c$ (точкой отмечен перегиб), групповой показатель преломления $N = c_0/v$ (точкой отмечен минимум), коэффициент дисперсии D_λ (точкой отмечен нуль)

Видно, что показатель преломления монотонно убывает с ростом длины волны и имеет перегиб при $\lambda_0 = 1{,}276$ мкм. На этой длине волны групповой

показатель преломления N имеет минимум, так что групповая скорость $v = c_0/N$ максимальна. Поскольку коэффициент дисперсии равен второй производной от n до λ_0, на этой длине волны он обращается в нуль. Нулевой коэффициент дисперсии означает наименьшее уширение импульса. На длинах волн меньше, чем 1,276 мкм, $D_\lambda < 0$, и среда демонстрирует нормальную дисперсию, в то время как в более длинноволновой области $D_\lambda > 0$ и дисперсия аномальная. Наличие длины волны с нулевым коэффициентом дисперсии дает значительные преимущества для проектирования волоконно-оптических систем связи, в которых оптические импульсы переносят информацию, как станет очевидным в разд. 9.3, 24.1 и 22.3. Волокна из кварцевого стекла в таких системах легируются и имеют нулевой коэффициент дисперсии вблизи 1,312 мкм.

5.7. ОПТИКА МАГНИТНЫХ МАТЕРИАЛОВ И МЕТАМАТЕРИАЛОВ

В данном разделе мы рассмотрим распространение волн в средах, обладающих поглощением и дисперсией как диэлектрической, так и магнитной природы; в таких средах как ε, так и μ являются комплексными и зависят от частоты.

Монохроматическая плоская волна с волновым вектором $\boldsymbol{k}$ имеет электрическое и магнитное поля $\boldsymbol{E}(\boldsymbol{r}) = \boldsymbol{E}_0 \exp(-j\boldsymbol{k}\cdot\boldsymbol{r})$ и $\boldsymbol{H}(\boldsymbol{r}) = \boldsymbol{H}_0 \exp(-j\boldsymbol{k}\cdot\boldsymbol{r})$ соответственно. Эти поля подчиняются уравнениям (5.65) и (5.66), которые мы снова приведем здесь для удобства:

$$\boldsymbol{k} \times \boldsymbol{H}_0 = -\omega\varepsilon\boldsymbol{E}_0; \tag{5.132}$$

$$\boldsymbol{k} \times \boldsymbol{E}_0 = \omega\mu\boldsymbol{H}_0. \tag{5.133}$$

Соответствующее волновое число и импеданс равны

$$k = \omega\sqrt{\varepsilon\mu}; \tag{5.134}$$

$$\eta = \frac{E_0}{H_0} = \frac{\omega\mu}{k}. \tag{5.135}$$

Если и ε, и μ действительны, то $k = nk_0$, где показатель преломления

$$n = \sqrt{\frac{\varepsilon}{\varepsilon_0}\frac{\mu}{\mu_0}} = c_0\sqrt{\varepsilon\mu}. \tag{5.136}$$

Для комплексных ε и μ, следуя (5.88), можно записать k в виде $nk_0 - \alpha/2$, так что

$$\boxed{nk_0 - j\frac{1}{2}\alpha = \omega\sqrt{\varepsilon\mu},} \tag{5.137}$$

где действительные числа n и α представляют собой показатель преломления и коэффициент ослабления соответственно. Мнимые части ε и μ, таким образом, дают вклад в коэффициент ослабления α. Дисперсионные свойства среды описываются частотной зависимостью n и α. Эти величины, в свою очередь, определяются частотными зависимостями диэлектрической и магнитной проницаемости $\varepsilon = \varepsilon(\omega)$ и $\mu = \mu(\omega)$, зависящими от динамики электрического и магнитного отклика среды на атомно-молекулярном уровне. Например, простому диэлектрику соответствует модель резонансной среды, описанная в разд. 5.5.3.

Недавно появился новый класс искусственно структурированных композитных материалов, называемых метаматериалами (см. гл. 7). Эти материалы изготовляются из структурных элементов макроскопических размеров, в отличие от природных материалов, структуру которых образуют атомы и молекулы. Электромагнитные свойства метаматериалов можно проектировать, обеспечивая заранее заданные электрические и магнитные параметры $\varepsilon(\nu)$ и $\mu(\nu)$ и, следовательно, дисперсионные свойства. Метаматериалы обычно представляют собой матрицы, образованные проводящими проволочками, решетками или кольцевыми металлическими петлями. Такие структуры обладают резонансным поведением типа описанного в разд. 5.5.3. Дважды периодическая структура из параллельных золотых наностержней, например, обладает отрицательным показателем преломления в ближней инфракрасной области с $n \approx -0{,}3$ при $\lambda = 1{,}5$ мкм. В этом случае отрицательный показатель преломления обязан своим происхождением плазмонным резонансам в парах наностержней как для электрической, так и для магнитной составляющей поля[7].

Дважды отрицательные метаматериалы

Диэлектрические и магнитные материалы с комплексными ε и μ поддерживают различные типы распространения волн в зависимости от значений действительной и мнимой частей этих комплексных параметров. Знаки действительных частей этих коэффициентов также играют решающую роль (для поглощающих сред мнимые части всегда отрицательны в соответствии с требованием принципа причинности). Необычная ситуация возникает, когда и ε и μ действительны и отрицательны, поскольку в этом случае самосогласованное и физически реализуемое решение уравнений Максвелла порождает отрицательный показатель преломления. Такие материалы называются **дважды отрицательными**.

Для дважды отрицательных материалов с действительными ε и μ, т. е. когда $\varepsilon = -|\varepsilon|$ и $\mu = -|\mu|$, уравнения (5.132) и (5.133) приобретают вид

$$\boldsymbol{k} \times \boldsymbol{H}_0 = \omega |\varepsilon| \boldsymbol{E}_0; \tag{5.138}$$

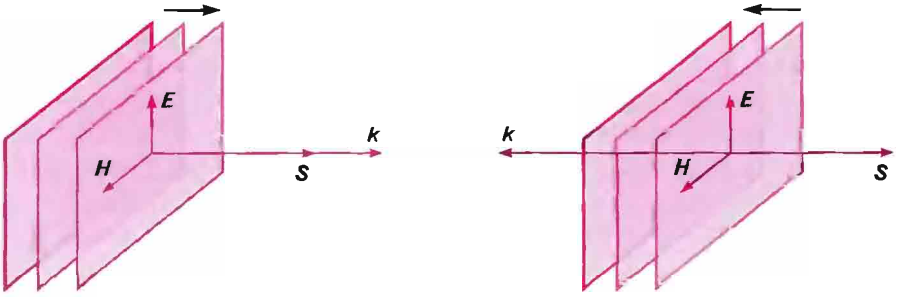
$$\boldsymbol{k} \times \boldsymbol{E}_0 = -\omega |\mu| \boldsymbol{H}_0, \tag{5.139}$$

[7] См. *Shalaev V.M., Cai W., Chettiar U.K., Yuan H.-K., Sarychev A.K., Drachev V.P., Kildishev A.V.* Negative Index of Refraction in Optical Metamaterials. Optics Letters. Vol. 30, 2005. P. 3356–3358.

соответственно. Обращение знаков в (5.138) и (5.139) [сравните с (5.132) и (5.133) для обычных материалов с положительными действительными ε и μ] эквивалентно тому, что электрическое и магнитное поля меняются ролями. В результате $\boldsymbol{E}_0$, $\boldsymbol{H}_0$ и $\boldsymbol{k}$ в дважды отрицательных материалах образуют левую тройку векторов, тогда как в обычных материалах эта тройка правая. Это приводит к кардинальным переменам, поскольку комплексный вектор Пойнтинга

$$\boldsymbol{S} = \frac{1}{2}\boldsymbol{E}_0 \times \boldsymbol{H}_0^*$$

становится антипараллельным волновому вектору $\boldsymbol{k}$. Это означает, что волновой фронт движется в направлении, противоположном потоку электромагнитной энергии. На рис. 5.29, *а* и *б* показаны направления потока энергии и движения волновых фронтов для нормальных и дважды отрицательных материалов соответственно.



Обычный материал *а*

Дважды отрицательный материал *б*

Рис. 5.29. Плоская волна в обычном (дважды положительном) материале (*а*). Векторы $\boldsymbol{E}$, $\boldsymbol{H}$ и $\boldsymbol{k}$ образуют правую тройку, и волновой фронт движется в том же направлении, что и поток энергии. Плоская волна в дважды отрицательном материале (*б*). Векторы $\boldsymbol{E}$, $\boldsymbol{H}$ и $\boldsymbol{k}$ образуют левую тройку, и волновой фронт движется в направлении, противоположном потоку энергии

Рассмотрим, например, плоскую волну, распространяющуюся вдоль оси z в дважды отрицательном материале с векторами $\boldsymbol{E}$ и $\boldsymbol{H}$, направленными по осям x и y соответственно:

$$\boldsymbol{E} = E_0 \exp(-jkz)\hat{\boldsymbol{x}}; \tag{5.140}$$

$$\boldsymbol{H} = H_0 \exp(-jkz)\hat{\boldsymbol{y}}. \tag{5.141}$$

Вектор Пойнтинга равен

$$\boldsymbol{S} = \frac{1}{2}E_0 H_0^* \hat{\boldsymbol{z}} = \frac{|E_0|^2}{2\eta}\hat{\boldsymbol{z}}.$$

Поскольку импеданс

$$\eta = \sqrt{\frac{|\mu|}{|\varepsilon|}}$$

в (5.135) положителен, в чем можно убедиться с помощью (5.138), вектор Пойнтинга указывает положительное направление оси z, а волновой вектор $\boldsymbol{k}$ — отрицательное. Мы приходим к выводу, что волновое число в (5.140) и (5.141) отрицательное. Эта любопытная ситуация соответствует выбору знака «минус» у квадратного корня в (5.136), что приводит к отрицательному показателю преломления

$$n = -c_0\sqrt{|\varepsilon||\mu|}. \tag{5.142}$$

Шансы на достижение ситуации, когда и ε и μ действительны и отрицательны, весьма малы. Например, если оба параметра описывать резонансной моделью, обсуждавшейся в подразд. 5.5.3, то во всей области, где действительная часть отрицательна, мнимая часть не может равняться нулю.

К счастью, условие действительности и отрицательности ε и μ *достаточно, но не необходимо* для того, чтобы среда была «левой», а показатель преломления — отрицательным. «Левизна» на самом деле оказывается совместимой с поглощением и возможна у материалов с комплексными ε и μ. Можно показать, что если действительные части и ε и μ отрицательны, материал будет обладать свойством «левизны» даже в присутствии поглощения (неисчезающей мнимой части)[8]. Например, если как ε, так и μ описывать резонансными моделями, то выше обеих резонансных частот существует частотная полоса, в которой оба необходимых условия выполняются (см. рис. 5.19). В таких материалах волна ослабляется и ее амплитуда убывает в направлении потока энергии (вектора Пойнтинга $\boldsymbol{S}$).

Более того, требование одновременной отрицательности действительных частей ε и μ, в свою очередь, также является *достаточным, но не необходимым* для достижения отрицательного показателя преломления. Класс «левых» сред оказывается шире, чем класс дважды отрицательных сред. Окончательным необходимым и достаточным условием «левизны» является[9]

$$\big(|\varepsilon| - \mathrm{Re}\{\varepsilon\}\big)\big(|\mu| - \mathrm{Re}\{\mu\}\big) > \mathrm{Im}\{\varepsilon\}\,\mathrm{Im}\{\mu\}. \tag{5.143}$$

Материалы, для которых оба параметра действительны, но только один из них отрицателен, не удовлетворяют (5.143) и, следовательно, не могут быть «левыми», но в них возможны затухающие волны, в чем можно убедиться с помощью (5.137). Не могут быть «левыми» и среды, у которых один из материальных параметров, ε или μ, действительный и положительный, какими бы ни были действительная и мнимая части другого. Поэтому ясно, что немагнитные материалы не могут быть «левыми».

---

[8] См, например, *McCall M.W.*, *Lakhtakia A.*, *Weiglhofer W.S.* The Negative Index of Refraction Demystified. European Journal of Physics. Vol. 23, 2002. P. 353–359.

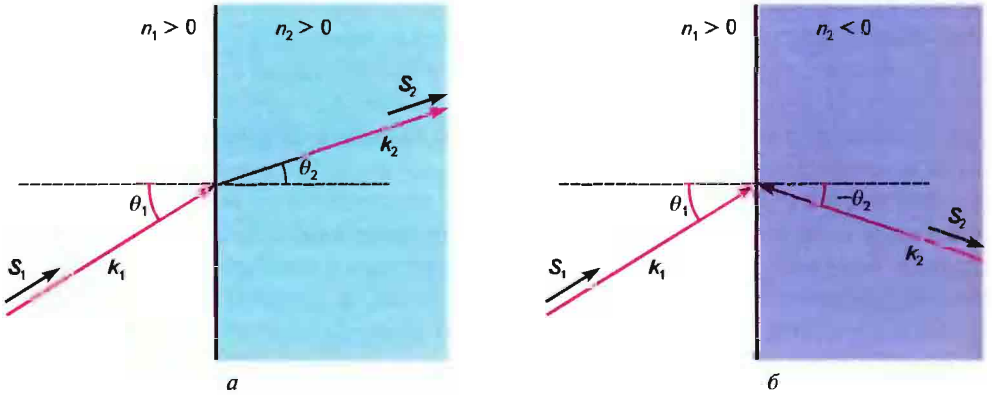
[9] Там же.

Оптика материалов с отрицательным показателем преломления

Многие обычные оптические явления выглядят совсем иначе в средах с отрицательным показателем преломления. Простым примером является преломление света, подчиняющееся закону Снелла, на границе раздела двух диэлектрических сред $n_1 \sin\theta_1 = n_2 \sin\theta_2$. Если одна из сред, например среда 2, имеет отрицательный показатель преломления, то этот закон принимает вид

$$n_1 \sin\theta_1 = -|n_2| \sin\theta_2. \tag{5.144}$$

Поскольку n_1 и θ_1 положительны, угол преломления θ_2 должен быть отрицательным, так что и падающий, и преломленный лучи оказываются с одной стороны от нормали к границе в точке падения. Две формы закона Снелла иллюстрируют рис. 5.30, *а* и *б* соответственно.



а *б*

Рис. 5.30. Преломление на границе раздела двух сред с положительным показателем преломления (*а*). Направления $\boldsymbol{S}_2$ и $\boldsymbol{k}_2$ совпадают. Преломление на границе раздела между средами с положительным и отрицательным показателем преломления (*б*). Направления $\boldsymbol{S}_2$ и $\boldsymbol{k}_2$ противоположны

В результате оптика плоских границ и линз значительно меняется. Например, выпуклая линза из материала с отрицательным показателем преломления ведет себя как вогнутая и наоборот. Еще более необычно, что плоская граница раздела между материалами с положительным и отрицательным показателями преломления действует как линза, что иллюстрирует рис. 5.31 для частного случая пластины из материала с $n = -1$ в свободном пространстве ($n = 1$). Далее, обе границы слоя из материала с отрицательным показателем преломления в этом случае обладают фокусирующим действием, так что второе изображение получается позади слоя. Более того, было показано, что такая система имеет замечательное свойство: ее коэффициент пропускания для плоских волн равен единице при любом угле наклона и независимо от поляризации[10]. С точ-

[10] См. *Pendry J.D.* Negative Refraction Makes a Perfect Lens. Physical Review Letters. Vol. 85, 2000. P. 3966—3969.

ки зрения Фурье-оптики, обсуждавшейся в разд. 4.4, это означает, что через слой проходят все пространственные гармоники изображения, включая нераспространяющиеся волны. В принципе слой ведет себя как «идеальная линза», которая передает информацию, превышая дифракционный предел.

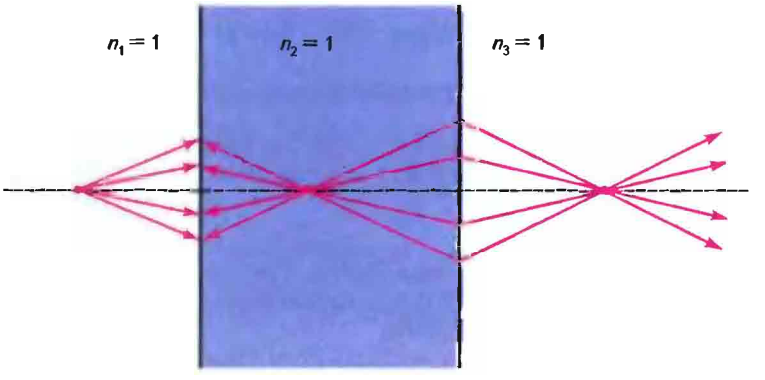


Рис. 5.31. Фокусировка лучей слоем с отрицательным показателем преломления в свободном пространстве. Каждая граница действует как линза, так что изображения формируются внутри и снаружи слоя

Рекомендуемая литература

ОБЩАЯ

См. также литературу к гл. 1.

Hayt W.H., Buck Jr. J.A. Engineering Electromagnetics. McGraw-Hill, 1958; 7th ed. 2006.

Sadiku M.N.O. Elements of Electromagnetics. Oxford University Press, 4th ed. 2006.

Kong J.A. Electromagnetic Wave Theory. Wiley, 2nd ed. 1990; EMW Publishing, 2005.

Lucarini V., Saarinen J.J., Peiponen K.-E., Vartiainen E.M. Kramers-Kronig Relations in Optical Materials Research. Springer-Verlag, 2005.

Narayana N. Rao Elements of Engineering Electromagnetics. Prentice Hall, 1977; 6th ed. 2004.

Ida N. Engineering Electromagnetics. Springer-Verlag, 2nd ed. 2004.

Lorrain P., Corson D.R., Lorrain F. Fundamentals of Electromagnetic Phenomena. Freeman, 2000.

Jackson J.D. Classical Electrodynamics. Wiley, 3rd ed. 1998.

Landau L.D., Lifshitz E.M. The Classical Theory of Fields. Butterworth-Heinemann, 4th revised ed. 1997.

Ramo S., Whinnery J.R., Van Duzer T. Fields and Waves in Communication Electronics. Wiley, 3rd ed. 1994.

Staelin D.H., Morgenthaler A.W., Kong J.A. Electromagnetic Waves. Prentice Hall, 1994.

Sarwate V.V. Electromagnetic Fields and Waves. Wiley, 1993.

Hopf F.A., Stegeman G.I. Applied Classical Electrodynamics. Wiley, 1985; Krieger, reissued 1992.

Kriezis E.E., Chrissoulidis D.P., Papagiannakis A.G. Electromagnetics and Optics. World Scientific, 1992.

Cheng D.K. Field and Wave Electromagnetics. Addison-Wesley, 1983, 2nd ed. 1989.

Haus H.A., Melcher J.R. Electromagnetic Fields and Energy. Prentice Hall, 1989.

Haus H.A. Waves and Fields in Optoelectronics. Prentice Hall, 1984.

Landau L.D., Lifshitz E.M., Pitaevskii L.P. Electrodynamics of Continuous Media. Pergamon. 2nd revised ed. 1984.

ОПТИЧЕСКИЕ КОНСТАНТЫ

Bass M., Van Stryland E.W., Williams D.R., Wolfe W.L., eds. Handbook of Optics. McGraw-Hill, 2nd ed. 1995.

Wolfe W.L., Zissis G.J., eds. The Infrared Handbook. Environmental Research Institute of Michigan. 1993.

Palik E.D., ed. Handbook of Optical Constants of Solids II. Academic Press, 1991.

Kingery W.D., Bowen H.K., Uhlmann D.R. Introduction to Ceramics. Wiley, 2nd ed. 1976.

ПЛАЗМОНЫ

Brongersma M.L., Kik P.G., eds. Surface Plasmon Nanophotonics. Springer-Verlag, 2006.

Tominaga J., Tsai D.P., eds. Optical Nanotechnologies: The Manipulation of Surface and Local Plasmons. Springer-Verlag, 2003.

Jutamulia S., ed. Selected Papers on Near-Field Optics. SPIE Optical Engineering Press (Milestone Series Volume 172), 2002.

Kawata S., Ohtsu M., Irie M., eds. Near-Field Optics and Surface Plasmon Polaritons. Springer-Verlag, 2001.

Pines D. Elementary Excitations in Solids: Lectures on Phonons, Electrons, and Plasmons. West-view, paperback ed. 1999.

Liebsch A. Electronic Excitations at Metal Surfaces. Springer-Verlag, 1997.

Raether H. Surface Plasmons on Smooth and Rough Surfaces and on Gratings. Springer-Verlag, 1988.

БЫСТРЫЙ И МЕДЛЕННЫЙ СВЕТ

Milonni P.W. Fast Light, Slow Light and Left-Handed Light. Institute of Physics, 2005.

Stenner M., Gauthier D.J., Neifeld M.A. Fast Causal Information Transmission in a Medium with a Slow Group Velocity. Physical Review Letters. Vol. 94, 2005. 053902.

Brunner N., Scarani V., Wegmüller M., Legre M., Gisin N. Direct Measurement of Superluminal Group Velocity and Signal Velocity in an Optical Fiber. Physical Review Letters. Vol. 93, 2004. 203902.

Boyd R.W., Gauthier D. «Slow» and «Fast» Light. In Progress in Optics. E. Wolf, ed. Elsevier. Vol. 43, 2002. P. 497—530.

Chiao R.Y., Steinberg A.M. Tunneling Times and Superluminality. In Progress in Optics. E. Wolf, ed. Elsevier. Vol. 37, 1997. P. 347—406.

Oughstun K.E., Sherman G.C. Electromagnetic Pulse Propagation in Causal Dielectrics. Springer-Verlag, 1994; corrected ed. 1997.

Brillouin L. Wave Propagation and Group Velocity. Academic Press, 1960.

ДВАЖДЫ ОТРИЦАТЕЛЬНЫЕ СРЕДЫ

Eleftheriades G.V., Balmain K.G., eds. Negative-Refraction Metamaterials: Fundamental Principles and Applications. Wiley, 2005.

Pendry J.B., Smith D.R. Reversing Light with Negative Refraction. Physics Today. Vol. 57, № 6, 2004. P. 37—43.

McCall M.W., Lakhtakia A., Weiglhofer W.S. The Negative Index of Refraction Demystified. European Journal of Physics. Vol. 23, 2002. P. 353—359.

Pendry J.B. Negative Refraction Makes a Perfect Lens. Physical Review Letters. Vol. 85, 2000. P. 3966—3969.

Veselago V.G. The Electrodynamics of Substances with Simultaneously Negative Values of ε and μ. Soviet Physics-Uspekhi. Vol. 10, 1968. P. 509—514.

Задачи

К РАЗДЕЛУ 5.1

1. **Электромагнитная волна.** Электромагнитная волна в свободном пространстве имеет вектор электрического поля

$$\mathcal{E} = f\left(t - \frac{z}{c_0}\right)\hat{x},$$

где $\hat{x}$ — единичный вектор направления x;

$$f(t) = \exp\left(-\frac{t^2}{\tau^2}\right),$$

где τ — постоянная. Опишите физическую природу этой волны и определите выражение для вектора магнитного поля.

К РАЗДЕЛУ 5.2

1. **Диэлектрические среды.** Определите тип сред, описываемых данными материальными уравнениями, по следующим признакам: линейность, дисперсия, пространственная дисперсия, однородность. Все среды считайте изотропными.

а. $\mathcal{P} = \varepsilon_0 \chi \mathcal{E} - a\nabla \times \mathcal{E}$;

б. $\mathcal{P} + a\mathcal{P}^2 = \varepsilon_0 \mathcal{E}$;

в. $a_1 \varepsilon^2 \mathcal{P}/\varepsilon t^2 + a_2 \varepsilon \mathcal{P}/\varepsilon t + \mathcal{P} = \varepsilon_0 \chi \mathcal{E}$;

г. $\mathcal{P} = \varepsilon_0 \{a_1 + a_2 \exp[-(x^2 + y^2)]\}$,

где χ, a, a_1, a_2 — постоянные.

К РАЗДЕЛУ 5.3

1. **Бегущая стоячая волна.** Вектор комплексной амплитуды электрического поля монохроматической волны с длиной волны λ_0 в свободном пространстве задан в виде $\boldsymbol{E}(\boldsymbol{r}) = E_0 \sin \beta y \exp(-j\beta z)\hat{x}$.

а. Найдите связь между β и λ_0.

б. Выведите выражение для вектора комплексной амплитуды магнитного поля $\boldsymbol{H}(\boldsymbol{r})$.

в. Определите направление потока энергии.

г. Эту волну можно рассматривать как сумму двух плоских ТЕМ-волн. Найдите направления их распространения.

К РАЗДЕЛУ 5.4

1. **Электрическое поле сфокусированного света.**

а. Свет мощностью 1 Вт фокусируется однородно на плоскую мишень размером 0,1 × 0,1 мм^2 в свободном пространстве. Найдите максимальное

значение электрического поля E_0 (В/м). Предполагается, что световая волна аппроксимируется плоской ТЕМ-волной в пределах площади мишени.

б. Определите электрическое поле в центре гауссова пучка (т. е. в его перетяжке на оси пучка), если мощность пучка равна 1 Вт, а радиус перетяжки $W_0 = 0{,}1$ мм (см. разд. 3.1).

К РАЗДЕЛУ 5.5

1. **Амплитудно-модулированная волна в диспергирующей среде.** Амплитудно-модулированная волна с комплексной волновой функцией, которая при $z = 0$ имеет вид $\mathcal{A}(t) = [1 + m\cos(2\pi f_s t)]\exp(j2\pi\nu_0 t)$, где $f_s \ll \nu_0$, распространяется на расстояние z в диспергирующей среде с постоянной распространения $\beta(\nu)$ и пренебрежимо малым поглощением. При условии, что $\beta(\nu_0) = \beta_0$, $\beta(\nu_0 - f_s) = \beta_1$, $\beta(\nu_0 + f_s) = \beta_2$ выведите выражение для комплексной огибающей прошедшей волны как функции β_0, β_1 и β_2. Покажите, что на некоторых расстояниях z волна имеет амплитудную модуляцию, но не имеет фазовой модуляции.

К РАЗДЕЛУ 5.6

1. **Дисперсия групповой скорости в среде, описываемой формулой Зельмейера.**

а. Получите выражение для группового показателя преломления N и коэффициента дисперсии групповой скорости D_1 в среде, показатель преломления которой описывается формулой Зельмейера (5.111).

б. Постройте на компьютере графики зависимости n, N и D_1 от длины волны для плавленого кварца в диапазоне от 0,25 до 3,5 мкм. Используйте параметры, приведенные в табл. 5.1 (и примере 5.2). Сравните с кривыми на рис. 5.28.

в. Постройте аналогичный набор графиков для GaAs в области от 1,5 до 10,5 мкм. Как отмечено в табл. 5.1, GaAs характеризуется трехчленной формулой Зельмейера с резонансными длинами волн 0, 0,4082 и 37,17 мкм, входящими с весом 3,5; 7,4969 и 1,9347 соответственно, в области длин волн от 1,4 до 11 мкм при комнатной температуре. Путем сравнения выявите различия в дисперсионных свойствах плавленого кварца и GaAs.

2. **Показатель преломления воздуха.** Показатель преломления воздуха может быть точно измерен с помощью интерферометра Майкельсона и перестраиваемого источника света. При атмосферном давлении и температуре 20 °С показатель преломления воздуха отличается от единицы на величину $n - 1 = 2{,}672 \cdot 10^{-4}$ при длине волны 0,76 мкм, на $n - 1 = 2{,}669 \cdot 10^{-4}$ при длине волны 0,81 мкм и на $n - 1 = 2{,}665 \cdot 10^{-4}$ при длине волны 0,86 мкм.

а. С помощью квадратичной аппроксимации определите зависимость групповой скорости от длины волны.

б. Получите выражение для коэффициента дисперсии D_λ в пс/(км нм) и сравните ваш результат с результатом для оптического волокна из плавленого кварца.

3. **Групповая скорость в металле.** Покажите, что в среде, описываемой моделью Друде (5.122), произведение фазовой и групповой скоростей равно c_0^2.

ГЛАВА 6 ПОЛЯРИЗАЦИОННАЯ ОПТИКА

Огюстен Жан Френель (1831—1879) выдвинул теорию, в которой свет представляет собой поперечные колебания. Уравнения, описывающие частичное отражение и пропускание света, носят его имя. Френель внес также важный вклад в теорию дифракции света.

Поляризация света в фиксированной точке пространства определяется временным поведением электрического вектора $\mathcal{E}(r, t)$. В простой среде этот вектор лежит в плоскости, касающейся волнового фронта в рассматриваемой точке. Для монохроматического света любые две ортогональные составляющие вектора комплексной амплитуды $E(r)$ в этой плоскости гармонически колеблются со временем, причем амплитуды и фазы этих колебаний, вообще говоря, различны, так что конец вектора $E(r)$ описывает эллипс. Поскольку в разных точках волновой фронт, вообще говоря, имеет разное направление, плоскость, ориентация и форма эллипса также могут меняться от точки к точке, как показано на рис. 6.1, *а*.

Для плоской волны волновые фронты представляют собой систему поперечных плоскостей и эллипс поляризации одинаков по всему фронты волны, как показано на рис. 6.1, *б*, хотя векторы поля не обязательно параллельны в любой момент времени. Плоская волна, таким образом, описывается единственным эллипсом; говорят, что она **эллиптически поляризована**. Ориентация и форма эллипса поляризации определяют состояние поляризации плоской волны, в то время как размер эллипса определяется интенсивностью света. Если эллипс вырождается в отрезок прямой или окружность, свет называется поляризованным **линейно** или **циркулярно (по кругу)** соответственно.

В параксиальной оптике свет распространяется вдоль направлений, лежащих внутри узкого конуса вокруг оптической оси (оси z). Волны приближенно являются поперечными TEM-электромагнитными волнами, векторы электрического поля приближенно лежат в поперечных плоскостях и имеют пренебрежимо малые продольные компоненты. С точки зрения поляризации параксиальные волны можно аппроксимировать плоскими волнами и описывать одним поляризационным эллипсом (отрезком прямой, окружностью).

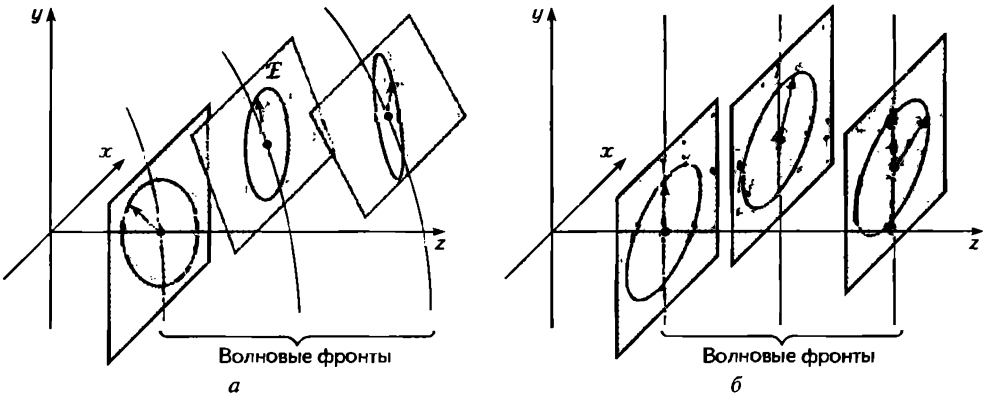


Рис. 6.1. Поведение вектора электрического поля монохроматической волны в нескольких точках: произвольной волны (*а*), плоской или параксиальной волны (*б*), бегущей вдоль оси z

Поляризация играет важную роль во взаимодействии света с веществом, что подтверждается следующими примерами.

- Доля света, отраженного поверхностью раздела между двумя материалами, зависит от поляризации падающего света.
- Поглощение света в некоторых материалах зависит от его поляризации.
- Рассеяние света веществом, как правило, чувствительно к его поляризации.
- Показатель преломления анизотропных материалов зависит от поляризации. Волны с разными поляризациями распространяются с разными скоростями и претерпевают разные фазовые сдвиги, так что эллипс поляризации изменяется по мере распространения волны (т. е. линейно поляризованный свет может быть преобразован в свет с круговой поляризацией). Это свойство используется при конструировании многих оптических приборов.
- Плоскость поляризации линейно поляризованного света поворачивается при прохождении через некоторые среды, включая оптические активные вещества, жидкие кристаллы и некоторые вещества в присутствии внешнего магнитного поля.

О данной главе

Эта глава посвящена описанию элементарных поляризационных явлений и множества их приложений. Эллиптически поляризованный свет вводится в разд. 6.1 с использованием матричного формализма, который удобен

для описания поляризационных устройств. В разд. 6.2 описывается влияние поляризации на отражение и преломление света на границах между диэлектрическими средами. Распространение света в анизотропных средах (кристаллах), оптически активных средах и жидких кристаллах является предметом разд. 6.3, 6.4 и 6.5 соответственно. Наконец, основные поляризационные устройства (поляризаторы, фазовые пластинки и вращатели плоскости поляризации) обсуждаются в разд. 6.6.

6.1. ПОЛЯРИЗАЦИЯ СВЕТА

6.1.1. Поляризация

Рассмотри монохроматическую волну с частотой ν и угловой частотой $\omega = 2\pi\nu$, распространяющуюся в направлении оси z со скоростью c. Электрическое поле лежит в плоскости x—y и в общем случае описывается выражением

$$\mathcal{E}(z,\ t) = \mathrm{Re}\left\{A \exp\left[j\omega\left(t - \frac{z}{c}\right)\right]\right\}, \tag{6.1}$$

где комплексная огибающая

$$A = A_x\hat{x} + A_y\hat{y} \tag{6.2}$$

представляет собой вектор с комплексными составляющими A_x и A_y. Для описания поляризации этой волны будем следить за движением конца вектора $\mathcal{E}(r, t)$ в каждой точке как за функцией времени.

Эллипс поляризации

Выражая A_x и A_y через модули и фазы, $A_x = a_x \exp(j\varphi_x)$ и $A_y = a_y \exp(j\varphi_y)$, и подставляя в (6.2) и (6.1), получаем

$$\mathcal{E}(z,\ t) = \mathcal{E}_x\hat{x} + \mathcal{E}_y\hat{y}, \tag{6.3}$$

где

$$\mathcal{E}_x = a_x \cos\left[\omega\left(t - \frac{z}{c}\right) + \varphi_x\right]; \tag{6.4а}$$

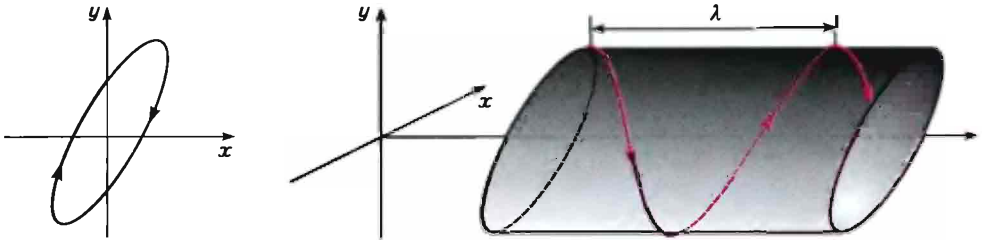
$$\mathcal{E}_y = a_y \cos\left[\omega\left(t - \frac{z}{c}\right) + \varphi_y\right] \tag{6.4б}$$

представляют собой x- и y-составляющие вектора электрического поля $\mathcal{E}(r, t)$. Компоненты $\mathcal{E}_x$ и $\mathcal{E}_y$ — периодические функции $(t - z/c)$, осциллирующие с частотой ν. Уравнения (6.4) — параметрические уравнения эллипса

$$\frac{\mathcal{E}_x^2}{a_x^2} + \frac{\mathcal{E}_y^2}{a_y^2} - 2\cos\varphi\frac{\mathcal{E}_x\mathcal{E}_y}{a_x a_y} = \sin^2\varphi, \tag{6.5}$$

где $\varphi = \varphi_y - \varphi_x$ — разность фаз.

При фиксированном значении z конец вектора электрического поля периодически вращается в плоскости x—y, описывая указанный эллипс. В фиксированный момент времени t конец электрического вектора лежит на винтовой линии в пространстве, которая лежит на поверхности эллиптического цилиндра (рис. 6.2). Электрическое поле вращается по мере распространения волны, повторяя свое движение периодически на расстоянии, соответствующем длине волны $\lambda = c/\nu$.



а *б*

Рис. 6.2. Вращение конца вектора электрического поля в плоскости x—y при фиксированном значении z (*а*). Вид траектории конца вектора электрического поля в фиксированный момент времени t (*б*)

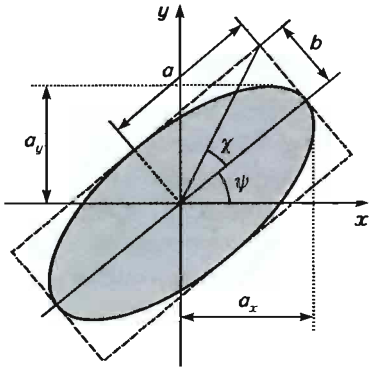


Рис. 6.3. Эллипс поляризации

Состояние поляризации волны определяется ориентацией и формой эллипса поляризации, характеризуемых двумя углами, определение которых ясно из рис. 6.3. Угол ψ определяет направление большой оси, а угол χ — степень эллиптичности, а именно, отношение малой и большой осей эллипса b/a. Зависимость этих углов от отношения модулей $r = a_y/a_x$ и разности фаз $\varphi = \varphi_y - \varphi_x$ определяется соотношениями

$$\operatorname{tg} 2\psi = \frac{2r}{1-r^2}\cos\varphi, \quad r = \frac{a_y}{a_x}; \tag{6.6}$$

$$\sin 2\chi = \frac{2r}{1+r^2}\sin\varphi, \quad \varphi = \varphi_y - \varphi_x. \tag{6.7}$$

Равенства (6.6) и (6.7) можно получить путем поиска такого угла поворота у координатной системы $\mathcal{E}_x$, $\mathcal{E}_y$, который обеспечивает обращение в нуль перекрестного члена в уравнении эллипса (6.5). Размер эллипса определяется интенсивностью волны, пропорциональной $|A_x|^2 + |A_y|^2 = |a_x|^2 + |a_y|^2$.

Линейно поляризованный свет

Если одна из компонент исчезает (напрмер, a_x), то свет линейно поляризован в направлении другой компоненты (направление y). Волна будет поля-

ризована линейно также в случае, когда разность фаз равна нулю или π, поскольку (6.4) в этом случае дает $\mathcal{E}_y = \pm(a_y/a_x)\mathcal{E}_x$, что представляет собой уравнение прямой линии с наклоном $\pm(a_y/a_x)$ (знаки «+» и «—» соответствуют $\varphi = 0$ и $\varphi = \pi$). В этих случаях эллиптический цилиндр на рис. 6.2, *б* вырождается в плоскость, как показано на рис. 6.4. Поэтому волна называется также **плоско поляризованной**. Если, например, $a_y = a_x$, то плоскость поляризации образует угол 45° с осью x. Если $a_x = 0$, то плоскостью поляризации является плоскость y—z.

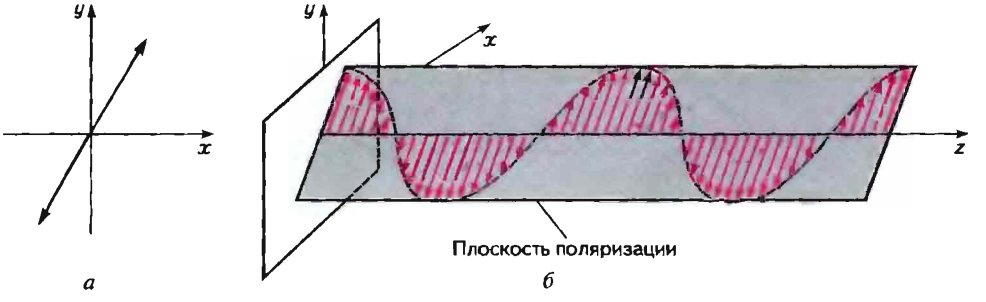


Рис. 6.4. Линейно поляризованный свет (называемый также плоско поляризованным). Изменение во времени при фиксированном положении z (*а*). Пространственная картина в фиксированный момент времени t (*б*)

Циркулярно поляризованный свет

Если $\varphi = \pm\pi/2$ и $a_y = a_x = a_0$, то (6.4) дает

$$\mathcal{E}_x = a_0 \cos\left[\omega\left(t - \frac{z}{c}\right) + \varphi_x\right]; \quad \mathcal{E}_y = \mp a_0 \sin\left[\omega\left(t - \frac{z}{c}\right) + \varphi_x\right],$$

откуда

$$\mathcal{E}_x^2 + \mathcal{E}_y^2 = a_0^2,$$

что представляет собой уравнение окружности. Эллиптический цилиндр на рис. 6.2, *б* становится круговым цилиндром, и волна называется поляризованной циркулярно (по кругу). В случае $\varphi = +\pi/2$ электрический вектор при фиксированном z вращается по часовой стрелке, если смотреть со стороны, откуда бежит волна. В этом случае говорят, что волна имеет **правую круговую поляризацию (ПКП)**. Случай $\varphi = -\pi/2$ соответствует вращению электрического вектора против часовой стрелки и **левой круговой поляризации (ЛКП)**[1] В случае правой круговой поляризации кривая положений конца электрического вектора в фиксированный момент времени представляет собой правую винтовую линию (правый винт смотрит в сторону распространения волны), как показано на рис. 6.5. Для левой круговой поляризации получается левый винт.

[1] Это соглашение принято в большинстве учебников по оптике. Противоположное определение часто используется в технической литературе.

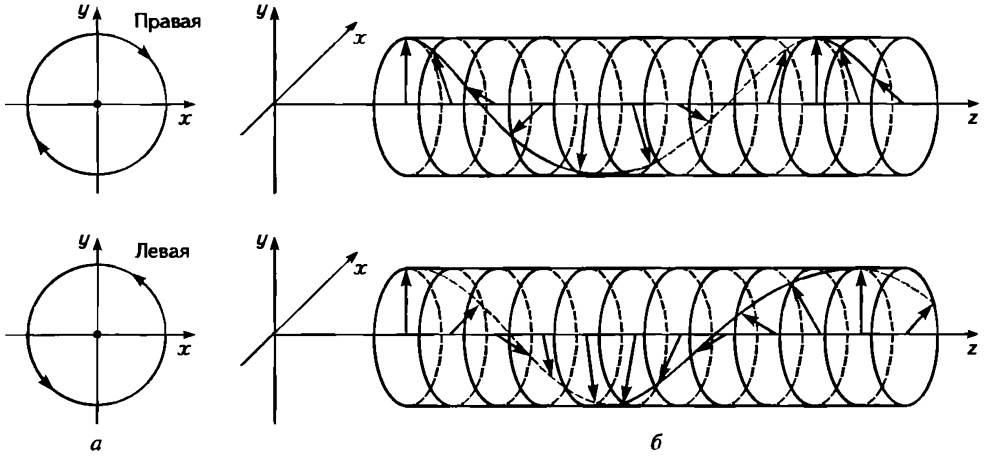


Рис. 6.5. Кривая положений конца вектора электрического поля для плоской волны, поляризованной по кругу:

а — при фиксированном z; *б* — в фиксированный момент времени t. Направления вращения на рисунках (*а*) и (*б*) противоположны, поскольку бегущая волна зависит от $(t - z/c)$

Сфера Пуанкаре и параметры Стокса

Как отмечалось ранее, состояние поляризации световой волны может быть описано двумя действительными параметрами: отношением модулей $r = a_y/a_x$ и разностью фаз $\varphi = \varphi_y - \varphi_x$ компонент поля. Иногда их объединяют в одно комплексное число $r\exp(j\varphi)$, которое называется комплексным поляризационным отношением. С другой стороны, можно характеризовать состояние поляризации двумя углами ψ и χ, которые представляют ориентацию и степень вытянутости поляризационного эллипса, соответственно, как определено на рис. 6.3.

Сфера Пуанкаре (рис. 6.6) — геометрическое построение, при котором каждое состояние поляризации изображается точкой на сфере единичного радиуса с координатами ($r = 1$, $\theta = 90° - 2\chi$, $\phi = 2\psi$) в сферической системе. Каждая точка сферы представляет поляризационное состояние. Например, точки экватора ($\chi = 0°$) соответствуют всевозможным состояниям линейной поляризации, причем две точки, $2\psi = 0°$ и $2\psi = 180°$, соответствуют поляризации вдоль осей x и y. Северный и Южный полюсы ($2\chi = \pm 90°$) представляют правую и левую круговые поляризации соответственно. Другие точки сферы отображают состояния эллиптической поляризации.

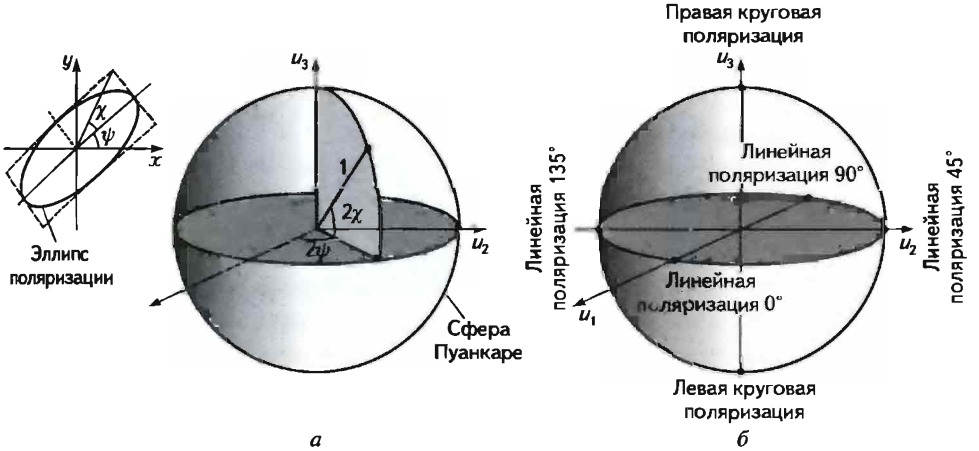
Две действительные величины (r, φ) или, что эквивалентно, углы (χ, ψ), описывают состояние поляризации, но не содержат информации об интенсивности волны. Другое представление, содержащее такую информацию, достигается с помощью **вектора Стокса**. Это набор из четырех действительных чисел (S_0, S_1, S_2, S_3), называемых **параметрами Стокса**. Первый из них, $S_0 = a_x^2 + a_y^2$, про-

порционален оптической интенсивности, а другие три являются декартовыми координатами точки на сфере Пуанкаре $(u_1, u_2, u_3) = (\cos 2\chi \cos 2\psi, \cos 2\chi \sin 2\psi, \sin 2\chi)$, умноженными на S_0, так что

$$S_1 = S_0 \cos 2\chi \cos 2\psi; \tag{6.8а}$$

$$S_2 = S_0 \cos 2\chi \sin 2\psi; \tag{6.8б}$$

$$S_3 = S_0 \sin 2\chi. \tag{6.8в}$$



а *б*

Рис. 6.6. Ориентация и вытянутость поляризационного эллипса геометрически представлены точкой на сфере Пуанкаре (*а*). Точки на сфере Пуанкаре, представляющие линейно поляризованный свет с различной ориентацией относительно оси x, а также право- и левополяризованный свет (*б*)

С помощью (6.6) и (6.7), а также известных тригонометрических тождеств, параметры Стокса в (6.8) можно выразить через параметры поля (a_x, a_y, φ) и компоненты комплексной огибающей:

$$S_0 = a_x^2 + a_y^2 = |A_x|^2 + |A_y|^2 \tag{6.9а}$$

$$S_1 = a_x^2 - a_y^2 = |A_x|^2 - |A_y|^2 \tag{6.9б}$$

$$S_2 = 2a_x a_y \cos\varphi = 2\,\mathrm{Re}\{A_x^* A_y\}; \tag{6.9в}$$

$$S_3 = 2a_x a_y \sin\varphi = 2\,\mathrm{Im}\{A_x^* A_y\}. \tag{6.9г}$$

Параметры Стокса

Поскольку

$$S_1^2 + S_2^2 + S_3^2 = S_0^2,$$

только три из четырех компонент вектора Стокса независимы; они полностью определяют интенсивность и состояние поляризации света. Обобщение параметров Стокса на случай частично поляризованного света представлено в разд. 11.4.

Мы приходим к заключению, что существует три эквивалентных представления для описания состояния поляризации оптического поля:

1) эллипс поляризации;
2) сфера Пуанкаре;
3) вектор Стокса.

Еще одно эквивалентное представление — вектор Джонса — вводится в следующем разделе.

6.1.2. Матричное представление

Вектор Джонса

Как отмечалось выше, монохроматическая плоская волна с частотой ν, распространяющаяся в направлении z, полностью характеризуется комплексными огибающими $A_x = a_x \exp(j\varphi_x)$ и $A_y = a_y \exp(j\varphi_y)$ x- и y-компонент вектора электрического поля. Эти комплексные величины можно записать в виде вектора-столбца, известного как **вектор Джонса**:

$$\boldsymbol{J} = \begin{bmatrix} A_x \\ A_y \end{bmatrix}. \tag{6.10}$$

Если задан $\boldsymbol{J}$, то мы можем определить полную интенсивность волны

$$I = \frac{|A_x|^2 + |A_y|^2}{2\eta}$$

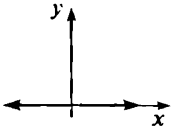
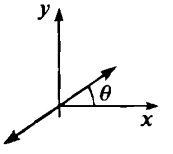
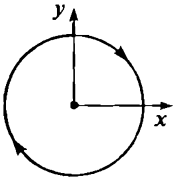
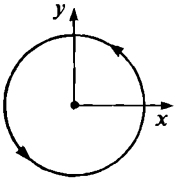
и использовать отношение

$$r = \frac{a_y}{a_x} = \frac{|A_y|}{|A_x|}$$

и разность фаз $\varphi = \varphi_y - \varphi_x = \arg\{A_y\} - \arg\{A_x\}$ для определения ориентации и формы эллипса поляризации, а также положения точки на сфере Пуанкаре и параметров Стокса.

Векторы Джонса для некоторых частных случаев поляризационных состояний приведены в табл. 6.1. Интенсивность в каждом случае нормирована так, чтобы $|A_x|^2 + |A_y|^2 = 1$, а фаза x-компоненты выбрана равной $\varphi_x = 0$.

Таблица 6.1. **Векторы Джонса света с линейной (ЛП), правой круговой (ПКП) и левой круговой (ЛКП) поляризацией**

ЛП вдоль x $\begin{bmatrix}1\\0\end{bmatrix}$ 	ЛП под углом θ $\begin{bmatrix}\cos\theta\\\sin\theta\end{bmatrix}$ 
ПКП $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\j\end{bmatrix}$ 	ЛКП $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-j\end{bmatrix}$ 

Ортогональные поляризации

Два поляризационных состояния, представленные векторами Джонса J_1 и J_2, называются ортогональными, если их скалярное произведение равно нулю. Скалярное произведение определяется как

$$(J_1,\ J_2) = A_{1x}A_{2x}^* + A_{1y}A_{2y}^*, \tag{6.11}$$

где A_{1x} и A_{1y} — элементы J_1; A_{2x} и A_{2y} — элементы J_2. Примерами ортогональных векторов Джонса являются случаи плоских волн, линейно поляризованных вдоль осей x и y, или вдоль любой пары взаимно перпендикулярных направлений. Другим примером являются волны с правой и левой круговыми поляризациями.

Разложение произвольной поляризации на две ортогональные поляризации

Произвольный вектор Джонса J всегда можно представить в виде взвешенной суперпозиции двух ортогональных векторов Джонса J_1 и J_2, образующих базис разложения: $J = a_1J_1 + a_2J_2$. Если J_1 и J_2 нормированы так, что $(J_1, J_1) = (J_2, J_2) = 1$, то коэффициенты разложения — скалярные произведения $a_1 = (J, J_1)$, $a_2 = (J, J_2)$.

Пример 6.1

Разложение в базисах линейно и циркулярно поляризованных состояний

При использовании в качестве базиса векторов $\begin{bmatrix}1\\0\end{bmatrix}$ и $\begin{bmatrix}0\\1\end{bmatrix}$, соответствующих линейной поляризации по x и y, для коэффициентов разложения вектора Джонса с компонентами A_x и A_y, удовлетворяющими условию нормировки $|A_x|^2 + |A_y|^2 = 1$,

по определении имеем $\alpha_1 = A_x$, $\alpha_2 = A_y$. Этот же вектор можно разложить и в других базисах.

• В базисе векторов линейной поляризации под углами 45 и 135°, т. е.

$$J_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}; \quad J_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\end{bmatrix},$$

коэффициенты разложения α_1 и α_2 имеют вид

$$A_{45} = \frac{1}{\sqrt{2}}\left(A_x + A_y\right); \quad A_{135} = \frac{1}{\sqrt{2}}\left(A_y - A_x\right). \tag{6.12}$$

• Аналогично, если в качестве базиса используются волны с правой $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\j\end{bmatrix}$ и левой $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-j\end{bmatrix}$ круговыми поляризациями, то коэффициенты α_1 и α_2 выражаются как

$$A_R = \frac{1}{\sqrt{2}}\left(A_x - jA_y\right); \quad A_L = \frac{1}{\sqrt{2}}\left(A_x + jA_y\right). \tag{6.13}$$

Например, линейно поляризованная волна с плоскостью поляризации, образующей угол θ с осью x (т. е. $A_x = \cos\theta$, $A_y = \sin\theta$), эквивалентна суперпозиции волн, поляризованных по правому и левому кругу с коэффициентами $\frac{1}{\sqrt{2}}e^{-j\theta}$ и $\frac{1}{\sqrt{2}}e^{j\theta}$ соответственно. Таким образом, линейно поляризованная волна равна взвешенной сумме волн с правой и левой круговыми поляризациями.

Упражнение 6.1

Измерение параметров Стокса

Покажите, что параметры Стокса, определенные формулами (6.9), для света с компонентами вектора Джонса A_x и A_y выражаются как

$$S_0 = |A_x|^2 + |A_y|^2; \tag{6.14а}$$

$$S_1 = |A_x|^2 - |A_y|^2 \tag{6.14б}$$

$$S_2 = |A_{45}|^2 - |A_{135}|^2; \tag{6.14в}$$

$$S_3 = |A_R|^2 - |A_L|^2, \tag{6.14г}$$

где A_{45} и A_{135} — коэффициенты разложения в базисе линейно поляризованных векторов, образующих с осью x углы 45° и 135° (6.12), а A_R и A_L — коэффициен-

ты разложения в базисе векторов с правой и левой круговыми поляризациями (6.13). Предложите метод измерения параметров Стокса произвольно поляризованного света.

Матричное представление поляризационных устройств

Рассмотрим прохождение плоской волны с произвольной поляризацией через оптическую систему, которая сохраняет волну плоской, но меняет ее поляризацию, как схематически показано на рис. 6.7. Система предполагается линейной, так что принцип суперпозиции оптических полей выполняет два примера таких систем — это отражение света от плоской границы между двумя средами и прохождение света через пластинку с анизотропными оптическими свойствами.

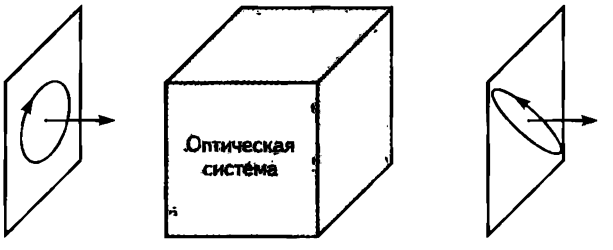


Рис. 6.7. Оптическая система, изменяющая поляризацию плоской волны

Комплексные огибающие двух компонент электрического поля входной (падающей) волны, A_{1x} и A_{1y}, и компоненты выходной (отраженной или прошедшей) волны, A_{2x} и A_{2y}, связаны посредством взвешенной суперпозиции

$$\begin{aligned} A_{2x} &= T_{11}A_{1x} + T_{12}A_{1y}; \\ A_{2y} &= T_{21}A_{1x} + T_{22}A_{1y}, \end{aligned} \tag{6.15}$$

где T_{11}, T_{12}, T_{21} и T_{22} — постоянные, описывающие устройство. Уравнения (6.15) — общие соотношения, которым должны удовлетворять все линейные оптические поляризационные устройства.

Линейные соотношения (6.15) удобно записать в матричном виде, определив матрицу **T** размерностью 2 × 2 с элементами T_{11}, T_{12}, T_{21} и T_{22}, тогда

$$\begin{bmatrix} A_{2x} \\ A_{2y} \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} A_{1x} \\ A_{1y} \end{bmatrix}. \tag{6.16}$$

Если входная и выходная волны описываются векторами Джонса $\boldsymbol{J}_1$ и $\boldsymbol{J}_2$ соответственно, то (6.16) можно записать в компактной форме

$$\boldsymbol{J}_2 = \mathbf{T}\boldsymbol{J}_1. \tag{6.17}$$

Матрица **T**, называемая матрицей Джонса, описывает оптическую систему, а векторы $\boldsymbol{J}_1$ и $\boldsymbol{J}_2$ — волны на входе и выходе.

Структура матрицы Джонса **T** данной оптической системы определяет ее действие на поляризационное состояние и интенсивность волны. Ниже приведены матрицы Джонса для некоторых систем с простыми характеристиками. Физические устройства с такими характеристиками будут впоследствии обсуждаться в данной главе.

Линейные поляризаторы

Система с матрицей Джонса

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{6.18}$$

Линейный поляризатор вдоль оси x

преобразует волну с компонентами (A_{1x}, A_{1y}) в волну с компонентами $(A_{1x}, 0)$, обращая в нуль y-компоненту, в результате на выходе волна поляризована вдоль x, как показано на рис. 6.8. Система является линейным поляризатором, ось пропускания которого направлена по оси x.

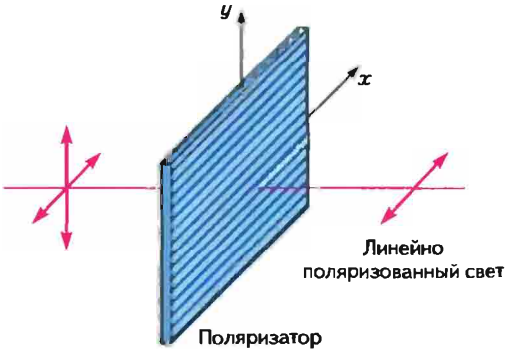


Рис. 6.8. Линейный поляризатор. Линиями показано направление поля, которое может пройти через систему

Фазовые пластинки

Система с матрицей Джонса

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-j\Gamma} \end{bmatrix} \tag{6.19}$$

Фазовая пластинка (быстрая ось вдоль оси x)

преобразует волну с компонентами (A_{1x}, A_{1y}) в волну с компонентами $(A_{1x}, e^{-j\Gamma}A_{1y})$, задерживая по фазе y-компоненту на величину Γ и оставляя x-компоненту без изменений. Такое устройство называется **фазовой пластинкой**. Оси x и y называются быстрой и медленной осями пластинки соответственно.

Простое применение матричной алгебры позволяет легко понять результаты, иллюстрируемые рис. 6.9.

• Когда $\Gamma = \pi/2$, пластинка (называемая **четвертьволновой**) преобразует линейно поляризованную волну $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ в волну с левой круговой поляризацией $\begin{bmatrix} 1 \\ -j \end{bmatrix}$, а волну с правой круговой поляризацией $\begin{bmatrix} 1 \\ j \end{bmatrix}$ — в линейно поляризованную волну $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

• Когда $\Gamma = \pi$, пластинка (называемая **полуволновой**) преобразует линейно поляризованную волну $\begin{bmatrix}1\\1\end{bmatrix}$ в линейно поляризованную волну $\begin{bmatrix}1\\-1\end{bmatrix}$, поворачивая плоскость поляризации на 90°. Полуволновая пластинка преобразует волну с правой круговой поляризацией $\begin{bmatrix}1\\j\end{bmatrix}$ в волну с левой круговой поляризацией $\begin{bmatrix}1\\-j\end{bmatrix}$.

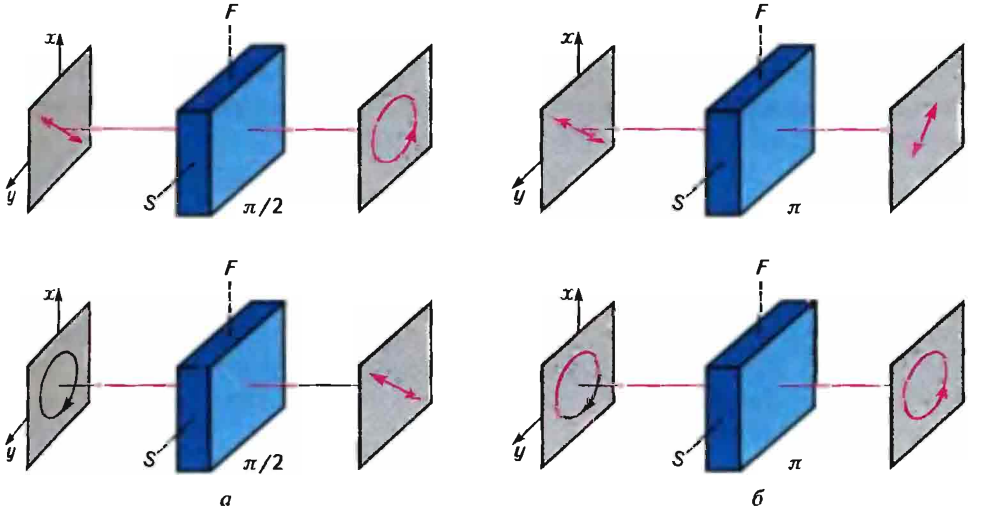


Рис. 6.9. Действие четвертьволновой ($\pi/2$) (*а*) и полуволновой (π) (*б*) пластинок на некоторые состояния поляризации. *F* и *S* обозначают быструю и медленную оси пластинок

Вращатели линейной поляризации

Если фазовые пластинки преобразуют тип поляризации волны, то **вращатель линейной поляризации** сохраняет линейность поляризации и лишь поворачивает плоскость поляризации на некоторый угол. Матрица Джонса

$$\mathbf{T} = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix} \qquad (6.20)$$

Вращатель плоскости поляризации

описывает устройство, которое преобразует линейно поляризованную волну $\begin{bmatrix}\cos\theta_1\\ \sin\theta_1\end{bmatrix}$ в другую линейно поляризованную волну $\begin{bmatrix}\cos\theta_2\\ \sin\theta_2\end{bmatrix}$, где $\theta_2 = \theta_1 + \theta$. Таким образом, оно поворачивает плоскость линейной поляризации волны на угол θ.

Последовательности поляризационных устройств

Действие последовательности оптических систем на поляризованный свет удобно описывается с помощью обычных правил перемножения матриц. Система, состоящая из последовательных подсистем с матрицами Джонса $\mathbf{T}_1$ и $\mathbf{T}_2$, эквивалентна системе с матрицей Джонса $\mathbf{T} = \mathbf{T}_2\mathbf{T}_1$. Матрица подсистемы, через которую свет проходит первым, должна стоять справа, поскольку она первой действует на вектор Джонса.

Упражнение 6.2

Каскад фазовых пластинок

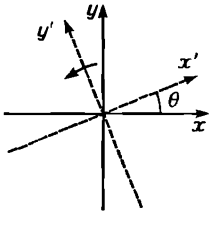
Покажите, что две последовательные четвертьволновые пластинки с параллельными быстрыми осями эквивалентны полуволновой пластинке. Каким будет результат, если быстрые оси ортогональны?

Преобразование координат

Элементы векторов и матриц Джонса зависят от выбора координатной системы. Однако, если эти элементы известны в одной координатной системе, то их можно определить в другой системе с использованием матричных методов. Если $\boldsymbol{J}$ — вектор Джонса в системе координат x—y, то в новой координатной системе x'—y', ось x' которой образует угол θ с осью x, вектор Джонса дается выражением

$$\boldsymbol{J}' = \boldsymbol{R}(\theta)\boldsymbol{J}, \tag{6.21}$$

где $\boldsymbol{R}(\theta)$ — матрица преобразования



$$\boldsymbol{R}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{6.22}$$

Матрица преобразования

Это нетрудно проверить, связывая компоненты электрического поля в двух координатных системах.

Матрица Джонса $\mathbf{T}$, характеризующая оптическую систему, преобразуется в $\mathbf{T}'$ в соответствии с матричными соотношениями

$$\mathbf{T}' = \boldsymbol{R}(\theta)\mathbf{T}\boldsymbol{R}(-\theta); \tag{6.23}$$

$$\mathbf{T} = \boldsymbol{R}(-\theta)\mathbf{T}'\boldsymbol{R}(\theta), \tag{6.24}$$

где $\boldsymbol{R}(-\theta)$ определяется выражением (6.22) с заменой θ на $-\theta$. Матрица $\boldsymbol{R}(-\theta)$ является обратной по отношению к $\boldsymbol{R}(\theta)$, так что $\boldsymbol{R}(-\theta)$ $\boldsymbol{R}(\theta)$ есть

единичная матрица. Уравнение (6.23) можно получить, используя соотношение $J_2 = \mathbf{T}J_1$ и преобразование $J_2' = R(\theta)J_2 = R(\theta)\mathbf{T}J_2$. Поскольку $J_1 = R(-\theta)J_1'$, имеем $J_2' = R(\theta)\mathbf{T}R(-\theta)J_1'$, а поскольку $J_2' = \mathbf{T}'J_1'$, получаем (6.23).

Упражнение 6.3

Матрица Джонса поляризатора

Покажите, что матрица Джонса линейного поляризатора с осью пропускания, образующей угол θ с осью x, равна

$$\mathbf{T} = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix}. \qquad (6.25)$$

Линейный поляризатор под углом θ

Выведите (6.25) с помощью (6.18), (6.22) и (6.24).

Нормальные моды

Нормальные моды поляризационной системы — это состояния поляризации, которые не меняются при прохождении волны через систему (см. приложение В). Эти состояния имеют векторы Джонса, удовлетворяющие уравнению

$$\mathbf{T}J = \mu J, \qquad (6.26)$$

где μ — постоянная. Нормальные моды, таким образом, являются собственными векторами матрицы Джонса $\mathbf{T}$, а значения μ — соответствующими собственными значениями. Поскольку матрица $\mathbf{T}$ имеет размерность 2×2, существует только две независимые нормальные моды $\mathbf{T}J_1 = \mu_1 J_1$ и $\mathbf{T}J_2 = \mu_2 J_2$. Если матрица $\mathbf{T}$ эрмитова, т. е. $T_{12} = T_{21}^*$, то нормальные моды ортогональны: $(J_1, J_2) = 0$. Нормальные моды обычно используются как базис для разложения, так что произвольная волна J на входе может быть разложена в суперпозицию нормальных мод $J = \alpha_1 J_1 + \alpha_2 J_2$. Отклик системы тогда легко определить, поскольку $\mathbf{T}J = \mathbf{T}(\alpha_1 J_1 + \alpha_2 J_2) = \alpha_1 \mathbf{T}J_1 + \alpha_2 \mathbf{T}J_2 = \alpha_1\mu_1 J_1 + \alpha_2\mu_2 J_2$ (см. приложение В).

Упражнение 6.4

Нормальные моды простых поляризационных систем

а. Покажите, что нормальные моды линейного поляризатора — линейно поляризованные волны.

б. Покажите, что нормальные моды фазовой пластинки — линейно поляризованные волны.

в. Покажите, что нормальные моды вращателя плоскости поляризации — волны с правой и левой круговой поляризацией.

Каковы собственные векторы перечисленных систем?

6.2. ОТРАЖЕНИЕ И ПРЕЛОМЛЕНИЕ

В данном разделе мы изучим отражение и преломление монохроматической плоской волны произвольной поляризации, падающей на плоскую границу между двумя диэлектрическими средами. Среды предполагаются линейными, однородными, изотропными с импедансами η_1 и η_2 и показателями преломления n_1 и n_2. Падающая, преломленная и отраженная волны обозначаются индексами 1, 2 и 3 соответственно, как показано на рис. 6.10.

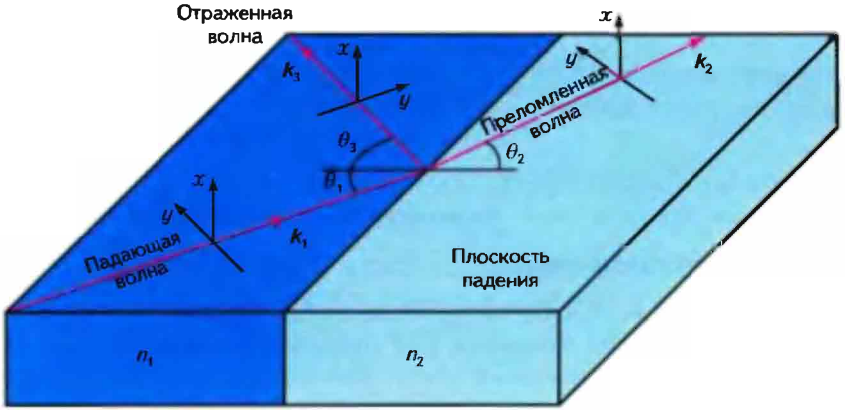


Рис. 6.10. Отражение и преломление на границе между двумя диэлектрическими средами

Как показано в разд. 2.4.1, волновые фронты этих волн согласованы на границе, если углы отражения и падения равны, $\theta_1 = \theta_3$, а угол преломления удовлетворяет закону Снелла

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{6.27}$$

Чтобы связать амплитуды и поляризации трех волн, свяжем с каждой из них систему координат x—y в плоскости, перпендикулярной направлению распространения (см. рис. 6.10). Огибающие электрического поля указанных волн описываются векторами Джонса

$$\boldsymbol{J}_1 = \begin{bmatrix} A_{1x} \\ A_{1y} \end{bmatrix}; \quad \boldsymbol{J}_2 = \begin{bmatrix} A_{2x} \\ A_{2y} \end{bmatrix}; \quad \boldsymbol{J}_3 = \begin{bmatrix} A_{3x} \\ A_{3y} \end{bmatrix}. \tag{6.28}$$

Перейдем к определению связи между $\boldsymbol{J}_2$ и $\boldsymbol{J}_1$ и между $\boldsymbol{J}_3$ и $\boldsymbol{J}_1$, которую будем искать в матричной форме $\boldsymbol{J}_2 = \mathbf{t}\boldsymbol{J}_1$ и $\boldsymbol{J}_3 = \mathbf{r}\boldsymbol{J}_1$, где $\mathbf{t}$ и $\mathbf{r}$ — матрицы Джонса размерностью 2×2, описывающие пропускание и отражение волны соответственно.

Элементы матриц пропускания и отражения можно определить путем наложения граничных условий, требуемых электромагнитной теорией, а именно, непрерывности тангенциальных составляющих $\boldsymbol{E}$ и $\boldsymbol{H}$ и нормальных составля-

ющих D и B. Электрическое поле, связанное с каждой из волн, ортогонально магнитному полю, отношение их огибающих есть характеристический импеданс, который равен η_1 для падающей и отраженной волн и η_2 — для преломленной волны. В результате получается система уравнений, которую нужно решить, чтобы получить связь между электрическими полями трех волн.

Вычисления существенно упрощаются, если заметить, что две нормальные моды этой системы представляют собой волны с линейной поляризацией вдоль осей x и y. Это можно доказать путем демонстрации того, что падающая, отраженная и преломленная волны, электрическое поле которых направлено вдоль оси x, удовлетворяют граничным условиям. То же самое нужно показать для волн, поляризованных вдоль оси y. Это в самом деле так. Тогда волны, поляризованные вдоль осей x и y, не связаны друг с другом.

Мода, поляризованная по x, называется **поперечной электрической (TE)** модой, а ее поляризация — **ортогональной**, поскольку электрическое поле ортогонально плоскости падения. Мода, поляризованная вдоль y, называется **поперечной магнитной (TM)** модой, поскольку ее магнитное поле ортогонально плоскости падения, а ее поляризация называется **параллельной**, поскольку электрический вектор параллелен плоскости падения. Ортогональная и параллельная поляризации обозначаются также буквами s и p от немецких слов *senkrecht* (перпендикулярный) и *parallel* (параллельный) соответственно. Оси y на рис. 6.10 по определению выбраны так, чтобы их проекции на границу раздела указывали в одном направлении.

Независимость x- и y-поляризаций подразумевает диагональность матриц Джонса **t** и **r**

$$\mathbf{t} = \begin{bmatrix} t_x & 0 \\ 0 & t_y \end{bmatrix}; \quad \mathbf{r} = \begin{bmatrix} r_x & 0 \\ 0 & r_y \end{bmatrix}, \tag{6.29}$$

так что

$$E_{2x} = t_x E_{1x}, \quad E_{2y} = t_y E_{1y}; \tag{6.30}$$

$$E_{3x} = r_x E_{1x}, \quad E_{3y} = r_y E_{1y}. \tag{6.31}$$

Кэффициенты t_x и t_y представляют собой комплексные амплитудные коэффициенты пропускания для TE- и TM-мод соответственно, а r_x и r_y — аналогичные комплексные амплитудные коэффициенты отражения.

Применяя граничные условия (т. е. приравнивая тангенциальные компоненты электрического и магнитного полей по обе стороны границы) отдельно для TE- и TM-волн, получаем следующие выражения для коэффициентов отражения и пропускания:

$$r_x = \frac{\eta_2 \sec\theta_2 - \eta_1 \sec\theta_1}{\eta_2 \sec\theta_2 + \eta_1 \sec\theta_1}, \quad t_x = 1 + r_x; \tag{6.32}$$

TE-поляризация

$$r_y = \frac{\eta_2 \cos\theta_2 - \eta_1 \cos\theta_1}{\eta_2 \cos\theta_2 + \eta_1 \cos\theta_1}, \quad t_y = (1 + r_y)\frac{\cos\theta_1}{\cos\theta_2}. \tag{6.33}$$

TM-поляризация

Отражение и пропускание

Характеристический импеданс

$$\eta = \sqrt{\frac{\mu}{\varepsilon}}$$

является комплексным, если ε и/или μ комплексное, как в случае проводящей среды или среды с потерями. Для немагнитных диэлектрических сред без потерь $\eta = \eta_0/n$ действительно, где

$$\eta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}};$$

n — показатель преломления. В этом случае из (6.32) и (6.33) для коэффициентов отражения и пропускания получаются формулы, известные как **формулы Френеля**:

$$r_x = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2}, \quad t_x = 1 + r_x; \tag{6.34}$$

ТЕ-поляризация

$$r_y = \frac{n_1 \sec\theta_1 - n_2 \sec\theta_2}{n_1 \sec\theta_1 + n_2 \sec\theta_2}, \quad t_y = (1 + r_y)\frac{\cos\theta_1}{\cos\theta_2}. \tag{6.35}$$

ТМ-поляризация

Формулы Френеля

При данных n_1, n_2 и θ_1 для нахождения коэффициентов отражения из формул Френеля надо прежде всего найти θ_2 из закона Снелла (6.27):

$$\cos\theta_2 = \sqrt{1 - \sin^2\theta_2} = \sqrt{1 - \frac{n_1}{n_2}\sin^2\theta_1}. \tag{6.36}$$

Поскольку величины под знаком квадратного корня в (6.36) могут быть отрицательными, коэффициенты отражения и пропускания в общем случае комплексные. Модули $|r_x|$, $|r_y|$ и фазовые сдвиги $\varphi_x = \arg\{r_x\}$ и $\varphi_y = \arg\{r_y\}$ показаны в виде графиков на рис. 6.11—6.14 для двух поляризаций как функции угла падения θ_1. Графики представлены для случаев как внешнего ($n_1 < n_2$), так и внутреннего ($n_1 > n_2$) отражения.

Поляризация ТЕ

Зависимость коэффициента отражения r_x от θ_1 для ТЕ-поляризованной волны дается формулой (6.34).

Внешнее отражение ($n_1 < n_2$)

Коэффициент отражения r_x всегда действительный и отрицательный, что соответствует сдвигу фазы $\varphi_x = \pi$. Модуль его равен

$$|r_x| = \frac{n_2 - n_1}{n_1 + n_2}$$

при $\theta_1 = 0°$ (нормальное падение) и возрастает до единицы при $\theta_1 = 90°$ (скользящее падение), как показано на рис. 6.11.

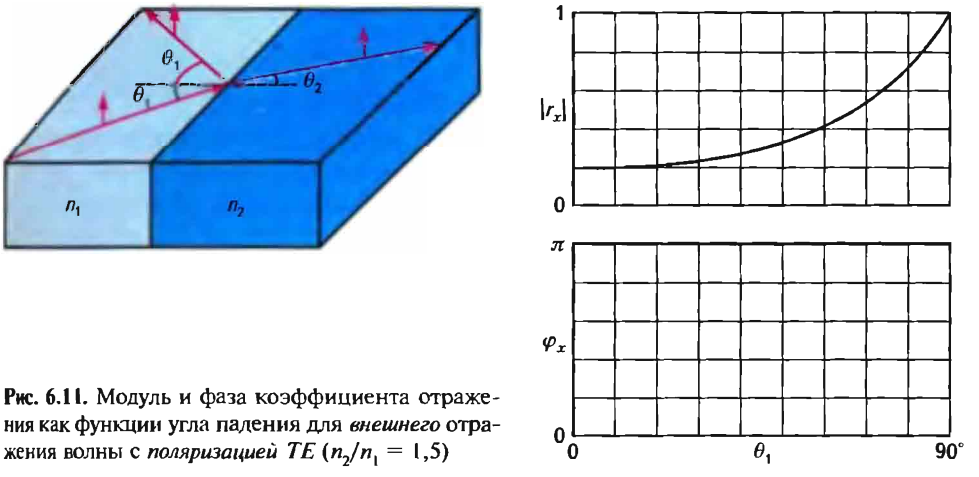


Рис. 6.11. Модуль и фаза коэффициента отражения как функции угла падения для *внешнего* отражения волны с *поляризацией TE* ($n_2/n_1 = 1{,}5$)

Внутреннее отражение ($n_1 > n_2$)

Для малых θ_1 коэффициент отражения действительный и положительный. Его величина равна $(n_1 - n_2)/(n_1 + n_2)$ при $\theta_1 = 0°$ и плавно растет до единицы при критическом значении угла θ_1, равном $\theta_c = \arcsin(n_2/n_1)$. При $\theta_1 > \theta_c$ модуль r_x остается равным единице, что соответствует полному внутреннему отражению. Это можно показать, используя (6.36) для получения[2] выражения

$$\cos\theta_2 = -\sqrt{1 - \frac{\sin^2\theta_1}{\sin^2\theta_c}} = -j\sqrt{\frac{\sin^2\theta_1}{\sin^2\theta_c} - 1}$$

с последующей подстановкой в (6.34). Полное внутреннее отражение сопровождается фазовым сдвигом $\varphi_x = \arg\{r_x\}$, который определяется соотношением

$$\boxed{\operatorname{tg}\frac{\varphi_x}{2} = \sqrt{\frac{\cos^2\theta_c}{\cos^2\theta_1} - 1}.} \qquad (6.37)$$

TE-отражение
Фазовый сдвиг

Фазовый сдвиг φ_x растет от 0 при $\theta_1 = \theta_c$ до π при $\theta_1 = 90°$, как показано на рис. 6.12. Этот сдвиг играет существенную роль в диэлектрических волноводах (см. разд. 8.2).

[2] Выбор знака «минус» перед квадратным корнем согласуется с выводом формул Френеля.

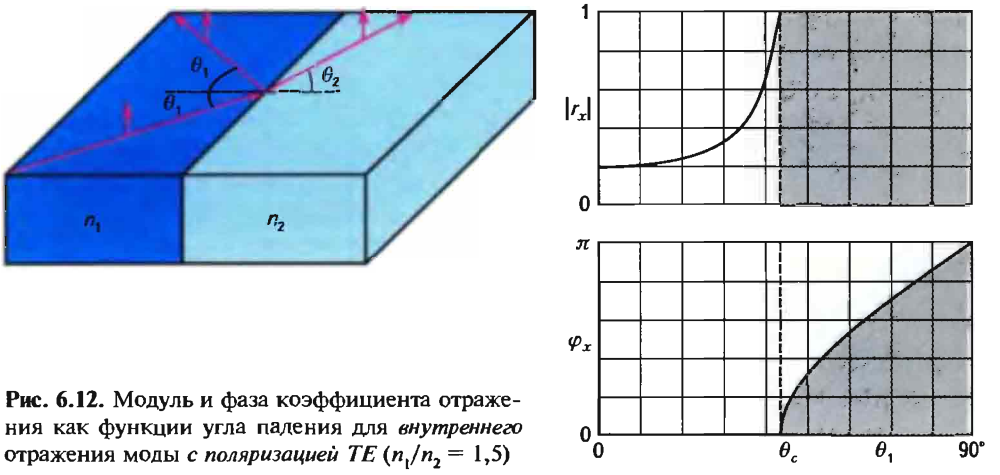


Рис. 6.12. Модуль и фаза коэффициента отражения как функции угла падения для *внутреннего* отражения моды *с поляризацией TE* ($n_1/n_2 = 1{,}5$)

Поляризация TM

Аналогичная зависимость коэффициента отражения r_y от θ_1 для TM-поляризованной волны дается формулой (6.35).

Внешнее отражение ($n_1 < n_2$)

Коэффициент отражения r_y всегда действительный. Он принимает отрицательное значение $(n_1 - n_2)/(n_1 + n_2)$ при $\theta_1 = 0°$ (нормальное падение). Затем его модуль убывает и обращается в нуль, когда $n_1 \sec\theta_1 = n_2 \sec\theta_2$ при $\theta_1 = \theta_B$, известном как **угол Брюстера**:

$$\theta_B = \operatorname{arctg}\left(\frac{n_2}{n_1}\right) \qquad (6.38)$$

Угол Брюстера

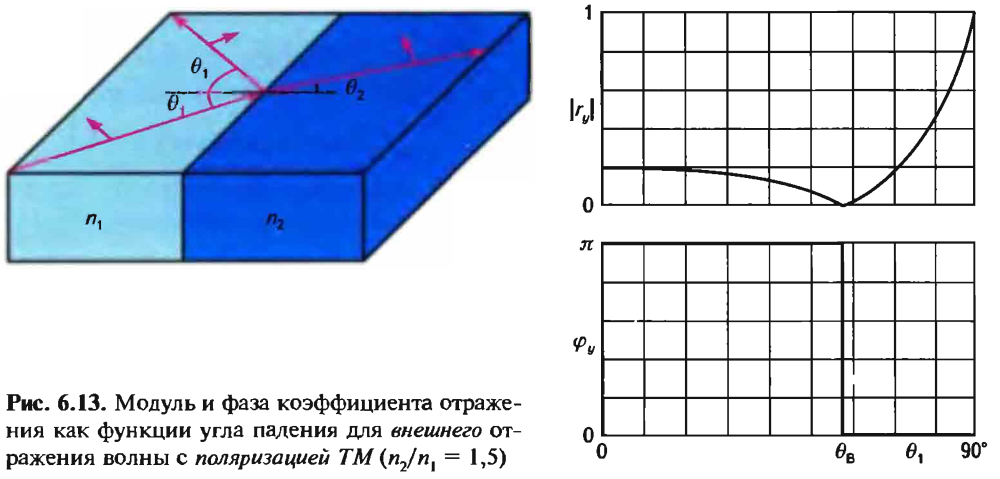


Рис. 6.13. Модуль и фаза коэффициента отражения как функции угла падения для *внешнего* отражения волны с *поляризацией TM* ($n_2/n_1 = 1{,}5$)

(см. задачу 1 к разд. 6.2 о других свойствах угла Брюстера). Для $\theta_1 > \theta_B$ коэффициент r_y меняет знак (φ_y меняется от π до 0), а его величина постепенно растет, пока не достигнет единицы при $\theta_1 = 90°$ (рис. 6.13). Отсутствие отражения волны TM под углом Брюстера применяется при изготовлении поляризаторов (см. разд. 6.6).

Внутреннее отражение ($n_1 > n_2$)

При $\theta_1 = 0°$ r_y положителен и равен $(n_1 - n_2)/(n_1 + n_2)$. С ростом θ_1 величина r_y убывает и обращается в нуль при падении под углом Брюстера $\theta_B = \text{arctg}\,(n_2/n_1)$. При $\theta_1 > \theta_B$ r_y становится отрицательным, а его модуль растет, достигая значения единицы при $\theta_1 = \theta_c$ (рис. 6.14). Для еще больших углов падения $\theta_1 > \theta_c$ волна претерпевает полное внутреннее отражение, сопровождаемое фазовым сдвигом $\varphi_x = \arg\{r_x\}$, который определяется соотношением

$$\text{tg}\frac{\varphi_y}{2} = \frac{-1}{\sin^2\theta_c}\sqrt{\frac{\cos^2\theta_c}{\cos^2\theta_1} - 1}. \qquad (6.39)$$

TM-отражение
Фазовый сдвиг

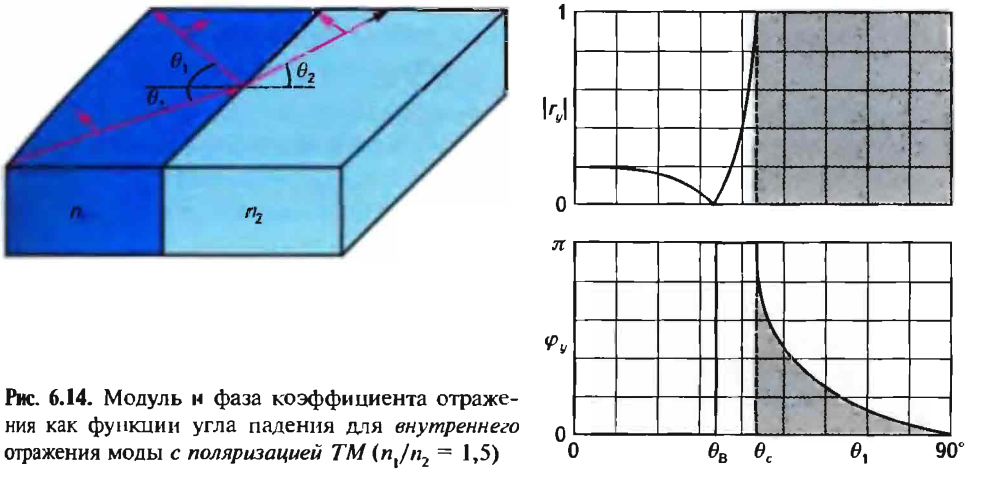


Рис. 6.14. Модуль и фаза коэффициента отражения как функции угла падения для *внутреннего* отражения моды *с поляризацией TM* ($n_1/n_2 = 1{,}5$)

При нормальном падении коэффициент отражения равен

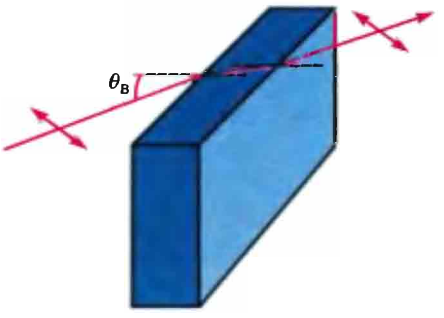
$$r = \frac{n_1 - n_2}{n_1 + n_2},$$

независимо от типа моды (TE или TM) и отражения (внешнее или внутреннее).

Упражнение 6.5

Брюстеровские окна

При каком угле падения пучок с TM-поляризацией проходит через стеклянную пластину с показателем преломления $n = 1{,}5$, расположенную в воз-



духе ($n = 1$), не испытывая отражения на обеих поверхностях пластины? Такие пластины, известные как окна Брюстера (рис. 6.15), используются в лазерах, как описано в подразд. 15.2.4.

Рис. 6.15. Окно Брюстера пропускает свет с TM-поляризацией без потерь на отражение

Отражение и пропускание энергии

Коэффициенты отражения r и пропускания t представляют собой отношения комплексных амплитуд. Коэффициенты отражения $\mathcal{R}$ и пропускания $\mathcal{T}$ по мощности определяются как отношение потоков энергии (по нормали к границе раздела) отраженной и прошедшей волн к потоку энергии падающей волны. Поскольку падающая и отраженная волны распространяются в одной и той же среде под одним и тем же углом к поверхности раздела, имеем

$$\mathcal{R} = |r|^2 \tag{6.40}$$

Независимо от типа моды (TE или TM) и отражения (внешнее или внутреннее) при нормальном падении

$$\boxed{\mathcal{R} = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2}. \tag{6.41}$$

Коэффициент отражения при нормальном падении

Например, на границе между стеклом ($n = 1{,}5$) и воздухом ($n = 1$) $\mathcal{R} = 0{,}04$, так что 4 % света при нормальном падении отражается. На границе между GaAs ($n = 3{,}6$) и воздухом ($n = 1$) $\mathcal{R} \approx 0{,}32$, так что при нормальном падении отражается 32 % света. Однако при наклонном падении коэффициент отражения может быть как значительно больше, так и значительно меньше 32 %, как видно на рис. 6.16.

Коэффициент пропускания $\mathcal{T}$ по мощности определяется законом сохранения энергии, так что в отсутствие поглощения он равен просто

$$\mathcal{T} = 1 - \mathcal{R}. \tag{6.42}$$

Важно, однако, заметить, что $\mathcal{T}$ в общем случае *не равен* $|t|^2$, поскольку поток энергии направлен под разными углами по разные стороны от границы, а также различны импедансы сред. Для волны, распространяющейся в среде с показателем преломления n под углом θ, поток энергии вдоль нормали к границе

$$\frac{|\mathcal{E}|^2}{2\eta}\cos\theta = \frac{|\mathcal{E}|^2}{2\eta_0} n\cos\theta.$$

Следовательно

$$\mathcal{T} = \frac{n_2 \cos\theta_2}{n_1 \cos\theta_1}|t|^2 \tag{6.43}$$

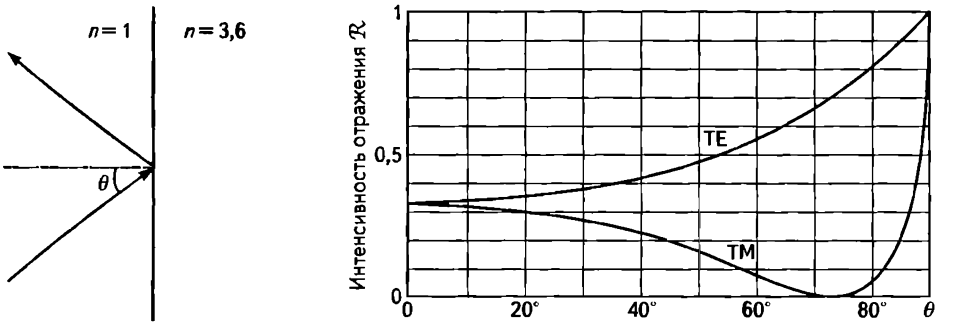


Рис. 6.16. Коэффициенты отражения по мощности для линейно поляризованных TE- и TM-волн для границы между воздухом ($n = 1$) и GaAs ($n = 3{,}6$) как функции угла падения

Отражение от пластины

Коэффициент отражения по мощности при нормальном падении для пластины с двумя поверхностями описывается выражением $\mathcal{R}(1 + \mathcal{T}^2)$, поскольку мощность, отраженная от дальней поверхности, два раза проходит через ближнюю поверхность. Для стеклянной пластины полный коэффициент отражения равен $\mathcal{R}(1 + \mathcal{T}^2) = 0{,}04[1 + (0{,}96)^2] \approx 0{,}077$, т. е. около 7,7 % мощности падающего света отражается. Однако этот расчет не учитывает эффекты интерференции, которые, впрочем, исчезают, если свет некогерентный (см. разд. 11.2), а также многократное отражение внутри пластины. Пропускание и отражение света множеством поверхностей в слоистых средах детально описывается в разд. 7.1.

Упражнение 6.6

Отражение от проводящей среды

Выражения для коэффициентов отражения (6.32) и (6.33) можно использовать для определения коэффициента отражения по интенсивности $\mathcal{R}$ на границе между диэлектрической и проводящей средами.

а. Покажите, что $\mathcal{R} \approx 1$, если удельная электропроводность среды бесконечна.

б. Покажите, что при нормальном падении в случае $\sigma \gg \varepsilon_0\omega$ можно получить формулу

$$\mathcal{R} = 1 - 2\sqrt{\frac{2\varepsilon_0\omega}{\sigma}},$$

называемую соотношением Хагена—Рубенса. С помощью этого соотношения определите коэффициент отражения меди на длинах волн

$\lambda_0 = 1{,}06$ и $10{,}6$ мкм. Удельную электропроводность меди принять равной $\sigma = 0{,}58 \cdot 10^8$ $(\text{Ом} \cdot \text{м})^{-1}$.

в. Покажите, что если проводящая среда описывается моделью Друде (5.122), то на частотах ниже плазменной частоты $\mathcal{R} = 1$.

6.3. ОПТИКА АНИЗОТРОПНЫХ СРЕД

Говорят, что диэлектрическая среда анизотропна, если ее макроскопические оптические свойства зависят от направления. Макроскопические свойства материала, конечно, неразрывно связаны с его микроскопическими свойствами: формой и ориентацией отдельных молекул, а также расположением их центров в пространстве. Оптические материалы имеют различные типы пространственного и ориентационного упорядочения, которые можно описать следующим образом (рис. 6.17).

- Если молекулы расположены в случайных точках пространства, сами по себе изотропны или ориентированы случайным образом, то среда изотропна. Таковы *газы*, *жидкости* и *аморфные твердые тела.*
- Если структура принимает вид отдельных кристаллических зерен, случайным образом ориентированных друг по отношению к другу, материал называется *поликристаллическим.* Отдельные зерна в общем случае анизотропны, однако их усредненное макроскопическое поведение изотропно.
- Если молекулы организованы в пространстве и образуют регулярную периодическую структуру и к тому же ориентированы в одном направлении, как в *кристаллах*, то среда, вообще говоря, анизотропна.
- Если молекулы анизотропны, а их ориентации не совсем случайны, среда анизотропна, даже если положения молекул в пространстве случайны. Этот случай реализуется в *жидких кристаллах*, в которых существует ориентационный порядок, тогда как положение центров молекул в пространстве упорядочено не полностью.

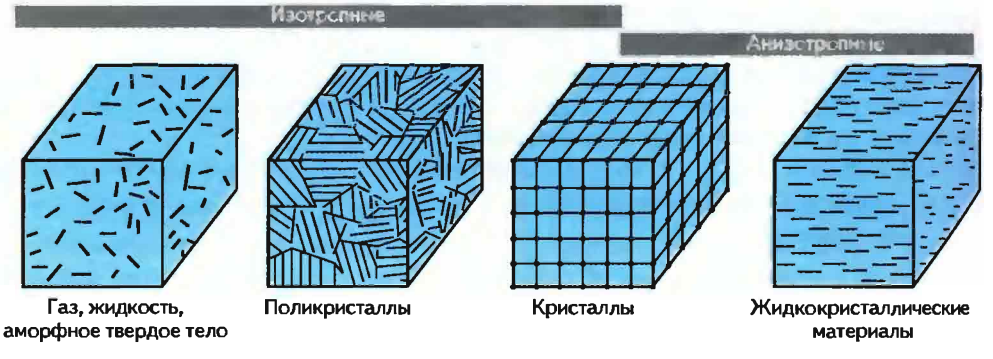


Рис. 6.17. Пространственное и ориентационное упорядочение в различных типах материалов

6.3.1. Показатели преломления

Тензор диэлектрической проницаемости

В линейной анизотропной диэлектрической среде (например, кристалле) каждая компонента вектора электрической индукции D представляет собой линейную комбинацию трех компонент вектора электрического поля

$$D_i = \sum_j \varepsilon_{ij} E_j. \tag{6.44}$$

Индексы $i, j = 1, 2, 3$ нумеруют компоненты x, y и z соответственно, как описано в подразд. 5.2.2. Таким образом, диэлектрические свойства среды характеризуются набором $\{\varepsilon_{ij}\}$ из 3×3 элементов, образующих **тензор диэлектрической проницаемости** второго ранга. **Материальное уравнение** (6.44) обычно записывается в символической форме

$$\boldsymbol{D} = \varepsilon \boldsymbol{E}. \tag{6.45}$$

Для большинства диэлектрических сред тензор диэлектрической проницаемости симметричен, т. е. $\varepsilon_{ij} = \varepsilon_{ji}$. Это значит, что соотношение между векторами $\boldsymbol{D}$ и $\boldsymbol{E}$ взаимное, т. е. их отношение остается тем же самым при обмене направлениями. Такая симметрия типична для диэлектрических немагнитных материалов, которые не обладают оптической активностью, а также в отсутствие внешнего магнитного поля (см. разд. 6.4). При такой симметрии среда характеризуется всего шестью независимыми числами в произвольной системе координат. Для кристаллов определенной симметрии это число может быть еще меньше, поскольку некоторые из компонент обращаются в нуль, а другие могут быть связаны друг с другом.

Геометрическое представление векторов и тензоров

Вектор, такой как, например, электрическое поле, описывает физическую переменную, обладающую величиной и направлением. *Геометрически* он изображается стрелкой, указывающей направление, а длина стрелки равна величине вектора (рис. 6.18, *а*). Вектор является тензором первого ранга и *численно* задается в виде набора из трех величин — проекций вектора на оси некоторой системы координат. Хотя эти проекции зависят от выбора системы координат, величина и направление вектора не зависят от такого выбора.

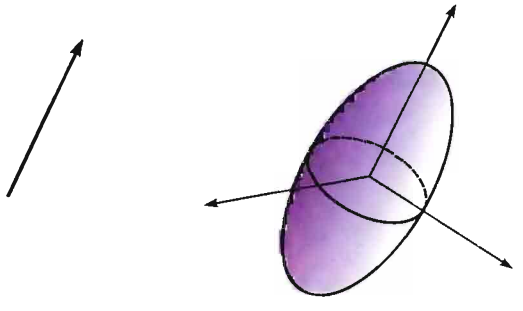
Тензор второго ранга определяет правило, связывающее два вектора. В заданной системе координат он представляется *численно* с помощью девяти чисел. При изменении системы координат мы получим другой набор из девяти чисел, однако физическая сущность правила остается неизменной. Полезное геометрическое представление (рис. 6.18, *б*) симметричного тензора второго ранга (например, тензора диэлектрической проницаемости ε) — поверхность второго порядка (эллипсоид), определяемая уравнением

$$\sum_{ij} \varepsilon_{ij} x_i x_j = 1. \tag{6.46}$$

такое представление называется **поверхностью второго порядка**. Эта поверхность не зависит от выбора системы координат; при повороте системы координат меняются как x_i, так и ε_{ij}, но эллипсоид остается на месте в физическом пространстве. Эллипсоид имеет шесть степеней свободы и несет всю информацию о симметричном тензоре второго ранга. В системе главных осей ε_{ij} диагонально, и уравнение эллипсоида принимает особенно простой вид

$$\varepsilon_1 x_1^2 + \varepsilon_2 x_2^2 + \varepsilon_3 x_3^2 = 1. \tag{6.47}$$

Главные оси эллипсоида являются главными осями тензора, длины полуосей составляют $1/\sqrt{\varepsilon_1}$, $1/\sqrt{\varepsilon_2}$ и $1/\sqrt{\varepsilon_3}$



а *б*

Рис. 6.18. Геометрическое представление вектора (*а*) и симметричного тензора второго ранга (*б*)

Главные оси и главные значения показателя преломления

Элементы тензора диэлектрической проницаемости зависят от того, как выбрана система координат относительно структуры кристалла. Однако всегда может быть найдена такая система координат, в которой недиагональные элементы ε_{ij} исчезают, так что

$$D_1 = \varepsilon_1 E_1, \quad D_2 = \varepsilon_2 E_2, \quad D_3 = \varepsilon_3 E_3, \tag{6.48}$$

где $\varepsilon_1 = \varepsilon_{11}$, $\varepsilon_2 = \varepsilon_{22}$ и $\varepsilon_3 = \varepsilon_{33}$. Согласно (6.44), $\boldsymbol{E}$ и $\boldsymbol{D}$ параллельны вдоль этих особых направлений, так что, например, если $\boldsymbol{E}$ направлено по оси x, то так же направлено и $\boldsymbol{D}$. Эта система координат определяет главные оси и главные плоскости кристалла. В оставшейся части данной главы координатные оси x, y и z, которые также обозначаются x_1, x_2 и x_3, предполагаются совпадающими с главными осями кристалла. Такой выбор упрощает весь анализ без потери общности. Диэлектрические проницаемости ε_1, ε_2 и ε_3 соответствуют показателям преломления

$$n_1 = \sqrt{\frac{\varepsilon_1}{\varepsilon_0}}; \quad n_2 = \sqrt{\frac{\varepsilon_2}{\varepsilon_0}}; \quad n_3 = \sqrt{\frac{\varepsilon_3}{\varepsilon_0}}, \tag{6.49}$$

где ε_0 — диэлектрическая проницаемость свободного пространства; эти показатели называются **главными значениями показателя преломления**.

Двуосные, одноосные и изотропные кристаллы

Кристаллы, в которых все три главных значения показателя преломления различны, называются **двуосными**. У некоторых кристаллов, обладающих определенной симметрией, например одной осью симметрии третьего, четвертого или шестого порядка, два из трех главных значений показателя преломления равны ($n_1 = n_2$), тогда кристалл называется **одноосным**. В этом случае обычно вводятся обозначения $n_1 = n_2 = n_o$ и $n_3 = n_н$, n_o называется **обыкновенным**, а $n_н$ — **необыкновенным** показателем преломления, при этом кристалл называется **положительным одноосным**, если $n_o < n_н$, и **отрицательным одноосным** — в противном случае. Ось z одноосного кристалла называется его **оптической осью**. В некоторых кристаллах с еще более высокой симметрией (например, в кристаллах с кубической решеткой) все три показателя равны и среда оптически изотропна.

Тензор диэлектрической непроницаемости

Соотношение $\boldsymbol{D} = \varepsilon \boldsymbol{E}$ можно обратить и записать в виде $\boldsymbol{E} = \varepsilon^{-1}\boldsymbol{D}$, где ε^{-1} — обратный тензор по отношению к ε. Полезно ввести тензор диэлектрической непроницаемости $\eta = \varepsilon_0\varepsilon^{-1}$ (не путать с импедансом среды η), так что $\varepsilon_0\boldsymbol{E} = \eta\boldsymbol{D}$. Так как тензор ε симметричен, симметричен и η. Оба тензора, ε и η, имеют общие главные оси. В системе главных осей тензор η диагонален, его главные значения равны

$$\frac{\varepsilon_0}{\varepsilon_1} = \frac{1}{n_1^2}; \quad \frac{\varepsilon_0}{\varepsilon_2} = \frac{1}{n_2^2}; \quad \frac{\varepsilon_0}{\varepsilon_3} = \frac{1}{n_3^2}.$$

Каждый из тензоров, ε или η, в отдельности полностью описывает оптические свойства кристалла.

Эллипсоид показателей преломления

Эллипсоид показателей преломления (называемый также **оптической индикатрисой**) — представление второго порядка тензора диэлектрической непроницаемости $\eta = \varepsilon_0\varepsilon^{-1}$:

$$\sum_{ij} \eta_{ij} x_i x_j = 1, \quad i, j = 1,\ 2,\ 3. \tag{6.50}$$

В главных осях его уравнение имеет вид

$$\boxed{\frac{x_1^2}{n_1^2} + \frac{x_2^2}{n_2^2} + \frac{x_3^2}{n_3^2} = 1} \tag{6.51}$$

Эллипсоид показателей преломления

с главными значениями $1/n_1^2$, $1/n_2^2$ и $1/n_3^2$ и полуосями n_1, n_2 и n_3.

Следовательно, оптические свойства кристалла (направления главных осей и главные значения показателя преломления) полностью описываются заданием эллипсоида показателей преломления (рис. 6.19). Для одноосных кристаллов эллипсоид показателей преломления является эллипсоидом вращения, для изотропных сред он становится сферой.

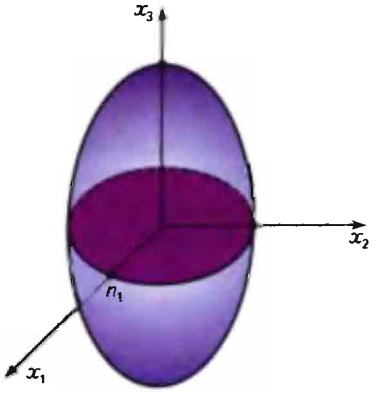


Рис. 6.19. Эллипсоид показателей преломления. Координаты (x_1, x_2, x_3) — главные оси; (n_1, n_2, n_3) — главные значения показателя преломления кристалла

6.3.2. Распространение вдоль главной оси

Правила, которым подчиняется распространение света в кристаллах, в общем случае весьма сложны. Однако они становятся сравнительно простыми, если свет представляет собой плоскую волну, бегущую вдоль одной из главных осей кристалла. Мы начнем именно с этого случая.

Нормальные моды

Пусть x—y—z — система координат, оси которой совпадают с главными осями кристалла. Плоская волна, бегущая вдоль оси z, линейно поляризована вдоль оси x и распространяется с фазовой скоростью c_0/n_1 (волновое число $k = n_1 k_0$) (рис. 6.20, *а*) без изменения поляризации. Последнее следует из того, что электрическое поле имеет всего одну компоненту E_1, направленную вдоль x, так что вектор $\boldsymbol{D}$ также имеет единственную компоненту $D_1 = \varepsilon_1 E_1$; волновое уравнение, вытекающее из уравнений Максвелла, обеспечивает скорость света

$$\frac{1}{\sqrt{\mu_0 \varepsilon_1}} = \frac{c_0}{n_1}.$$

Аналогично, плоская волна, бегущая вдоль оси z и поляризованная в направлении y (рис. 6.20, *б*), распространяется со скоростью c_0/n_2, соответствующей показателю преломления n_2. Таким образом, нормальные моды, распространяющиеся вдоль оси z, — это волны, линейно поляризованные вдоль направлений x и y. Эти волны называются нормальными модами потому, что их скорости и поляризации сохраняются в процессе распространения (см. приложение В). Другие случаи распространения вдоль главной оси волн, поляризованных вдоль другой главной оси, рассматриваются аналогично (рис. 6.20, *в*).

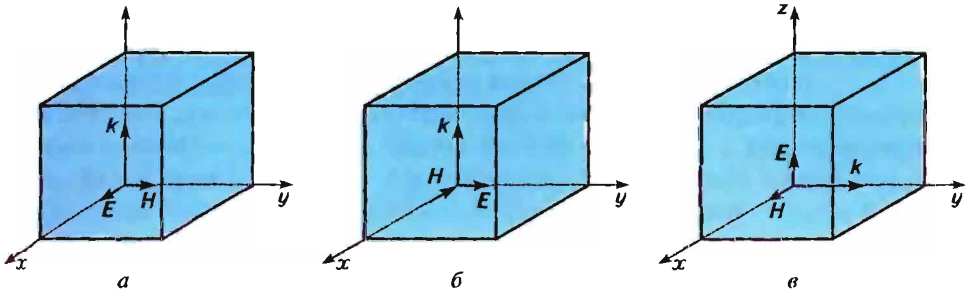


Рис. 6.20. Волна, распространяющаяся вдоль главной оси и поляризованная вдоль другой главной оси, имеет фазовую скорость c_0/n_1, c_0/n_2 или c_0/n_3, когда электрический вектор направлен по осям x, y или z соответственно:
$а$ — $k = n_1k_0$; $б$ — $k = n_2k_0$; $в$ — $k = n_3k_0$

Поляризация в произвольном направлении

Рассмотрим теперь волну, бегущую вдоль одной из главных осей (например, z) и линейно поляризованную в произвольном направлении в плоскости x—y. Для рассмотрения этого случая нужно представить волну как сумму нормальных мод, т. е. волн, поляризованных вдоль осей x и y. Эти две компоненты распространяются с различными фазовыми скоростями c_0/n_1 и c_0/n_2 соответственно. Следовательно, они приобретают различные фазовые сдвиги $\varphi_x = n_1k_0d$ и $\varphi_y = n_2k_0d$, соответственно, после прохождения расстояния d. Разность фаз между ними получается равной $\varphi = \varphi_y - \varphi_x = (n_1 - n_2)k_0d$. Сложение этих волн дает эллиптически поляризованную волну, как пояснялось в разд. 6.1 и показано на рис. 6.21. Такой кристалл может служить **фазовой пластинкой** — устройством, в котором две ортогонально поляризованные волны распространяются с различной фазовой скоростью, так, что одна из волн замедляется по отношению к другой (см. рис. 6.8).

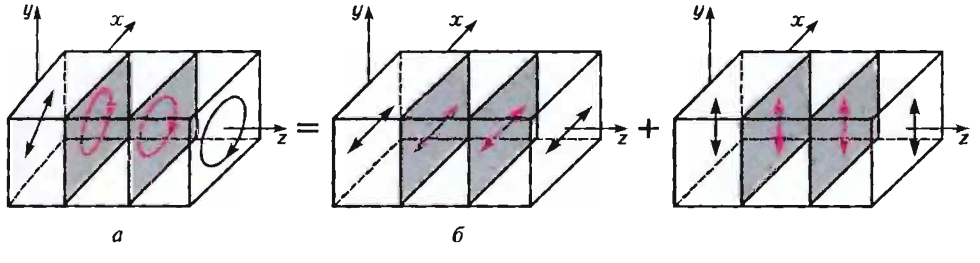


Рис. 6.21. Волна, линейно поляризованная под углом 45° в плоскости $z = 0$ ($а$), разлагается на сумму двух компонент, линейно поляризованных по осям x и y, которые распространяются со скоростями c_0/n_1 и c_0/n_2 [($б$) и ($в$) соответственно]. В результате фазовой задержки волна из плоско поляризованной преобразуется в эллиптически поляризованную ($а$). Ясно, таким образом, что исходная линейно поляризованная волна не является нормальной модой системы

6.3.3. Распространение в произвольном направлении

Теперь рассмотрим общий случай плоской волны, распространяющейся в анизотропном кристалле в произвольном направлении, заданном единичным вектором $\hat{\boldsymbol{u}}$. Покажем, что две нормальные моды — линейно поляризованные волны. Показатели преломления n_a и n_b и направления поляризации этих мод можно определить с помощью процедуры, основанной на использовании эллипсоида показателей преломления.

Построение эллипсоида показателей преломления для определения нормальных мод

На рис. 6.22 показано геометрическое построение для определения поляризаций и показателей преломления n_a и n_b нормальных мод волны, распространяющейся в направлении единичного вектора $\hat{\boldsymbol{u}}$ в анизотропном материале, характеризуемом эллипсоидом показателей преломления

$$\frac{x_1^2}{n_1^2}+\frac{x_2^2}{n_2^2}+\frac{x_3^2}{n_3^2}=1.$$

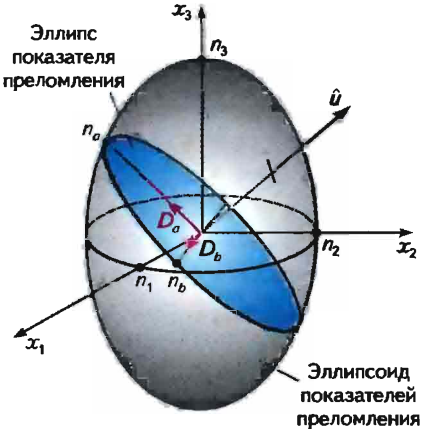


Рис. 6.22. Определение нормальных мод из эллипсоида показателей преломления

- Строим плоскость, проходящую через центр эллипсоида показателей преломления и перпендикулярную $\hat{\boldsymbol{u}}$. Пересечение плоскости с эллипсоидом дает **эллипс показателей преломления.**
- Большая и малая полуоси эллипса показателей преломления равны показателям преломления n_a и n_b двух нормальных мод.
- Направления большой и малой оси этого же эллипса — направления векторов $\boldsymbol{D}_a$ и $\boldsymbol{D}_b$ нормальных мод. Эти направления ортогональны.
- Векторы $\boldsymbol{E}_a$ и $\boldsymbol{E}_b$ можно найти из $\boldsymbol{D}_a$ и $\boldsymbol{D}_b$ с помощью (6.48).

Доказательство построения эллипсоида показателей преломления для нахождения нормальных мод

Для определения нормальных мод (см. подраздел 6.1.2) плоской волны, распространяющейся в направлении $\hat{\boldsymbol{u}}$, обратимся к уравнениям Максвелла (5.45)—(5.48) и материальному уравнению $\boldsymbol{D}=\varepsilon\boldsymbol{E}$ (6.45), чтобы сформу-

лировать задачу на собственные значения. Поскольку предполагается, что все поля изменяются в пространстве по закону $\exp(-j\boldsymbol{k}\cdot\boldsymbol{r})$, где $\boldsymbol{k} = k\hat{\boldsymbol{u}}$, то уравнения Максвелла (5.65) и (5.66) сводятся к виду

$$\boldsymbol{k}\times\boldsymbol{H} = -\omega\boldsymbol{D}; \tag{6.52}$$

$$\boldsymbol{k}\times\boldsymbol{E} = \omega\mu_0\boldsymbol{H}. \tag{6.53}$$

Подстановка (6.53) в (6.52) приводит к соотношению

$$\boldsymbol{k}\times(\boldsymbol{k}\times\boldsymbol{E}) = -\omega^2\mu_0\boldsymbol{D}. \tag{6.54}$$

Используя $\boldsymbol{E} = \varepsilon^{-1}\boldsymbol{D}$, получаем

$$\boldsymbol{k}\times(\boldsymbol{k}\times\varepsilon^{-1}\boldsymbol{D}) = -\omega^2\mu_0\boldsymbol{D}. \tag{6.55}$$

Это уравнение на собственные значения, которому должен удовлетворять вектор $\boldsymbol{D}$. Работать с $\boldsymbol{D}$ удобно, потому что мы знаем, что он лежит в плоскости, нормальной по отношению к направлению распространения волны $\hat{\boldsymbol{u}}$.

Теперь упростим (6.55) с использованием соотношений

$$\eta = \varepsilon_0\varepsilon^{-1};\quad \boldsymbol{k} = k\hat{\boldsymbol{u}};\quad n = \frac{k}{k_0};\quad k_0^2 = \omega^2\mu_0\varepsilon_0,$$

что дает

$$-\hat{\boldsymbol{u}}\times(\hat{\boldsymbol{u}}\times\eta\boldsymbol{D}) = \frac{1}{n^2}\boldsymbol{D}. \tag{6.56}$$

Операция $-\hat{\boldsymbol{u}}\times(\hat{\boldsymbol{u}}\times\eta\boldsymbol{D})$ может быть интерпретирована как проектирование вектора $\eta\boldsymbol{D}$ на плоскость, перпендикулярную $\hat{\boldsymbol{u}}$. Таким образом, перепишем (6.56) в виде

$$\boldsymbol{P}_u\eta\boldsymbol{D} = \frac{1}{n^2}\boldsymbol{D}, \tag{6.57}$$

где $\boldsymbol{P}_u$ — оператор проектирования. Уравнение (6.57) — это уравнение на собственные значения для оператора $\boldsymbol{P}_u\eta$ с собственным значением $1/n^2$ и собственным вектором $\boldsymbol{D}$. Два собственных значения $1/n_a^2$ и $1/n_b^2$ и два собственных вектора $\boldsymbol{D}_a$ и $\boldsymbol{D}_b$ представляют две нормальные моды.

Задача на собственные значения (6.57) имеет простую геометрическую интерпретацию. Тензор η геометрически представляется квадратичным представлением — эллипсоидом показателей преломления. Оператор $\boldsymbol{P}_u\eta$ осуществляет проектирование на плоскость, нормальную по отношению к $\hat{\boldsymbol{u}}$. Решение задачи на собственные значения, таким образом, эквивалентно нахождению главных осей эллипса, образованного пересечением плоскости, нормальной к $\hat{\boldsymbol{u}}$, с эллипсоидом показателей преломления. Это и есть построение для нахождения нормальных мод, показанное на рис. 6.22.

Специальный случай: одноосный кристалл

В одноосном кристалле ($n_1 = n_2 = n_o$, $n_3 = n_н$) эллипсоид показателей преломления, показанный на рис 6.22, является эллипсоидом вращения. Для волны, у которой направление распространения $\hat{u}$ образует угол θ с оптической осью, эллипс показателей преломления имеет полуоси n_o и $n(\theta)$, где

$$\frac{1}{n^2(\theta)} = \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_н^2}, \qquad (6.58)$$

Показатель преломления необыкновенной волны

так что нормальным модам соответствуют показатели преломления $n_b = n_o$ и $n_a = n(\theta)$. Первая мода, называемая **обыкновенной волной**, имеет показатель преломления n_o независимо от θ. В соответствии с рис. 6.23 вторая мода, называемая **необыкновенной волной**, имеет показатель преломления $n(\theta)$, который меняется от n_o при $\theta = 0°$ до $n_н$ при $\theta = 90°$ Вектор $\boldsymbol{D}$ обыкновенной волны перпендикулярен плоскости, образованной оптической осью (осью z) и направлением распространения $\boldsymbol{k}$, причем векторы $\boldsymbol{D}$ и $\boldsymbol{E}$ параллельны. У необыкновенной же волны вектор $\boldsymbol{D}$ перпендикулярен $\boldsymbol{k}$ и лежит в плоскости k—z, при этом $\boldsymbol{D}$ и $\boldsymbol{E}$ не параллельны, что видно на рис. 6.23.

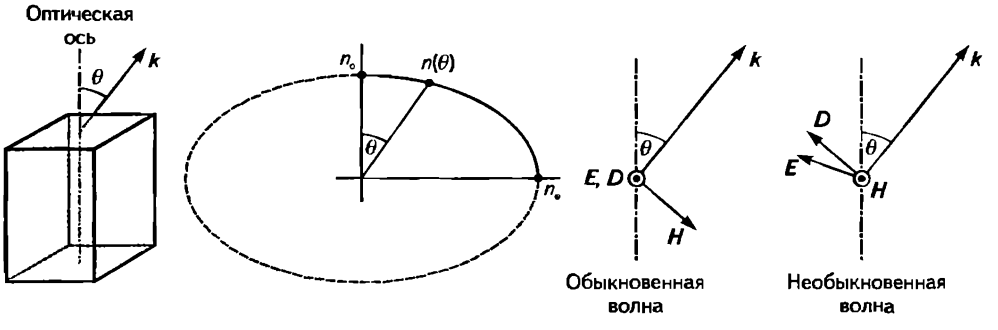


Рис. 6.23. Изменение показателя преломления $n(\theta)$ необыкновенной волны с ростом θ (угол между направлением распространения и оптической осью) в одноосном кристалле и направления электрического и магнитного полей в обыкновенной и необыкновенной волнах. Кружок с точкой в центре означает вектор, направленный перпендикулярно плоскости чертежа в сторону читателя

6.3.4. Дисперсионные соотношения, лучи, волновые фронты и перенос энергии

Перейдем к исследованию других свойств волн в анизотропных средах, включающих дисперсионные соотношения (связь между ω и $\boldsymbol{k}$).

Световая волна характеризуется волновым вектором $\boldsymbol{k}$, векторами полей $\boldsymbol{E}$, $\boldsymbol{D}$, $\boldsymbol{H}$ и $\boldsymbol{B}$, а также комплексным вектором Пойнтинга $\boldsymbol{S} = (\boldsymbol{E} \times \boldsymbol{H}^*)/2$ (плотнос-

тью потока энергии). Эти векторы связаны соотношениями (6.52), (6.53). Из (6.52) следует, что вектор $\boldsymbol{D}$ перпендикулярен как $\boldsymbol{k}$, так и $\boldsymbol{H}$. Из (6.53) аналогичным образом следует, что вектор $\boldsymbol{H}$ перпендикулярен как $\boldsymbol{k}$, так и $\boldsymbol{E}$. Эти геометрические условия иллюстрирует рис. 6.24, на котором показан также вектор Пойнтинга $\boldsymbol{S}$, перпендикулярный $\boldsymbol{E}$ и $\boldsymbol{H}$. Таким образом, $\boldsymbol{D}$, $\boldsymbol{E}$, $\boldsymbol{k}$ и $\boldsymbol{S}$ лежат в одной плоскости, по отношению к которой $\boldsymbol{H}$ и $\boldsymbol{B}$ перпендикулярны. В этой плоскости $\boldsymbol{D} \perp \boldsymbol{k}$ и $\boldsymbol{S} \perp \boldsymbol{E}$, но вектор $\boldsymbol{D}$ не обязательно параллелен $\boldsymbol{E}$, а $\boldsymbol{S}$ не обязательно параллелен $\boldsymbol{k}$.

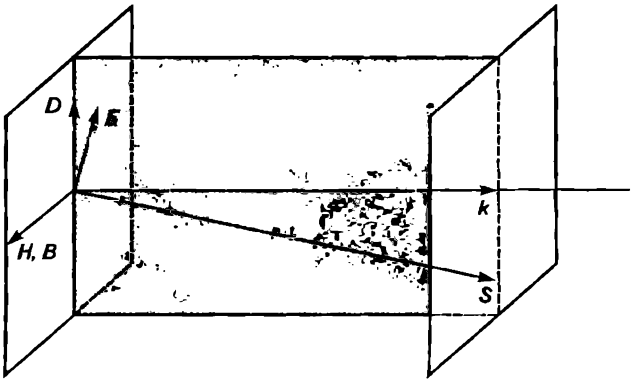


Рис. 6.24. Векторы $\boldsymbol{D}$, $\boldsymbol{E}$, $\boldsymbol{k}$ и $\boldsymbol{S}$ лежат в одной плоскости, по отношению к которой векторы $\boldsymbol{H}$ и $\boldsymbol{B}$ перпендикулярны: $\boldsymbol{D} \perp \boldsymbol{k}$ и $\boldsymbol{S} \perp \boldsymbol{E}$

Используя соотношение $\boldsymbol{D} = \varepsilon \boldsymbol{E}$ в (6.54), получаем

$$\boldsymbol{k} \times (\boldsymbol{k} \times \boldsymbol{E}) + \omega^2 \mu_0 \varepsilon \boldsymbol{E} = 0. \tag{6.59}$$

Это векторное уравнение, которому должен удовлетворять вектор $\boldsymbol{E}$, можно записать как систему трех линейных однородных уравнений для составляющих E_1, E_2 и E_3 вдоль главных осей в матричном виде

$$\begin{bmatrix} n_1^2 k_0^2 - k_2^2 - k_3^2 & k_1 k_2 & k_1 k_3 \\ k_2 k_1 & n_2^2 k_0^2 - k_1^2 - k_3^2 & k_2 k_3 \\ k_3 k_1 & k_3 k_2 & n_3^2 k_0^2 - k_1^2 - k_2^2 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{6.60}$$

где (k_1, k_2, k_3) — компоненты $\boldsymbol{k}$; $k_0 = \omega/c_0$; (n_1, n_2, n_3) — главные значения показателя преломления (6.49). Условие нетривиальной совместности этой системы состоит в равенстве нулю ее определителя. В результате получается уравнение, связывающее ω с k_1, k_2, k_3 и имеющее вид $\omega = \omega(k_1, k_2, k_3)$, где $\omega(k_1, k_2, k_3)$ — нелинейная функция. Это соотношение, называемое **дисперсионным соотношением**, представляет собой уравнение поверхности в пространстве с координатами k_1, k_2, k_3, которая называется **поверхностью нормалей**, или k-поверхностью. Пересечение направления $\hat{\boldsymbol{u}}$ с k-поверхностью определяет вектор $\boldsymbol{k}$, длина которого $k = n\,\omega/c_0$ дает показатель преломления n. Существует два таких пересечения, соответствующих двум нормальным модам, связанным с каждым направлением.

k-Поверхность является центрально-симметричной поверхностью, состоящей из двух листов, каждый из которых соответствует решению (нормальной моде). Можно показать, что k-поверхность пересекает каждую из главных плоскостей по эллипсу и по окружности, как показано на рис. 6.25. Для двуосных кристаллов ($n_1 < n_2 < n_3$) два листа пересекаются в четырех точках, определяющих две оптические оси. В случае одноосного кристалла ($n_1 = n_2 = n_o$, $n_3 = n_н$) листы k-поверхности превращаются в сферу и эллипсоид вращения, которые касаются друг друга только в двух точках, определяющих единственную оптическую ось (ось z). В изотропном случае ($n_1 = n_2 = n_3 = n$) два листа вырождаются в одну сферу.

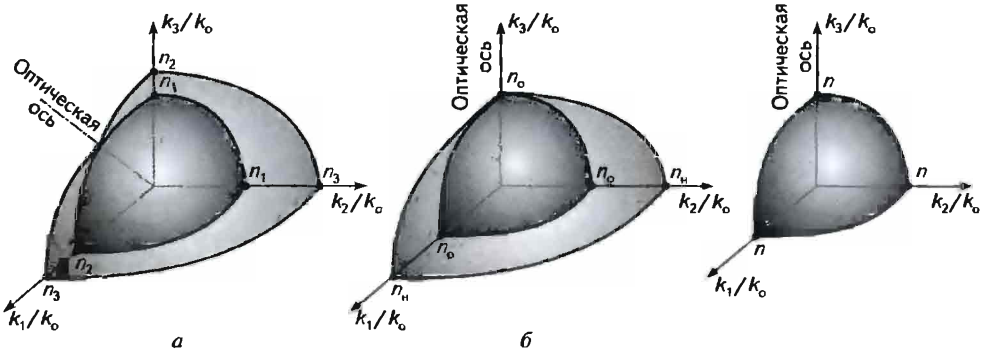


Рис. 6.25. Один октант k-поверхности для двуосного ($n_1 < n_2 < n_3$) (*а*), одноосного ($n_1 = n_2 = n_o$, $n_3 = n_н$) (*б*) и изотропного ($n_1 = n_2 = n_3 = n$) (*в*) кристаллов

Пересечение направления $\acute{u} = (u_1, u_2, u_3)$ с k-поверхностью соответствует волновому числу, которое удовлетворяет уравнению

$$\sum_{j=1,\,2,\,3} \frac{u_j^2 k^2}{k^2 - n_j^2 k_0^2} = 1. \tag{6.61}$$

Это уравнение четвертого порядка по k (или второго порядка по k^2). Оно имеет четыре решения, $\pm k_a$ и $\pm k_b$, из которых только два положительных имеют смысл, поскольку отрицательным соответствует обратное направление распространения. Таким образом, задача решена: волновые числа нормальных мод равны k_a и k_b, а показатели преломления $n_a = k_a/k_0$ и $n_b = k_b/k_0$.

Для определения направлений поляризации нормальных мод определим компоненты $(k_1, k_2, k_3) = (ku_1, ku_2, ku_3)$ и элементы матрицы, входящей в (6.60) для каждого из волновых чисел $k = k_a$ и $k = k_b$. Затем решим два из трех уравнений (6.60) и найдем отношения E_1/E_3 и E_2/E_3, из которых определим направление соответствующего электрического поля $\boldsymbol{E}$.

Природа волн в анизотропных средах наилучшим образом объясняется при изучении k-поверхности $\omega = \omega(k_1, k_2, k_3)$, получаемой путем приравнивания к нулю определителя матрицы (6.60), как показано на рис. 6.25. Изменение фа-

зовой скорости $c = \omega/k$ в зависимости от направления $\hat{\boldsymbol{u}}$ можно определить с помощью $\boldsymbol{k}$-поверхности: расстояние от начала координат до $\boldsymbol{k}$-поверхности в направлении $\hat{\boldsymbol{u}}$ обратно пропорционально фазовой скорости.

Групповую скорость также можно определить с помощью $\boldsymbol{k}$-поверхности. По аналогии с групповой скоростью $v = \mathrm{d}\omega/\mathrm{d}k$, характеризующей распространение световых импульсов (волновых пакетов), обсуждавшееся в разд. 5.6, групповая скорость для лучей (локализованных пучков или пространственных волновых пакетов) есть вектор $\boldsymbol{v} = \nabla_k \omega(\boldsymbol{k})$ градиента ω по $\boldsymbol{k}$. Поскольку $\boldsymbol{k}$-поверхность есть поверхность $\omega(k_1, k_2, k_3) = \mathrm{const}$, вектор $\boldsymbol{v}$ должен быть перпендикулярен к этой поверхности. Таким образом, лучи распространяются по нормали к $\boldsymbol{k}$-поверхности. Волновые фронты перпендикулярны вектору $\boldsymbol{k}$, поскольку фаза волны содержит $\boldsymbol{k} \cdot \boldsymbol{r}$. Следовательно, нормали к волновым фонтам параллельны волновому вектору $\boldsymbol{k}$.

Комплексный вектор Пойнтинга $\boldsymbol{S} = (\boldsymbol{E} \times \boldsymbol{H}^*)/2$ также направлен по нормали к $\boldsymbol{k}$-поверхности. Это можно показать, выбрав некоторое значение ω и рассматривая два вектора $\boldsymbol{k}$ и $\boldsymbol{k} + \Delta\boldsymbol{k}$, концы которых лежат на $\boldsymbol{k}$-поверхности. Беря дифференциал от (6.52) и (6.53) и используя векторные тождества, можно показать, что $\Delta\boldsymbol{k} \cdot \boldsymbol{S} = 0$, так что $\boldsymbol{S}$ перпендикулярен $\boldsymbol{k}$-поверхности. Следовательно, $\boldsymbol{S}$ также параллелен вектору групповой скорости $\boldsymbol{v}$.

Если $\boldsymbol{k}$-поверхность представляет собой сферу, как в изотропных средах, то векторы $\boldsymbol{v}$, $\boldsymbol{S}$ и $\boldsymbol{k}$ параллельны. В этом случае лучи параллельны волновому вектору $\boldsymbol{k}$, и поток энергии течет в том же направлении, как показано на рис. 6.26, *а*. Если же $\boldsymbol{k}$-поверхность не перпендикулярна по отношению к волновому вектору $\boldsymbol{k}$, как показано на рис 6.26, *б*, то лучи и направления распространения энергии не ортогональны по отношению к волновым фронтам. Лучи в этом случае обладают «необыкновенным» свойством: они направлены наклонно по отношению к волновым фронтам (см. рис. 6.26, *б*).

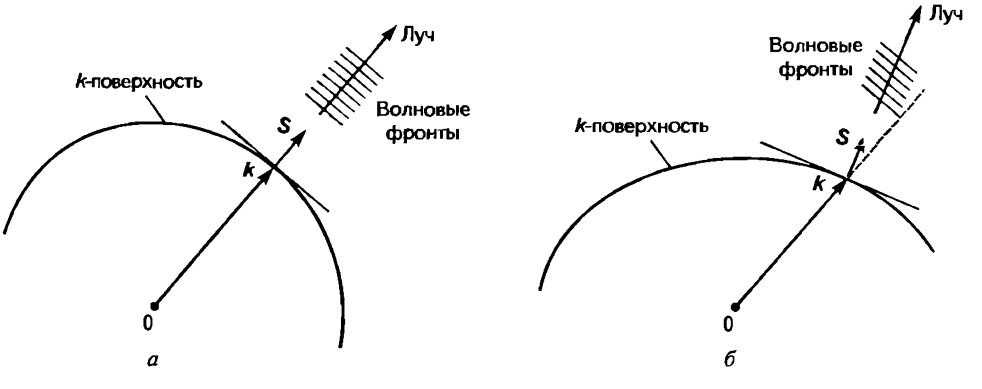


Рис. 6.26. Лучи и волновые фронты для сферической (*а*) и для несферической (*б*) $\boldsymbol{k}$-поверхностей:

а — обыкновенный луч; *б* — необыкновенный луч

Особый случай: одноосные кристаллы

В одноосных кристаллах ($n_1 = n_2 = n_o$, $n_3 = n_н$) уравнение k-поверхности $\omega = \omega(k_1, k_2, k_3)$ сводится к упрощенному виду

$$\left(k^2 - n_0^2 k_0^2\right)\left(\frac{k_1^2 + k_2^2}{n_н^2} + \frac{k_3^2}{n_o^2} - k_0^2\right) = 0. \tag{6.62}$$

Это уравнение имеет два решения: сфера, соответствующая обращению в нуль первого сомножителя

$$k = n_o k_0 \tag{6.63}$$

и эллипсоид вращения, соответствующий обращению в нуль второго сомножителя

$$\frac{k_1^2 + k_2^2}{n_н^2} + \frac{k_3^2}{n_o^2} = k_0^2. \tag{6.64}$$

В силу аксиальной симметрии относительно оси z (оптической оси) без потери общности можно предположить, что вектор $\boldsymbol{k}$ лежит в плоскости y—z. Его направление тогда характеризуется углом наклона θ по отношению к оптической оси. Удобно изобразить сечение $\boldsymbol{k}$-поверхности плоскостью y—z, имеющее вид окружности и эллипса (рис. 6.27).

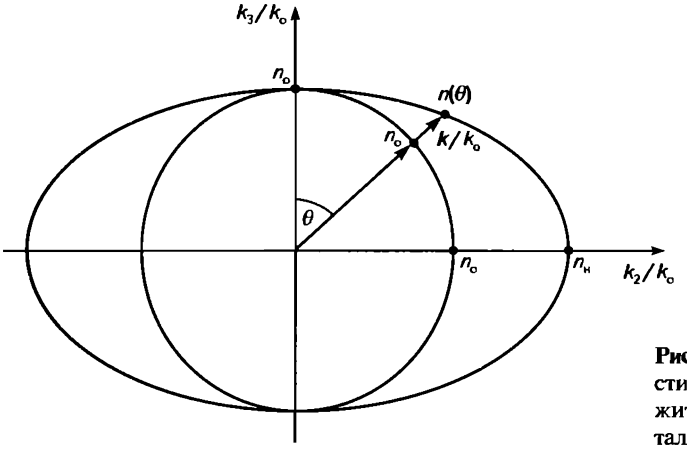


Рис. 6.27. Сечение $\boldsymbol{k}$-поверхности плоскостью y—z для положительного одноосного кристалла ($n_н > n_o$)

При заданном направлении $\hat{\boldsymbol{u}}$ вектора $\boldsymbol{k}$ волновое число k определяется положением точки пересечения с листами $\boldsymbol{k}$-поверхности. Два решения определяют две нормальные моды — обыкновенную и необыкновенную волны. Обыкновенная волна имеет волновое число $k = n_o k_0$, независимо от направления $\boldsymbol{u}$, тогда как необыкновенная волна имеет волновое число $n(\theta)k_0$, где $n(\theta)$ определяется уравнением (6.58). Таким образом, подтверждаются результаты, полученные ранее с помощью геометрического построения для эллипсоида

показателей преломления. Направления лучей, волновых фронтов, потока энергии и векторов поля $\boldsymbol{E}$ и $\boldsymbol{D}$ для обыкновенной и необыкновенной волн в одноосном кристалле показаны на рис. 6.28.

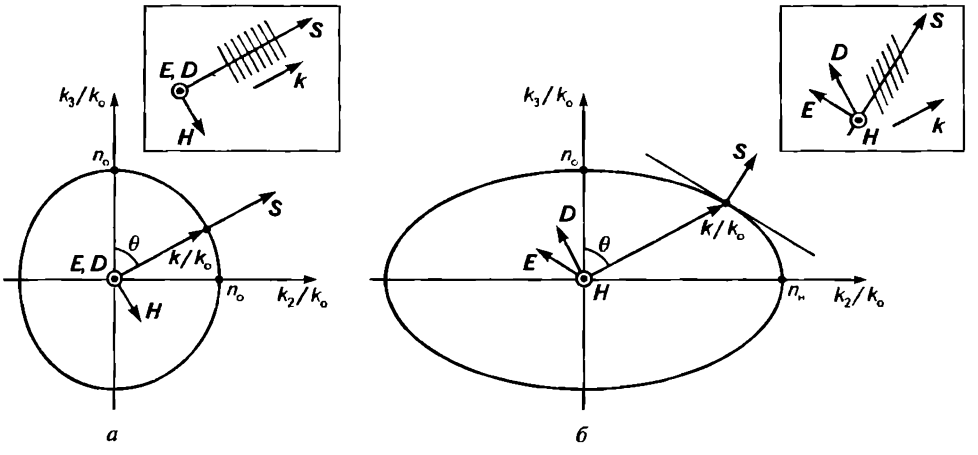


Рис. 6.28. Нормальные моды для плоской волны, распространяющейся в направлении $\boldsymbol{k}$, образующем угол θ с оптической осью z в одноосном кристалле, представляют собой:

а — обыкновенную волну с показателем преломления n_o, поляризованную в направлении, перпендикулярном плоскости k—z; *б* — необыкновенную волну с показателем преломления $n(\theta)$ [определяемым из уравнения (6.58)], поляризованную в плоскости k—z вдоль направления, касательного к эллипсу (k-поверхности) в точке ее пересечения с k. Эта волна является «необыкновенной» по следующим признакам: D не параллельно E, но оба вектора лежат в плоскости k—z; S не параллельно k, так что энергия не распространяется вдоль направления k; лучи, следовательно, не ортогональны волновым фронтам, так что волна распространяется «наискосок»

6.3.5. Двулучепреломление

Преломление плоских волн

Перейдем к исследованию преломления плоской волны на границе раздела между изотропной средой (например, воздухом, $n = 1$) и анизотропной средой (кристаллом). Основной принцип, которому подчиняется преломление волн в данной ситуации, — необходимость согласования волновых фронтов падающей и преломленной волн. Поскольку в анизотропной среде существует две нормальные моды с различными фазовыми скоростями и, следовательно, разными показателями преломления, падающая волна создает две преломленные волны с разными направлениями распространения и разными поляризациями. Это явление известно как **двулучепреломление**.

Фазовый синхронизм требует, чтобы выполнялся закон Снелла, т. е.

$$k_0 \sin\theta_1 = k \sin\theta, \tag{6.65}$$

где θ_1 и θ — углы падения и преломления соответственно. В анизотропной среде волновое число $k = n(\theta)k_0$ само является функцией θ, так что

$$\sin\theta_1 = n(\theta_a + \theta)\sin\theta, \tag{6.66}$$

где θ_a — угол между оптической осью и нормалью к поверхности раздела, так что $\theta_a + \theta$ представляет собой угол между преломленным лучом и оптической осью. Уравнение (6.66) представляет собой модифицированный закон Снелла. Для решения (6.65) построим сечение k-поверхности плоскостью падения и будем искать угол θ, для которого (6.65) удовлетворяется. Ожидается получить два решения, соответствующие двум нормальным модам. Поляризационное состояние падающего света даст распределение энергии между двумя преломленными волнами.

Возьмем, например, одноосный кристалл и пусть плоскость падения параллельна оптической оси. Пересечение k-поверхности с плоскостью падения даст окружность и эллипс (рис. 6.29). Две преломленные волны, удовлетворяющие условиям фазового синхронизма, определяются из (6.66):

- обыкновенная волна с ортогональной поляризацией (TE) под углом $\theta = \theta_o$, для которого

$$\sin\theta_1 = n_o \sin\theta_o; \tag{6.67}$$

- необыкновенная волна с параллельной поляризацией (TM) под углом $\theta = \theta_н$, для которого

$$\sin\theta_1 = n(\theta_a + \theta_н)\sin\theta_н, \tag{6.68}$$

где $n(\theta)$ определяется уравнением (6.58).

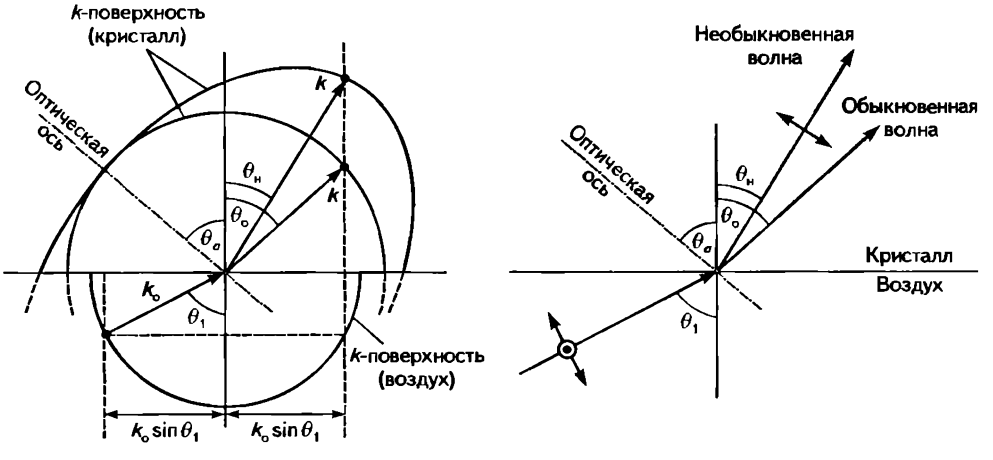


Рис. 6.29. Определение углов преломления путем согласования проекций векторов в воздухе и в одноосном кристалле

Если в падающей волне присутствуют две поляризации, то появятся две преломленные волны, как показано на рис. 6.29.

Преломление лучей

Только что проведенный анализ относился к преломлению плоских волн. Преломление лучей в анизотропной среде происходит иначе, поскольку лучи не обязательно направлены по нормали к волновым фронтам. В воздухе, прежде чем войти в кристалл, волновые фронты перпендикулярны лучам. Преломленная волна должна иметь волновой вектор, удовлетворяющий условию фазового синхронизма, так что можно применять закон Снелла (6.66) и угол преломления θ определяет направление $\boldsymbol{k}$. Однако, поскольку направление $\boldsymbol{k}$ не есть направление луча, закон Снелла неприменим к лучам в анизотропной среде.

Особенно яркий пример отклонения от закона Снелла — нормальное падение на одноосный кристалл, оптическая ось которого не параллельна и не перпендикулярна поверхности кристалла. Волновой вектор падающей волны $\boldsymbol{k}$ перпендикулярен поверхности. Для обеспечения фазового синхронизма волновые векторы преломленных волн должны иметь то же направление. Пересечение выбранного направления с $\boldsymbol{k}$-поверхностью дает две точки, соответствующие двум волнам. Обыкновенный луч параллелен $\boldsymbol{k}$. Однако необыкновенный луч направлен по нормали к $\boldsymbol{k}$-поверхности, т. е. под углом θ_s к нормали, проведенной к поверхности кристалла, как показано на рис. 6.30. Таким образом, нормальное падение создает наклонное преломление. При этом, однако, принцип фазового синхронизма работает, и волновые фронты обеих преломленных волн параллельны поверхности кристалла и волновому фронту падающего луча.

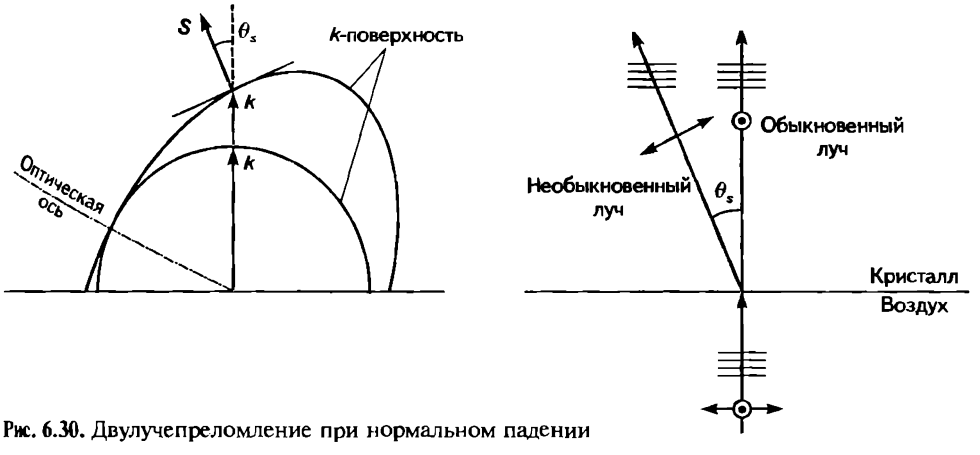


Рис. 6.30. Двулучепреломление при нормальном падении

При пропускании световых лучей через пластину из анизотропного материала, как описано выше, два луча, преломленные входной поверхностью, снова преломляются на второй поверхности и приобретают первоначальное направление, однако оказываются сдвинутыми друг относительно друга в поперечном направлении (рис. 6.31). Эти лучи имеют взаимно ортогональные линейные поляризации.

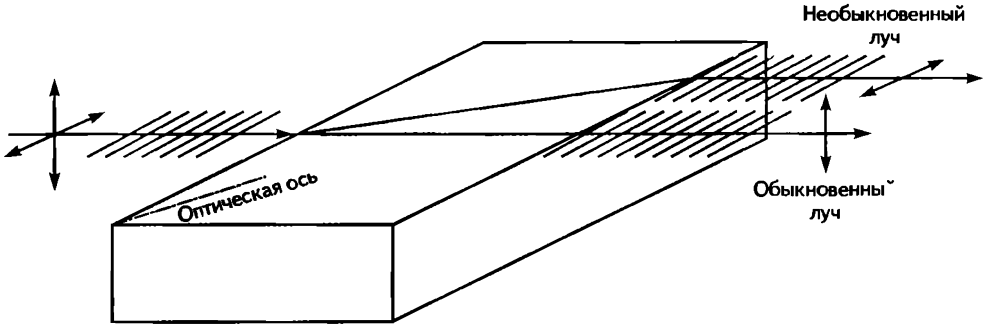


Рис. 6.31. Двулучепреломление при прохождении анизотропной пластины. Пластина может служить поляризующим делителем луча

6.4. ОПТИЧЕСКАЯ АКТИВНОСТЬ И МАГНИТООПТИКА

6.4.1. Оптическая активность

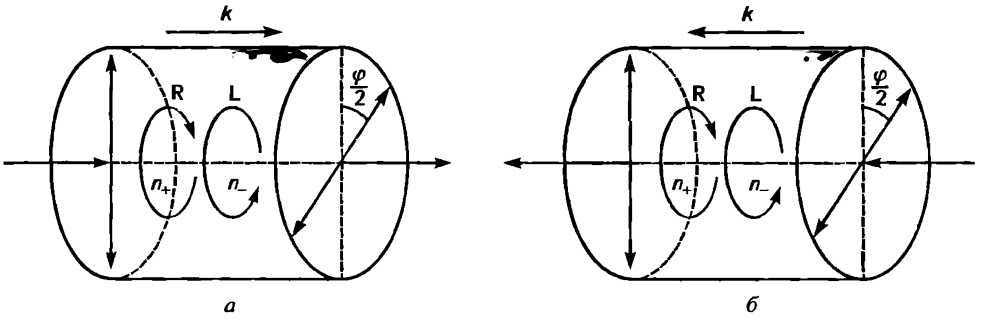
Некоторые материалы действуют как естественные вращатели плоскости поляризации. Это их свойство называется оптической активностью. Их нормальные моды представляют собой не линейно, а циркулярно поляризованные волны, причем волны с правой и левой круговыми поляризациями распространяются с различными фазовыми скоростями.

Ниже мы покажем, что оптически активная среда с фазовыми скоростями c_0/n_+ и c_0/n_- волн, поляризованных по правому и левому кругу, соответственно, действует как вращатель плоскости поляризации с углом поворота $\pi(n_- - n_+)d/\lambda_0$, пропорциональным толщине слоя среды d. Вращательная способность (угол поворота на единицу длины) оптически активной среды, называемая также **удельным вращением**, равна

$$\boxed{\rho = \frac{\pi}{\lambda_0}(n_- - n_+).} \qquad (6.69)$$

Удельное вращение

Направление поворота плоскости поляризации совпадает с направлением вращения поля у циркулярной компоненты с большей фазовой скоростью (меньшим показателем преломления). Если $n_- > n_+$, то ρ положительно, и поворот плоскости поляризации происходит в направлении вращения электрического вектора волны с правой круговой поляризацией (по часовой стрелке, если смотреть против хода волны, как показано на рис. 6.32, *а*). Такие материалы называются **правовращающими**, а те, у которых $n_- < n_+$, — **левовращающими**.



а *б*

Рис. 6.32. Вращение плоскости поляризации в оптически активной среде возникает из-за разности скоростей двух волн с круговыми поляризациями. На данной иллюстрации волна с правой круговой поляризацией (R) распространяется быстрее, чем с левой (L), то есть $n_- > n_+$, так что r положительно и материал правовращающий (*а*). При отражении волны плоскость поляризации начинает вращаться в противоположном направлении, так что волна повторяет свой ход в обратном порядке (*б*):

а — прямая волна; *б* — обратная волна

Вывод выражения для удельного вращения

Выражение (6.69) можно получить, если разложить падающую линейно поляризованную волну на две волны с противоположными круговыми поляризациями и одинаковыми амплитудами (см. упражнение 6.2)

$$\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} = \frac{1}{2}e^{-j\theta}\begin{bmatrix}1\\ j\end{bmatrix} + \frac{1}{2}e^{j\theta}\begin{bmatrix}1\\ -j\end{bmatrix}, \tag{6.70}$$

где θ — первоначальный угол ориентации плоскости поляризации. После прохождения расстояния d в среде набег фазы у волн, поляризованных по правому и левому кругу, составляет

$$\varphi_+ = 2\pi n_+ \frac{d}{\lambda_0}; \quad \varphi_- = 2\pi n_- \frac{d}{\lambda_0},$$

соответственно, и вектор Джонса получается равным

$$\frac{1}{2}e^{-j\theta}e^{-j\varphi_+}\begin{bmatrix}1\\ j\end{bmatrix} + \frac{1}{2}e^{j\theta}e^{-j\varphi_-}\begin{bmatrix}1\\ -j\end{bmatrix} = e^{-j\varphi_0}\begin{bmatrix}\cos\left(\theta-\frac{\varphi}{2}\right)\\ \sin\left(\theta-\frac{\varphi}{2}\right)\end{bmatrix}, \tag{6.71}$$

где

$$\varphi_0 = \frac{\varphi_+ + \varphi_-}{2}; \quad \varphi = 2\pi\left(n_- - n_+\right)\frac{d}{\lambda_0}.$$

Этот вектор Джонса представляет линейно поляризованную волну с плоскостью поляризации, повернутой на угол

$$\frac{\varphi}{2} = \pi(n_- - n_+)\frac{d}{\lambda_0},$$

что совпадает с (6.69).

Оптическая активность присуща материалам с внутренней спиральной структурой. Примерами служат селен, теллур, окись теллура (TeO_2), кварц (α-SiO_2) и киноварь (HgS). Оптически активные жидкости состоят из так называемых киральных молекул, которые существуют в виде правых и левых зеркальных форм (изомеров). Многие органические соединения, такие как аминокислоты и сахара, обладают оптической активностью. Почти все аминокислоты левовращающие, тогда как обычные сахара встречаются в обеих формах: D-глюкоза и фруктоза, например, являются, соответственно, право- и левовращающими (как подразумевают их вторые названия *dextrose* и *levulose*). Удельное вращение и направление поворота в растворах таких веществ чувствительны как к концентрации раствора, так и к структуре растворителя. Для измерения оптической активности растворов сахара служит прибор, называемый сахариметром, в котором по измеренной оптической активности определяется концентрация сахара в растворе.

Материальные уравнения

Переменное магнитное поле с индукцией $\boldsymbol{B}$, приложенное к оптически активной структуре, индуцирует циркулирующий ток, который, благодаря винтовому характеру структуры, вызывает появление электрического дипольного момента (и, следовательно, поляризации), пропорционального $j\omega\boldsymbol{B} = -\nabla \times \boldsymbol{E}$. Следовательно, оптически активная среда обладает пространственной дисперсией, т. е. связь между $\boldsymbol{D}(\boldsymbol{r})$ и $\boldsymbol{E}(\boldsymbol{r})$ нелокальная. Вектор $\boldsymbol{D}(\boldsymbol{r})$ в некоторой точке среды $\boldsymbol{r}$ определяется не только $\boldsymbol{E}(\boldsymbol{r})$, но и $\boldsymbol{E}(\boldsymbol{r}')$ в точках $\boldsymbol{r}'$, находящихся в непосредственной близости от $\boldsymbol{r}$, поскольку он зависит от пространственных производных, входящих в $\nabla \times \boldsymbol{E}(\boldsymbol{r})$. Для плоской волны имеем $\boldsymbol{E}(\boldsymbol{r}) = \boldsymbol{E}\exp(-j\boldsymbol{k}\cdot\boldsymbol{r})$ и $\nabla \times \boldsymbol{E} = -j\boldsymbol{k} \times \boldsymbol{E}$, так что тензор диэлектрической проницаемости зависит от волнового вектора $\boldsymbol{k}$. Пространственная дисперсия аналогична временной дисперсии, которая происходит от немгновенности отклика среды (см. разд. 5.2). Подобно тому, как диэлектрическая проницаемость среды с временной дисперсией зависит от частоты, в случае пространственной дисперсии она зависит от волнового вектора $\boldsymbol{k}$.

Оптически активная среда описывается зависящим от материальным уравнением

$$\boldsymbol{D} = \varepsilon\boldsymbol{E} + j\varepsilon_0\xi\boldsymbol{k} \times \boldsymbol{E}, \tag{6.72}$$

где ξ — величина (называемая псевдоскаляром), которая меняет знак при замене правой тройки координатных осей на левую. Это соотношение является

приближением первого порядка для зависимости тензора диэлектрической проницаемости от $\boldsymbol{k}$ при соответствующих условиях симметрии[3]. Первый член описывает отклик изотропной диэлектрической среды, а второй, как будет показано ниже, учитывает оптическую активность. Это соотношение между $\boldsymbol{D}$ и $\boldsymbol{E}$ часто пишут в виде

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} + j\varepsilon_0 \boldsymbol{G} \times \boldsymbol{E}, \tag{6.73}$$

где $\boldsymbol{G} = \xi \boldsymbol{k}$ называется вектором гирации. В таких средах вектор $\boldsymbol{D}$, очевидно, не параллелен $\boldsymbol{E}$, поскольку вектор $\boldsymbol{G} \times \boldsymbol{E}$ в (6.73) перпендикулярен $\boldsymbol{E}$.

Нормальные моды оптически активной среды

Перейдем к доказательству того, что нормальные моды среды, описываемой уравнением (6.73), представляют собой волны с круговой поляризацией, и выразим скорости c_0/n_+ и c_0/n_- через постоянную $G = \xi k$.

Предположим, что волна распространяется в направлении z, так что $\boldsymbol{k} = (0, 0, k)$ и, таким образом, $\boldsymbol{G} = (0, 0, G)$. Уравнение (6.73) можно записать в матричном виде

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \varepsilon_0 \begin{bmatrix} n^2 & -jG & 0 \\ jG & n^2 & 0 \\ 0 & 0 & n^2 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}, \tag{6.74}$$

где $n^2 = \varepsilon/\varepsilon_0$. Диагональные элементы в (6.74) соответствуют распространению в изотропной среде с показателем преломления n, а недиагональные элементы, пропорциональные G, описывают оптическую активность.

Для доказательства того, что нормальные моды поляризованы по кругу, рассмотрим две циркулярно поляризованные волны с векторами электрического поля $\boldsymbol{E} = (E_0, \pm jE_0, 0)$. Знаки «+» и «−» соответствуют правой и левой круговым поляризациям. Подстановка в (6.74) дает $\boldsymbol{D} = (D_0, \pm jD_0, 0)$, где $D_0 = \varepsilon_0(n^2 \pm G)E_0$. Следовательно,

$$\boldsymbol{D} = \varepsilon_0 n_\pm^2 \boldsymbol{E},$$

где

$$\boxed{n_\pm = \sqrt{n^2 \pm G}.} \tag{6.75}$$

Следовательно, для каждой из двух волн с круговой поляризацией вектор $\boldsymbol{D}$ параллелен вектору $\boldsymbol{E}$. Уравнение (6.54) удовлетворяется, если волновое число $k = n_\pm k_0$. Таким образом, право- и левополяризованная волны распространяются без изменения состояния их поляризации с показателями преломления n_+ и n_- соответственно. Это означает, что они являются нормальными модами данной среды.

---

[3] См., например, *Ландау Л.Д., Лифшиц Е.М., Питаевский Л.П.* Электродинамика сплошных сред. М.: Наука, 1982.

Упражнение 6.7

Удельное вращение оптически активной среды

Покажите, что если $G \ll n$, то удельное вращение оптически активной среды (угол поворота плоскости поляризации на единицу длины) приближенно выражается как

$$\rho \approx -\frac{\pi G}{\lambda_0 n}. \tag{6.76}$$

Удельное вращение сильно зависит от длины волны. Поскольку величина G пропорциональна k, как следует из (6.73), она обратно пропорциональна длине волны λ_0. Таким образом, удельное вращение в (6.76) обратно пропорционально λ_0^2. Более того, показатель преломления n сам зависит от длины волны. Например, удельное вращение ρ кварца составляет ≈31 град/мм при $\lambda_0 = 500$ нм и ≈22 град/мм при $\lambda_0 = 600$ нм; для $AgGaS_2$ величина ρ составляет ≈700 град/мм при $\lambda_0 = 490$ нм и ≈500 град/мм при $\lambda_0 = 500$ нм.

6.4.2. Магнитооптика: эффект Фарадея

Многие материалы обладают свойством вращать плоскость поляризации в присутствии постоянного магнитного поля. Это явление называется **эффектом Фарадея**. Угол поворота в этом случае пропорционален толщине слоя материала, а удельное вращение ρ (угол поворота на единицу длины) пропорционально составляющей вектора индукции магнитного поля B в направлении распространения волны

$$\rho = \mathfrak{D}B, \tag{6.77}$$

где $\mathfrak{D}$ называется **постоянной Верде**.

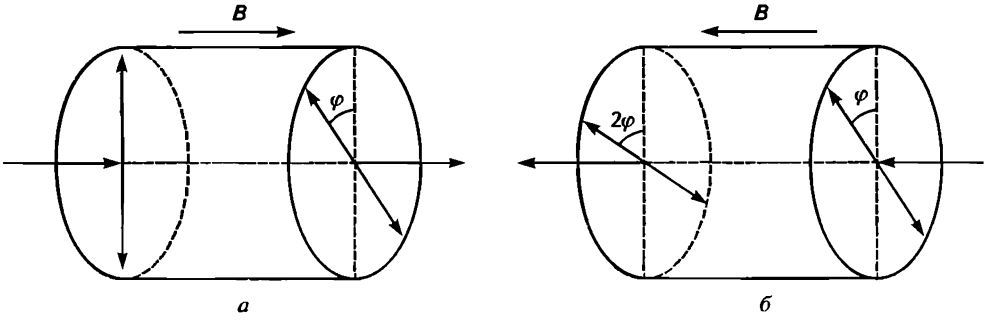


Рис. 6.33. Поворот плоскости поляризации в среде с эффектом Фарадея (*а*). Направление поворота инвариантно относительно направления распространения волны (*б*)

Направление вращения определяется направлением магнитного поля: для $\mathfrak{D} > 0$ вращение происходит по правому винту, направленному вдоль магнитного поля (рис. 6.33, *а*). В отличие от оптической активности, направление

вращения не меняется при изменении направления распространения волны на противоположное. Поэтому когда волна проходит через фарадеевский вращатель туда и обратно, поворот плоскости поляризации удваивается (рис. 6.33, *б*). К числу материалов, в которых наблюдается эффект Фарадея, относятся стекла, железоиттриевый гранат (YIG), тербий-галлиевый гранат (TGG) и тербий-алюминиевый гранат (TbAlG). Постоянная Верде TbAlG равна $\mathfrak{D} \approx 1{,}16$ мин/(эрстед · см) при $\lambda_0 = 500$ нм. Пленки из этих ферромагнитных материалов используются для создания компактных устройств.

Материальные уравнения

У магнитооптических материалов тензор диэлектрической проницаемости $\boldsymbol{\varepsilon}$ меняется при наложении *статического* магнитного поля $\boldsymbol{H}$, так что $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\boldsymbol{H})$. Этот эффект происходит из-за взаимодействия статического магнитного поля с электронами вещества, движущимися под действием *оптического* электрического поля $\boldsymbol{E}$. В частности, для эффекта Фарадея материальное уравнение имеет вид

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} + j\varepsilon_0 \boldsymbol{G} \times \boldsymbol{E}, \tag{6.78}$$

где

$$\boldsymbol{G} = \gamma \boldsymbol{B}. \tag{6.79}$$

Здесь $\boldsymbol{B} = \mu \boldsymbol{H}$ — статическая индукция магнитного поля; γ — постоянная среды, называемая **коэффициентом магнитогирации**.

Уравнение (6.78) идентично (6.73), так что вектор $\boldsymbol{G} = \gamma \boldsymbol{B}$ в эффекте Фарадея играет ту же роль, что и вектор гирации $\boldsymbol{G} = \xi \boldsymbol{k}$ в оптически активных средах. Однако в эффекте Фарадея $\boldsymbol{G}$ не зависит от $\boldsymbol{k}$, так что обращение направления распространения волны не обращает направление вращения плоскости поляризации. Это свойство полезно при конструировании оптических изоляторов, рассмотренном в подразд. 6.6.3.

В рамках указанной аналогии и с использованием (6.76) можно прийти к выводу, что удельное вращение фарадеевской среды:

$$\rho \approx -\frac{\pi G}{\lambda_0 n} = -\frac{\pi \gamma B}{\lambda_0 n},$$

откуда постоянная Верде (удельное вращение на единицу магнитной индукции) равна

$$\boxed{\mathfrak{D} = -\frac{\pi \gamma}{\lambda_0 n}.} \tag{6.80}$$

Постоянная Верде, очевидно, является функцией длины волны λ_0.

6.5. ОПТИКА ЖИДКИХ КРИСТАЛЛОВ

Жидкие кристаллы

Жидкий кристалл представляет собой структуру из ориентированных органических молекул, обычно имеющих сигарообразную форму. Эти молекулы не упорядочены по положению их центров в пространстве (как в жидкости), но упорядочены по ориентации (как в кристалле). Существует три типа жидких кристаллов, показанные на рис. 6.34.

- В **нематических** жидких кристаллах ориентация молекул стремится к постоянной, но положения их центров в пространстве совершенно случайны.
- В **смектических** жидких кристаллах ориентации молекул одинаковы, а центры расположены параллельными слоями, в пределах которых положение молекул неупорядоченно. Таким образом, позиционный порядок существует только в одном измерении.
- **Холестерический** жидкий кристалл представляет собой скрученный нематик, у которого ориентация молекул поворачивается вокруг оси от слоя к слою.

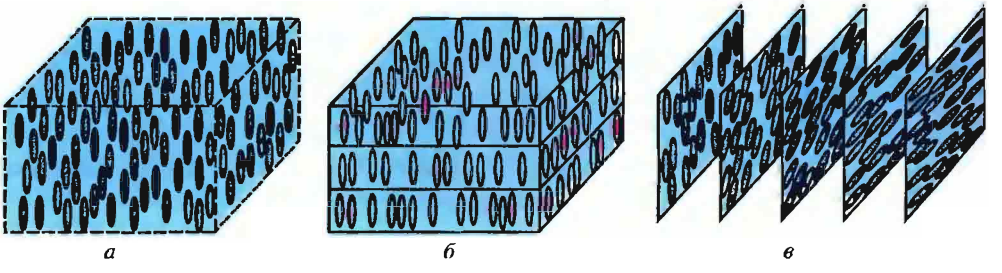


Рис. 6.34. Расположение молекул в различных типах жидких кристаллов:
а — нематик; *б* — смектик; *в* — холестерик

Жидкокристаллическое состояние представляет собой *текучее* состояние вещества. Молекулы могут менять свою ориентацию под действием внешней силы. Когда тонкий слой жидкого кристалла помещается между двумя параллельными стеклянными пластинами, при трении пластин друг о друга молекулы ориентируются в направлении трения.

Твист-нематические жидкие кристаллы — это нематические жидкие кристаллы, у которых скручивание (аналогичное естественно существующему у холестерических жидких кристаллов) навязано извне. Этого можно добиться, например, помещая тонкий слой нематического жидкого кристалла между двумя стеклянными пластинами, отполированными в перпендикулярных направлениях, как схематически показано на рис. 6.35. Этот раздел посвящен обсуждению оптических свойств твист-нематических жидких кристаллов, которые широко используются в фотонике, например в жидкокристаллических дисплеях. Электрооптические свойства твист-нематических жидких кристаллов и их использование в качестве оптических модуляторов и переключателях описаны в гл. 20.

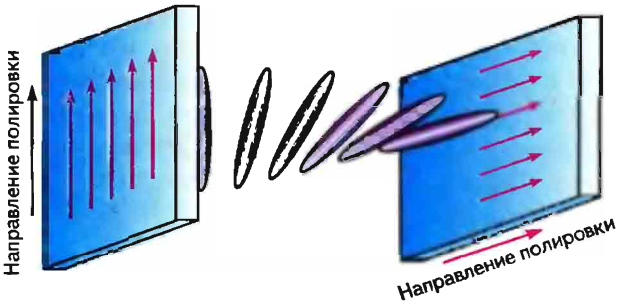


Рис. 6.35. Ориентация молекул в твист-нематическом жидком кристалле

Оптические свойства твист-нематических жидких кристаллов

Твист-нематический жидкий кристалл — оптически *неоднородная* и *анизотропная* среда, локально ведущая себя как одноосный кристалл, оптическая ось которого параллельна направлению выстраивания длинных осей молекул. Оптические свойства такого материала удобно анализировать, рассматривая его как совокупность тонких слоев, перпендикулярных оси кручения, каждый из которых рассматривается как одноосный кристалл; при этом оптическая ось постепенно поворачивается от слоя к слою вокруг оси кручения (рис. 6.36). Затем рассчитывается накапливающееся влияние этих слоев на проходящую волну. Покажем, что при некоторых условиях твист-нематический жидкий кристалл действует как вращатель, в котором плоскость поляризации света поворачивается вместе с ориентацией молекул.

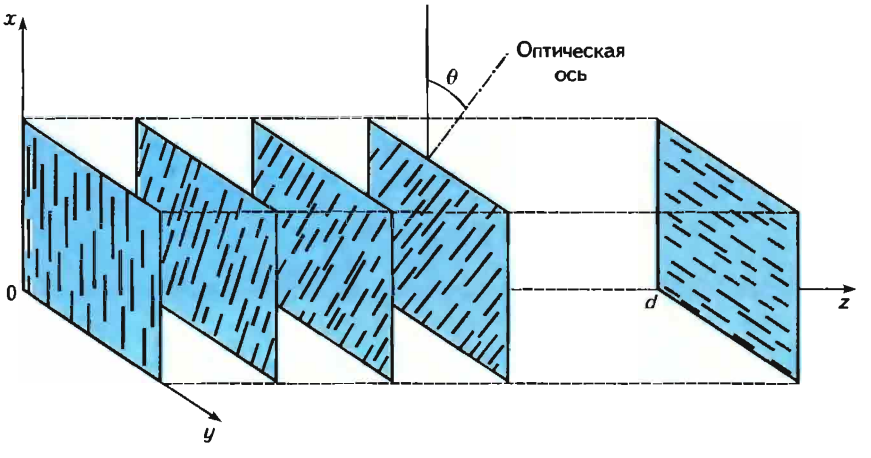


Рис. 6.36. Распространение света в твист-нематическом жидком кристалле. На схеме угол кручения равен 90°

Рассмотрим распространение света вдоль оси кручения (оси z) в твист-нематическом жидком кристалле и предположим, что угол закручивания θ линейно растет с ростом z:

$$\theta = \alpha z, \tag{6.81}$$

где α — коэффициент кручения (в градусах на единицу длины). Следовательно, оптическая ось параллельна плоскости x—y и образует угол θ с направлением оси x. Обыкновенный и необыкновенный показатели преломления равны n_o и $n_н$, соответственно (обычно $n_o < n_н$), и коэффициент фазового запаздывания (запаздывание по фазе на единицу длины) равен

$$\beta = (n_н - n_o)k_0. \tag{6.82}$$

Жидкокристаллическая ячейка полностью характеризуется коэффициентом кручения α и коэффициентом фазового запаздывания β.

На практике $\beta \gg \alpha$, поэтому прежде чем оптическая ось заметно повернется, набег фазы успевает достичь многих периодов. Мы покажем, что если это условие выполнено и при $z = 0$ волна линейно поляризована в направлении оси x, то она сохраняет состояние линейной поляризации, плоскость которой поворачивается вместе с ориентацией молекул от слоя к слою, так что угол поворота составляет $\theta = \alpha z$. Полный поворот плоскости поляризации в жидком кристалле толщиной d равен αd. Жидкокристаллическая ячейка в этом случае действует как вращатель плоскости поляризации с удельным вращением α. Это свойство твист-нематических жидких кристаллов применяется при изготовлении отображающих устройств (дисплеев), как объясняется в разд. 20.3.

Доказательство того, что твист-нематический жидкий кристалл ведет себя как вращатель плоскости поляризации

Рассмотрим случай $\beta \gg \alpha$. Разделим полную толщину ячейки d на N тонких слоев равной толщины $\Delta z = d/N$. Слой с номером m, расположенный на расстоянии $z = z_m = m\Delta z$, $m = 1, 2, ..., N$, есть фазовая пластинка, медленная ось которой (оптическая ось) образует угол $\theta_m = m\Delta\theta$ с осью x, где $\Delta\theta = \alpha\Delta z$. Следовательно, она имеет матрицу Джонса [см. (6.24)]

$$\mathbf{T}_m = \mathbf{R}(-\theta_m)\mathbf{T}_r\mathbf{R}(\theta_m), \tag{6.83}$$

где $\mathbf{T}_r$ — матрица Джонса фазовой пластинки, ось которой направлена вдоль оси x,

$$\mathbf{T}_r = \begin{bmatrix} \exp(-jn_н k_0 \Delta z) & 0 \\ 0 & \exp(-jn_o k_0 \Delta z) \end{bmatrix}; \tag{6.84}$$

$\mathbf{R}(\theta)$ — матрица поворота координат, входящая в (6.22).

Удобно выразить $\mathbf{T}_r$ через коэффициент фазовой задержки β:

$$\mathbf{T}_r = \exp(-j\varphi\Delta z)\begin{bmatrix} \exp\left(-\frac{j\beta\Delta z}{2}\right) & 0 \\ 0 & \exp\left(\frac{j\beta\Delta z}{2}\right) \end{bmatrix}, \tag{6.85}$$

где

$$\varphi = \frac{(n_o + n_н)k_0}{2}.$$

Поскольку умножение вектора Джонса на постоянный фазовый множитель не меняет состояния поляризации, мы отбросим множитель $\exp(-j\varphi\Delta z)$ в (6.85).

Полная матрица Джонса всего устройства выражается произведением

$$\mathbf{T} = \prod_{m=N}^{1} \mathbf{T}_m = \prod_{m=N}^{1} \mathbf{R}(-\theta_m)\mathbf{T}_r\mathbf{R}(\theta_m). \tag{6.86}$$

Используя (6.83) и замечая, что $\mathbf{R}(\theta_m)\mathbf{R}(-\theta_{m-1}) = \mathbf{R}(\theta_m - \theta_{m-1}) = \mathbf{R}(\Delta\theta)$, получим

$$\mathbf{T} = \mathbf{R}(-\theta_N)\left[\mathbf{T}_r\mathbf{R}(\Delta\theta)\right]^{N-1}\mathbf{T}_r\mathbf{R}(\theta_1). \tag{6.87}$$

Подставляя сюда (6.85) и (6.22), находим

$$\mathbf{T}_r\mathbf{R}(\Delta\theta) = \begin{bmatrix} \exp\left(-\dfrac{j\beta\Delta z}{2}\right) & 0 \\ 0 & \exp\left(\dfrac{j\beta\Delta z}{2}\right) \end{bmatrix}\begin{bmatrix} \cos\alpha\Delta z & \sin\alpha\Delta z \\ -\sin\alpha\Delta z & \cos\alpha\Delta z \end{bmatrix}. \tag{6.88}$$

С помощью (6.87) и (6.88) полную матрицу Джонса можно выразить через параметры α, β и $d = N\Delta z$.

При $\alpha \ll \beta$ можно принять матрицу бесконечно малого поворота приблизительно равной единичной матрице, после чего

$$\mathbf{T} \approx \mathbf{R}(-\theta_N)\left[\mathbf{T}_r\right]^N \mathbf{R}(\theta_1) = \tag{6.89}$$

$$= \mathbf{R}(-\alpha N\Delta z)\begin{bmatrix} \exp\left(-\dfrac{j\beta\Delta z}{2}\right) & 0 \\ 0 & \exp\left(\dfrac{j\beta\Delta z}{2}\right) \end{bmatrix}^N = \mathbf{R}(-\alpha N\Delta z)\begin{bmatrix} \exp\left(-\dfrac{j\beta\Delta z}{2}\right) & 0 \\ 0 & \exp\left(\dfrac{j\beta\Delta z}{2}\right) \end{bmatrix},$$

так что

$$\mathbf{T} = \mathbf{R}(-\alpha d)\begin{bmatrix} \exp\left(-\dfrac{j\beta d}{2}\right) & 0 \\ 0 & \exp\left(\dfrac{j\beta d}{2}\right) \end{bmatrix}. \tag{6.90}$$

Эта матрица Джонса описывает фазовую пластинку с задержкой фазы βd, медленная ось которой направлена вдоль оси x, за которой следует вращатель плоскости поляризации, обеспечивающий поворот на αd. Если исходная волна

поляризована линейно вдоль оси x, то фазовая пластинка вносит только сдвиг фазы, и устройство затем просто поворачивает плоскость поляризации на угол αd, равный углу кручения. Волна, линейно поляризованная вдоль оси y, поворачивается на тот же угол.

6.6. ПОЛЯРИЗАЦИОННЫЕ УСТРОЙСТВА

Этот раздел содержит краткое описание ряда устройств, используемых для изменения состояния поляризации света. Основные принципы действия этих устройств сформулированы в данной главе.

6.6.1. Поляризаторы

Линейный поляризатор — устройство, пропускающее компоненту электрического поля, направленную вдоль оси пропускания, и блокирующее ортогональную компоненту. Блокирующее действие может быть обеспечено селективным поглощением, селективным отражением от границы изотропной среды либо селективным отражением/преломлением в анизотропных средах.

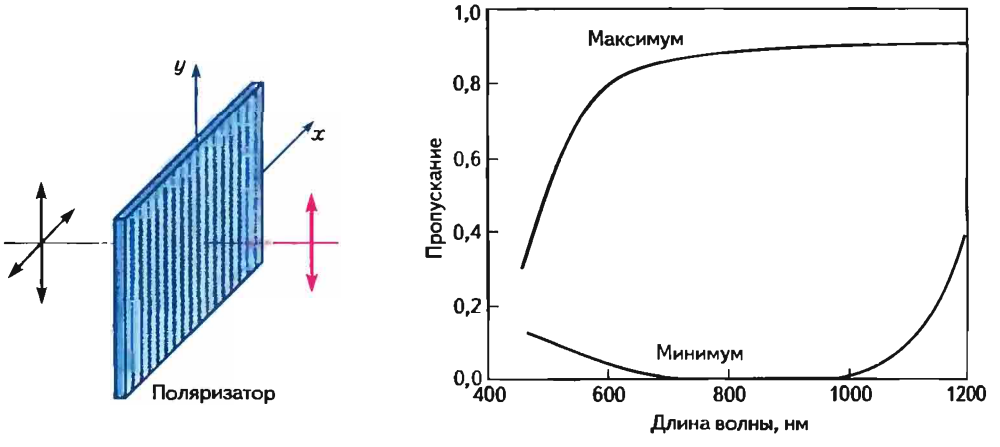


Рис. 6.37. Коэффициенты пропускания по мощности для типичного дихроичного поляризатора для плоскости поляризации света, ориентированной в направлении максимального и минимального поглощения

Поляризация за счет селективного поглощения (дихроизм)

Поглощение света некоторыми анизотропными средами, которые называются **дихроичными материалами**, зависит от направления падающего электрического поля (рис. 6.37). Эти материалы обычно имеют анизотропную молекулярную структуру, отклик которой чувствителен к направлению электри-

ческого поля. Наиболее распространенный дихроичный материал — **поляроид** (Polaroid H-sheet), изобретенный в 1938 г. и до сих пор имеющий широкое применение. Он изготовляется на основе пленки из поливинилового спирта, в которую вкраплены микрокристаллы йода. Пленка нагревается и растягивается в одном направлении. Аналогичное устройство в инфракрасном диапазоне — поляризатор в виде плоской решетки из тонких близко расположенных проволочек, вытянутых в одном направлении. Продольная по отношению к проволочкам компонента падающего электрического поля поглощается, а перпендикулярная проходит через решетку.

Поляризация за счет селективного отражения

Коэффициент отражения света на границе между двумя изотропными диэлектрическими материалами зависит от его поляризации, как обсуждалось в разд. 6.2. В частности, при падении под углом Брюстера волна с поляризацией TM не отражается, а полностью проходит (см. рис. 6.13). Под этим углом отражается только волна с поляризацией TE, так что такой отражатель служит поляризатором.

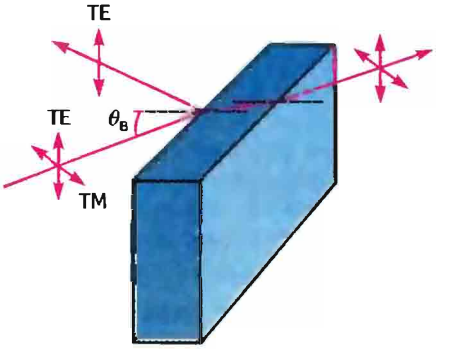


Рис. 6.38. Брюстеровский поляризатор

Поляризация за счет селективного преломления (поляризующие светоделители)

При входе света в анизотропный кристалл обыкновенная и необыкновенная волны преломляются под разными углами и постепенно отделяются друг от друга (см. подраздел 6.3.5 и рис. 6.31). Это дает эффективный способ получения поляризованного света из неполяризованного, который часто находит применение. Подобные устройства обычно состоят из двух склеенных призм из анизотропных (одноосных) материалов, часто с различной ориентацией осей, как показано на рис. 6.39. Такие призмы служат в качестве **поляризующих светоделителей**.

Достоинством призмы Глана—Томпсона является большой угол разделения выходящих пучков.

6.6.2. Фазовые пластинки

Фазовая пластинка служит для преобразования волны с одним типом поляризации в волну с другим типом. Она характеризуется фазовым запаздывани-

ем Г, а также направлением быстрой и медленной осей (см. подразд. 6.1.2). Нормальные моды пластинки — волны, линейно поляризованные вдоль указанных осей. Скорости двух волн различны, поэтому прохождение через пластинку вносит разность фаз Г между модами.

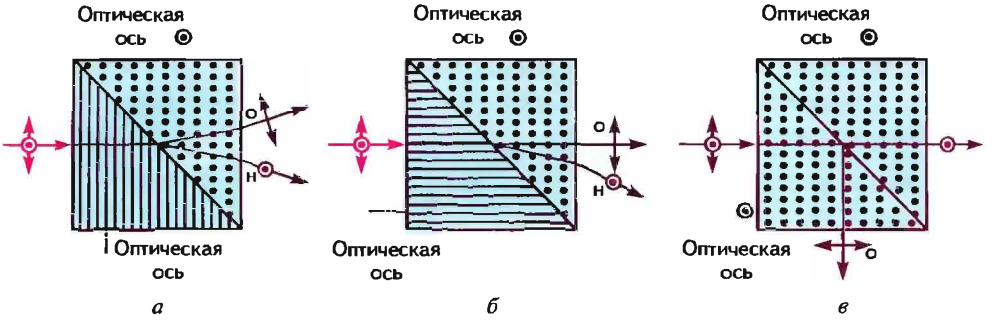


Рис. 6.39. Поляризующие светоделители. Направления и поляризации выходящих волн различны у трех призм. На данной иллюстрации кристаллы отрицательные одноосные (например, кальцит):

а — призма Волластона; *б* — призма Рошона; *в* — призма Глана—Томпсона

Фазовые пластинки часто изготавливаются из анизотропных кристаллов. Как поясняется в подразд. 6.3.2, при распространении света вдоль одной из главных осей (скажем, z) нормальными модами являются волны, линейно поляризованные вдоль двух других осей (x и y). Для этих мод показатели преломления — это главные значения n_1 и n_2, которым соответствуют фазовые скорости c_0/n_1 и c_0/n_2. Если $n_1 < n_2$, ось x является быстрой осью. При толщине пластинки d фазовая задержка равна

$$\Gamma = (n_2 - n_1)k_0 d = 2\pi(n_2 - n_1)\frac{d}{\lambda_0}.$$

Она прямо пропорциональна толщине пластинки d и обратно пропорциональна длине волны λ_0 (заметим, однако, что и сама разность $n_2 - n_1$ зависит от длины волны).

Например, показатели преломления тонкой пластинки слюды равны 1,599 и 1,594 при $\lambda_0 = 633$ нм, так что $\Gamma/d \approx 15{,}8\pi$ рад/мм. Пластинка толщиной 63,3 мкм дает $\Gamma \approx \pi$ и обеспечивает задержку в полволны.

Управление интенсивностью света при помощи фазовой пластинки и двух поляризаторов

Рассмотрим фазовую пластинку с фазовой задержкой Г, помещенную между двумя скрещенными линейными поляризаторами, оси которых ориентированы под углом 45° к осям пластинки, как показано на рис. 6.40. Коэффициент пропускания по интенсивности (энергии) такой системы равен

$$\mathcal{T} = \sin^2\left(\frac{\Gamma}{2}\right), \qquad (6.91)$$

что можно обосновать с помощью матриц Джонса или с помощью анализа эллипса поляризации света на выходе из пластинки в зависимости от Г с последующим нахождением компоненты вдоль оси выходного поляризатора, как показано на рис. 6.40. Если Г = 0, свет не проходит через систему, поскольку поляризаторы ортогональны. С другой стороны, если Г = π, то свет полностью проходит через систему, поскольку фазовая пластинка поворачивает плоскость поляризации на 90°, после чего свет оказывается поляризованным вдоль оси второго поляризатора.

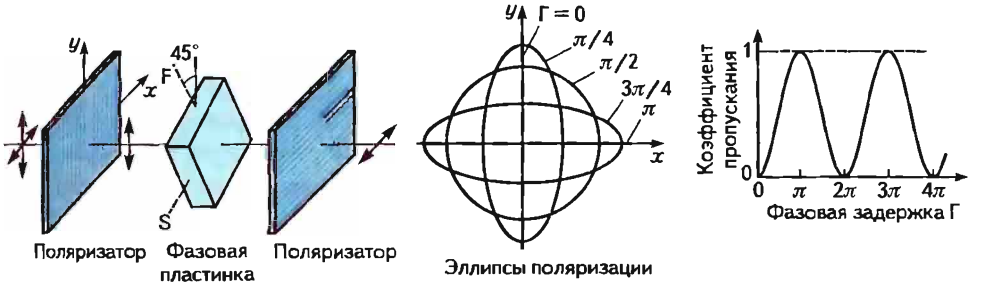


Рис. 6.40. Управление интенсивностью с помощью фазовой пластинки с варьируемой задержкой фазы Г между двумя скрещенными поляризаторами

Таким образом, интенсивностью прошедшего света можно легко управлять посредством изменения фазовой задержки Г. Для этого можно, например, направленно менять показатели преломления n_1 и n_2 с помощью внешнего постоянного электрического поля. В этом состоит основной принцип действия электрооптического модулятора, обсуждаемого в гл. 20.

Далее, поскольку Г зависит от d, малые изменения толщины образца можно отслеживать по интенсивности прошедшего света. Более того, так как Г зависит от длины волны, коэффициент пропускания системы чувствителен к частоте. Хотя систему и можно использовать как фильтр, его селективность невысока. Для построения узкополосных фильтров пропускания можно использовать другие схемы на основе фазовых пластинок и поляризаторов.

6.6.3. Вращатели плоскости поляризации

Вращатель плоскости поляризации служит для обеспечения поворота плоскости поляризации на заданный угол, при этом поляризация остается линейной. Как говорилось в разд. 6.4, для этой цели можно использовать оптически активные вещества или эффект Фарадея. Твист-нематические жидкие кристаллы при определенных условиях также действуют как вращатели плоскости поляризации (см. разд. 6.5).

Если вращатель плоскости поляризации поместить между двумя поляризаторами, то доля пропущенного света будет зависеть от угла вращения. Таким образом можно управлять интенсивностью (модулировать ее), если влиять на

угол поворота каким-либо внешним воздействием (например, меняя величину магнитного поля в эффекте Фарадея или меняя ориентацию молекул жидкого кристалла внешним электрическим полем). Электрооптическая модуляция света и жидкокристаллические дисплеи обсуждаются в гл. 20.

6.6.4. Невзаимные поляризационные устройства

Устройство, действие которого на состояние поляризации инвариантно относительно обращения направления распространения, называется **взаимным**. Если волна проходит через такое устройство в одном направлении, а затем в обратном направлении, то она будет испытывать все изменения поляризационного состояния в обратном порядке и выйдет назад в том же поляризационном состоянии, которое имела первоначально. Устройства, не обладающие этим свойством, называются **невзаимными**. Все устройства, рассмотренные в данной главе, являются взаимными, кроме фарадеевского вращателя (см. подразд. 6.4.2). Множество полезных невзаимных устройств получается путем комбинирования фарадеевского вращателя с другими взаимными поляризационными элементами.

Оптический изолятор

Оптический изолятор — устройство, которое пропускает свет только в одном направлении, действуя в этом смысле как односторонний клапан. Оптические изоляторы нужны, чтобы предотвратить возвращение отраженного света в источник. Такая обратная связь может оказывать вредное действие на работу некоторых устройств, таких как полупроводниковые лазеры.

Оптический изолятор изготавливается путем помещения фарадеевского вращателя между двумя поляризаторами, оси которых образуют друг с другом угол 45°. Вектор индукции магнитного поля, приложенного к вращателю, выбирается так, чтобы последний поворачивал плоскость поляризации на 45° в направлении правого винта вдоль оси z (рис. 6.41, *а*). Свет, идущий через систему в прямом направлении (слева направо), проходит поляризатор *A*, испытывает поворот плоскости поляризации на 45° и, следовательно, полностью проходит через поляризатор *B*. Свет, линейно поляризованный под углом 45°, но идущий через систему в обратном направлении (справа налево на рис. 6.41, *б*), успешно проходит через поляризатор *B*. Однако фарадеевский вращатель поворачивает его плоскость поляризации еще на 45°, после чего он блокируется поляризатором *A*. Поскольку обратная волна может возникать из-за отражения прямой волны последующими поверхностями, изолятор служит для защиты источника от отраженного света.

Заметим, что фарадеевский вращатель является обязательной составной частью оптического изолятора. Оптически активная среда или жидкий кристалл не могут заменить его, поскольку являются *взаимными* элементами, в которых при проходе волны назад поляризационные состояния повторяются в обратном порядке, так что отраженная волна вернется назад в исходном поляризационном состоянии и полностью пройдет через поляризатор *A* к источнику.

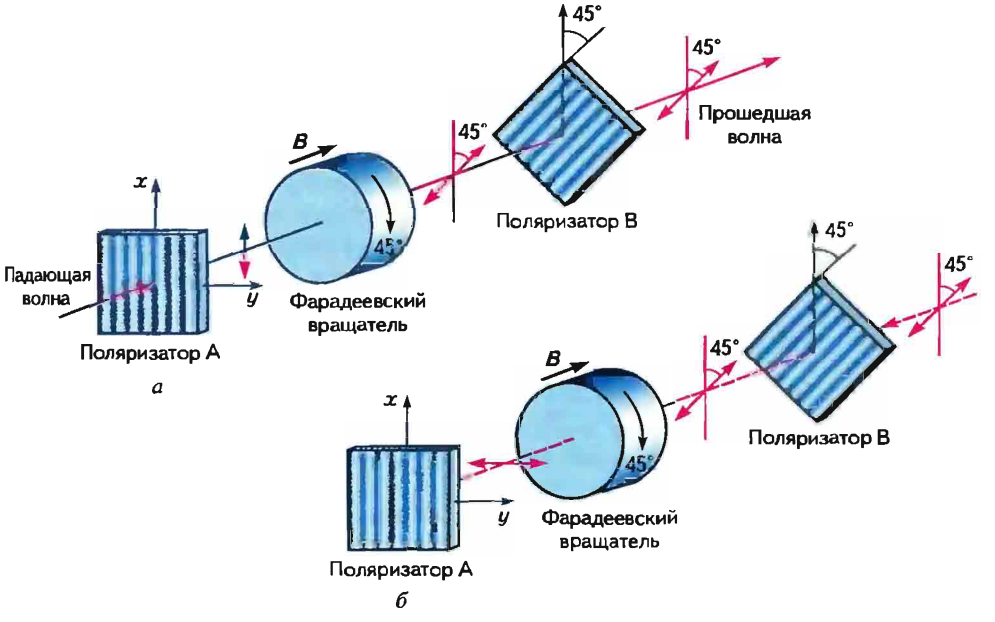


Рис. 6.41. Оптический изолятор с использованием фарадеевского вращателя пропускает свет в одном направлении. Волна, бегущая в прямом направлении, проходит (*а*). Волна, бегущая в обратно направлении, блокируется (*б*)

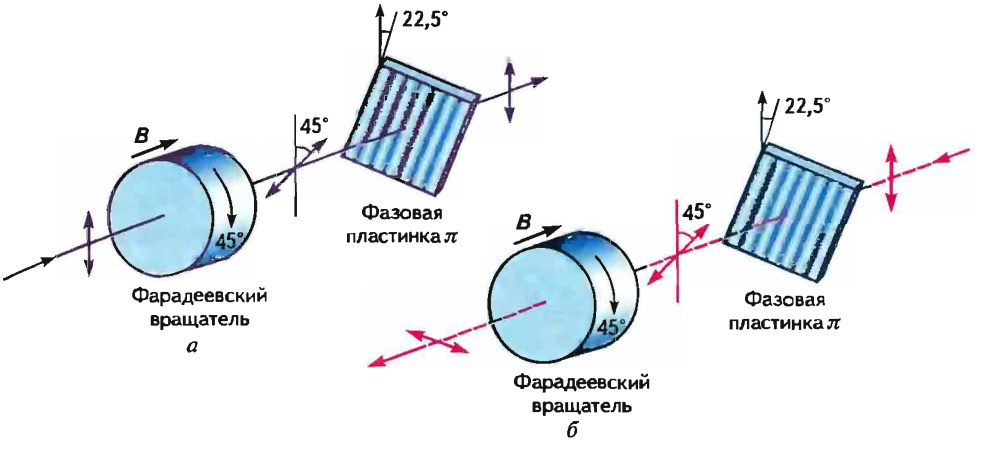


Рис. 6.42. Невзаимное устройство, сохраняющее поляризацию волны в прямом направлении (*а*) и поворачивающее на 90° в обратном (*б*)

Невзаимное вращение плоскости поляризации

Комбинация 45° фарадеевского вращателя и полуволновой фазовой пластинки — другое полезное невзаимное устройство. Как показано на рис. 6.42, *а*, состояние поляризации прямой линейно поляризованной волны с плоскостью

поляризации, ориентированной под углом 22,5° к быстрой оси фазовой пластинки, поддерживается при распространении через устройство (поскольку фарадеевский вращатель поворачивает плоскость поляризации на 45°, а затем фазовая пластинка на −45°). Однако для обратной волны плоскость поляризации повернется на 45° + 45° = 90°, как легко видеть на рис. 6.42, *б*. Таким образом, устройство можно использовать в комбинации с поляризующим светоделителем, чтобы направить обратную волну от источника прямой волны и иметь к ней независимый доступ. Система может оказаться полезной в оптических соединительных устройствах, таких как **оптические циркуляторы**, описанные в разд. 23.1.

Рекомендуемая литература

ОБЩАЯ

См. также литературу к гл. 1 и 5.

Damask J.N. Polarization Optics in Telecommunications. Springer-Verlag, 2004.

Goldstein D.H. Polarized Light. Marcel Dekker, 2nd ed. 2003.

Yariv A., Yeh P. Optical Waves in Crystals: Propagation and Control of Laser Radiation. Wiley; reprinted 2003.

Nye J.F. Physical Properties of Crystals: Their Representation by Tensors and Matrices. Oxford University Press, 1957. Reprinted with corrections and new material, 2001.

Sugano S., Kojima N., eds. Magneto-Optics. Springer-Verlag, 2000.

Brosseau C. Fundamentals of Polarized Light: A Statistical Optics Approach. Wiley, 1998.

Clarke D., Grainger J.F. Polarized Light and Optical Measurement. Pergamon, 1971. Reprinted 1996.

Huard S. Polarization of Light. Wiley, 1996.

Collett E. Polarized Light: Fundamentals and Applications. Marcel Dekker, 1993.

Kliger D.S., Lewis J.W., Randall C.E. Polarized Light in Optics and Spectroscopy. Academic Press, 1990.

Azzam R.M.A., Bashara N.M. Ellipsometry and Polarized Light. North-Holland, 1977. Reprinted 1989.

Gay P. An Introduction to Crystal Optics. Longmans, 1967. Paperback ed. 1982.

Robson B.A. The Theory of Polarization Phenomena. Clarendon, 1974.

Shurcliff W.A. Polarized Light: Production and Use. Harvard University Press, 1962. Reprinted 1966.

Velluz L., Le Grand M., Grosjean M. Optical Circular Dichroism: Principles, Measurements, and Applications. Academic Press, 1965.

Shurcliff W.A., Ballard S.S. Polarized Light. Van Nostrand, 1964.

КНИГИ О ЖИДКИХ КРИСТАЛЛАХ

Oswald P., Pieranski P. Nematic and Cholesteric Liquid Crystals: Concepts and Physical Properties Illustrated by Experiments. CRC Pressrraylor & Francis, 2005.

Vicari L. Optical Applications of Liquid Crystals. Institute of Physics, 2003.

Collings P.J. Liquid Crystals: Nature's Delicate Phase of Matter. Princeton University Press, 2nd ed. 2002.

Yeh P., Gu C. Optics of Liquid Crystal Displays. Wiley, 1999.

Chigrinov V.G. Liquid Crystal Devices: Physics and Applications. Artech House, 1999.

de Gennes P.G. The Physics of Liquid Crystals. Clarendon Press, 1974; Oxford University Press, 2nd ed. 1995.

Chandrasekhar S. Liquid Crystals. Cambridge University Press, 2nd ed. 1992.

Ericksen J.L., Kinderlehrer D., eds. Theory and Applications of Liquid Crystals. Springer-Verlag, 1987.

Blinov L.M. Electro-Optical and Magneto-Optical Properties of Liquid Crystals. Wiley, 1983.

de Jeu W.H. Physical Properties of Liquid Crystalline Materials. Gordon and Breach, 1980.

Meier G., Sackmann E., Grabmaier J.G. Applications of Liquid Crystals. Springer-Verlag, 1975.

СТАТЬИ

Ando K., Challener W., Gambino R., Levy M., eds. Magneto-Optical Materials for Photonics and Recording. Materials Research Society Symposium Proceedings. Vol. 834. Materials Research Society, 2005.

Mansuripur M. The Faraday Effect, Optics & Photonics News. Vol. 10, № 11, 1999. P. 32–36.

Billings B.H., ed. Selected Papers on Applications of Polarized Light. SPIE Optical Engineering Press (Milestone Series. Vol. 57), 1992.

Jacobs S.D., ed. Selected Papers on Liquid Crystals for Optics. SPIE Optical Engineering Press (Milestone Series. Vol. 46), 1992.

Billings B.H., ed. Selected Papers on Polarization. SPIE Optical Engineering Press (Milestone Series. Vol. 23), 1990.

Lakhtakia A., ed. Selected Papers on Natural Optical Activity. SPIE Optical Engineering Press (Milestone Series. Vol. 15), 1990.

Ginzburg V.L. On Crystal Optics with Spatial Dispersion. In Physics Reports. Vol. 194, 1990. P. 245–251.

Bennett J.M., Bennett H.E. Polarization. In Handbook of Optics. W.G. Driscoll, ed. McGraw-Hill, 1978.

Swindell W., ed. Benchmark Papers in Optics: Polarized Light, Dowden. Hutchinson & Ross, 1975.

v. Agranovich M., Ginzburg V.L. Crystal Optics with Spatial Dispersion. E. Wolf, ed. In Progress in Optics. Vol. 9. North-Holland, 1971.

Задачи

К РАЗДЕЛУ 6.1

1. **Ортогональные поляризации.** Покажите, что если два состояния эллиптической поляризации ортогональны, то большие оси их эллипсов перпендикулярны, а направления вращения противоположны.

2. **Поворот вращателя плоскости поляризации.** Покажите, что матрица Джонса вращателя плоскости поляризации инвариантна относительно поворотов системы координат.

3. **Полуволновая пластинка.** Рассмотрим линейно поляризованный свет, проходящий через полуволновую пластинку. В случае, когда плоскость поляризации образует угол θ с быстрой осью пластинки, покажите, что прошедший свет поляризован линейно под углом $-\theta$, т. е. поворачивается на угол 2θ. Почему полуволновая пластинка не эквивалентна вращателю плоскости поляризации?

4. **Тандем из фазовых пластинок.** Запишите матрицы Джонса:

а) для фазовой пластинки $\pi/2$ с быстрой осью вдоль направления x;

б) фазовой пластинки π с быстрой осью под 45° к оси x;

в) фазовой пластинки $\pi/2$ с быстрой осью вдоль направления y.

Покажите, что если эти три пластинки поставить друг за другом [сначала (а), потом (б), затем (в)], то получающееся устройство будет производить поворот на 90° Что произойдет, если пластинки расположить в обратном порядке?

5. **Отражение света, поляризованного по кругу.** Покажите, что свет, поляризованный по кругу, меняет правую поляризацию на левую и, наоборот, при отражении от зеркала.

6. **Экран, защищающий от бликов.** Самосветящийся объект наблюдается через стеклянное окно. Для исключения бликов, вызванных отражением фонового излучения от поверхностей окна, применяется специальный экран. Покажите, что такой экран можно сделать путем комбинации линейного поляризатора и четвертьволновой пластинки, оси которых расположены под углом 45° относительно оси пропускания поляризатора. Можно ли экран рассматривать как оптический изолятор?

К РАЗДЕЛУ 6.2

1. **Вывод формул Френеля.** Выведите формулу отражения (6.32), из которой выводится формула Френеля (6.34) в случае поляризации TE. Что вы будете делать для получения коэффициента отражения, если падающий свет имеет вид узкого пучка, а не плоской волны?

2. **Коэффициент отражения стекла.** Плоская волна падает из воздуха ($n = 1$) на стеклянную пластинку ($n = 1{,}5$) под углом 45° Определите коэффициент отражения по мощности для TE- и TM-мод. Чему равен средний коэффициент отражения для неполяризованного света (суммы TE- и TM-мод с равными интенсивностями)?

3. **Преломление под углом Брюстера.** С помощью условия $n_1 \sec\theta_1 = n_2 \sec\theta_2$ и закона Снелла $n_1 \sin\theta_1 = n_2 \sin\theta_2$ выведите выражение (6.38) для угла Брюстера. Покажите также, что при падении под углом Брюстера $\theta_1 + \theta_2 = 90°$, так что направления отраженной и преломленной волн ортогональны и, следовательно, электрическое поле преломленной TM-волны параллельно направлению распространения отраженной волны. Отражение света можно трактовать как процесс рассеяния, в котором преломленная волна действует как источник излучения, генерирующий отраженную волну. При падении под углом Брюстера этот источник колеблется в направлении, параллельном направлению распространения отраженной волны, так что излучение не происходит и TM-волна не отражается.

4. **Задержка фазы при полном внутреннем отражении.** Определите задержку фазы между TE- и TM-волнами, вносимую полным внутренним отражением на границе между стеклом ($n = 1{,}5$) и воздухом ($n = 1$) при угле падения $\theta = 1{,}2\theta_{кр}$, где $\theta_{кр}$ — критический угол.

5. **Сдвиг Гуса—Хенхена.** Рассмотрим две плоские TE-волны, испытывающие полное внутреннее отражение при падении под углами θ и $\theta + d\theta$, где $d\theta$ — малый угол. Выразив вносимое при этом запаздывание по фазе между отраженными волнами в виде $d\varphi = \xi d\theta$, найдите выражение для коэффициента ξ. Изобразите картину интерференции двух падающих и двух отраженных волн и

убедитесь, что они смещены вдоль границы на расстояние, пропорциональное ξ. Если падающая волна представляет собой пучок (состоящий из множества плоских волн), то отраженный пучок смещен вдоль границы на расстояние, пропорциональное ξ. Это смещение называется сдвигом Гуса—Хенхена.

6. **Отражение от поглощающей среды.** С помощью уравнений Максвелла и соответствующих граничных условий покажите, что комплексный амплитудный коэффициент отражения на границе между свободным пространством и средой с показателем преломления n и коэффициентом поглощения α при нормальном падении равен

$$r = \left[\left(n - \frac{j\alpha c}{2\omega}\right) - 1\right]\left[\left(n - \frac{j\alpha c}{2\omega}\right) + 1\right].$$

К РАЗДЕЛУ 6.3

1. **Максимальная фазовая задержка в кварце.** Кварц — положительный одноосный кристалл с $n_o = 1{,}544$ и $n_н = 1{,}553$.
   а. Определите задержку фазы на 1 мм при $\lambda_0 = 633$ нм, если кристалл ориентирован так, что эта задержка максимальна.
   б. При какой толщине кристалла он действует как пластинка в четверть волны?

2. **Максимальное проявление необыкновенности.** Определите направление распространения в кварце ($n_o = 1{,}544$ и $n_н = 1{,}553$), при котором угол между волновым вектором $\boldsymbol{k}$ и вектором Пойнтинга $\boldsymbol{S}$ (направлением луча) максимален.

3. **Двулучепреломление.** Неполяризованная плоская волна падает из свободного пространства на кристалл кварца ($n_o = 1{,}544$ и $n_н = 1{,}553$) под углом 30° Оптическая ось лежит в плоскости падения и перпендикулярна направлению распространения падающей волны до того, как она вошла в кристалл. Определите направления волновых векторов и лучей двух преломленных волн.

4. **Боковой сдвиг при двулучепреломлении.** Какова оптимальная геометрия, обеспечивающая наибольший сдвиг вдоль поверхности между обыкновенным и необыкновенным отраженными пучками в положительном одноосном кристалле? Укажите все необходимые углы и направления.

5. **Прохождение через пластину $LiNbO_3$.** Исследуйте прохождение неполяризованного пучка He—Ne-лазера ($\lambda_0 = 633$ нм) при нормальном падении на пластину $LiNbO_3$ ($n_o = 2{,}20$ и $n_н = 2{,}29$) толщиной 2 см, вырезанную так, что ее оптическая ось образует угол 45° с нормалью к пластине. Определите поперечный сдвиг на выходе из пластины и фазовую задержку между обыкновенным и необыкновенным пучками.

6. **Коническое преломление.** Когда волновой вектор направлен вдоль оптической оси двуосного кристалла, возникает необычная ситуация. Два листа $\boldsymbol{k}$-поверхности пересекаются, и поверхность можно приближенно считать конической. Рассмотрим луч, нормально падающий на поверхность двуосного кристалла, у которого одна из оптических осей также направлена по нормали к поверхности. Покажите, что в этом случае возникает множественное преломм-

ление, при котором преломленные лучи образуют конус. Этот эффект известен как коническое преломление. Что происходит, когда конические лучи преломляются на параллельной границе при выходе из кристалла в воздух?

К РАЗДЕЛУ 6.6

1. **Круговой дихроизм.** Некоторые материалы имеют различные коэффициенты поглощения для волн, поляризованных по правому и левому кругу. Это свойство называется **круговым дихроизмом.** Определите матрицу Джонса устройства, преобразующего свет с любым состоянием поляризации в свет, поляризованный по правому кругу.

2. **Вращение плоскости поляризации последовательностью линейных поляризаторов.** Волна, линейно поляризованная в направлении оси x, проходит через последовательность N линейных поляризаторов, оси пропускания которых отклонены на углы $m\theta$ ($m = 1, 2, ..., N$; $\theta = \pi/2N$) по отношению к оси x. Покажите, что прошедший свет линейно поляризован в направлении оси y, но его амплитуда уменьшена на множитель $\cos^N \theta$. Что происходит в пределе при $N \to \infty$?

Указание. Используйте матрицы Джонса и соотношение

$$\mathbf{R}[(m + 1)\theta]\mathbf{R}(-m\theta) = \mathbf{R}(\theta),$$

где $\mathbf{R}(\theta)$ — матрица преобразования координат.

ГЛАВА 7

ОПТИКА ФОТОННЫХ КРИСТАЛЛОВ

Феликс Блох (1905—1983) разработал теорию, описывающую электронные волны в периодической структуре твердого тела.

Эли Яблонович (род. 1946) — один из авторов концепции фотонной запрещенной зоны; изготовил первый фотонный кристалл с запрещенной зоной.

Сейджив Джон (род. 1957) разработал теорию локализации фотонов, один из авторов идеи фотонной запрещенной зоны.

Распространение света в однородных средах, его отражение и преломление на границах их раздела — основной предмет интереса для оптики, рассмотренный в предыдущих главах. Устройства, используемые в фотонике, часто содержат многослойные структуры из различных материалов, предназначенные, например, для подавления или усиления отражения либо для изменения поляризационных или спектральных характеристик света. Многослойные структуры обнаруживаются в естественных физических и биологических системах и, например, отвечают за выраженную окраску некоторых насекомых и крыльев бабочек. Многослойные среды, в частности, могут быть периодическими, т. е. состоять из идентичных повторяющихся диэлектрических элементов, расположенных в виде одномерной, двумерной или трехмерной пространственных периодических структур, как показано на рис. 7.1. Одномерные периодические структуры состоят из повторяющихся плоскопараллельных многослойных элементов. Они часто используются в качестве дифракционных

решеток, отражающих свет при падении под определенными углами, или фильтров, селективно отражающих волны определенных частот. Двумерные периодические структуры состоят из параллельных стержней или цилиндрических отверстий, например, таких, которые используются для модификации характеристик оптических волокон (микроструктурные волокна, см. гл. 9). Трехмерные периодические структуры включают упорядоченные наборы из кубов, сфер или отверстий различной формы, расположение которых напоминает решетки природных кристаллов.

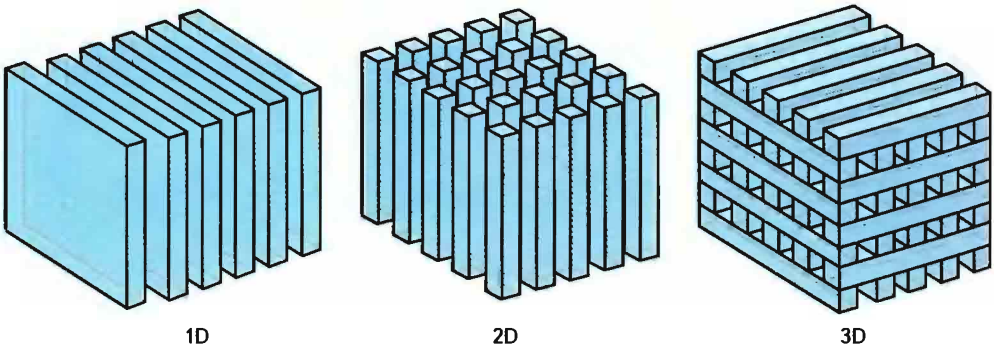


Рис. 7.1. Одномерные, двумерные и трехмерные периодические структуры, используемые в фотонике

Оптические волны, которые по своей природе периодичны, взаимодействуют с периодическим структурами уникальным образом, особенно если масштаб структуры того же порядка, что и длина волны света. Например, возникают спектральные полосы, в которых световые волны не могут распространяться в среде без жесткого ослабления. Волны с частотами, лежащими в этих так называемых **фотонных запрещенных зонах**, ведут себя примерно так, как при полном внутреннем отражении, но во всех направлениях. Подавление проходящей волны происходит в результате деструктивной интерференции волн, рассеянных элементами периодической структуры в прямом направлении. Замечательно, что этот эффект распространяется на целые сплошные полосы частот, а не на отдельные дискретные частоты.

Это явление аналогично свойствам электронов в кристаллических твердых телах, таких как полупроводники. Связанная с электроном периодическая волна распространяется в периодической решетке кристалла, и при этом, как правило, обнаруживаются запрещенные энергетические зоны. По причине этой аналогии периодические структуры в фотонике получили название **фотонных кристаллов**. Фотонные кристаллы уже нашли многочисленные применения, включающие использование в качестве специальных фильтров, волноводов и резонаторов. Многие их другие применения ожидаются в перспективе.

Для описания оптических свойств неоднородных сред, таких как многослойные и периодические среды, обычно требуется теоретический аппарат элек-

тромагнитной оптики. В неоднородных диэлектрических средах, как мы знаем из подразд. 5.2.2, диэлектрическая проницаемость $\varepsilon(\boldsymbol{r})$ меняется в пространстве, и волновые уравнения принимают общий вид (5.31), (5.32). Для гармонической волны с частотой ω это приводит к обобщенным уравнениям Гельмгольца для электрического и магнитного полей вида

$$\eta(\boldsymbol{r})\nabla\times(\nabla\times\boldsymbol{E})=\frac{\omega^2}{c_0^2}\boldsymbol{E}; \tag{7.1}$$

$$\nabla\times\left[\eta(\boldsymbol{r})\nabla\times\boldsymbol{H}\right]=\frac{\omega^2}{c_0^2}\boldsymbol{H}. \tag{7.2}$$

Обобщенные уравнения Гельмгольца

где $\eta(\boldsymbol{r}) = \varepsilon_0/\varepsilon(\boldsymbol{r})$ — диэлектрическая непроницаемость (см. подразд. 6.3.1). Одно из этих уравнений можно решить относительно электрического либо магнитного поля, после чего другое поле можно найти непосредственно из уравнений Максвелла. Обратите внимание на то, что (7.1) и (7.2) имеют вид задач на собственные значения: дифференциальный оператор, действуя на функцию поля, дает ту же функцию, умноженную на число. Собственными значениями являются ω^2/c_0^2, а собственные функции дают пространственное распределение для мод распространяющегося поля (см. приложение В). По причинам, которые будут объяснены ниже, мы будем работать с уравнением (7.2) для магнитного поля, а не с уравнением для электрического поля (7.1).

Для многослойных сред $\varepsilon(\boldsymbol{r})$ — кусочно-постоянная функция, т. е. постоянная внутри каждого слоя и меняющаяся скачком на границе раздела слоев. Поэтому при описании распространения волн можно использовать результаты, полученные для однородных сред, в сочетании с соответствующими граничными условиями, диктующими законы отражения и пропускания.

Периодические диэлектрические структуры характеризуются периодическими зависимостями $\varepsilon(\boldsymbol{r})$ и $\eta(\boldsymbol{r})$. Эта периодичность накладывает определенные условия на оптическую волну. Например, зависимость постоянной распространения от угловой частоты ω отличается от простой прямой пропорциональности, как в случае однородной среды. Если в однородной среде распространяющимися модами являются плоские волны вида $\exp(-j\boldsymbol{k}\cdot\boldsymbol{r})$, то моды периодической среды, называемые **модами Блоха**, представляют собой бегущие волны, промодулированные стоячими волнами.

О данной главе

В предыдущих главах внимание было сосредоточено на отделенных друг от друга тонких оптических элементах, таких как тонкие линзы, плоские дифракционные решетки или несущие изображение пленки, через которые проходит свет. Предметом этой главы являются *объемные* среды, включающие повторяющиеся диэлектрические слои, одномерные, двумерные и трехмерные периодические структуры. Раздел 7.1, в котором рассматриваются слоистые

среды, с меняющимися в одном измерении свойствами, служит прелюдией к рассмотрению периодических сред и фотонных кристаллов. Матричный подход обеспечивает систематическое описание многократных отражений, происходящих от множества границ раздела в таких средах. В разд. 7.2 вводится понятие фотонных кристаллов в их простейшей форме одномерных периодических структур. Матричные методы использованы для определения дисперсионных соотношений и зонной структуры. Представлен также альтернативный подход, основанный на представлении периодических функций, связанных со средой и волной, в виде рядов Фурье. Эти результаты обобщаются в разд. 7.3 на двумерные и трехмерные фотонные кристаллы.

Во всей главе различные среды предполагаются изотропными и поэтому описываются скалярной диэлектрической проницаемостью ε, однако отражение и преломление на границах имеют присущие им поляризационно-чувствительные характеристики.

Фотонные кристаллы в других главах

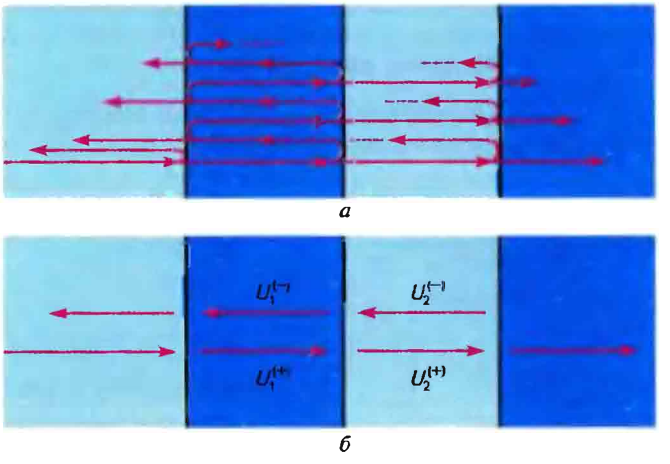
Благодаря своему свойству одинаково отражать свет по всем направлениям фотонные кристаллы могут быть использованы как «идеальные» диэлектрические зеркала. Слой однородного вещества, погруженный в фотонный кристалл, можно использовать для направления света посредством многократного отражения от границ. Применения оптических волноводов описаны в разд. 8.4. Аналогично, свет можно направлять по оптическому волокну с однородной сердцевиной, окруженной оболочкой из того же материала, но с цилиндрическими отверстиями, параллельными оси волокна. Такие «дырчатые» волокна, описанные в разд. 9.4, имеют ряд замечательных свойств, отсутствующих у обычных волокон. Полость в толще фотонного кристалла может служить оптическим резонатором, поскольку она имеет идеально отражающие стенки на частотах, лежащих в фотонной запрещенной зоне. Фотонно-кристаллические микрорезонаторы будут кратко описаны в подразд. 10.4.4 и 17.4.3.

7.1. ОПТИКА МНОГОСЛОЙНЫХ ДИЭЛЕКТРИЧЕСКИХ СРЕД

7.1.1. Матричная теория многослойной оптики

Плоская волна, нормально падающая на слоистую среду, частично отражается на каждой границе каждого слоя, а частично проходит через нее. Получающиеся при этом волны испытывают бесконечную последовательность отражений и пропусканий через границы, как показано на рис. 7.2, *а*. Комплексные амплитуды прошедшей и отраженной волн можно найти по формулам Френеля на каждой границе (см. разд. 6.2). Полный коэффициент отражения и пропускания среды можно в принципе вычислить путем суперпозиции этих отдельных волн. Эта методика использовалась в подразд. 2.5.2 для определения пропускания интерферометра Фабри—Перо.

При большом числе слоев отслеживание бесконечного числа микроотражений и пропусканий превращается в утомительную процедуру. Альтернативный макроподход основан на том факте, что в каждом слое существует всего два типа волн, а именно, волны, направленные вперед и назад. Суммируясь, эти волны в каждой точке образуют одну волну $U^{(+)}$, направленную вперед, и одну волну $U^{(-)}$, бегущую назад (рис. 7.2, *б*). Таким образом, описание распространения волны в слоистой среде эквивалентно определению амплитуд этой пары волн во всех ее точках. Комплексные амплитуды четырех волн с двух сторон каждой границы можно связать, накладывая соответствующие граничные условия или просто используя формулы Френеля для отражения и пропускания.



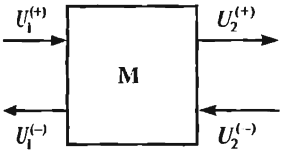
а

б

Рис. 7.2. Отражения одной волны от границ раздела в многослойной среде (*а*). В каждом слое волны, бегущие вперед, собираются в одну прямую волну $U^{(+)}$, а волны, бегущие назад, — в одну обратную волну $U^{(-)}$ (*б*)

Волновая матрица передачи

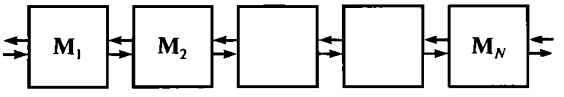
Расчет комплексных амплитуд прямой и обратной волн в многослойной среде облегчается при использовании матричного метода. Рассмотрим две произвольные плоскости 1 и 2 внутри данной оптической системы. Амплитуды суммарной прямой и суммарной обратной волн в этих плоскостях обозначим $U_1^{(\pm)}$ и $U_2^{(\pm)}$ соответственно и представим в виде матриц-столбцов. Линейная связь между ними может быть записана в матричном виде



$$\begin{bmatrix} U_2^{(+)} \\ U_2^{(-)} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} U_1^{(+)} \\ U_1^{(-)} \end{bmatrix}. \qquad (7.3)$$

Матрица **M** с элементами A, B, C и D называется волновой матрицей передачи, или матрицей пропускания. Она зависит от оптических свойств слоистой среды между двумя плоскостями.

Многослойную среду удобно представлять в виде последовательности базовых элементов, описываемых известными волновыми матрицами передачи $\mathbf{M}_1$, $\mathbf{M}_2$, ..., $\mathbf{M}_N$. Амплитуды прямой и обратной волн на двух границах всей среды тогда связаны одной матрицей, равной произведению матриц элементов

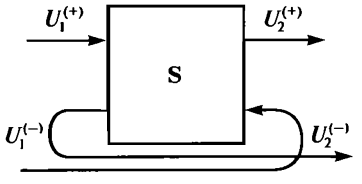


$$\mathbf{M} = \mathbf{M}_N ... \mathbf{M}_2 \mathbf{M}_1, \qquad (7.4)$$

где элементы 1, 2, нумеруются слева направо, как показано на рисунке. Формула каскада волновых матриц передачи (7.4) идентична формуле каскада матриц передачи лучей (1.49), что оказывается очень полезным.

Матрица рассеяния

Альтернативным способом связать четыре комплексные амплитуды $U_{1,2}^{(+)}$ на двух границах слоистой среды служит матрица рассеяния, или S-матрица. Она часто используется для описания линий передачи, микроволновых схем и рассеивающих систем. В этом случае выходящие волны выражаются через входящие:



$$\begin{bmatrix} U_2^{(+)} \\ U_1^{(-)} \end{bmatrix} = \begin{bmatrix} t_{12} & r_{21} \\ r_{12} & t_{21} \end{bmatrix} \begin{bmatrix} U_1^{(+)} \\ U_2^{(-)} \end{bmatrix}, \qquad (7.5)$$

где элементы S-матрицы обозначены t_{12}, r_{21}, r_{12} и t_{21}. В отличие от элементов волновой матрицы передачи эти элементы имеют прямой физический смысл. Величины t_{12} и r_{21} представляют собой амплитудные коэффициенты пропускания и отражения для прямой волны, т. е. волны, падающей слева, а r_{12} и t_{21} — амплитудные коэффициенты пропускания и отражения для обратной волны, т. е. волны, падающей справа. Индекс 12, например, означает, что волна падает из среды 1 в среду 2. В этом легко убедиться, замечая, что если в плоскости 2 нет обратной волны, так что $U_2^{(-)} = 0$, мы получаем

$$U_2^{(+)} = t_{12} U_1^{(+)}; \quad U_1^{(-)} = r_{12} U_1^{(+)}$$

Аналогично, если нет прямой волны в плоскости 1, так что $U_1^{(+)} = 0$, мы получаем

$$U_2^{(+)} = r_{21} U_2^{(-)}; \quad U_1^{(-)} = t_{21} U_2^{(-)}.$$

Явным преимуществом формализма S-матрицы является то, что ее элементы непосредственно связаны с физическими параметрами системы. С другой стороны, недостатком является то, что S-матрица каскада элементов не есть матричное произведение S-матриц отдельных элементов. Полезная систематиче-

ческая процедура для анализа каскадных систем включает оба подхода: используется перемножение матриц **M**, после чего результирующая матрица преобразуется в матрицу **S** для определения полного коэффициента отражения и пропускания.

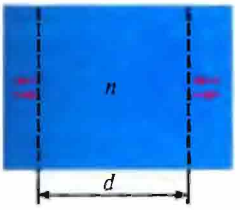
Пример 7.1

Распространение в однородной среде

Для однородного слоя толщины d с показателем преломления n комплексные амплитуды суммарных волн, показанных стрелками, связаны соотношениями

$$U_2^{(+)} = e^{-j\varphi}U_1^{(+)}; \quad U_1^{(-)} = e^{-j\varphi}U_2^{(-)},$$

где $\varphi = nk_0d$, так что в этом случае волновая матрица и матрица рассеяния имеют вид



$$\mathbf{M} = \begin{bmatrix} \exp(-j\varphi) & 0 \\ 0 & \exp(j\varphi) \end{bmatrix}; \qquad \mathbf{S} = \begin{bmatrix} \exp(-j\varphi) & 0 \\ 0 & \exp(-j\varphi) \end{bmatrix}, \quad \varphi = nk_0d. \tag{7.6}$$

Связь между матрицей рассеяния и волновой матрицей передачи

Элементы матриц **M** и **S** связаны через определяющие уравнения (7.3) и (7.3), из которых получаются следующие формулы преобразования:

$$\mathbf{M} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{1}{t_{21}} \begin{bmatrix} t_{12}t_{21} - r_{12}r_{21} & r_{21} \\ -r_{12} & 1 \end{bmatrix}; \tag{7.7}$$

$$\mathbf{S} = \begin{bmatrix} t_{12} & r_{21} \\ r_{12} & t_{21} \end{bmatrix} = \frac{1}{D} \begin{bmatrix} AD - BC & B \\ -C & 1 \end{bmatrix}. \tag{7.8}$$

Формулы преобразования

Эти формулы не справедливы в предельных случаях, когда $t_{21} = 0$ или $D = 0$.

Резюме

Матричная волновая оптика предоставляет систематическую процедуру для определения амплитудных коэффициентов пропускания и отражения стопы диэлектрических слоев с заданными показателями преломления и толщинами:

- Стопа делится на последовательность элементов, включающих границы с однородными слоями между ними.

• Для каждого элемента определяется матрица M. Для этого используются формулы Френеля, получаются коэффициенты пропускания и отражения, строится S-матрица и конвертируется в M-матрицу с помощью (7.7).

• Полная M-матрица многослойной среды получается перемножением M-матриц отдельных элементов в соответствии с формулой (7.4).

• Матрица рассеяния всей среды получается преобразованием (7.8) из M-матрицы. Элементы полученной S-матрицы непосредственно дают амплитудные коэффициенты пропускания и отражения для всей многослойной среды.

Каскад из двух систем: формулы Эйри

Матричный метод можно использовать для выражения элементов матрицы рассеяния составной системы через элементы матриц рассеяния ее составных частей. Рассмотрим прохождение волны через систему, описываемую S-матрицей с элементами t_{12}, r_{21}, r_{12} и t_{21}, а затем через другую систему с элементами S-матрицы t_{23}, r_{32}, r_{23} и t_{32},. Путем перемножения двух M-матриц с последующим преобразованием результата в S-матрицу получаются полные коэффициенты пропускания и отражения:

$$t_{13} = \frac{t_{12}t_{23}}{1 - r_{21}r_{23}}; \quad r_{13} = r_{12} + \frac{t_{12}t_{21}r_{23}}{1 - r_{21}r_{23}}. \tag{7.9}$$

Если две последовательные системы разделены участком однородной среды, как показано на рис. 7.3, то нужно вставить волновую матрицу передачи однородной среды (7.6) с фазой $\varphi = nk_0d$, где d — длина участка однородной среды; n — ее показатель преломления. Это приводит к выражениям, известным как **формулы Эйри**:

$$\boxed{t_{13} = \frac{t_{12}t_{23}\exp(-j\varphi)}{1 - r_{21}r_{23}\exp(-j2\varphi)}; \quad r_{13} = r_{12} + \frac{t_{12}t_{21}r_{23}\exp(-j2\varphi)}{1 - r_{21}r_{23}\exp(-j2\varphi)}.} \tag{7.10}$$

Формулы Эйри

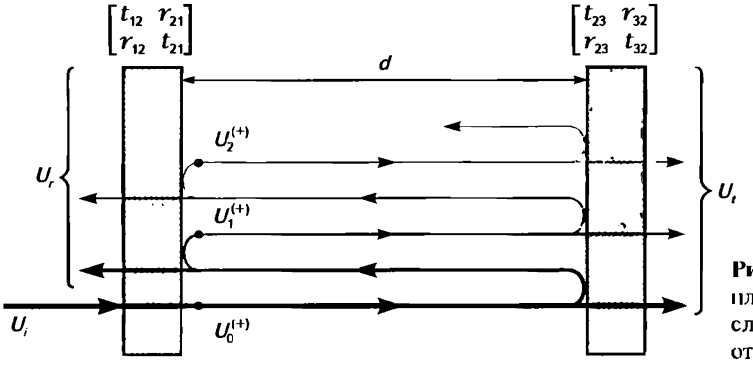


Рис. 7.3. Прохождение плоской волны через последовательность двух отдельных систем

Формулы Эйри можно вывести также путем отслеживания многократных отражений и пропусканий, испытываемых падающей волной между двумя подсистемами и сложения комплексных амплитуд, как показано на рис. 7.3. Плоская волна с комплексной амплитудой U_i, падающая на первую подсистему, производит начальную внутреннюю волну с амплитудой $U_0^{(+)} = t_{12}U_i$, которая отражается туда и обратно между двумя подсистемами, порождая дополнительные внутренние волны $U_1^{(+)}$, $U_2^{(+)}$, ... , все бегущие в прямом направлении. Амплитуда полной прошедшей волны U_t связана с полной внутренней амплитудой $U^{(+)} = U_0^{(+)} + U_1^{(+)} +$ соотношением $U_t = t_{23}e^{-j\varphi}U^{(+)}$, где $\varphi = nk_0d$. Следовательно, полный амплитудный коэффициент пропускания равен

$$t_{13} = \frac{U_t}{U_i} = t_{12}t_{23}e^{-j\varphi}\frac{U^{(+)}}{U_0^{(+)}}.$$

Поскольку

$$U^{(+)} = U_0^{(+)}(1 + h + h^2 + ...) = \frac{U_0^{(+)}}{1 - h},$$

где $h = r_{21}r_{23}e^{-j2\varphi}$ — мультипликативный фактор обхода, для полного амплитудного коэффициента пропускания t_{13} получается формула Эйри (7.10)

Следствия закона сохранения энергии для сред без потерь

Если среда между плоскостями 1 и 2 не вносит потерь, то входящая и выходящая оптические мощности должны быть равны. Далее, если среды в плоскостях входа и выхода имеют одинаковые импедансы и показатели преломления, то эти мощности характеризуются квадратами модулей комплексных амплитуд $|U_{1,2}^{(\pm)}|^2$. В этом случае закон сохранения энергии требует, чтобы

$$\left|U_1^{(+)}\right|^2 + \left|U_2^{(-)}\right|^2 = \left|U_2^{(+)}\right|^2 + \left|U_1^{(-)}\right|^2$$

для любой комбинации амплитуд входящих волн. Выбирая амплитуды входящих волн $U_1^{(+)}$ и $U_2^{(-)}$ в виде (1, 0), (0, 1) и (1, 1), из вышеприведенного закона сохранения получаем три уравнения, связывающие элементы матрицы рассеяния S, с помощью которых можно доказать следующие формулы:

$$|t_{12}| = |t_{21}| \equiv |t|, \quad |r_{12}| = |r_{21}| \equiv |r|, \quad |t|^2 + |r|^2 = 1; \tag{7.11}$$

$$\frac{t_{12}}{t_{21}^*} = -\frac{r_{12}}{r_{21}^*}. \tag{7.12}$$

Равенства (7.11) связывают модули элементов S-матрицы для сред без потерь, у которых во входной и выходной плоскостях одинаковый показатель преломления, а (7.12) связывает их аргументы.

Формулы (7.11) и (7.12) можно преобразовать в соотношения для элементов — матрицы **M**:

$$|D| = |A|, \quad |C| = |B|, \quad |A|^2 - |B|^2 = 1; \tag{7.13}$$

$$\det \mathbf{M} = \frac{C}{B^*} = \frac{A}{D^*} = \frac{t_{12}}{t_{21}}, \quad |\det \mathbf{M}| = 1. \tag{7.14}$$

Эти результаты можно получить путем подстановки формул (7.11) и (7.12) в формулы преобразования M- и S-матриц (7.13) и (7.14).

Взаимные системы без потерь

Для систем без потерь, обладающих свойством взаимности, т. е. систем, коэффициенты пропускания и отражения которых не зависят от направления прохождения волны, имеем: $t_{12} = t_{21} \equiv t$, $r_{21} = r_{12} \equiv r$. В этом случае из (7.11) и (7.12) следует

$$|t|^2 + |r|^2 = 1; \quad \frac{t}{r} = -\left(\frac{t}{r}\right)^* \quad \arg\{t\} - \arg\{r\} = \pm\frac{\pi}{2}, \tag{7.15}$$

откуда видно, что фазы, связанные с пропусканием и отражением, различаются на $\pi/2$. При этих условиях элементы M-матрицы удовлетворяют следующим соотношениям:

$$A = D^*; \quad B = C^* \cdot \quad |A|^2 - |B|^2 = 1; \quad \det \mathbf{M} = 1. \tag{7.16}$$

Матрицы **S** и **M** тогда принимают простой вид:

$$\boxed{\mathbf{S} = \begin{bmatrix} t & r \\ r & t \end{bmatrix}; \quad \mathbf{M} = \begin{bmatrix} \dfrac{1}{t^*} & \dfrac{r}{t} \\ \dfrac{r^*}{t^*} & \dfrac{1}{t} \end{bmatrix},} \tag{7.17}$$

Взаимная система без потерь

и система описывается двумя комплексными числами t и r, связанными условиями (7.15).

Пример 7.2

Частично отражающее зеркало (светоделитель)

Частично отражающее зеркало без потерь, помещенное в однородную среду, представляет собой взаимную систему, S-матрица которой дается формулой (7.17). Предполагая, что фаза $\arg\{t\} = 0$, получаем, что (7.15) диктует условие $\arg\{r\} = \mp\pi/2$, так что $r = \mp j|r|$. Используя знак «+», получаем модель матрицы рассеяния для светоделителя в виде

$$\mathbf{S} = \begin{bmatrix} |t| & j|r| \\ j|r| & |t| \end{bmatrix}; \quad |t|^2 + |r|^2 = 1. \tag{7.18}$$

Соответствующая M-матрица:

$$\mathbf{M} = \frac{1}{|t|}\begin{bmatrix} 1 & j|r| \\ -j|r| & 1 \end{bmatrix}. \tag{7.19}$$

Идеальное зеркало имеет S-матрицу (7.18) с $|r| = 1$ и $|t| = 1$. В этом предельном случае (7.19) неприменимо, и M-матрица не обеспечивает должного представления, поскольку две стороны идеального зеркала изолированы и независимы.

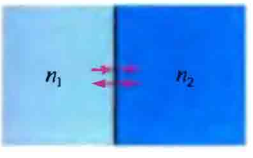
Пример 7.3

Одна граница между диэлектриками

В этом примере система представляет собой одну границу. В соответствии с формулами Френеля (см. разд. 6.2) коэффициенты пропускания и отражения на границе раздела сред с показателями преломления n_1 и n_2 определяются S-матрицей

$$\mathbf{S} = \begin{bmatrix} t_{12} & r_{21} \\ r_{12} & t_{21} \end{bmatrix} = \frac{1}{n_1 + n_2}\begin{bmatrix} 2n_1 & n_2 - n_1 \\ n_2 - n_1 & 2n_2 \end{bmatrix}. \tag{7.20}$$

Подстановка в (7.7) дает M-матрицу

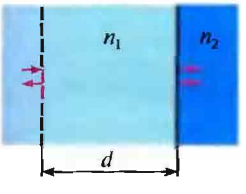


$$\mathbf{M} = \frac{1}{2n_2}\begin{bmatrix} n_2 + n_1 & n_2 - n_1 \\ n_2 - n_1 & n_2 + n_1 \end{bmatrix}. \tag{7.21}$$

Пример 7.4

Распространение в однородной среде + граница

Матрица M однородного слоя среды толщиной d, за которым следует граница, есть матрица границы (7.21), умноженная на матрицу однородного слоя (7.6)

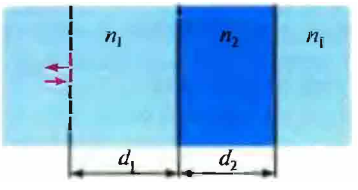


$$\mathbf{M} = \frac{1}{2n_2}\begin{bmatrix} (n_2 + n_1)e^{-j\varphi} & (n_2 - n_1)e^{j\varphi} \\ (n_2 - n_1)e^{-j\varphi} & (n_2 + n_1)e^{j\varphi} \end{bmatrix}, \quad \varphi = n_1 k_0 d. \tag{7.22}$$

Пример 7.5

Распространение в однородной среде + прохождение через слой

Система представляет собой каскад из двух подсистем, обе из которых рассмотрены в примере 7.6. В первой подсистеме свет проходит из среды с показателем преломления n_1 в среду с показателем n_2. Согласно (7.4), полная матрица **M** есть произведение составляющих **M**-матриц, причем матрицы перемножаются в обратном порядке:



$$\mathbf{M} = \frac{1}{4n_1n_2}\begin{bmatrix}(n_1+n_2)e^{-j\varphi_2} & (n_1-n_2)e^{j\varphi_2}\\ (n_1-n_2)e^{-j\varphi_2} & (n_1+n_2)e^{j\varphi_2}\end{bmatrix}\times \times\begin{bmatrix}(n_2+n_1)e^{-j\varphi_1} & (n_2-n_1)e^{j\varphi_1}\\ (n_2-n_1)e^{j\varphi_1} & (n_2+n_1)e^{j\varphi_1}\end{bmatrix}. \qquad (7.23)$$

Здесь $\varphi_1 = n_1k_0d_1$ и $\varphi_2 = n_2k_0d_2$, где d_1 и d_2 — толщины двух слоев соответственно. Тогда элементы матрицы **M**

$$A = D^* = \frac{1}{t^*} = \frac{1}{4n_1n_2}\left[(n_1+n_2)^2e^{-j\varphi_2} - (n_2-n_1)^2e^{j\varphi_2}\right]e^{-j\varphi_1}; \qquad (7.24)$$

$$B = C^* = \frac{r}{t} = \frac{1}{4n_1n_2}\left(n_2^2-n_1^2\right)^2\left[e^{-j\varphi_2} - e^{j\varphi_2}\right]e^{j\varphi_1} \qquad (7.25)$$

удовлетворяют свойствам взаимных систем без потерь [соотношения (7.16)].

Из (7.24) и (7.25) можно определить выражения для t и r. Так

$$t = \exp(-j\varphi_1)\frac{4n_1n_2\exp(-j\varphi_2)}{(n_1+n_2)^2 - (n_1-n_2)^2\exp(-j2\varphi_2)}. \qquad (7.26)$$

Это выражение можно также непосредственно вывести, рассматривая систему как комбинацию двух границ, разделенных участком распространения через однородную среду, из формулы Эйри (7.10) при

$$t_{12} = t_{32} = \frac{2n_1}{n_1+n_2}, \quad t_{21} = t_{23} = \frac{2n_2}{n_1+n_2},$$

$$r_{12} = r_{32} = -r_{21} = -r_{23} = \frac{n_1-n_2}{n_1+n_2}.$$

Упражнение 7.1

Четвертьволновая пленка как антиотражательное покрытие

Специально разработанные тонкие пленки часто применяются для того, чтобы исключить или уменьшить отражение от границы между средами с

различными показателями преломления. Рассмотрим тонкую пленку с показателем преломления n_2 и толщиной d, заключенную между двумя средами с показателями преломления n_1 и n_3 (рис. 7.4). Получите выражение для элемента B матрицы M для этой многослойной среды. Покажите, что свет, падающий из среды 1, испытывает нулевой коэффициент отражения, если $d = \lambda/4$ и $n_2 = \sqrt{n_1 n_3}$, где $\lambda = \lambda_0/n_2$.

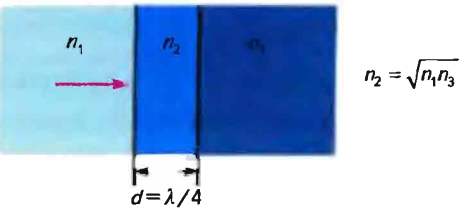
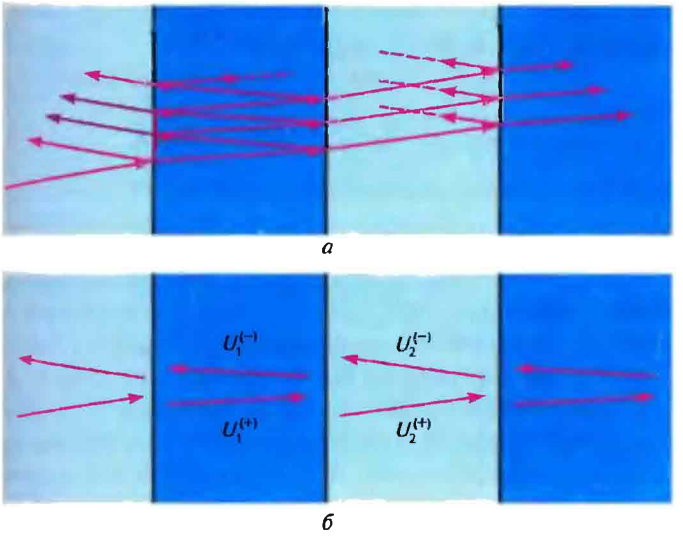


Рис. 7.4. Антиотражательное покрытие

Внеосевые волны в слоистых средах

При наклонном падении волны на слоистую среду отраженные и прошедшие волны в ходе последующих отражений и прохождений многократно отражаются туда и обратно в каждом слое, как показано на рис. 7.5, *а*. Законы преломления и отражения гарантируют параллельность направлений распространения всех прямых и всех обратных волн в каждом слое. Более того, в каждом слое угол наклона всех волн одинаков, если для прямых волн отсчитывать его от направления $+z$, а для обратных волн — от $-z$.



а

б

Рис. 7.5. Отражения одной наклонно падающей волны от границ в многослойной среде (*а*). В каждом слое все прямые волны складываются в одну общую прямую волну; то же самое происходит с обратными волнами (*б*)

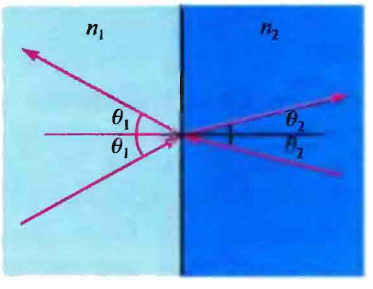
«Макроскопический» подход, ранее использованный для нормального падения волн, можно применять и в случае наклонного падения. Разница состоит лишь в том, что френелевские коэффициенты пропускания и отражения на границе t_{12}, r_{21}, r_{12} и t_{21} становятся зависящими от угла наклона и поляризации волн.

Простейший пример — распространение на расстояние d в однородной среде с показателем преломления n под углом θ к оси z. Волновая матрица передачи **M** дается формулой (7.6), где фаза теперь равна $\varphi = nk_0 d\cos\theta$. Два других примера приведены ниже.

Пример 7.6

Одна граница: наклонная TE-волна

Волна, проходящая плоскую границу между средами с показателями преломления n_1 и n_2 при углах θ_1 и θ_2, удовлетворяющих закону Снелла ($n_1\sin\theta_1 = n_2\sin\theta_2$), описывается **S**-матрицей, определяемой из формул Френеля (6.34) и (6.35) и соответствующей **M**-матрицей



$$\mathbf{S} = \begin{bmatrix} t_{12} & r_{21} \\ r_{12} & t_{21} \end{bmatrix} = \frac{1}{\tilde{n}_1 + \tilde{n}_2}\begin{bmatrix} 2a_{12}\tilde{n}_1 & \tilde{n}_2 - \tilde{n}_1 \\ \tilde{n}_1 - \tilde{n}_2 & 2a_{21}\tilde{n}_2 \end{bmatrix}; \qquad (7.27)$$

$$\mathbf{M} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{1}{2a_{21}\tilde{n}_2}\begin{bmatrix} \tilde{n}_1 + \tilde{n}_2 & \tilde{n}_2 - \tilde{n}_1 \\ \tilde{n}_2 - \tilde{n}_1 & \tilde{n}_1 + \tilde{n}_2 \end{bmatrix}. \qquad (7.28)$$

Эти выражения применимы как для TE-, так и для TM-волн при следующих определениях:

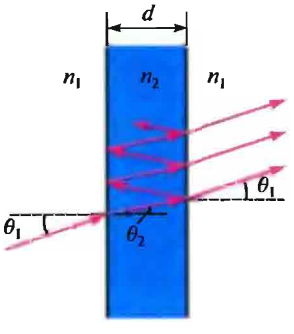
$$\text{TE:}\quad \tilde{n}_1 = n_1\cos\theta_1,\quad \tilde{n}_2 = n_2\cos\theta_2,\quad a_{12} = a_{21} = 1;$$

$$\text{TM:}\quad \tilde{n}_1 = n_1\sec\theta_1,\quad \tilde{n}_2 = n_2\sec\theta_2,\quad a_{12} = \frac{\cos\theta_1}{\cos\theta_2} = \frac{1}{a_{21}}.$$

Пример 7.7

Распространение в однородной среде с последующим прохождением слоя: внеосевая волна

В этом примере наклонная волна проходит через систему, рассмотренную в примере 7.5: слой толщиной d_2 с показателем преломления n_2 в среде с показателем преломления n_1. Волна проходит расстояние d_1 в окружающей среде перед входом в слой. Волновая матрица передачи для наклонной волны получается обобщением результата, полученного при нормальном падении:



$$\mathbf{M} = \frac{1}{4\tilde{n}_1\tilde{n}_2}\begin{bmatrix} (\tilde{n}_1 + \tilde{n}_2)e^{-j\tilde{\varphi}_2} & (\tilde{n}_1 - \tilde{n}_2)e^{j\tilde{\varphi}_2} \\ (\tilde{n}_1 - \tilde{n}_2)e^{-j\tilde{\varphi}_2} & (\tilde{n}_1 + \tilde{n}_2)e^{j\tilde{\varphi}_2} \end{bmatrix} \times \begin{bmatrix} (\tilde{n}_2 + \tilde{n}_1)e^{-j\tilde{\varphi}_1} & (\tilde{n}_2 - \tilde{n}_1)e^{j\tilde{\varphi}_1} \\ (\tilde{n}_2 - \tilde{n}_1)e^{-j\tilde{\varphi}_1} & (\tilde{n}_2 + \tilde{n}_1)e^{j\tilde{\varphi}_1} \end{bmatrix}, \qquad (7.29)$$

где $\tilde{\varphi}_1 = n_1 k_0 d_1 \cos\theta_1$ и $\tilde{\varphi}_2 = n_2 k_0 d_2 \cos\theta_2$. Как и в примере 7.8, для ТЕ-поляризации $\tilde{n}_1 = n_1 \cos\theta_1$, $\tilde{n}_2 = n_2 \cos\theta_2$, для ТМ-поляризации — $\tilde{n}_1 = n_1 \sec\theta_1$, $\tilde{n}_2 = n_2 \sec\theta_2$.

Выражение для матрицы **M** (7.29) совпадает с (7.23), описывающим аксиальную систему, за исключением того, что параметры n_1, n_2, φ_1 и φ_2 нужно заменить на зависящие от угла и поляризации параметры $\tilde{n}_1$, $\tilde{n}_2$ и зависящие от угла $\tilde{\varphi}_1$, $\tilde{\varphi}_2$. Заметим, что множители a_{12} и a_{21}, появляющиеся в (7.28) на каждой границе, сокращаются, поскольку $a_{12}a_{21} = 1$. Подставляя полученные результаты в выражение (7.26) из примера (7.5), получаем его обобщение на случай внеосевых волн, где появляется зависимость от поляризации:

$$t = \exp(-j\tilde{\varphi}_1)\frac{4\tilde{n}_1\tilde{n}_2\exp(-j\tilde{\varphi}_2)}{(\tilde{n}_1+\tilde{n}_2)^2 - (\tilde{n}_1-\tilde{n}_2)^2\exp(-j2\tilde{\varphi}_2)}. \qquad (7.30)$$

7.1.2. Эталон Фабри—Перо

Эталон Фабри—Перо был введен в подразд. 2.5.2 как интерферометр, образованный двумя параллельными зеркалами с высоким коэффициентом отражения, пропускающий свет только для особого набора равноотстоящих пространственных частот, определяемых оптической длиной пути между зеркалами. Он используется как фильтр и как спектроанализатор, управляемый посредством изменения указанной длины, т. е. посредством смещения одного из зеркал. Он применяется также в качестве оптического резонатора (см. разд. 10.1). В данном разделе мы изучим его как многослойную систему при помощи развитого выше матричного метода.

Зеркальный эталон Фабри—Перо

Рассмотрим два частично отражающих зеркала без потерь с амплитудными коэффициентами пропускания t_1, t_2 и отражения r_1, r_2, расстояние между которыми d заполнено средой с показателем преломления n. Вся система описывается произведением матриц

$$\mathbf{M} = \begin{bmatrix} \frac{1}{t_1^*} & \frac{r_1}{t_1} \\ \frac{r_1^*}{t_1^*} & \frac{1}{t_1} \end{bmatrix} \begin{bmatrix} \exp(-j\varphi) & 0 \\ 0 & \exp(j\varphi) \end{bmatrix} \begin{bmatrix} \frac{1}{t_2^*} & \frac{r_2}{t_2} \\ \frac{r_2^*}{t_2^*} & \frac{1}{t_2} \end{bmatrix}, \qquad (7.31)$$

где $\varphi = nk_0 d$. Поскольку система взаимная и не вносит потерь, матрица **M** принимает упрощенный вид (7.17), тогда амплитудный коэффициент пропускания t равен обратной величине матричного элемента D матрицы **M**, так что

$$t = \frac{t_1 t_2 \exp(-j\varphi)}{1 - r_1 r_2 \exp(-j2\varphi)}. \qquad (7.32)$$

Это соотношение можно вывести также путем непосредственного применения формулы Эйри (7.18).

В результате коэффициент пропускания эталона по интенсивности равен

$$\mathcal{T} = |t|^2 = \frac{|t_1 t_2|^2}{|1 - r_1 r_2 \exp(-j2\varphi)|^2}. \tag{7.33}$$

Данное выражение аналогично (2.61) для интенсивности бесконечного числа волн с равными разностями фаз и амплитудами, убывающими по закону геометрической прогрессии, как описано в подразд. 2.5.2. Предполагая, что $\arg\{r_1 r_2\} = 0$, это выражение можно записать виде[1]

$$\mathcal{T} = |t|^2 = \frac{\mathcal{T}_{\max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2 \varphi}, \tag{7.34}$$

где

$$\mathcal{T}_{\max} = \frac{|t_1 t_2|^2}{(1 - |r_1 r_2|)^2} = \frac{(1 - |r_1|^2)(1 - |r_2|^2)}{(1 - |r_1 r_2|)^2}; \tag{7.35}$$

$$\boxed{\mathcal{F} = \frac{\pi\sqrt{|r_1 r_2|}}{1 - |r_1 r_2|}}. \tag{7.36}$$

Параметр резкости

Параметр $\mathcal{F}$, называемый **параметром резкости**, есть монотонно возрастающая функция произведения коэффициентов отражения $r_1 r_2$ и является мерой качества эталона. Например, если $r_1 r_2 = 0{,}99$, то $\mathcal{F} \approx 313$.

Как описано в подразд. 2.5.2, коэффициент пропускания $\mathcal{T}$ является периодической функцией φ с периодом π. Он достигает максимального значения $\mathcal{T}_{\max}$, которое равно единице, если $|r_1| = |r_2|$, когда φ — целое кратное π. Когда параметр резкости велик (т. е. $|r_1 r_2| \approx 1$), зависимость $\mathcal{T}$ от φ имеет острые пики шириной приблизительно $\pi/\mathcal{F}$. Таким образом, чем выше параметр резкости $\mathcal{F}$, тем острее пики пропускания в зависимости от фазы φ.

Фаза $\varphi = nk_0 d = (\omega/c)d$ пропорциональна частоте, так что условие $\varphi = \pi$ соответствует условию, $\omega = \omega_F$ или $\nu = \nu_F$, где

$$\boxed{\nu_F = \frac{c}{2d}; \quad \omega_F = \frac{\pi c}{d}} \tag{7.37}$$

Область дисперсии интерферометра

называется **областью дисперсии** интерферометра. Получается, что коэффициент пропускания является периодической функцией частоты $\mathcal{T}(\nu)$ с периодом ν_F:

$$\boxed{\mathcal{T}(\nu) = \frac{\mathcal{T}_{\max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2(\pi\nu/\nu_F)},} \tag{7.38}$$

Коэффициент пропускания (Эталон Фабри—Перо)

[1] Это выражение воспроизводит (2.63) при замене набегом фазы за полный обход.

что иллюстрирует рис. 7.6. Коэффициент пропускания достигает максимального значения $\mathcal{T}_{max}$ на резонансных частотах $\nu_q = q\nu_F$, где q — целое число. Когда параметр резкости велик, $\mathcal{F} \gg 1$, $\mathcal{T}(\nu)$ резко спадает с удалением от резонансной частоты ν_q, так что график функции $\mathcal{T}(\nu)$ напоминает гребень с острыми зубьями. Спектральная ширина каждой из областей высокого пропускания

$$\delta\nu = \frac{\nu_F}{\mathcal{F}} \tag{7.39}$$

в $\mathcal{F}$ раз меньше расстояния между соседними резонансными частотами.

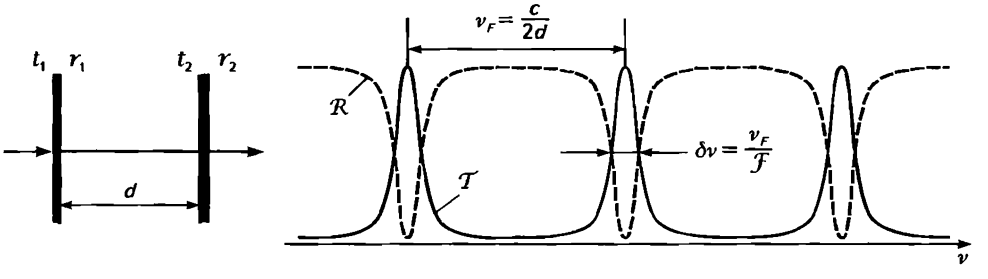


Рис. 7.6. Коэффициенты пропускания и отражения по интенсивности $\mathcal{T}$ и $\mathcal{R}$ эталона Фабри—Перо в зависимости от частоты

Эталон Фабри—Перо можно использовать как остро настраиваемый оптический фильтр или анализатор спектра. Однако ввиду периодичности спектрального отклика полоса частот исследуемого света не должна превышать ширину области дисперсии $\nu_F = c/2d$ во избежание неоднозначности. Фильтр можно настраивать (т. е. сдвигать резонансные частоты), меняя расстояние d между зеркалами. Малое приращение Δd сдвигает резонансную частоту $\nu_q = qc/2d$ на сравнительно большую величину

$$\Delta\nu_q = -\frac{qc}{2d^2}\Delta d = -\frac{\nu_q \Delta d}{d}.$$

Хотя частотное расстояние ν_F также меняется, но на значительно меньшую величину $-\nu_F \Delta d/d$. Например, расстояние между зеркалами d =1,5 см дает область дисперсии ν_F = 10 ГГц при n = 1. Для типичной оптической частоты $\nu = 10^{14}$ Гц, что соответствует $q = 10^4$, уменьшение d на 10^{-4} от его величины (Δd = 1,5 мкм) сдвигает частоту пика на $\Delta\nu_q$ = 10 Ггц, тогда как ширина области дисперсии меняется лишь на 1 МГц и становится равной 9,999 ГГц.

Применения эталона Фабри—Перо в качестве резонатора описаны в разд. 10.1.

Диэлектрический слой как эталон Фабри—Перо

На основе (7.26) коэффициент пропускания диэлектрического слоя толщины d с показателем преломления n_2, находящегося в среде с показателем преломления n_1, можно записать в виде

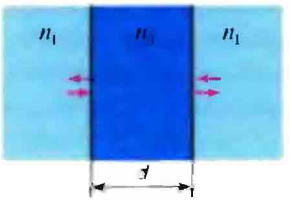
$$t = \frac{4n_1 n_2 \exp(-j\varphi)}{(n_1 + n_2)^2 - (n_1 - n_2)^2 \exp(-j2\varphi)}, \tag{7.40}$$

где $\varphi = n_2 k_0 d$. Это выражение воспроизводит (7.32), описывающее эталон Фабри—Перо, если выполнить подстановки

$$t_1 t_2 = \frac{4n_1 n_2}{n_1 + n_2}; \quad r_1 r_2 = \frac{(n_1 - n_2)^2}{(n_1 + n_2)^2}.$$

Отсюда следует, что выражения для коэффициента пропускания по интенсивности (7.34) и (7.38), описывающие зеркальный эталон, можно применять к диэлектрическому слою. Из (7.36) параметр резкости слоя получается равным

$$\mathcal{F} = \frac{\pi}{4} \frac{\left| n_2^2 - n_1^2 \right|}{n_1 n_2}. \tag{7.41}$$



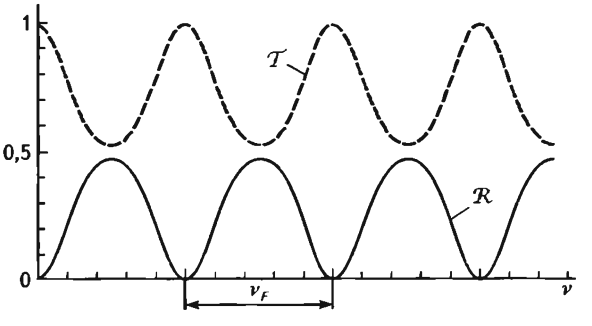


Рис. 7.7. Частотная зависимость коэффициентов пропускания $\mathcal{T}$ и отражения $\mathcal{R}$ по интенсивности у слоя с показателем преломления $n_2 = 3{,}5$ (показатель преломления Si) в среде с показателем преломления $n_1 = 1{,}5$ (показатель преломления SiO_2)

Обычно большие значения $\mathcal{F}$ не достигаются в эталонах в виде диэлектрического слоя. Например, для $n_1 = 1{,}5$ (показатель преломления SiO_2) и $n_2 = 3{,}5$ (показатель преломления Si) получается $\mathcal{F} = 1{,}68$. Как видно на рис. 7.7, частотная зависимость $\mathcal{T}$ в этом случае не имеет острых пиков, присущих эталонам с хорошо отражающими зеркалами, как на рис. 7.6. Для повышения добротности на поверхности слоя нужно нанести покрытие, усиливающее внутреннее отражение.

Внеосевое пропускание эталона Фабри—Перо

Для волны, падающей под углом θ к оси зеркального эталона, амплитудный коэффициент пропускания дается формулой (7.32), где фазу φ нужно заменить на $\tilde{\varphi} = nk_0 d\cos\theta$. Отсюда следует, что обобщением выражения (7.38) для коэффициента пропускания по интенсивности на внеосевой случай является

$$\mathcal{T}(\nu) = \frac{\mathcal{T}_{\max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2(\pi\cos\theta\nu/\nu_F)}. \tag{7.42}$$

Максимальное пропускание происходит на частотах, для которых

$$\boxed{\nu = q\nu_F \sec\theta}, \quad q = 1,\ 2, \qquad \nu_F = \frac{c}{2d}. \tag{7.43}$$

Резонансное условие

Если параметр резкости эталона велик, то пропускание происходит на этих частотах и почти отсутствует на всех других. Соответствующий этому соотношению график, приведенный на рис. 7.8, *в* (см. цв. вклейку), показывает, что для каждого угла θ существует конечный набор дискретных частот, на которых возможно пропускание. Аналогично, на каждой частоте ν волна может распространяться через эталон под вполне определенными углами, набор которых также дискретен и конечен. Поэтому конус падающего широкополосного (белого) света, проходя через эталон, порождает набор концентрических радужных колец, показанный на рис. 7.8, *б* (см. цв. вклейку). Если спектр падающего света более узкий, чем ширина области дисперсии ν_F, то каждая частотная компонента соответствует одному и только одному углу, так что эталон можно использовать как спектральный прибор.

7.1.3. Решетка Брэгга

Решетка Брэгга была введена в упражнении 2.11 как набор равноотстоящих параллельных частично отражающих плоских зеркал. Такая структура обладает угловой и частотной селективностью, полезной во многих приложениях. В этом разделе мы обобщим определение решетки Брэгга, включив в него наборы из N равноотстоящих идентичных многослойных сегментов, и разовьем для него теорию отражения света на основе матричной волновой оптики.

Упрощенная теория

Коэффициент отражения решетки Брэгга был определен в упражнении 2.11 при двух допущениях:

1) зеркала считались слабо отражающими, так что интенсивность падающей волны не терялась при ее распространении;

2) вторичным отражением (отражением уже отраженных волн) можно было пренебречь.

В этих приближениях коэффициент отражения $\mathcal{R}_N$ решетки из N зеркал связан с коэффициентом отражения одного зеркала R соотношением[2]

$$\mathcal{R}_N = \frac{\sin^2 N\varphi}{\sin^2 \varphi} R. \tag{7.44}$$

Как описано в подразд. 2.5.2, множитель $\sin^2 N\varphi/\sin^2 \varphi$ представляет собой квадрат модуля суммы комплексных векторов с единичной амплитудой и разностью фаз 2φ. Эта функция имеет максимальное значение, равное N^2, при выполнении условий Брэгга, т. е. когда 2φ равно $q2\pi$, где $q = 0, 1, 2,$ Она быстро спадает при удалении от этих значений, причем ширина пиков обратно пропорциональна N. В этой упрощенной модели полная интенсивность отраженной волны превышает интенсивность волны, отраженной от одного слоя, не более чем в N^2 раз.

Для решетки Брэгга, образованной частично отражающими зеркалами, отстоящими друг от друга на расстояние Λ, набег фазы за полный обход равен $2\varphi = 2k\Lambda \cos\theta$, где θ — угол падения. Следовательно, максимальное отражение происходит, когда $2k\Lambda \cos\theta = q2\pi$ или

$$\boxed{\cos\theta = q\frac{\lambda}{2\Lambda} = q\frac{\omega_B}{\omega} = q\frac{\nu_B}{\nu},} \tag{7.45}$$

Условие Брэгга

где

$$\boxed{\nu_B = \frac{c}{2\Lambda}; \quad \omega_B = \frac{\pi c}{\Lambda}} \tag{7.46}$$

Брэгговская частота

есть брэгговская частота.

При нормальном падении ($\theta = 0°$) пиковое отражение достигается на частотах, равных целому кратному брэгговской частоты, т. е. $\nu = q\nu_B$. На частотах $\nu < \nu_B$ условие Брэгга не может быть выполнено ни при каких значениях угла θ. На частотах $\nu_B < \nu < 2\nu_B$ условие Брэгга выполняется при единственном значении угла

$$\theta = \arccos\frac{\lambda}{2\Lambda} = \arccos\frac{\nu_B}{\nu}.$$

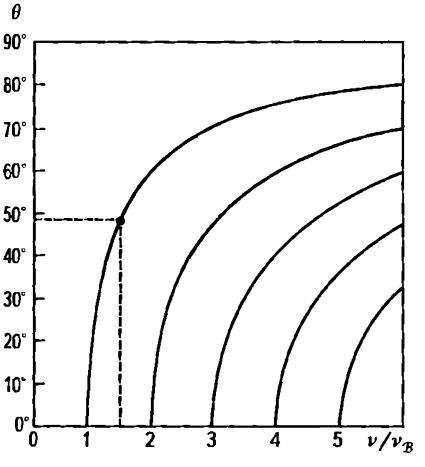
Угол, дополнительный к этому, $\theta_B = \pi/2 - \theta$, есть угол Брэгга [см. (2.8) и рис. 2.29]

$$\boxed{\theta_B = \arcsin\left(\frac{\lambda}{2\Lambda}\right).} \tag{7.47}$$

Угол Брэгга

[2] Напомним, что в упражнении 2.11 φ обозначает разность фаз между двумя последовательными векторами на комплексной плоскости, тогда как теперь эта величина обозначена через, поскольку она описывает 2φ набег фазы за полный обход.

На частотах $\nu \geq 2\nu_B$ условие Брэгга выполняется при более чем одном значении угла. На рис. 7.9 проиллюстрированы спектральная и угловая зависимости отражения от брэгговской решетки, вытекающие из упрощенной теории.



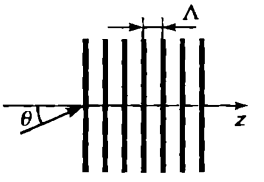


Рис. 7.9. Геометрическое место частот ν и углов θ, при которых выполняется условие Брэгга. Например, если $\nu = 1{,}5\nu_B$, то $\theta = 48{,}2°$ Это соответствует брэгговскому углу $\theta_B = 41{,}8°$ (отсчитываемому от плоскости решетки)

Матричная теория

Используем теперь матричный подход, введенный в предыдущем разделе, для развития точной теории брэгговского отражения, включая как многократные отражения, так и ослабление падающей волны в процессе распространения. Оказывается, совместное действие многократных отражений приводит к усилению полной отраженной волны и к тому, что отражение происходит не на отдельных частотах, являющихся целыми кратными $\nu_B/\cos\theta$, а в протяженных спектральных полосах, окружающих эти частоты.

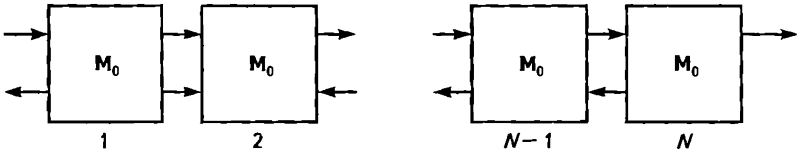


Рис. 7.10. Решетка Брэгга, образованная N сегментами, каждый из которых описывается матрицей $\mathbf{M}_0$

Рассмотрим решетку, образованную стопой из N идентичных первичных элементов (рис. 7.10), каждый из которых описывается унимодулярной волновой матрицей передачи $\mathbf{M}_0$ для консервативной взаимной системы без потерь, а именно

$$\mathbf{M}_0 = \begin{bmatrix} \dfrac{1}{t^*} & \dfrac{r}{t} \\ \dfrac{r^*}{t^*} & \dfrac{1}{t} \end{bmatrix}, \tag{7.48}$$

где t и r — комплексные амплитудные коэффициенты пропускания и отражения, удовлетворяющие условиям (7.15); тогда $\mathcal{T} = |t|^2$ и $\mathcal{R} = |r|^2$ — соответствующие коэффициенты пропускания и отражения по интенсивности.

В соответствии с (7.4) волновая матрица передачи $\mathbf{M}$ для N сегментов есть просто произведение $\mathbf{M} = \mathbf{M}_0^N$ Так как $\mathbf{M}_0$ унимодулярна, т. е. $\det \mathbf{M}_0 = 1$, она обладает свойством

$$\mathbf{M}_0^N = \Psi_N \mathbf{M}_0 - \Psi_{N-1}\mathbf{I}, \tag{7.49}$$

где

$$\boxed{\Psi_N = \frac{\sin N\Phi}{\sin \Phi}}; \tag{7.50}$$

$$\boxed{\cos\Phi = \mathrm{Re}\left\{\frac{1}{t}\right\}} \tag{7.51}$$

и $\mathbf{I}$ — единичная матрица. Равенство (7.49) можно доказать по индукции, т. е. показать, что если оно верно для $(N-1)$ сегментов, то оно будет верно и для N сегментов. Последнее нетрудно сделать с помощью прямой подстановки и применения тригонометрических тождеств.

Так как совокупная N-сегментная система также является взаимной и не имеет потерь, ее матрица может быть записана в виде

$$\mathbf{M}_0^N = \begin{bmatrix} \dfrac{1}{t_N^*} & \dfrac{r_N}{t_N} \\ \dfrac{r_N^*}{t_N^*} & \dfrac{1}{t_N} \end{bmatrix}, \tag{7.52}$$

где t_N и r_N — амплитудные коэффициенты пропускания и отражения системы из N сегментов соответственно. Подставляя (7.48) и (7.52) в (7.49) и сравнивая диагональные и недиагональные элементы матриц в обеих частях равенства, получаем

$$\frac{1}{t_N} = \Psi_N \frac{1}{t} - \Psi_{N-1}; \tag{7.53}$$

$$\frac{r_N}{t_N} = \Psi_N \frac{r}{t}. \tag{7.54}$$

Эти два равенства выражают t_N и r_N через t и r.

Коэффициент пропускания по интенсивности $\mathcal{T}_N = |t_N|^2$ получается путем взятия квадрата модуля (7.54) и использования соотношения $\mathcal{R} = 1 - \mathcal{T}$:

$$\mathcal{T}_N = \frac{\mathcal{T}}{\mathcal{T} + \Psi_N^2 (1 - \mathcal{T})}. \tag{7.55}$$

Отсюда следует, что коэффициент отражения по интенсивности $\mathcal{R}_N = 1 - \mathcal{T}_N$ выражается как

$$\boxed{\mathcal{R}_N = \frac{\Psi_N^2 \mathcal{R}}{1 - \mathcal{R} + \Psi_N^2 \mathcal{R}}.} \qquad (7.56)$$

Коэффициент отражения решетки Брэгга

Резюме

Коэффициент отражения $\mathcal{R}_N$ среды, состоящей из N идентичных сегментов, связан с коэффициентом отражения одного сегмента $\mathcal{R}$ нелинейным соотношением (7.56), содержащим множитель Ψ_N, который возникает благодаря интерференции волн, многократно отраженных от N сегментов решетки. Определяемый соотношением (7.50) фактор Ψ_N зависит от числа сегментов N и другого параметра Ф, который связан с комплексным амплитудным коэффициентом пропускания одного сегмента t формулой (7.51).

Зависимость $\mathcal{R}_N$ от $\mathcal{R}$, описываемая (7.56), упрощается в предельных случаях. Если коэффициент отражения одного сегмента очень мал, т. е. $\mathcal{R} \ll 1$, и Ψ_N^2 не очень велико, так что $\Psi_N^2 \mathcal{R} \ll 1$, то (7.56) приближенно записывается как

$$\mathcal{R}_N \approx \Psi_N^2 \mathcal{R} = \frac{\sin^2 N\Phi}{\sin^2 \Phi} \mathcal{R}. \qquad (7.57)$$

Это выражение совпадает по виду с приближенным выражением (7.44), где Ф играет роль фазы φ.

В противоположном случае $\Psi_N^2 \gg 1$ коэффициент отражения принимает вид

$$\mathcal{R}_N = \frac{\Psi_N^2 \mathcal{R}}{1 + \Psi_N^2 \mathcal{R}}.$$

Эта нелинейная связь отражает насыщение и типична для систем с обратной связью, которая в данном случае является результатом многократных внутренних отражений от границ сегментов. В конечном счете если $\Psi_N^2 \gg 1$, то $\mathcal{R}_N$ достигает максимального значения — единицы, так что структура из N сегментов ведет себя как идеальное зеркало, несмотря на то, что каждый из сегментов отражает лишь частично. Большой интерференционный фактор Ψ_N ускоряет приближение $\mathcal{R}_N$ к единице по мере роста $\mathcal{R}$.

Интерференционный фактор Ψ_N, зависящий от $\Phi = \arccos(\mathrm{Re}\{1/t\})$ через (7.50), ведет себя по двум различным сценариям:

1) нормальный режим, при котором Ф действительно и решетка проявляет частичное отражение/пропускание (включая нулевое отражение, или полное пропускание);

2) аномальный режим, когда Ф комплексное и Ψ_N может быть предельно большим, что соответствует полному отражению.

Режим частичного или нулевого отражения

Этот режим определяется условием $|\mathrm{Re}\{1/t\}| \le 1$, которое гарантирует действительность $\Phi = \arccos(\mathrm{Re}\{1/t\})$. В этом случае $\mathcal{R}_N$ зависит от $\mathcal{R}$ и Ψ_N в соответствии с (7.50) и (7.56). Максимальное отражение происходит, когда Ψ_N достигает своего наибольшего значения, равного N. В этом случае

$$\mathcal{R}_N = \frac{N^2\mathcal{R}}{1-\mathcal{R}+N^2\mathcal{R}}.$$

Следовательно, $\mathcal{R}_N$ не может точно равняться единице, кроме случая точного равенства $\mathcal{R} = 1$. Например, для $N = 10$, $\mathcal{R} = 1$ максимальное значение $\mathcal{R}_N$ составляет ≈0,99.

Нулевое отражение, или **полное пропускание** возможно даже в том случае, когда коэффициент отражения отдельного сегмента R значителен. Это происходит когда $\Psi_N = 0$, т. е. когда $\sin N\Phi = 0$ или $\Phi = q\pi/N$ для $q = 0, 1, \ldots, (N-1)$. N частот, на которых имеет место такая полная прозрачность, являются резонансными частотами решетки. Это явление представляет собой один из видов туннелирования через сегменты, каждый из которых частично отражает свет.

Режим полного отражения

В этом режиме $|\mathrm{Re}\{1/t\}| = |\cos\Phi| > 1$, так что Φ — комплексная переменная, $\Phi = \Phi_R + j\Phi_I$. Используя тождество

$$\cos(\Phi_R + j\Phi_I) = \cos\Phi_R \,\mathrm{ch}\,\Phi_I - j\sin\Phi_R \,\mathrm{sh}\,\Phi_I$$

и приравнивая действительные и мнимые части в обеих частях (7.51), получаем $\sin\Phi_R = 0$, так что $\Phi_R = m\pi$, и $\cos\Phi_R = +1$ или -1, когда m — четное или нечетное число соответственно, что дает

$$\boxed{\mathrm{ch}\,\Phi_I = \left|\mathrm{Re}\left\{\frac{1}{t}\right\}\right|.} \qquad (7.58)$$

Режим полного отражения

Фактор $\Psi_N = \sin N\Phi/\sin\Phi$ тогда приобретает вид

$$\boxed{\Psi_N = \pm\frac{\mathrm{sh}\,N\Phi_I}{\mathrm{sh}\,\Phi_I},} \qquad (7.59)$$

Режим полного отражения

где знак «±» определяется знаком множителя $\cos(Nm\pi)/\cos(m\pi)$. Поскольку sh ( ) растет экспоненциально при больших N, $|\Psi_N|$ может быть много больше, чем N. В этом случае, в соответствии с (7.56), коэффициент отражения $\mathcal{R}_N \approx 1$, и решетка действует как полный отражатель. Прямые волны становятся нераспространяющимися и не проникают в глубь структуры, что в значительной мере напоминает полное внутреннее отражение.

Поскольку Ф зависит от коэффициента t, который зависит от частоты ν, два режима, как хорошо видно из следующих ниже примеров, соответствуют различным частотным полосам. Спектральные полосы, связанные с режимом полного отражения, называются **полосами заграждения**, так как в них пропускание света почти полностью блокируется. Второй режим соответствует **полосам пропускания.** Внутри этих полос на отдельных резонансных частотах имеет место полное пропускание (нулевое отражение).

Пример 7.8

Пакет частично отражающих зеркал

Рассмотрим решетку из N одинаковых частично отражающих зеркал, расстояние между которыми Λ заполнено однородной средой с показателем преломления n (рис. 7.11, *а*). Отдельный сегмент решетки состоит из слоя однородного вещества толщиной Λ и частично отражающего зеркала с амплитудными коэффициентами пропускания и отражения t и r соответственно.

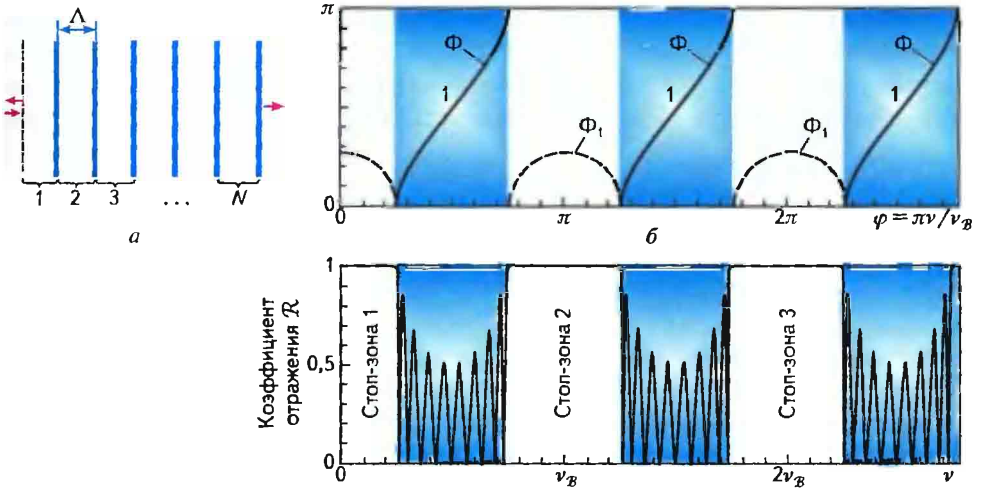


Рис. 7.11. Решетка Брэгга из $N = 10$ идентичных зеркал, каждое из которых имеет коэффициент отражения по интенсивности $|r|^2 = 0{,}5$ (*а*). Зависимость Ф от фазовой задержки $\varphi = nk_0\Lambda$ между соседними зеркалами. В областях, не отмеченных фоном, Ф комплексное, и его мнимая часть Φ_I показана штриховыми линиями (*б*). Коэффициент отражения $\mathcal{R}$ как функция частоты (в единицах частоты Брэгга $\nu_B = c/2\Lambda$) (*в*). В полосах заграждения коэффициент отражения примерно равен единице

Волновая матрица передачи **M** такого сегмента определяется умножением матрицы (7.19) на матрицу (7.6):

$$\mathbf{M}_0 = \frac{1}{|t|}\begin{bmatrix} e^{-j\varphi} & j|r|e^{j\varphi} \\ -j|r|e^{-j\varphi} & e^{j\varphi} \end{bmatrix}, \quad \varphi = nk_0\Lambda = \frac{\pi\nu}{\nu_B}, \tag{7.60}$$

где ν_B — частота Брэгга. Отсюда получается $t = |t|e^{j\varphi}$ и, следовательно, Ф:

$$\cos\Phi = \frac{1}{|t|}\cos\varphi \quad \text{при} \quad |\cos\varphi| \le |t|; \tag{7.61}$$

$$\cosh\Phi_I = \frac{1}{|t|}\cos\varphi \quad \text{при} \quad |\cos\varphi| > |t|. \tag{7.62}$$

Связи между Ф и φ, а также между Φ_I и φ являются нелинейными и необычными, как видно из рис. 7.11, *б*. Соответствующая зависимость коэффициента отражения по интенсивности $\mathcal{R}_N$ от φ показана на рис. 7.11, *в*. В нормальном режиме (области более темного фона) Ф действительно, а коэффициент отражения имеет многочисленные пики с нулями между ними. Ни один из пиков не достигает единицы, несмотря на то, что Ψ_N достигает своего максимального значения $N = 10$.

Ситуация совершенно иная в режиме полного отражения (светлые области), где Ф становится комплексным. Фактор Ψ_N достигает значения ≈3000 в центре полосы ($\varphi = \pi$), когда $|t|^2 = 0{,}5$. Эти области соответствуют значениям φ, при которых происходит полное отражение ($\mathcal{R}_N \approx 1$). Поскольку φ пропорционально частоте ν, на рис. 7.11, *в* фактически изображена спектральная зависимость коэффициента отражения, и светлые области соответствуют полосам заграждения.

Пример 7.9

Диэлектрическая решетка Брэгга

Решетка состоит из N идентичных диэлектрических слоев толщиной d_2 каждый с показателем преломления n_2, разделенных промежутками толщиной d_1, заполненными веществом с показателем преломления n_1, как показано на рис. 7.12. Ее можно рассматривать как пакет из плотно сложенных двойных слоев, каждый из которых описан в примере 7.7. Элемент $A = 1/t^*$ волновой матрицы передачи $\mathbf{M}_0$ дается формулой (7.24), откуда

$$\mathrm{Re}\left\{\frac{1}{t}\right\} = \frac{(n_1+n_2)^2}{4n_1n_2}\cos(\varphi_1+\varphi_2) - \frac{(n_2-n_1)^2}{4n_1n_2}\cos(\varphi_1-\varphi_2), \tag{7.63}$$

где $\varphi_1 = n_1k_0d_1$ и $\varphi_2 = n_2k_0d_2$ — фазовые сдвиги, вносимые каждым из слоев сегмента. Этот результат можно использовать в сочетании с (7.50), (7.51), (7.56), (7.58) и (7.59) для получения коэффициента отражения решетки.

Спектральную зависимость коэффициента отражения можно вычислить, если заметить, что

$$\varphi_1 + \varphi_2 = k_0(n_1d_1 + n_2d_2) = \frac{\pi\nu}{\nu_B},$$

где

$$\nu_B = \frac{c_0/\bar{n}}{2\Lambda};$$

$\bar{n}$ — средний показатель преломления,

$$\bar{n} = \frac{n_1 d_1 + n_2 d_2}{\Lambda}.$$

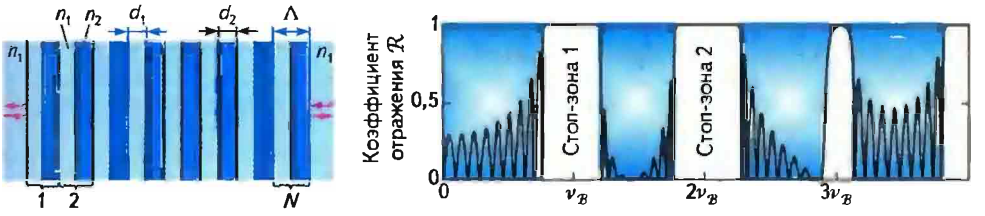


Рис. 7.12. Коэффициент отражения по интенсивности как функция частоты для диэлектрической решетки Брэгга из $N = 10$ сегментов, каждый из которых имеет два слоя толщиной $d_1 = d_2$ с показателями преломления $n_1 = 1{,}5$ и $n_2 = 3{,}5$. Решетка находится в среде с согласованным показателем преломления n_1. Коэффициент отражения примерно равен единице внутри полос заграждения (*светлые участки*), центры которых находятся на частотах, кратных $\nu_B = c/2\Lambda$, где $c = c_0/\bar{n}$, а $\bar{n}$ — средний показатель преломления

Частота Брэгга ν_B — это частота, на которой фазовый набег за полный обход одного сегмента $2k_0(n_1 d_1 + n_2 d_2) = 2\pi$. Разность фаз

$$\varphi_1 - \varphi_2 = \frac{\zeta \pi \nu}{\nu_B},$$

где величина

$$\zeta = \frac{n_1 d_1 - n_2 d_2}{n_1 d_1 + n_2 d_2}$$

также пропорциональна частоте. На рис. 7.12 (справа) представлен пример спектральной зависимости коэффициента отражения.

Пример 7.10

Диэлектрическая решетка Брэгга: наклонное падение

Результаты примера 7.9 можно обобщить на случай наклонного падения волн под углом θ_1 в среде 1, которому в среде 2 соответствует угол θ_2, причем $n_1 \sin\theta_1 = n_2 \sin\theta_2$. В этом случае (7.63) приобретает вид

$$\mathrm{Re}\left\{\frac{1}{t}\right\} = \frac{(\tilde{n}_1 + \tilde{n}_2)^2}{4\tilde{n}_1\tilde{n}_2}\cos(\tilde{\varphi}_1 + \tilde{\varphi}_2) - \frac{(\tilde{n}_2 - \tilde{n}_1)^2}{4\tilde{n}_1\tilde{n}_2}\cos(\tilde{\varphi}_1 - \tilde{\varphi}_2), \qquad (7.64)$$

где $\tilde{\varphi}_1 = n_1 k_0 d_1 \cos\theta_1$, $\tilde{\varphi}_2 = n_2 k_0 d_2 \cos\theta_2$; $\tilde{n}_1 = n_1 \cos\theta_1$, $\tilde{n}_2 = n_2 \cos\theta_2$ для ТЕ-поляризации; $\tilde{n}_1 = n_1 \sec\theta_1$, $\tilde{n}_2 = n_2 \sec\theta_2$ для ТМ-поляризации. Это соотношение можно использовать для нахождения спектральной зависимости коэффициента отражения при любом угле падения. На рис. 7.13 показана зависимость коэффициента отражения по интенсивности $\mathcal{R}_N$ от частоты и угла падения для ТЕ- и

ТМ-поляризаций у высококонтрастной решетки. Область углов, в которой коэффициент отражения близок к единице, расширяется с ростом отношения n_2/n_1, определяющего контраст решетки.

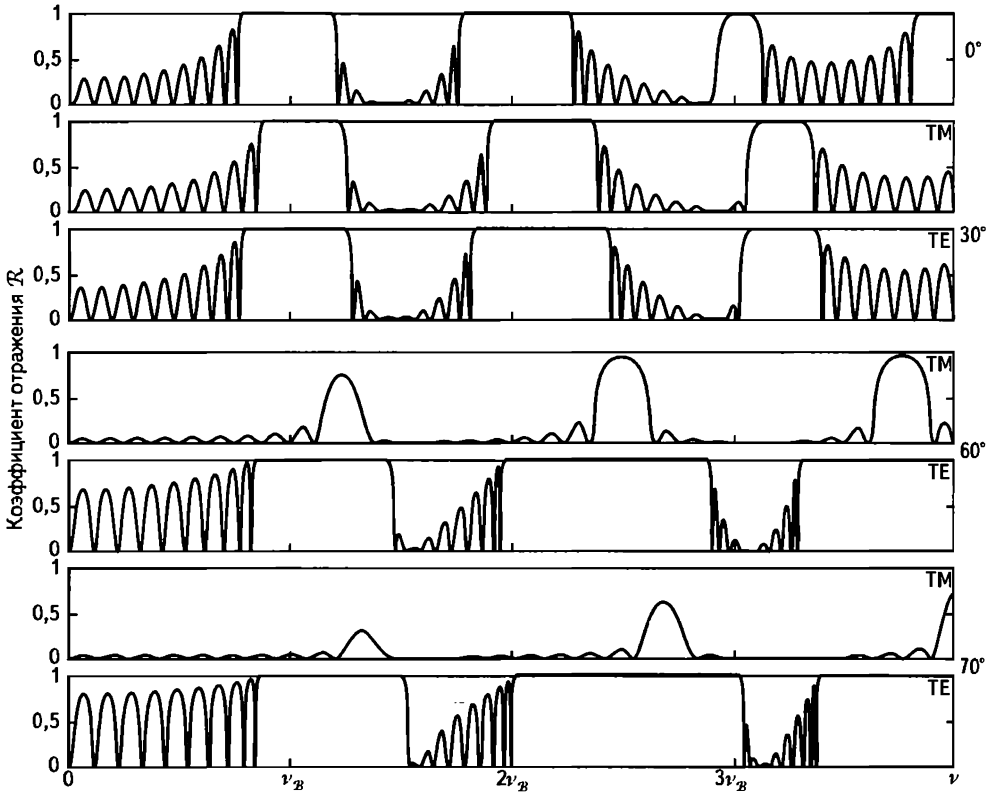


Рис. 7.13. Спектральная зависимость коэффициента отражения $\mathcal{R}$ для решетки Брэгга из 10 сегментов, показанной на рис. 7.12, при нескольких значениях угла падения θ_1 для ТЕ- и ТМ-поляризаций

Решетка Брэгга в несогласованной среде

При анализе, проведенном выше, предполагалось, что решетка из N идентичных сегментов находится в согласованной среде. Это означает, что если каждый сегмент состоит из нескольких диэлектрических слоев, то показатель преломления окружающей среды равен показателю преломления переднего слоя, так что свет не испытывает дополнительного преломления на передней границе системы и отражается на задней границе так, как если бы далее следовал очередной слой решетки. Решетка, описанная в примере 7.9, как раз и удовлетворяет этим условиям.

В большинстве приложений решетка находится в несогласованной среде, такой как воздух, и необходимо учитывать краевые эффекты. Для этого нужно

записать волновую матрицу передачи **M** составной системы, включающей все границы, а затем преобразовать ее в соответствующую матрицу рассеяния **S**. Зная **S**, легко определить коэффициент отражения полной системы.

Пусть $\mathbf{M}_0^N$ — волновая матрица передачи решетки из N сегментов в среде, согласованной с передним слоем, тогда полная волновая матрица передачи может быть записана в виде

$$\mathbf{M} = \mathbf{M}_e \mathbf{M}_0^{N-1} \mathbf{M}_i, \tag{7.65}$$

где $\mathbf{M}_i$ — волновая матрица передачи входной границы; $\mathbf{M}_e$ — волновая матрица передачи N-го сегмента с выходом в несогласованную среду.

Пример 7.11

Коэффициент отражения диэлектрической решетки Брэгга в несогласованной среде

Решетка Брэгга, состоящая из N пар чередующихся слоев диэлектрика с показателями преломления n_1 и n_2 и толщинами d_1 и d_2, помещена в диэлектрическую среду с показателем преломления n_0. Найдем коэффициент отражения для волны, падающей под углом θ_0 из внешней среды (рис. 7.14). Этому углу соответствуют углы θ_1 и θ_2 в первом и втором слоях каждого сегмента, определяемые законом Снелла ($n_1 \sin\theta_1 = n_2 \sin\theta_2$).

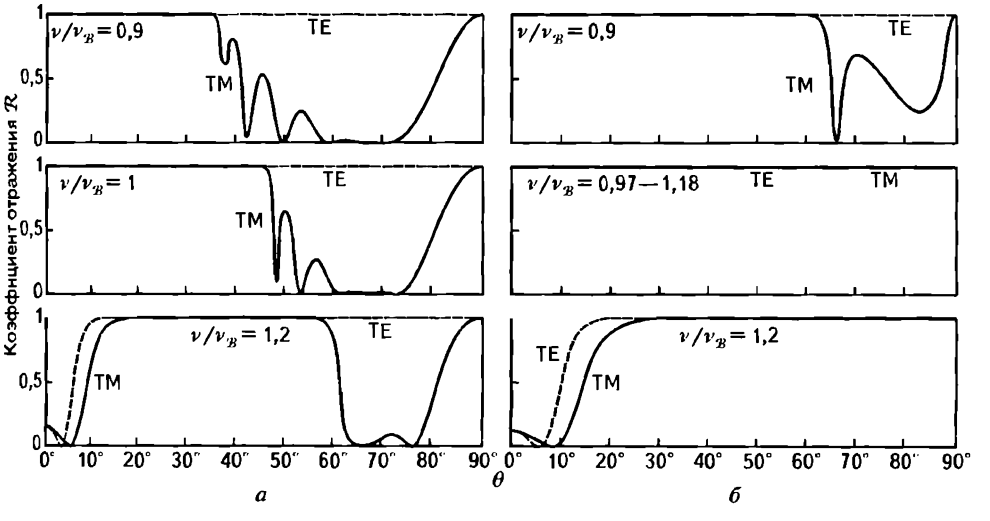


Рис. 7.14. Коэффициент отражения по интенсивности как функция угла падения θ при фиксированных частотах для решетки, изображенной на рис. 7.12:

а — решетка находится в согласованной среде ($n = n_1$); *б* — решетка находится в воздухе ($n = 1$). В воздухе решетка имеет единичный коэффициент отражения при всех углах как для ТЕ-, так и для ТМ-поляризаций, в полосе частот от $0{,}97\nu_B$ до $1{,}18\nu_B$

В этом случае можно использовать (7.65) со следующими волновыми матрицами передачи:

1) $\mathbf{M}_i$ представляет границу раздела между средами с показателями преломления n_0 и n_1, как описано в примере 7.6;

2) $\mathbf{M}_0$ представляет один сегмент решетки, как описано в примере 7.7;

3) $\mathbf{M}_e$ представляет распространение на расстояние d_1 в среде с показателем преломления n_1, затем в слое толщиной d_2 с показателем преломления n_2, затем прохождение границы между средами с n_2 и n_0.

После нахождения матрицы M используется соотношение (7.8) для определения соответствующей матрицы рассеяния S. Полный коэффициент отражения есть элемент r_{12} в (7.8).

7.2. ОДНОМЕРНЫЕ ФОТОННЫЕ КРИСТАЛЛЫ

Одномерные (1D) фотонные кристаллы — это диэлектрические структуры, оптические свойства которых периодически меняются в одном направлении, называемом осью периодичности, и постоянны в ортогональных направлениях. Эти структуры обладают уникальными оптическими свойствами, особенно когда их период того же порядка, что и длина волны. Если принять ось периодичности за ось z, то такие оптические свойства, как диэлектрическая проницаемость $\varepsilon(z)$ и непроницаемость $\eta(z) = \varepsilon_0/\varepsilon(z)$, являются периодическими функциями z, удовлетворяющими соотношению

$$\eta(z + \Lambda) = \eta(z) \tag{7.66}$$

для всех z, где Λ — период. Распространение волн в таких периодических средах можно исследовать посредством решения обобщенных уравнений Гельмгольца (7.2) для периодического $\eta(z)$.

Для любой волны, распространяющейся вдоль оси z и поляризованной в направлении x, компоненты электрического и магнитного полей E_x и H_y являются функциями z и не зависят от x и y, поэтому (7.2) принимает вид

$$-\frac{\mathrm{d}}{\mathrm{d}z}\left[\eta(z)\frac{\mathrm{d}}{\mathrm{d}z}\right]H_y = \frac{\omega^2}{c_0^2}H_y. \tag{7.67}$$

Для внеосевой волны, т. е. волны, распространяющейся в произвольном направлении в плоскости x—z, обобщенное уравнение Гельмгольца имеет более сложную форму. Например, для внеосевой волны с поляризацией TM магнитное поле направлено вдоль оси y и (7.2) дает:

$$\left\{-\frac{\partial}{\partial z}\left[\eta(z)\frac{\partial}{\partial z}\right] + \eta(z)\frac{\partial^2}{\partial x^2}\right\}H_y = \frac{\omega^2}{c_0^2}H_y. \tag{7.68}$$

Заметим, что (7.67) и (7.68) имеют вид задач на собственные значения и собственные функции, определяющих моды $H_y(x, z)$.

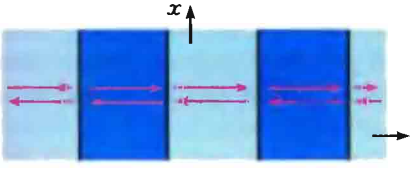
Прежде чем перейти к поиску решений этих задач на собственные значения, исследуем условия, накладываемые на распространяющиеся моды трансляционной симметрией, связанной с периодичностью.

7.2.1. Моды Блоха

Рассмотрим вначале *однородную среду*, которая инвариантна относительно произвольного сдвига системы координат. Для этой среды оптическая мода представляет собой волну, которая не меняется при таком сдвиге; ее изменение сводится к умножению на постоянную, модуль которой равен единице (фазовый множитель). Такой модой является плоская волна $\exp(-jkz)$; при сдвиге на расстояние d она приобретает вид $\exp[-jk(z+d)] = \exp(-jkd)\exp(-jkz)$. Фазовый множитель $\exp(-jkd)$ является собственным значением оператора сдвига, что обсуждается в приложении В.

Аксиальные моды Блоха

Теперь рассмотрим *одномерную периодическую среду*, которая инвариантна относительно сдвига на расстояние Λ вдоль оси периодичности. Ее оптические моды — это волны, сохраняющие свою форму при таком сдвиге с точностью до фазового множителя. Эти моды должны иметь вид



$$\boxed{U(z) = p_K(z)\exp(-jKz),} \qquad (7.69)$$

Мода Блоха

где U представляет собой любую из компонент полей E_x, E_y, H_x или H_y; K — постоянная, а $p_K(z)$ — периодическая функция с периодом Λ. Такой вид удовлетворяет условию, чтобы при сдвиге на Λ волна приобретала лишь фазовый множитель $\exp(-jK\Lambda)$, поскольку периодическая ее часть не меняется вообще. Эта оптическая волна называется **модой Блоха**, а параметр K, определяющий моду и связанную с ней периодическую функцию $p_K(z)$, называется **волновым числом Блоха**.

Таким образом, мода Блоха представляет собой плоскую волну $\exp(-jKz)$ с постоянной распространения K, промодулированную периодической функцией $p_K(z)$, имеющей характер стоячей волны, что видно по ее действительной части, ход которой показан на рис. 7.15, *а*. Поскольку периодическую функцию с периодом Λ можно разложить в ряд Фурье и представить в виде суперпозиции гармоник вида $\exp(-jmgz)$, $m = 0, \pm1, \pm2, \ldots$, где

$$\boxed{g = \frac{2\pi}{\Lambda},} \qquad (7.70)$$

то, следовательно, волна Блоха есть суперпозиция плоских волн с пространственными частотами $K + mg$. Фундаментальная пространственная частота g периодической структуры и ее гармоники mg в сумме с волновым числом Блоха K образуют пространственный спектр волны Блоха, показанный на рис. 7.15, *б*. Сдвиг пространственной частоты, вносимый периодической средой, аналоги-

чен сдвигу временной частоты (доплеровскому сдвигу), вносимому отражением от движущегося объекта.

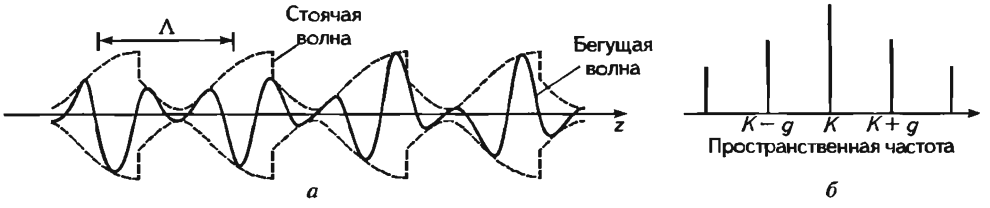


Рис. 7.15. Мода Блоха (*а*) и ее пространственный спектр (*б*)

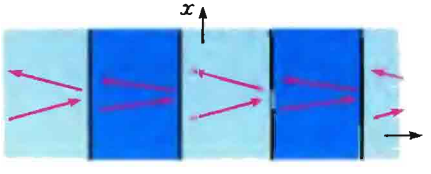
Две моды с волновыми числами Блоха K и $K' = K + g$ эквивалентны, поскольку они соответствуют одному и тому же фазовому множителю

$$\exp(-jK'\Lambda) = \exp(-jK\Lambda)\exp(-j2\pi) = \exp(-jK\Lambda).$$

Это очевидно также и потому, что множитель $\exp(-jgz)$ сам по себе периодичен и может быть объединен в одну периодическую функцию с $p_K(z)$. Следовательно, для полного задания всех мод нужно рассматривать только значения K в интервале пространственных частот шириной $g = 2\pi/\Lambda$. На практике часто используется интервал $[-g/2, g/2] = [-\pi/\Lambda, \pi/\Lambda]$, называемый первой **зоной Бриллюэна**.

Внеосевые моды Блоха

Внеосевые оптические моды, распространяющиеся под некоторыми углами в плоскости x—z, имеют вид функций Блоха



$$U(x, y, z) = p_K(z)\exp(-jKz)\exp(-jk_x x). \quad (7.71)$$

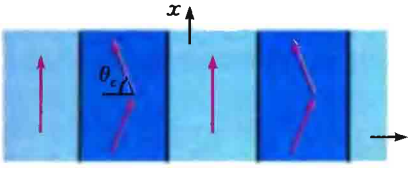
Внеосевая мода Блоха

Однородность среды в направлении x сводит зависимость оптической моды от x к гармоническому виду $\exp(-jk_x x)$, не накладывая иных ограничений на поперечную компоненту волнового вектора k_x. В точке, где показатель преломления равен n, имеем $k_x = nk_0 \sin\theta$, где θ — угол наклона волны по отношению к оси z. По мере распространения волны через различные слои неоднородной среды этот угол меняется, однако в силу закона Снелла величина $n\sin\theta$ и, следовательно, k_x не меняются.

Блоховские моды, нормальные к оси

Когда угол падения в наиболее плотной среде больше критического угла, моды не распространяются вдоль оси периодичности (направления z). Они являются нормальными к этой оси модами, распространяющимися в по-

перечном направлении x, для которых блоховский вид дается формулой (7.71) при $K = 0$



$$U(x,\ y,\ z) = p_0(z)\exp(-jk_x x), \quad (7.72)$$

Мода Блоха, нормальная к оси

где $p_0(z)$ — периодическая функция, представляющая собой стоячую волну вдоль оси периодичности.

Задача на собственные значения, дисперсионное соотношение и фотонные запрещенные зоны

Теперь, когда мы установили общий вид мод, соответствующий требованиям трансляционной симметрии периодической среды, перейдем к решению задачи на собственные значения и собственные функции, описываемой обобщенным уравнением Гельмгольца. Для моды с блоховским волновым вектором K собственные значения ω^2/c_0^2 порождают дискретное множество частот ω. С помощью этих значений строится дисперсионное соотношение между K и ω. Собственные функции рассматриваемой задачи позволяют определить блоховские периодические функции $p_K(z)$ для каждого значения ω, связанного с данным K.

Дисперсионное соотношение между K и ω представляет собой периодическую многозначную функцию переменной K с периодом g, являющимся фундаментальной пространственной частотой периодической структуры; графически она часто изображается в пределах первой зоны Бриллюэна [$-g/2 < K \le g/2$], как схематически показано на рис. 7.16, *а*. Будучи представлена как монотонно возрастающая функция K, она кусочно-непрерывна и имеет конечные разрывы при значениях K, равных целому кратному $g/2$. Эти разрывы соответствуют фотонным запрещенным зонам, т. е. полосам частот, в которых не существует дисперсионной кривой, а значит и распространяющихся мод.

Происхождение разрывов дисперсионной зависимости связано с особой симметрией, которая возникает, например, при $K = g/2$, т. е. когда пространственный период распространяющейся волны точно равен половине периода среды. Рассмотрим две моды с $K = \pm g/2$ и периодическими функциями Блоха $p_K(z) = p_{\pm g/2}(z)$. Поскольку эти моды распространяются с одинаковым волновым числом, но в противоположных направлениях, т. е. «видят» среду обращенной, для них $p_{-g/2}(z) = p_{g/2}(-z)$. Но фактически эти две моды — одна и та же мода, поскольку их волновые векторы Блоха различаются на g. Отсюда следует, что на краю зоны Бриллюэна имеются две периодические функции Блоха, являющиеся инвертированными по отношению друг к другу. Поскольку среда неоднородна или кусочно-однородна в пределах элементарной ячейки, две эти различные

функции по-разному взаимодействуют со средой и, следовательно, имеют два различных собственных значения, т. е. две разные частоты ω. Этот факт объясняет появление разрыва при пересечении непрерывной дисперсионной кривой $\omega(K)$ границы первой зоны Бриллюэна. Аналогичные соображения объясняют появление разрывов при других значениях K, равных целому кратному $g/2$.

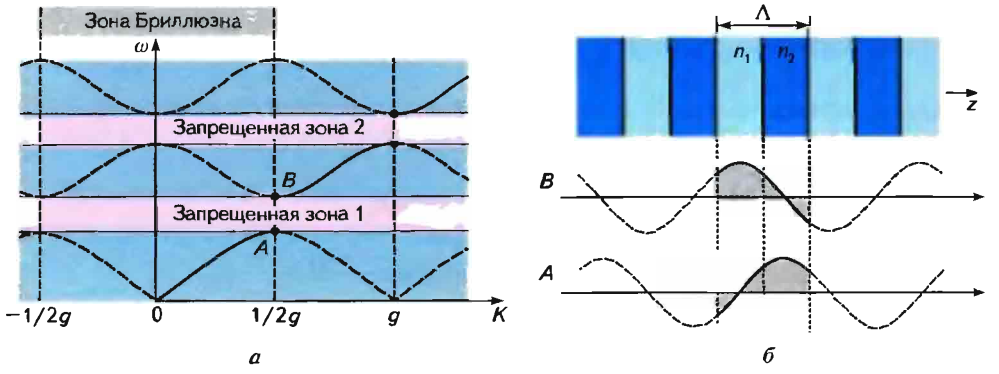


Рис. 7.16. Дисперсионное соотношение определяет многозначную периодическую функцию с периодом $g = 2\pi/\Lambda$; выбор монотонно возрастающей ветви приводит к кусочно-непрерывной функции с разрывами при K, равном целому кратному $g/2$ (*а*). Функции Блоха в точках A и B на краю зоны Бриллюэна для среды из чередующихся диэлектрических слоев с $n_2 > n_1$ (*б*)

Для объяснения некоторых свойств собственных функций полезно воспользоваться вариационным принципом (см. приложение В). Исходя из этого принципа, собственные функции эрмитова оператора представляют собой ортогональные пространственные распределения, минимизирующие вариационную энергию. Вариационной энергией, ассоциированной с линейным оператором $\mathcal{L}$ в уравнении на собственные значения (7.2), называется величина

$$E_v = \frac{1}{2}\frac{(\boldsymbol{H},\ \mathcal{L}\boldsymbol{H})}{(\boldsymbol{H},\ \boldsymbol{H})}.$$

С использованием уравнений Максвелла можно показать, что

$$(\boldsymbol{H},\ \mathcal{L}\boldsymbol{H}) = \left(\boldsymbol{H},\ \nabla\times\left[\eta(\boldsymbol{r})\nabla\times\boldsymbol{H}\right]\right) = \int\frac{|\boldsymbol{D}(\boldsymbol{r})|^2}{\varepsilon(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}},$$

так что наименьшее значение E_v достигается для распределений поля, у которых большие значения электрической индукции $\boldsymbol{D}(\boldsymbol{r})$ расположены в области низких значений $1/\varepsilon(\boldsymbol{r})$, т. е. большего показателя преломления. Например, если периодическая среда состоит из чередующихся диэлектрических слоев, как показано на рис. 7.16, *б*, то в точке разрыва у собственной функции с меньшей частотой электрическая индукция концентрируется в слоях с большим показателем преломления, а у собственной функции с большей частотой — в слоях с меньшим показателем преломления.

Теперь основной задачей является найти решение задачи на собственные значения, связанной с уравнением Гельмгольца. Для этого существует два подхода.

• Первый подход основан на разложении периодической функции $\eta(z)$ для среды и периодической части моды Блоха $p_K(z)$ для поля в ряды Фурье, в результате чего уравнение Гельмгольца превращается в систему алгебраических уравнений, записываемую в виде матричной задачи на собственные значения. Эти уравнения решаются численно. Этот подход называется **Фурье-оптическим.**

• Второй подход применим к многослойным (кусочно-однородным) средам с плоскими границами. Вместо решения уравнения Гельмгольца непосредственно используются законы распространения в слоях, отражения и преломления на границах, являющиеся следствием уравнений Максвелла. Для этого используются матричные методы, разработанные в подразд. 7.1.1 для многослойных сред и примененные к решеткам Брэгга в подразд. 7.1.3. Этот **матрично-оптический** подход приводит к задаче на собственные значения для матрицы 2×2, откуда определяются закон дисперсии и моды Блоха.

Ниже рассматривается матрично-оптический подход, а Фурье-оптическому подходу посвящен подразд. 7.2.3.

7.2.2. Матричная оптика периодических сред

Одномерная периодическая среда состоит из идентичных сегментов, называемых **элементарными ячейками**, которые периодически повторяются вдоль одного направлении (оси z) с периодом Λ (рис. 7.17). Каждая из элементарных ячеек содержит последовательность диэлектрических слоев без потерь или частично-отражающих зеркал, расположенных в определенном порядке и образующих взаимную систему, описываемую порождающей унимодулярной волновой матрицей передачи

$$\mathbf{M}_0 = \begin{bmatrix} \dfrac{1}{t^*} & \dfrac{r}{t} \\ \dfrac{r^*}{t^*} & \dfrac{1}{t} \end{bmatrix}, \tag{7.73}$$

где t и r — комплексные амплитудные коэффициенты пропускания и отражения, удовлетворяющие условиям, сформулированным при получении (7.17). Соответствующие коэффициенты пропускания и отражения по интенсивности равны $\mathcal{T} = |t|^2$ и $\mathcal{R} = |r|^2$. Среда представляет собой решетку Брэгга, подобную той, которая описана в подразд. 7.1.3, с бесконечным числом сегментов. Волна, распространяющаяся через среду, претерпевает многократные пропускания и отражения, дающие вклад в одну прямую и одну обратную волну в каждой плоскости. Применим матричный метод, развитый в подразд. 7.1.1, к нахождению мод Блоха.

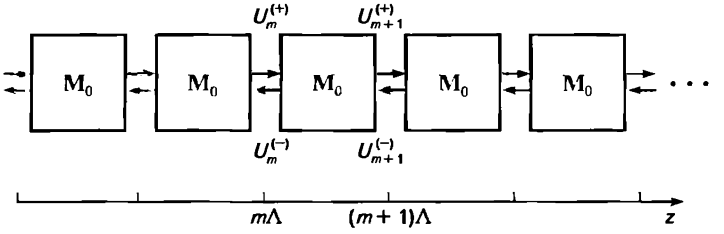


Рис. 7.17. Матричное представление периодической среды

Пусть $\{U_m^{(\pm)}\}$ — комплексные амплитуды прямой и обратной волн в начальной точке $z = m\Lambda$ m-й элементарной ячейки. Зная эти амплитуды, можно определить амплитуды в любой другой точке с помощью соответствующих волновых матриц передачи, как описано в разд. 7.1. Рассмотрим динамику изменения амплитуд $\{U_m^{(\pm)}\}$ от ячейки к ячейке. Эта динамика описывается рекуррентными соотношениями

$$\begin{bmatrix} U_{m+1}^{(+)} \\ U_{m+1}^{(-)} \end{bmatrix} = \mathbf{M}_0 \begin{bmatrix} U_m^{(+)} \\ U_m^{(-)} \end{bmatrix}, \tag{7.74}$$

с помощью которых можно определить амплитуды в данной ячейке, если они известны в предыдущей ячейке.

Задача на собственные значения и моды Блоха

По определению моды периодической среды — это самовоспроизводящиеся волны, для которых

$$\begin{bmatrix} U_{m+1}^{(+)} \\ U_{m+1}^{(-)} \end{bmatrix} = e^{-j\Phi} \begin{bmatrix} U_m^{(+)} \\ U_m^{(-)} \end{bmatrix}, \quad m = 1,\ 2, \tag{7.75}$$

После прохождения расстояния Λ (в данном случае — элементарной ячейки) модули прямой и обратной волн остаются неизменными, а фазы меняются на одну и ту же величину Φ, называемую **фазой Блоха**. Соответствующее волновое число Блоха равно $K = \Phi/\Lambda$, так что

$$\boxed{\Phi = K\Lambda.} \tag{7.76}$$

Фаза Блоха

Определение комплексных амплитуд $U_m^{(\pm)}$ и фазы $\Phi = K\Lambda$, удовлетворяющих условию самовоспроизведения (7.75), можно свести к задаче на собственные значения. Положим в (7.74) $m = 0$, тогда из (7.75) получим

$$\mathbf{M}_0 \begin{bmatrix} U_0^{(+)} \\ U_0^{(-)} \end{bmatrix} = e^{-j\Phi} \begin{bmatrix} U_0^{(+)} \\ U_0^{(-)} \end{bmatrix}. \tag{7.77}$$

Это задача на собственные значения для двумерной матрицы $\mathbf{M}_0$ элементарной ячейки. Множитель $e^{-j\Phi}$ представляет собой собственное значение, а вектор с компонентами $U_0^{(+)}$ и $U_1^{(-)}$ — собственный вектор.

Собственные значения определяются приравниванием определителя матрицы $\mathbf{M}_0 - e^{-j\Phi}\mathbf{I}$ к нулю. Замечая, что $|t|^2 + |r|^2 = 1$, находим решение получающегося квадратного уравнения в виде

$$e^{-j\Phi} = \frac{1}{2}\left(\frac{1}{t} + \frac{1}{t^*}\right) \pm j\left[1 - \frac{1}{2}\left(\frac{1}{t} + \frac{1}{t^*}\right)^2\right]^{1/2}$$

откуда

$$\boxed{\cos\Phi = \mathrm{Re}\left\{\frac{1}{t}\right\}.} \tag{7.78}$$

Формула (7.78) идентична формуле (7.51) для решетки Брэгга. Этот результат вызывает законное удовлетворение, поскольку рассматриваемая периодическая среда и есть решетка Брэгга, но с бесконечным числом сегментов.

Поскольку $\mathbf{M}_0$ — матрица размерностью 2×2, она имеет два собственных значения. Следовательно, только два из множества решений (7.78) независимы. Поскольку функция arccos четная, два решения в интервале $[-\pi, \pi]$ имеют одинаковые модули и противоположные знаки. Они соответствуют модам Блоха, распространяющимся в прямом и обратном направлениях. Другие решения, получаемые добавлением целого числа 2π, не являются независимыми, так как они не соответствуют фазовому множителю $e^{-j\Phi}$

Соответствующие собственные векторы матрицы $\mathbf{M}_0$ имеют вид

$$\begin{bmatrix} U_0^{(+)} \\ U_0^{(-)} \end{bmatrix} \propto \begin{bmatrix} \dfrac{r}{t} \\ e^{-j\Phi} - \dfrac{1}{t^*} \end{bmatrix}, \tag{7.79}$$

в чем можно убедиться, действуя на правую часть (7.79) матрицей $\mathbf{M}_0$. В результате снова получается правая часть (7.79) с точностью до постоянной.

Периодическую функцию $p_K(z)$, связанную с модой Блоха, можно определить, рассматривая изменение амплитуд $U_0^{(+)}$ и $U_1^{(-)}$ при прохождении элементарной ячейки. Например, если начальный слой ячейки — однородная среда толщиной d_1 с показателем преломления n_1, то на глубине z внутри этого слоя волна имеет вид

$$p_K(z)\exp(-jKz) = U_0^{(+)}e^{-jn_1k_0z} + U_0^{(-)}e^{jn_1k_0z}, \quad 0 < z < d_1. \tag{7.80}$$

Используя (7.79) и (7.76) получаем из (7.80):

$$p_K(z) \propto \left[-re^{-jn_1k_0z} + (e^{-jK\Lambda} - 1)e^{jn_1k_0z}\right]e^{jKz}, \quad 0 < z < d_1. \tag{7.81}$$

Распространение волн в (7.81) в последующие слои в пределах ячейки можно описать с помощью соответствующих матриц М.

Дисперсионное соотношение и фотонная зонная структура

Дисперсионное соотношение — это уравнение, связывающее угловую частоту ω с волновым числом Блоха K. Уравнение (7.78), определяющее собственное значение $e^{-j\Phi}$ матрицы элементарной ячейки, является исходным для дисперсионного соотношения в случае одномерной периодической среды. Фаза $\Phi = K\Lambda$ пропорциональна K, а $t = t(\omega)$ связано с ω через фазовую задержку при распространении на одну элементарную ячейку, так что (7.78), записанное в виде

$$\cos\left(2\pi\frac{K}{g}\right) = \mathrm{Re}\left\{\frac{1}{t(\omega)}\right\}, \qquad (7.82)$$

Дисперсионное соотношение

есть дисперсионное соотношение между ω и K. Здесь $g = 2\pi/\Lambda$ — фундаментальная пространственная частота периодической среды.

Функция $\cos(2\pi K/g)$ — периодическая функция K с периодом $g = 2\pi/\Lambda$, поэтому для данного ω уравнение (7.82) имеет множество решений. Однако решения, отличающиеся на период g, не являются независимыми, поскольку они приводят к одинаковым волнам Блоха. По этой причине принято ограничивать область определения дисперсионного соотношения одним периодом, так что K принимает значения в интервале $[-g/2, g/2] = [-\pi/\Lambda, \pi/\Lambda]$, который представляет собой первую зону Бриллюэна. Это соответствует изменениям фазы в пределах интервала $[-\pi, \pi]$. Далее, поскольку $\cos(2\pi K/g)$ — четная функция K, при каждом значении ω внутри первой зоны Бриллюэна существует два независимых значения K с равными модулями и противоположными знаками. Они соответствуют независимым волнам Блоха, распространяющимся в прямом и обратном направлениях.

Данному дисперсионному соотношению соответствует множество частотных полос, которые делятся на два типа.

- **Режим распространения.** Спектральные полосы, внутри которых K имеет действительные значения, соответствуют распространяющимся модам. Определяемые условием

$$\left|\mathrm{Re}\left\{\frac{1}{t(\omega)}\right\}\right| \le 1,$$

эти полосы нумеруются индексами 1, 2, , начиная с полосы с наименьшей частотой.

- **Режим фотонной запрещенной зоны.** Спектральные полосы, в которых K принимает комплексные значения, соответствуют нераспространяющимся волнам, которые быстро затухают. Определяемые условием

$$\left|\mathrm{Re}\left\{\frac{1}{t(\omega)}\right\}\right| > 1,$$

эти полосы соответствуют полосам заграждения дифракционной решетки, обсужденным в подразд. 7.1.3. Их называют также **фотонными запрещенными зонами**, поскольку в них не существует распространяющихся мод.

При построении дисперсионных кривых K часто выражают в единицах фундаментальной пространственной частоты периодической структуры $g = 2\pi/\Lambda$, а ω — в единица частоты Брэгга: $\omega_B = \pi c/\Lambda$, где $c = c_0/\bar{n}$, а $\bar{n}$ — средний показатель преломления периодической среды. Отношение

$$\frac{\omega_B}{g/2} = c$$

определяет наклон дисперсионной кривой $\omega = cK$ для распространения в однородной среде.

Пример 7.12

Периодический пакет частично-отражающих зеркал

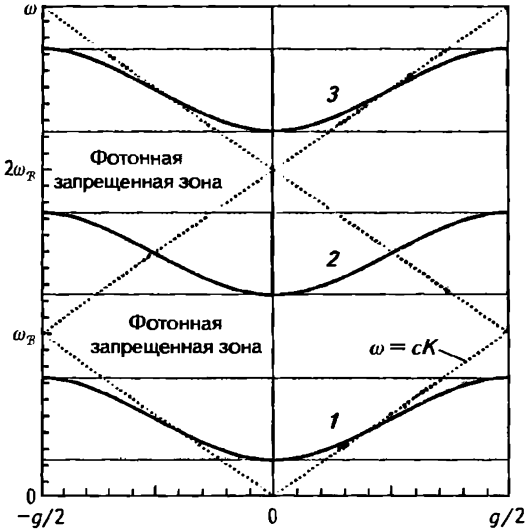
Дисперсионное соотношение для волны, распространяющейся вдоль оси периодического пакета идентичных частично-отражающих зеркал без потерь с коэффициентами отражения $|r|^2$ и пропускания $|t|^2 = 1 - |r|^2$ по интенсивности, расположенных на расстоянии Λ друг от друга, было найдено в примере 7.8. Используя полученные там результаты, а именно, $t = |t|e^{j\varphi}$, где

$$\varphi = nk_0\Lambda = \frac{\omega}{c}\Lambda,$$

в сочетании с (7.78), приходим к дисперсионному соотношению

$$\boxed{\cos\left(2\pi\frac{K}{g}\right) = \frac{1}{|t|}\cos\left(\pi\frac{\omega}{\omega_B}\right)}, \tag{7.83}$$

где $g = 2\pi/\Lambda$; $\omega_B = \pi c/\Lambda$ — частота Брэгга. Результат изображен графически на рис. 7.18.



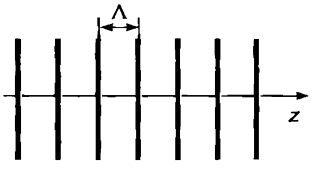


Рис. 7.18. Дисперсионные кривые для периодического набора зеркал, каждое из которых имеет коэффициент пропускания по интенсивности $|t|^2 = 0{,}5$, разделенных расстоянием Λ. Здесь $g = 2\pi/\Lambda$, $\omega_B = \pi c/\Lambda$. Прямые пунктирные линии соответствуют распространению в однородной среде, у которой $\omega/K = \omega_B/(g/2) = c$

Фотонные запрещенные зоны, соответствующие областям частот, в которых (7.83) не имеет действительных решений, имеют центральные частоты ω_B, $2\omega_B$, В этих частотных областях запрещены распространяющиеся моды, они соответствуют полосам заграждения на рис. 7.12, в которых коэффициент отражения равен единице. В данной системе низшая фотонная запрещенная зона начинается с $\omega = 0$.

Пример 7.13

Чередующиеся диэлектрические слои

Периодическая среда состоит из чередующихся слоев диэлектрика толщиной d_1 и d_2 с показателями преломления n_1 и n_2, так что период структуры составляет $\Lambda = d_1 + d_2$. Эта система — диэлектрическая решетка Брэгга, описанная в примере 7.9, при $N \to \infty$. Для волны, бегущей вдоль оси периодичности, $\mathrm{Re}\{1/t\} = \mathrm{Re}\{A\}$ дается формулой (7.63). Пользуясь соотношениями

$$\varphi_1 + \varphi_2 = k_0(n_1 d_1 + n_2 d_2) = \frac{\pi\omega}{\omega_B}; \quad \varphi_1 - \varphi_2 = \frac{\zeta\pi\nu}{\nu\omega/\omega_B},$$

где ω_B — частота Брэгга,

$$\omega_B = \frac{c_0}{\bar{n}}\frac{\pi}{\Lambda};$$

$\bar{n}$ — средний показатель преломления,

$$\bar{n} = \frac{n_1 d_1 + n_2 d_2}{\Lambda}; \quad \zeta = \frac{n_1 d_1 - n_2 d_2}{n_1 d_1 + n_2 d_2}.$$

Из (7.78) получаем дисперсионное соотношение

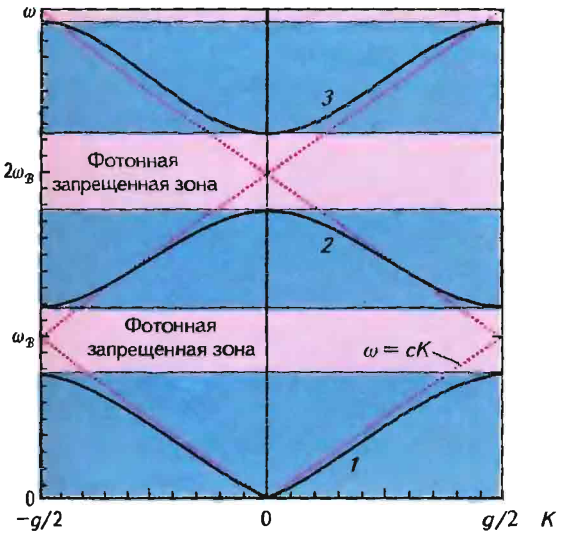
$$\boxed{\cos\left(2\pi\frac{K}{g}\right) = \frac{1}{t_{12}t_{21}}\left[\cos\left(\pi\frac{\omega}{\omega_B}\right) - |r_{12}|^2\cos\left(\pi\zeta\frac{\omega}{\omega_B}\right)\right],} \tag{7.84}$$

где

$$t_{12}t_{21} = \frac{4n_1n_2}{(n_1+n_2)^2}; \quad |r_{12}|^2 = \frac{(n_2-n_1)^2}{(n_1+n_2)^2}.$$

Пример соответствующих дисперсионных кривых приведен на рис. 7.19 для диэлектрических материалов с $n_1 = 1{,}5$, $n_2 = 3{,}5$ и $d_1 = d_2$. Как и для периодического пакета частично-отражающих зеркал, рассмотренного в примере 7.12, фотонные запрещенные зоны имеют центры на частотах, кратных ω_B. Эти зоны возникают либо на краях ($K = g/2$), либо в центре ($K = 0$) первой зоны Бриллюэна. Однако в рассматриваемом здесь случае область частот вблизи $\omega = 0$ отвечает распространяющимся волнам и не является фотонной запрещенной зо-

ной. Диэлектрические материалы с меньшей разницей в показателях преломления обладают более узкими фотонными запрещенными зонами, которые, однако, всегда существуют, как бы низок не был контраст.



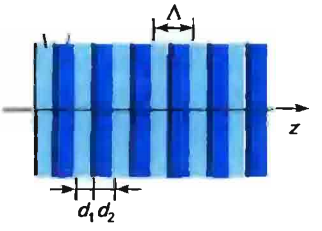


Рис. 7.19. Дисперсионные кривые для среды из периодически чередующихся слоев диэлектрика с $n_1 = 1{,}5$, $n_2 = 3{,}5$ и $d_1 = d_2$. Здесь $\omega_B = \pi c_0/\Lambda\bar{n}$ и $g = 2\pi/\Lambda$. Пунктирными прямыми показан случай распространения в однородной среде со средним показателем преломления $\bar{n}$, когда $\omega/K = \omega_B/(g/2) = c_0/\bar{n} = c$

Фазовая и групповая скорости

Постоянная распространения K соответствует фазовой скорости ω/K и эффективному показателю преломления $n_{эфф} = c_0K/\omega$. Групповая скорость $v = d\omega/dK$, определяющая распространение импульса в среде, связана с эффективным групповым показателем преломления $N_{эфф} = c_0 dK/d\omega$ (см. разд. 5.6). Эти показатели преломления можно найти в любой точке дисперсионной кривой, определяя наклон касательной $d\omega/dK$ и отношение ω/K, т. е. наклон прямой, соединяющей данную точку с началом координат. На рис 7.20 схематически показаны дисперсионные кривые для периодической структуры из чередующихся диэлектрических слоев, а также эффективный и групповой показатели преломления в диапазоне частот, покрывающем две фотонные зоны с фотонной запрещенной зоной между ними.

На низких частотах в пределах первой фотонной зоны $n_{эфф}$ приблизительно равен среднему показателю преломления $\bar{n}$. Это ожидаемый результат, поскольку при больших длинах волн материал ведет себя как однородная среда со средним показателем преломления. С ростом частоты $n_{эфф}$ растет по сравнению с $\bar{n}$, достигая наибольшего значения на краю зоны. На дне следующей зоны $n_{эфф}$ меньше, чем $\bar{n}$, затем он растет и приближается к $\bar{n}$ в середине зоны.

Такой спад $n_{эфф}$ от значения больше среднего в верхней части первой зоны к значению меньше среднего в нижней части второй зоны можно отнести на счет существенного различия пространственных распределений поля у соот-

ветствующих мод Блоха, которые ортогональны друг другу. Мода в верхней части нижней зоны имеет большую энергию в слоях диэлектрика с большим показателем преломления, поэтому ее эффективный показатель преломления выше среднего. У моды в нижней части верхней зоны, напротив, энергия сосредоточена в слоях диэлектрика с меньшим показателем преломления, и эффективный показатель преломления меньше среднего.

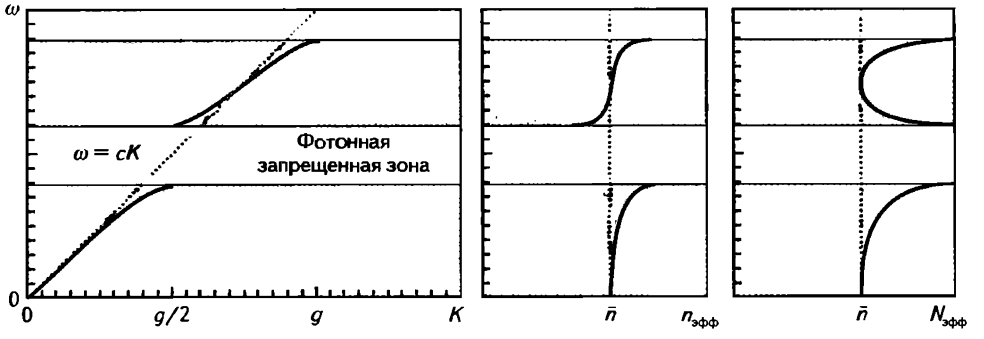


Рис. 7.20. Частотная зависимость эффективного показателя преломления $n_{эфф}$, который определяет фазовую скорость, и эффективного группового показателя $N_{эфф}$, который определяет групповую скорость

Частотная зависимость эффективного группового показателя преломления $N_{эфф}$ выглядит иначе, чем показано на рис. 7.20. По мере приближения к краям зоны, снизу или сверху, этот показатель существенно увеличивается, так что групповая скорость значительно уменьшается, т. е. вблизи краев зоны импульсы распространяются очень медленно.

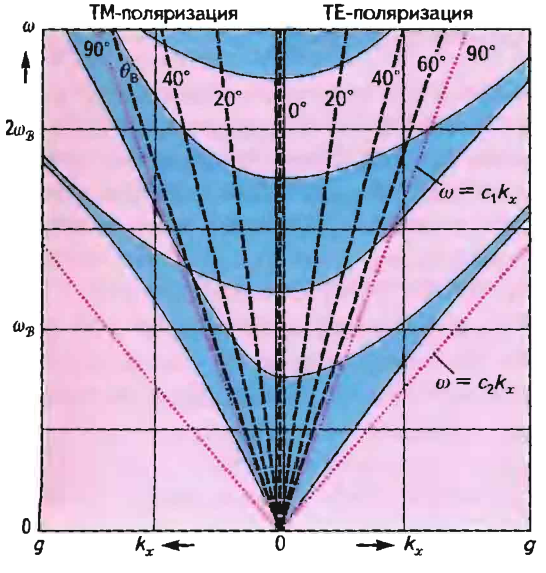
Дисперсионное соотношение и зонная структура в случае внеосевых волн

Для внеосевых волн дисперсионное соотношение определяется тем же самым уравнением $\cos(K\Lambda) = \mathrm{Re}\{1/t(\omega)\}$, где, однако, теперь $\mathrm{Re}\{1/t(\omega)\}$ зависит от углов падения внутри слоев каждого сегмента и от состояния поляризации (TE или TM). Например, для периодической среды из чередующихся диэлектрических слоев $\mathrm{Re}\{1/t(\omega)\}$ имеет более общий вид (7.64).

Поскольку одна и та же поперечная компонента k_x волнового вектора определяет углы падения в двух слоях ($k_x = n_1 k_0 \sin\theta_1 = n_2 k_0 \sin\theta_2$), удобнее выражать дисперсионное соотношение не через угол θ, а через k_x, и изображать его в виде трехмерной поверхности $\omega = \omega(K, k_x)$. Каждое значение k_x порождает дисперсионную диаграмму с разрешенными и запрещенными фотонными зонами, аналогичные приведенным на рис. 7.19.

Более простым представлением трехмерной поверхности $\omega(K, k_x)$ является **проекционная дисперсионная диаграмма**, представляющая собой двумерное изображение краев зон в зависимости от k_x для TE- и TM-поляризаций и показан-

ная на рис. 7.21. Такая диаграмма строится путем проецирования фотонных разрешенных и запрещенных зон на каждую вертикаль, отвечающую определенному значению k_x. На рис. 7.21 голубым показаны разрешенные, а розовым — запрещенные фотонные зоны.



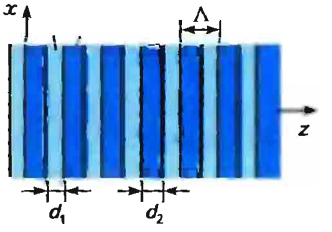


Рис. 7.21. Проекционная дисперсионная диаграмма для периодической среды из чередующихся слоев диэлектрика с $n_1 = 1{,}5$, $n_2 = 3.5$, $d_1 = d_2 = \Lambda/2$. Здесь $\omega_B = \pi c_0/\Lambda\bar{n}$ и $g = 2\pi/\Lambda$. Фотонные разрешенные зоны показаны голубым цветом. Штриховые линии соответствуют фиксированным углам падения θ_1 в слое 1, включая угол Брюстера $\theta_B = 66{,}8°$. Точки области, ограниченной световыми линиями $\omega = c_1k_x$ и $\omega = c_2k_x$, представляют моды, нормальные к оси

На этой диаграмме каждому углу падения соответствует прямая линия, проходящая через начало координат. Например, угол падения θ_1 в слое 1 соответствует линии $k_x = (\omega/c_1)\sin\theta_1$, т. е. $\omega = (c_1/\sin\theta_1)k_x$, где $c_1 = c_0/n_1$. Линия $\omega = c_1k_x$, называемая **световой линией**, соответствует $\theta_1 = 90°$. Аналогичные линии можно построить для углов падения в среде 2, из них на рис. 7.21 показана только световая линия $\omega = c_2k_x$ в предположении, что $n_1 < n_2$, т. е. $c_2 < c_1$. Точки области, ограниченной двумя световыми линиями, представляют моды, нормальные по отношению к оси, которые распространяются в поперечном направлении, претерпевая полное отражение в более плотной среде (среде 2).

Возникает вопрос: существует ли такая область частот, в которой распространение волн запрещено при любых углах падения θ_1 и θ_2 и для обеих поляризаций? Это было бы возможно, если бы запрещенные зоны для всех значений k_x между линиями $k_x = 0$ и $k_x = \omega/c_2$ для обеих поляризаций расположились таким образом, чтобы образовалась общая фотонная запрещенная зона. В примере, показанном на рис. 7.21, ясно видно, что это не так. Оказывается, что такая ситуация вообще невозможна в одномерных периодических структурах. Она возможна только в двумерных и трехмерных периодических структурах, как будет показано в разд. 7.3.

Существует особый случай, когда фотонная запрещенная зона вообще не может возникнуть. Это случай наклонного падения ТМ-волны под углом Брю-

стера $\theta_B = \operatorname{arctg}(n_2/n_1)$ в слое 1. Как показано на рис. 7.21, световая линия, соответствующая этому углу, не пересекает запрещенных зон. Это не удивительно, так как при таком угле падения коэффициент отражения элементарной ячейки равен нулю, прямая и обратная волны не связаны, и коллективный эффект, приводящий к полному отражению, отсутствует.

7.2.3. Фурье-оптика периодических сред

Матричный анализ периодических сред, представленный в предыдущем разделе, применим к слоистым (т. е. кусочно-однородным) средам. Более общий подход, пригодный для описания произвольных периодических сред, основан на разложении периодических функций в ряды Фурье с преобразованием уравнения Гельмгольца в систему алгебраических уравнений, решение которой дает дисперсионное соотношение и моды Блоха. Этот подход можно обобщить на двух- и трехмерные периодические среды, как будет показано в разд. 7.3.

Волна, распространяющаяся вдоль оси одномерной периодической среды (оси z) и поляризованная в направлении x, описывается обобщенным уравнением Гельмгольца (7.67). Поскольку функция $\eta(z)$ периодична с периодом Λ, ее можно разложить в ряд Фурье

$$\eta(z) = \sum_{\ell=-\infty}^{\infty} \eta_\ell \exp(-j\ell g z), \tag{7.85}$$

где $g = 2\pi/\Lambda$ — пространственная частота (рад/мм) периодической структуры; η_ℓ — коэффициент Фурье для ℓ-й гармоники. Диэлектрическая непроницаемость $\eta(z)$ действительна, поэтому $\eta_{-\ell} = \eta_\ell^*$.

Периодическая часть волны Блоха $p_K(z)$ в (7.69) также может быть разложена в ряд Фурье

$$p_K(z) = \sum_{m=-\infty}^{\infty} C_m \exp(-jmgz), \tag{7.86}$$

в силу чего магнитное поле волны Блоха приобретает вид

$$H_y(z) = \sum_{m=-\infty}^{\infty} C_m \exp\left[-j(K+mg)z\right]. \tag{7.87}$$

Для краткости зависимость коэффициентов Фурье $\{C_m\}$ от волнового числа Блоха K не указана. Подставляя эти разложения в уравнение Гельмгольца (7.67) и приравнивая коэффициенты при одинаковых гармониках, получаем

$$\sum_{\ell=-\infty}^{\infty} F_{m\ell} C_\ell = \frac{\omega^2}{c_0^2} C_m; \quad F_{m\ell} = (K+mg)(K+\ell g)\eta_{m-\ell}, \tag{7.88}$$

где $m = 0, \pm 1, \pm 2,$

Дифференциальное уравнение (7.67) преобразовалось в систему линейных уравнений (7.88) для неизвестных коэффициентов Фурье $\{C_m\}$. Эти уравнения

можно записать в виде матричной задачи на собственные значения. Для каждого K собственные значения ω^2/c_0^2 соответствуют множеству значений ω, что позволяет построить дисперсионное соотношение между ω и K. Собственные векторы представляют собой наборы коэффициентов $\{C_m\}$, определяющих периодическую функцию моды Блоха $p_K(z)$ для каждого K.

Поставленная как задача на собственные значения для матрицы $\mathbf{F}$ с элементами $F_{m\ell}$, эта система уравнений может быть решена с применением стандартных численных методов. Поскольку $\eta_{m-\ell} = \eta^*_{\ell-m}$, матрица $\mathbf{F}$ эрмитова, т. е. $F_{m\ell} = F^*_{\ell m}$. Заметим, что если бы мы использовали уравнение Гельмгольца не для магнитного, а для электрического поля, мы получили бы другое матричное представление задачи на собственные значения, у которого матрица неэрмитова, что затрудняет решение. Вот почему мы выбрали для работы уравнение Гельмгольца для магнитного поля[3].

Приближенное решение задачи на собственные значения

В уравнениях (7.88) гармоники оптической волны связаны друг с другом через гармоники периодической среды. Гармоника оптической волны с пространственной частотой $K + \ell g$ смешивается с гармоникой среды на частоте $(m - \ell)g$ и вносит вклад в гармонику оптической волны на частоте $(K + \ell g) + (m - \ell)g = K + \ell g$.

Условия, при которых возникает сильная связь, можно определить посредством выделения m-го члена из (7.88), что приводит к выражению

$$C_m = \sum_{\ell \neq m} \frac{\eta_{m-\ell}}{\eta_0} \frac{(K + mg)(K + \ell g)}{(\bar{n}\omega/c_0)^2 - (K + mg)^2} C_\ell, \quad m = 0, \pm 1, \pm 2, \tag{7.89}$$

где $\bar{n}$ — средний показатель преломления среды,

$$\bar{n} = \frac{1}{\sqrt{\eta_0}}.$$

Сильная связь между m-й гармоникой волны и другими гармониками существует, если знаменатель в (7.89) мал, т. е.

$$\frac{\omega \bar{n}}{c_0} \approx |K + mg|. \tag{7.90}$$

Это уравнение выражает резонансное условие взаимодействия гармоник. Его можно рассматривать также как **условие фазового синхронизма**.

На рис. 7.22 показан график (7.90) при точном равенстве левой и правой частей. Для каждого значения m график зависимости ω от K имеет вид V-образной кривой. Точки пересечения таких кривых определяют общие значения ω и K, при которых (7.90) одновременно выполняется для двух гармоник. Пересечения между кривой $m = 0$ (штриховая линия) и кривыми для $m = -1, -2,$

[3] Можно показать, что сам дифференциальный оператор в обобщенном уравнении Гельмгольца (7.2) эрмитов, а в уравнении для электрического поля — неэрмитов.

отмеченные черными кружками, соответствуют запрещенным зонам низших порядков *1*, *2*, ... соответственно. В каждой точке пересечения K равно целому кратному от $g/2$, а ω — целому кратному частоты Брэгга:

$$\omega_B = \frac{c_0}{\bar{n}}\frac{g}{2} \quad \text{или} \quad \frac{\omega_B}{2\pi} = \nu_B = \frac{c_0/\bar{n}}{2\Lambda}.$$

Она соответствует брэгговской длине волны $\lambda_B = 2\Lambda$ в среде и, следовательно, полному отражению. Неотмеченные пересечения на рис. 7.22 не соответствуют независимым решениям, поскольку для каждого из них ω имеет то же значение, что и у одного из отмеченных пересечений, а значение K отличается на постоянную обратной решетки g.

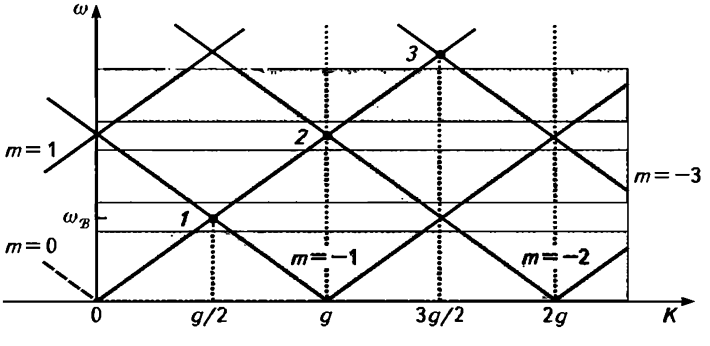


Рис. 7.22. График (7.90) при точном равенстве левой и правой частей для различных значений m. Кривая $m = 0$ показана штриховой линией. Сильная связь между гармониками оптической волны и гармониками среды имеет место в точках пересечения *1*, *2*, ..., которые соответствуют запрещенным зонам низших порядков

Самая нижняя запрещенная зона соответствует пересечению кривых $m = 0$ и $m = -1$ (точка *1* на рис. 7.22). В этом случае сильно связаны только коэффициенты C_0 и C_{-1}, так что из (7.89) получаются два уравнения:

$$C_0 = \frac{\eta_1}{\eta_0}\frac{(K-g)K}{(\bar{n}\omega/c_0)^2 - K^2}C_1; \tag{7.91}$$

$$C_{-1} = \frac{\eta_1^*}{\eta_0}\frac{K(K-g)}{(\bar{n}\omega/c_0)^2 - (K-g)^2}C_0, \tag{7.92}$$

где $\eta_{-1} = \eta_1^*$. Эти уравнения совместны, если

$$\boxed{\frac{|\eta_1|^2}{\eta_0^2}K^2(K-g)^2 = \left[\omega^2\frac{\bar{n}^2}{c_0^2} - K^2\right]\left[\omega^2\frac{\bar{n}^2}{c_0^2} - (K-g)^2\right]}. \tag{7.93}$$

Дисперсионное соотношение

График этого соотношения (рис. 7.23) изображает дисперсионное соотношение вблизи края запрещенной зоны, где это уравнение верно. При $K = g/2$ (7.93) дает две частоты

$$\omega_\pm = \omega_B\sqrt{1 \pm \frac{|\eta_1|}{\eta_0}}, \tag{7.94}$$

определяющих края первой фотонной запрещенной зоны. Центр зоны находится на частоте Брэгга

$$\omega_B = \frac{c_0}{\bar{n}}\frac{g}{2} = \frac{\pi}{\Lambda}\frac{c_0}{\bar{n}}.$$

Отношение ширины запрещенной зоны к ее центральной частоте, называемое **относительной шириной зоны**, растет с увеличением контраста структуры $|\eta_1|/\eta_0$.

Аналогичная процедура может быть использована для фотонных запрещенных зон более высокого порядка. Ширина m-й зоны определяется формулой (7.94), в которой отношение $|\eta_1|/\eta_0$ нужно заменить на $|\eta_m|/\eta_0$, поскольку высшие запрещенные зоны определяются высшими пространственными гармониками периодической функции $\eta(z)$.

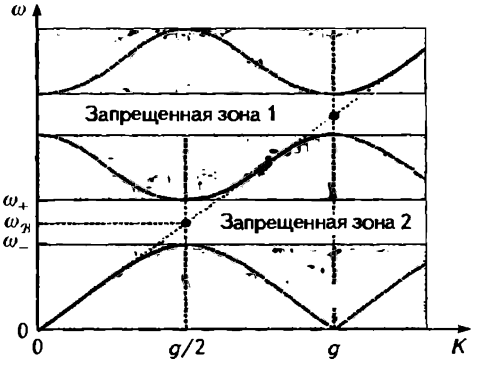


Рис. 7.23. Дисперсионное соотношение в окрестности фотонных запрещенных зон

Внеосевые волны

Дисперсионное соотношение для внеосевых волн можно определить с помощью того же метода разложения Фурье. Для внеосевой волны TM, распространяющейся в произвольном направлении в плоскости x—z, уравнение Гельмгольца имеет вид (7.68). Волна Блоха представляет собой обобщение (7.87), получаемое с использованием (7.71)

$$H_y(z) = \sum_{m=-\infty}^{\infty} C_m \exp\left[-j(K+mg)z\right]\exp(-jk_x x). \qquad (7.95)$$

Производя вычисления, аналогичные случаю аксиальной волны, приходим к системе алгебраических уравнений для коэффициентов C_m:

$$\sum_{\ell=-\infty}^{\infty} F_{m\ell}C_\ell = \frac{\omega^2}{c_0^2}C_m; \quad F_{m\ell} = \left[(K+mg)(K+\ell g)+k_x^2\right]\eta_{m-\ell}. \qquad (7.96)$$

Уравнение (7.96) является обобщением (7.88) на случай внеосевой волны. Дисперсионное соотношение для набора частот ω, связанных с каждой парой значений K и k_x, можно получить путем решения этой матричной задачи на собственные значения.

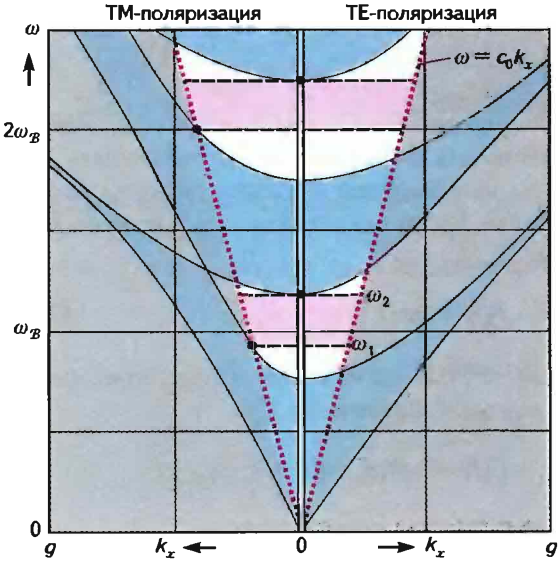
7.2.4. Границы между периодическими и однородными средами

До сих пор исследование распространения света в периодических средах было ограничено дисперсионными соотношениями и связанной с ними зонной структурой, а также оценками фазовой и групповой скорости волн. По

определению периодическая среда простирается неограниченно во всех направлениях. Следующий шаг состоит в изучении отражения и пропускания на границах между периодической и однородной средой. Вначале рассмотрим отражение от одной границы, а затем — слой периодической среды, находящийся в однородной среде. Другие конфигурации однородных структур, такие как однородные слои или отверстия в протяженной периодической среде, описаны в разд. 9.4 и подразд. 10.4.4.

Всенаправленное отражение от одной границы

Исследуем прохождение и отражение оптической волны на границе между полубесконечной однородной средой и полубесконечной периодической средой (рис. 7.24). Покажем, что при определенных условиях и в определенном диапазоне угловых частот периодическая среда действует как идеальное зеркало, полностью отражающее свет, падающий под любым углом и с любой поляризацией.



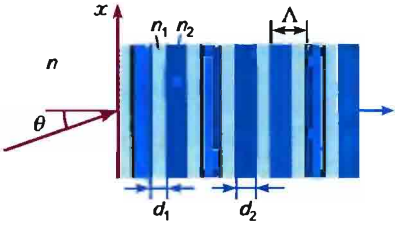


Рис. 7.24. Проекционная дисперсионная диаграмма для среды с чередованием диэлектрических слоев с $n_1 = 1{,}5$, $n_2 = 3{,}5$, $d_1 = d_2 = \Lambda/2$. Пунктирные красные линии — световые линии для однородной среды с показателем преломления $n = 1$. В спектральной полосе между ω_1 и ω_2 среда действует как идеальный всенаправленный отражатель для всех поляризаций. Аналогичная полоса показана на более высоких частотах

Прохождение и отражение волны на границе раздела двух сред определяется условием фазового синхронизма. Например, на границе между двумя однородными средами поперечная компонента волнового вектора k_x должна быть одинаковой по обе стороны границы. Поскольку $k_x = k \sin\theta = (\omega/c_0)n \sin\theta$, это условие означает, что произведение $n \sin\theta$ инвариантно. Как объясняется в подразд. 2.4.1, отсюда вытекает закон преломления Снелла.

Аналогично, для волны, падающей из однородной среды на границу одномерной периодической среды, k_x должно оставаться неизменным. Если падающая волна имеет угловую частоту ω и угол падения θ, то $k_x = (\omega/c_0)n \sin\theta$, где

n — показатель преломления однородной среды. Зная k_x и ω, можно использовать дисперсионное соотношение $\omega = \omega(K, k_x)$ для периодической среды при соответствующей поляризации света, чтобы определить волновое число Блоха K. Если угловая частота ω лежит в запрещенной зоне при данном значении k_x, то падающая волна не будет распространяться в периодической среде, а полностью отразится от нее. Эта процедура должна быть повторена для всех частот, углов падения и поляризаций падающей волны.

Рассмотрим теперь возможность того, что граница действует как всенаправленный отражатель (идеальное зеркало). Для этой цели используем проекционную дисперсионную диаграмму, которая показывает запрещенные зоны для каждого значения k_x, как в примере на рис. 7.24. На такой диаграмме очертим красной штриховой линией область ω—k_x, которая доступна волнам, входящим из однородной среды. Эта область определяется уравнением $k_x = (\omega/c_0)n \sin\theta$, которое требует, чтобы выполнялось условие $k_x < (\omega/c_0)n$ или $\omega > (c_0/n)k_x$. Следовательно, данная область ограничена линией $\omega = (c_0/n)k_x$, известной как световая линия. Эта линия соответствует углу $\theta = 90^\circ$ в окружающей среде.

На рис. 7.24 воспроизведен рис. 7.21 с добавлением световых линий и цветовым выделением разрешенных областей ω—k_x внутри светового конуса. Волны, падающие из однородной среды под любыми углами и с обеими поляризациями, представляются точками этих областей; точки вне данных областей недоступны волнам, падающим из однородной среды, независимо от их угла падения и поляризации. Спектральная полоса между частотами ω_1 и ω_2, показанная на рис. 7.24, представляет особый интерес, поскольку все точки в этой полосе попадают в фотонную запрещенную зону. Следовательно, в этой полосе никакая падающая волна, каковы бы ни были ее угол падения и поляризация, не может быть согласована с распространяющейся волной в периодической среде, граница которой ведет себя как идеальный всенаправленный отражатель. На рис. 7.24 показана и еще одна спектральная полоса на более высоких частотах, которая ведет себя аналогичным образом.

Слой периодического материала в однородной среде

Слой одномерного периодического материала, погруженный в однородную среду, есть не что иное, как одномерная решетка Брэгга с конечным числом сегментов. Отражение и пропускание света такой решеткой уже рассматривалось в подразд. 7.1.3.

Можно ожидать, что решетка Брэгга с большим, но конечным числом сегментов N приобретает свойства периодической среды из аналогичных элементарных ячеек. Это на самом деле так, поскольку полосы пропускания и полосы заграждения решетки Брэгга математически идентичны фотонным разрешенным и запрещенным зонам протяженной периодической среды. Однако спектральная зависимость коэффициентов пропускания и отражения решетки Брэгга, осциллирующий характер которой чувствителен к размеру решетки и наличию у нее границ, не имеет аналога в протяженной периодической среде.

Аналогично, фазовая и групповая скорости и связанные с ними эффективные показатели преломления, определяемые из дисперсионного уравнения в протяженной периодической среде, не имеют прямых аналогов в случае решетки Брэгга конечного размера. Тем не менее такие параметры можно ввести и для решетки, если определить ее комплексный амплитудный коэффициент пропускания $t(\omega)$ и приписать его слою однородной среды той же толщины с помощью соотношения $\arg\{t_N\} = (\omega/c_0)n_{эфф}d$. Затем можно ввести эффективный групповой показатель преломления

$$N_{эфф} = n_{эфф} + \omega \frac{dn_{эфф}}{d\omega}$$

[см. (5.125)]. Зависимость этих эффективных показателей преломления от частоты отличается от показанной на рис. 7.20 для протяженной периодической среды тем, что она имеет колебания внутри полос пропускания. Однако при больших N, скажем, больше 100, эти колебания сглаживаются и эффективные показатели преломления становятся примерно такими же, как и для протяженной периодической среды.

Другая интересная конфигурация — слой однородной среды, погруженный в периодическую среду. В этой конфигурации свет может пленяться слоем за счет всенаправленного идеального отражения от границ окружающей периодической среды, так что слой становится оптическим волноводом. Эта конфигурация обсуждается в разд. 8.4.

7.3. ДВУМЕРНЫЕ И ТРЕХМЕРНЫЕ ФОТОННЫЕ КРИСТАЛЛЫ

Понятия, введенные в разд. 7.2 для описания распространения волн в одномерных периодических средах, легко обобщаются на двумерные и трехмерные периодические структуры. Сюда входят волны Блоха как собственные моды периодической среды и дисперсионные соотношения ω—K с фотонными разрешенными и запрещенными зонами. В отличие от одномерных структур, двумерные фотонные кристаллы имеют двумерно-полные фотонные запрещенные зоны, т. е. общие запрещенные зоны для волн обеих поляризаций, распространяющихся в любом направлении в плоскости периодичности. Однако трехмерно-полные запрещенные зоны, т. е. общие запрещенные зоны для всех направлений и поляризаций, могут существовать только в трехмерных фотонных кристаллах. Математическое описание двумерных и трехмерных периодических сред более сложно и графическое представление дисперсионных диаграмм более затруднительно ввиду появления дополнительных степеней свободы, однако основные концепции существенно те же, что и для одномерных периодических сред. Данный раздел начинается с простого рассмотрения двумерных структур, затем более подробно рассматривается трехмерный случай.

7.3.1. Двумерные фотонные кристаллы

Двумерные периодические структуры

Рассмотрим двумерную периодическую структуру, например набор одинаковых параллельных стержней, трубок или жил, погруженных в однородную базовую среду (рис. 7.25, *а*) и расположенных в виде прямоугольной решетки, как показано на рис. 7.25, *б*. Диэлектрическая непроницаемость $\eta(x, y)$ является периодической функцией поперечных координат x, y и однородна в аксиальном направлении z. Если a_1 и a_2 — периоды в направлениях x и y, то $\eta(x, y)$ удовлетворяет условию трансляционной симметрии

$$\eta(x + m_1 a_1,\ y + m_2 a_2) = \eta(x,\ y) \tag{7.97}$$

для любых целых m_1 и m_2. Эта периодическая функция представляется двумерным разложением в ряд Фурье:

$$\eta(x,\ y) = \sum_{\ell_1=-\infty}^{\infty} \sum_{\ell_2=-\infty}^{\infty} \eta_{\ell_1\ell_2} \exp(-j\ell_1 g_1 z) \exp(-j\ell_2 g_2 z), \tag{7.98}$$

где $g_1 = 2\pi/a_1$ и $g_2 = 2\pi/a_2$ — фундаментальные пространственные частоты (рад/мм) в направлениях x и y, а $\ell_1 g_1$ и $\ell_2 g_2$ — их гармоники. Коэффициенты $\eta_{\ell_1\ell_2}$ зависят от профиля периодической функции $\eta(x, y)$, т. е. от размера стержней.

Двумерный Фурье-образ периодической функции состоит из точек (дельта-функций), образующих прямоугольную решетку, как показано на рис. 7.25, *в*. Эта решетка в Фурье-пространстве известна в физике твердого тела как **обратная решетка**.

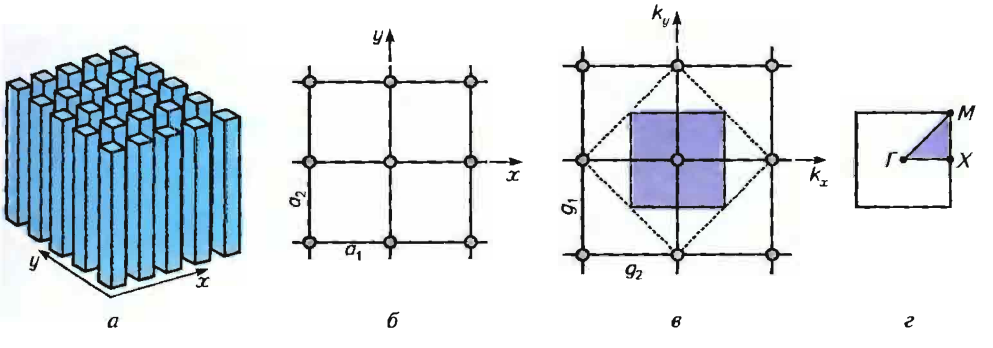


Рис. 7.25. Двумерная периодическая структура из параллельных стержней (*а*). Прямоугольная решетка, в узлах которой расположены стержни (*б*). Двумерный Фурье-образ точек решетки — новый набор точек, образующих обратную решетку с периодами $g_1 = 2\pi/a_1$ и $g_2 = 2\pi/a_2$ (*в*). Сиреневым отмечена первая зона Бриллюэна. Для квадратной решетки ($a_1 = a_2 = a$) неприводимая зона Бриллюэна — треугольник *ГМХ* (*г*)

Каковы оптические моды системы с такой симметрией? Ответ на этот вопрос представляет собой простое обобщение одномерного случая, описываемо-

го формулой (7.69). Для волн, распространяющихся параллельно плоскости x—y, эти моды представляют собой двумерные волны Блоха

$$U(x, y) = p_{K_x K_y}(x, y)\exp(-jK_x x)\exp(-jK_y y), \qquad (7.99)$$

где $p_{K_x K_y}(x, y)$ — периодическая функция с теми же периодами, что и у среды. Волна характеризуется парой волновых чисел Блоха (K_x, K_y). Другая волна с волновыми числами Блоха $(K_x + g_1, K_y + g_2)$ не является новой модой. Как показано на рис. 7.25, *в*, полный набор мод на Фурье-плоскости характеризуется точками, лежащими внутри прямоугольника $[-g_1/2 < K_x \le g_1/2|$, $[-g_2/2 < K_y \le g_2/2|$, который представляет собой первую зону Бриллюэна.

Для сокращения набора независимых волновых векторов Блоха в пределах первой зоны Бриллюэна могут быть использованы другие симметрии. Когда учтены все имеющиеся симметрии, получающаяся зона Бриллюэна называется **неприводимой**. Например, вращательная симметрия, которой обладает квадратная решетка, дает неприводимую зону Бриллюэна в виде треугольника, как видно на рис. 7.25, *г*.

Двумерные косопериодические структуры

Примером другого класса двумерных периодических структур является набор цилиндрических отверстий, расположенных в виде треугольной решетки, как показано на рис. 7.26, *а*. Поскольку точки решетки расположены косо (не по осям x и y), для построения решетки введем два основных *вектора* трансляции $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ (рис. 7.26, *б*), тогда каждая точка решетки имеет радиус вектор $\boldsymbol{R} = m_1\boldsymbol{a}_1 + m_2\boldsymbol{a}_2$, где m_1 и m_2 — целые числа. Введем также радиус-вектор $\boldsymbol{r}_T = (x, y)$, так что периодическая функция $\varepsilon(\boldsymbol{r}_T) = \varepsilon(x, y)$ удовлетворяет условию трансляционной симметрии $\varepsilon(\boldsymbol{r}_T) = \varepsilon(\boldsymbol{r}_T + \boldsymbol{R})$ (индекс «Т» означает поперечность).

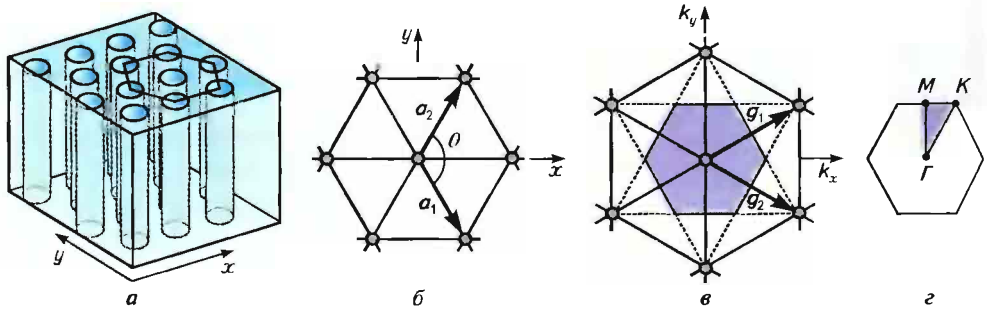


Рис. 7.26. Двумерная периодическая структура из параллельных цилиндрических отверстий (*а*). Треугольная решетка, в узлах которой расположены отверстия. На схеме векторы $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ равны по величине, $a_1 = a_2 = a$, и образуют угол $\theta = 120°$ (*б*). Обратная решетка; сиреневым помечена первая зона Бриллюэна, представляющая собой правильный шестиугольник (*в*). Неприводимая зона Бриллюэна — треугольник ΓMK (*г*)

Двойное разложение Фурье для такой функции порождает набор точек обратной решетки, определяемой векторами g_1 и g_2, которые ортогональны a_1 и a_2, соответственно, и имеют величины

$$g_1 = \frac{2\pi}{a_1 \sin\theta}; \quad g_2 = \frac{2\pi}{a_2 \sin\theta},$$

где θ — угол между векторами a_1 и a_2. Двумерная обратная решетка тоже треугольная и генерируется вектором $G = \ell_1 g_1 + \ell_2 g_2$, где ℓ_1 и ℓ_2 — целые числа (рис. 7.26, *в*).

Для волн, распространяюшихся параллельно плоскости x—y, моды Блоха имеют вид

$$U(r_T) = p_{K_T}(r_T)\exp(-jK_T \cdot r_T), \tag{7.100}$$

где $K_t = (K_x, K_y)$ — волновой вектор Блоха, а $p_{K_T}(r_T)$ — двумерно-периодическая функция на той же решетке. Две моды Блоха с волновыми векторами K_Γ и $K_T + G$ эквивалентны. Полный набор волновых векторов Блоха включает только векторы, концы которых лежат внутри первой зоны Бриллюэна, показанной на рис. 7.26, *в*.

Дисперсионное соотношение можно получить из условия, чтобы волна Блоха (7.99) или (7.100) удовлетворяла обобщенному уравнению Гельмгольца. Вычисления выполняются с помощью разложений в ряды Фурье, как делалось в одномерном случае и как будет описано в более общем виде для трехмерного случая.

Пример 7.14

Треугольная решетка из цилиндрических отверстий

Двумерный фотонный кристалл состоит из однородной среды (n = 3,6) с заполненными воздухом отверстиями радиусом 0,48a, центры которых образуют треугольную решетку с постоянной a. Рассчитанная дисперсионная диаграмма (рис. 7.27) для TE- и TM-волн, распространяющихся в плоскости периодичности (k_z = 0), характеризуется наличием двумерно-полной фотонной запрещенной зоны на частотах вблизи угловой частоты $\omega_0 = \pi c_0/a$[4]. Как и в одномерном случае, запрещенная зона может быть расширена за счет использования материалов с большей разницей показателей преломления. Вообще, для большинства геометрий периодические структуры демонстрируют фотонные запрещенные зоны при достаточном контрасте показателей преломления.

Такая фотонно-кристаллическая структура находит применение в «дырчатых» оптических волокнах, имеющих ряд перспективных приложений.

---

[4] См. *Johnson S.G., Joannopoulos J.D.* Block-Iterative Frequency-Domain Methods for Maxwell's Equations in a Planewave Basis. Optics Express. Vol. 8, 2001. P. 173—190.

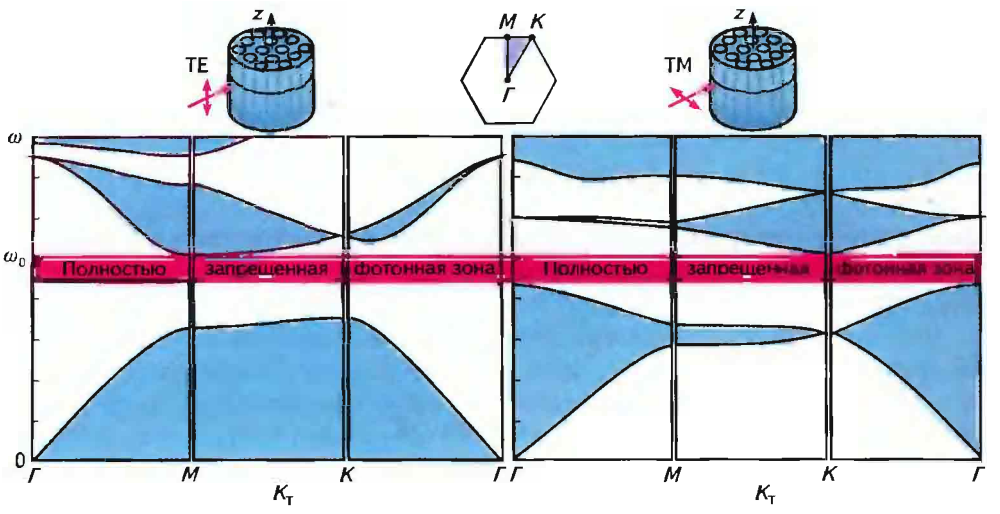


Рис. 7.27. Рассчитанная зонная структура двумерного фотонного кристалла, состоящего из однородной среды ($n = 3{,}6$) с заполненными воздухом цилиндрическими отверстиями радиусом $0{,}48a$, расположенными в виде треугольной решетки с постоянной решетки a. Вдоль оси абсцисс развернуто движение конца волнового вектора Блоха по периметру неприводимой зоны Бриллюэна — треугольнику ΓMK. Ось ординат проградуирована в единицах $\omega_0 = \pi c_0/a$. Волна распространяется в плоскости периодичности и имеет поляризацию TE (*слева*) и TM (*справа*)

Для наклонной волны, распространяющейся под углом к плоскости x—y, волна Блоха (7.100) превращается в

$$U(\boldsymbol{r}_\text{T}) = p_{\boldsymbol{K}}(\boldsymbol{r}_\text{T})\exp(-j\boldsymbol{K}_\text{T}\cdot\boldsymbol{r}_\text{T})\exp(-jk_z z), \qquad (7.101)$$

где k_z постоянно. Зонная структура принимает вид набора поверхностей, описываемых уравнением $\omega = \omega(\boldsymbol{K}_\text{T}, k_z)$.

Трехмерно-полная запрещенная зона — это область частот ω, которая не пересекается ни одной из этих поверхностей, т. е. множество значений ω, которые не получаются ни при какой комбинации действительных $\boldsymbol{K}_\text{T}$ и k_z. Если, как видно из примера на рис. 7.27, при $k_z = 0$ в двумерном кристалле существует двумерно-полная фотонная запрещенная зона, то для всех наклонных волн полностью запрещенная зона в двумерных периодических структурах существовать не может.

7.3.2. Трехмерные фотонные кристаллы

Структура кристалла

Трехмерный фотонный кристалл генерируется помещением одинаковых копий базовой диэлектрической структуры, такой как шар или куб, в точки трехмерной решетки, генерируемой вектором решетки $\boldsymbol{R} = m_1\boldsymbol{a}_1 + m_2\boldsymbol{a}_2 + m_3\boldsymbol{a}_3$,

где m_1, m_2, m_3 — целые числа, а $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ — основные векторы, определяющие элементарную ячейку. Структура в целом периодична, и ее оптические свойства, такие как диэлектрическая проницаемость $\varepsilon(\boldsymbol{r})$ и непроницаемость $\eta(\boldsymbol{r}) = \varepsilon_0/\varepsilon(\boldsymbol{r})$, инвариантны относительно трансляции на $\boldsymbol{R}$, например,

$$\eta(\boldsymbol{r} + \boldsymbol{R}) = \eta(\boldsymbol{r}) \tag{7.102}$$

для всех точек $\boldsymbol{r}$. Следовательно, эти периодические функции можно разложить в тройной ряд Фурье

$$\eta(\boldsymbol{r}) = \sum_{\boldsymbol{G}} \eta_{\boldsymbol{G}} \exp(-j\boldsymbol{G}\cdot\boldsymbol{r}), \tag{7.103}$$

где $\boldsymbol{G} = \ell_1\boldsymbol{g}_1 + \ell_2\boldsymbol{g}_2 + \ell_3\boldsymbol{g}_3$ — вектор, определяемый основными векторами $\boldsymbol{g}_1$, $\boldsymbol{g}_2$, $\boldsymbol{g}_3$ **обратной решетки**; ℓ_1, ℓ_2, ℓ_3 — целые числа. Векторы $\boldsymbol{g}$ выражаются через векторы $\boldsymbol{a}$ формулами

$$\boldsymbol{g}_1 = 2\pi\frac{\boldsymbol{a}_2\times\boldsymbol{a}_3}{\boldsymbol{a}_1\cdot(\boldsymbol{a}_2\times\boldsymbol{a}_3)};\quad \boldsymbol{g}_2 = 2\pi\frac{\boldsymbol{a}_3\times\boldsymbol{a}_1}{\boldsymbol{a}_2\cdot(\boldsymbol{a}_3\times\boldsymbol{a}_1)};\quad \boldsymbol{g}_3 = 2\pi\frac{\boldsymbol{a}_1\times\boldsymbol{a}_2}{\boldsymbol{a}_3\cdot(\boldsymbol{a}_1\times\boldsymbol{a}_2)}, \tag{7.104}$$

так что $\boldsymbol{g}_1\cdot\boldsymbol{a}_1 = 2\pi$, $\boldsymbol{g}_1\cdot\boldsymbol{a}_2 = 0$, $\boldsymbol{g}_1\cdot\boldsymbol{a}_3 = 0$, т. е. $\boldsymbol{g}_1$ ортогонален $\boldsymbol{a}_2$ и $\boldsymbol{a}_3$, а его длина обратно пропорциональна длине $\boldsymbol{a}_1$. Такими же свойствами обладают $\boldsymbol{g}_2$ и $\boldsymbol{g}_3$. Можно показать также, что $\boldsymbol{G}\cdot\boldsymbol{R} = 2\pi$.

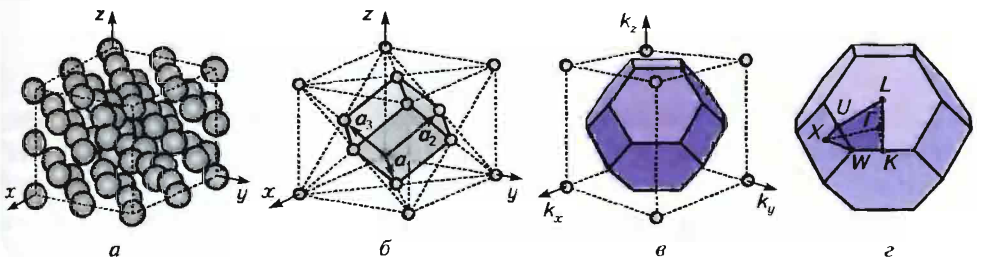


Рис. 7.28. Трехмерная периодическая структура из диэлектрических сфер (*а*). Сферы расположены в узлах решетки типа алмаза (кубической гранецентрированной) с

$$\boldsymbol{a}_1 = \frac{a}{\sqrt{2}}(\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}});\quad \boldsymbol{a}_2 = \frac{a}{\sqrt{2}}(\hat{\boldsymbol{y}} + \hat{\boldsymbol{z}});\quad \boldsymbol{a}_3 = \frac{a}{\sqrt{2}}(\hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}),$$

где a — постоянная решетки (*б*). Соответствующая обратная решетка — кубическая объемно-центрированная, с зоной Бриллюэна, отмеченной сиреневым цветом и известной как ячейка Вигнера—Зейца (*в*). Неприводимая зона Бриллюэна представляет собой многогранник, вершины которого помечены кристаллографическими символами *ГXULKW* (*г*)

Если $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ и $\boldsymbol{a}_3$ взаимно ортогональны, то $\boldsymbol{g}_1$, $\boldsymbol{g}_2$ и $\boldsymbol{g}_3$ также взаимно ортогональны, а их величины $g_1 = 2\pi/a_1$, $g_2 = 2\pi/a_2$ и $g_3 = 2\pi/a_3$ представляют собой пространственные частоты, связанные с периодичностью в каждом из трех направлений. Пример трехмерной кристаллической решетки и соответствующей обратной решетки показан на рис. 7.28.

Моды Блоха

Моды трехмерной периодической среды — это волны, сохраняющие свою форму при трансляции на вектор решетки $\boldsymbol{R}$ и изменяющиеся при этом лишь на множитель с модулем, равным единице. Эти моды имеют блоховский вид $p_{\boldsymbol{K}}(\boldsymbol{r})\exp(-j\boldsymbol{K}\cdot\boldsymbol{r})\hat{\boldsymbol{e}}$, где $p_{\boldsymbol{K}}(\boldsymbol{r})$ — функция, периодическая в трех измерениях; ее периодичность характеризуется тем же вектором $\boldsymbol{R}$; $\boldsymbol{K}$ — **волновой вектор Блоха**; $\hat{\boldsymbol{e}}$ — единичный вектор направления поляризации. Мода Блоха — бегущая плоская волна $\exp(-j\boldsymbol{K}\ \boldsymbol{r})$, модулированная периодической функцией $p_{\boldsymbol{K}}(\boldsymbol{r})$. Сдвиг на $\boldsymbol{R}$ приводит к умножению на фазовый множитель $\exp(-j\boldsymbol{K}\cdot\boldsymbol{R})$, зависящий от $\boldsymbol{K}$.

Две моды с волновыми векторами Блоха $\boldsymbol{K}$ и $\boldsymbol{K}' = \boldsymbol{K} + \boldsymbol{G}$ эквивалентны, поскольку $\exp(-j\boldsymbol{K}'\cdot\boldsymbol{R}) = \exp(-j\boldsymbol{K}\cdot\boldsymbol{R})$, т. е. трансляция на $\boldsymbol{R}$ для них эквивалентна умножению на один и тот же фазовый множитель. Это происходит потому, что $\exp(-j\boldsymbol{G}\cdot\boldsymbol{R}) = \exp(-j2\pi) = 1$. Следовательно, для полного описания всех мод достаточно рассматривать значения $\boldsymbol{K}$ внутри конечного объема обратной решетки — зоны Бриллюэна. Зона Бриллюэна представляет собой область, точки которой находятся ближе к одной особой точке обратной решетки (началу зоны, обозначаемому Г), чем к любым другим точкам этой решетки. Наличие дополнительных симметрий решетки кристалла позволяет дополнительно сократить указанный объем до неприводимой зоны Бриллюэна, пример которой показан на рис. 7.28.

Структура фотонных зон

Определение дисперсионного соотношения между ω и $\boldsymbol{K}$ для трехмерной среды начнем с задачи на собственные значения, описываемой обобщенным уравнением Гельмгольца (7.2). Один из подходов к решению этой задачи состоит в обобщении метода Фурье, введенного в подразд. 7.2.3 для структур, периодических в одном измерении. Путем разложения периодических функций $\eta(\boldsymbol{r})$ и $p_{\boldsymbol{K}}(\boldsymbol{r})$ в ряды Фурье дифференциальное уравнение (7.2) сводится к системе алгебраических уравнений, приводящей к матричной задаче на собственные значения, решаемой численными методами. Как указывалось в подразд. 7.2.3, мы работаем с магнитным полем, чтобы гарантировать эрмитовость матричного представления.

Раскладывая периодическую функцию $p_{\boldsymbol{K}}(\boldsymbol{r})$, входящую в выражение для волны Блоха, в тройной ряд Фурье

$$p_{\boldsymbol{K}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} C_{\boldsymbol{G}} \exp(-j\boldsymbol{G}\cdot\boldsymbol{r}), \tag{7.105}$$

запишем вектор магнитного поля в блоховском виде

$$\boldsymbol{H}(\boldsymbol{r}) = p_{\boldsymbol{K}}(\boldsymbol{r})\exp(-j\boldsymbol{K}\cdot\boldsymbol{r})\hat{\boldsymbol{e}} = \sum_{\boldsymbol{G}} C_{\boldsymbol{G}} \exp\left[-j(\boldsymbol{K}+\boldsymbol{G})\cdot\boldsymbol{r}\right]\hat{\boldsymbol{e}}. \tag{7.106}$$

Для простоты обозначений зависимость коэффициентов Фурье C_G от волнового вектора Блоха $\boldsymbol{K}$ явно не указана. Поставляя (7.103) и (7.106) в (7.2), используя равенство

$$\nabla \times \exp(-j\boldsymbol{K}\cdot\boldsymbol{r})\hat{e} = -j(\boldsymbol{K}\times\hat{e})\exp(-j\boldsymbol{K}\cdot\boldsymbol{r})$$

и приравнивая члены с одинаковой пространственной частотой, получаем

$$-\sum_{G'}(\boldsymbol{K}+\boldsymbol{G})\times\left[(\boldsymbol{K}+\boldsymbol{G}')\times\hat{e}\right]\eta_{G-G'}C_{G'} = \frac{\omega^2}{c_0^2}C_G\hat{e}. \qquad (7.107)$$

Скалярное умножение обеих частей этого векторного равенства на $\hat{e}$ и использование векторного тождества $\boldsymbol{A}\cdot(\boldsymbol{B}\times\boldsymbol{C}) = -(\boldsymbol{B}\times\boldsymbol{A})\cdot\boldsymbol{C}$ дает

$$-\sum_{G'}F_{GG'}C_{G'} = \frac{\omega^2}{c_0^2}C_G; \quad F_{GG'} = \left[(\boldsymbol{K}+\boldsymbol{G})\times\hat{e}\right]\cdot\left[(\boldsymbol{K}+\boldsymbol{G}')\times\hat{e}\right]\eta_{G-G'}. \qquad (7.108)$$

Дифференциальное уравнение Гельмгольца преобразовалось в систему линейных уравнений для коэффициентов Фурье $\{C_G\}$. Поскольку $\eta(\boldsymbol{r})$ действительно, $\eta_{G-G'} = \eta^*_{G-G'}$, и матрица $F_{GG'}$ эрмитова. Таким образом, (7.108) представляет задачу на собственные значения для эрмитовой матрицы. Для каждого волнового вектора Блоха $\boldsymbol{K}$ собственные значения ω^2/c_0^2 дают множество значений ω, определяющих ω—$\boldsymbol{K}$ — диаграмму и структуру фотонных зон. Собственные векторы $\{C_G\}$ определяют периодическую функцию $p_{\boldsymbol{K}}(\boldsymbol{r})$, входящую в волну Блоха.

Примеры

Решетка алмаза из сферических отверстий

Примером трехмерного фотонного кристалла, у которого имеется полностью запрещенная во всех трех измерениях фотонная зона, является структура из сферических воздушных полостей в материале с высоким показателем преломления, расположенных в узлах решетки алмаза (см. рис. 7.28). Радиусы сфер достаточно велики, так что они перекрываются и образуют пересекающиеся промежутки. Рассчитанная зонная структура, показанная на рис. 7.29, имеет сравнительно широкую полностью запрещенную в трех измерениях фотонную зону между двумя нижними разрешенными зонами[5]. Фотонные кристаллы с использованием сферических отверстий в кремнии были изготовлены путем выращивания кремния в пустотах образца из плотно упакованных кварцевых сфер, соединенных небольшими перемычками, которые формируются в процессе спекания, с последующим удалением кварцевого шаблона[6].

---

[5] См. *Johnson S.G., Joannopoulos J.D.* Block-Iterative Frequency-Domain Methods for Maxwell's Equations in a Planewave Basis. Optics Express. Vol. 8, 2001. P. 173—190.

[6] См. *Blanco A. et al.*, Large-Scale Synthesis of a Silicon Photonic Crystal with a Complete Three-Dimensional Bandgap Near 1.5 Micrometers. Nature. Vol. 405, 2000. P. 437—440.

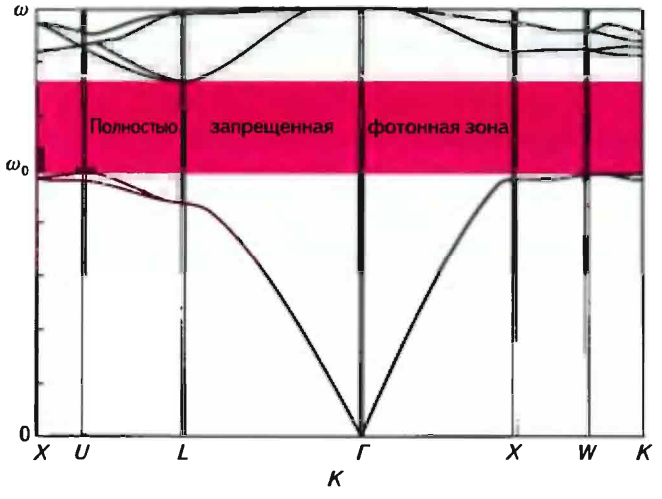


Рис. 7.29. Рассчитанная зонная структура трехмерного фотонного кристалла с решеткой типа алмаза, постоянная решетки a. Структура состоит из полых сфер радиусом $0{,}325a$ в однородном материале с показателем преломления $n = 3{,}6$. Фотонная запрещенная зона занимает интервал примерно от $\omega_0 = \pi c_0/a$ до $1{,}32\omega_0$

Яблоновит

Первое экспериментальное наблюдение трехмерной полностью запрещенной фотонной зоны было выполнено Эли Яблоновичем в 1991 г. с использованием искусственной решетки со структурой алмаза, которая теперь называется «яблоновит». Эта косопористая структура изготавливается путем сверления периодического набора цилиндрических отверстий под специальными углами в диэлектрическом слое. В каждой точке двумерной треугольной решетки на поверхности слоя сверлится три отверстия, направления отверстий параллельны трем осям решетки кристалла типа алмаза, как показано на рис. 7.30, *а*. Эта структура обладает полностью запрещенной фотонной зоной с относительной шириной 0,19 при показателе преломления слоя 3,6.

Поленница

Еще один фотонный кристалл, более простой в изготовлении, выполняется в виде периодической в одном измерении пачки чередующихся слоев, каждый из которых, в свою очередь, является двумерным фотонным кристаллом. Например, «поленница», изображенная на рис. 7.30, *б*, состоит из слоев параллельных брусков, повторяющаяся через каждые четыре слоя. Ориентация брусков в соседних слоях различается на 90°, а через каждые два слоя бруски сдвинуты на полшага. Получающаяся структура имеет симметрию гранецентрированной тетрагональной кристаллической решетки. Изготовленная с применением кремниевой технологии с минимальным размером структурных элементов 180 нм, эта структура демонстрирует полностью запрещенную в трех измерениях фотонную зону в области длин волн 1,35—1,95 мкм [7]

[7] См. *Fleming J.G., Lin S.Y.* Three-Dimensional Photonic Crystal with a Stop Band from 1,35 to 1,95 mm, Optics Letters. Vol. 24, 1999. P. 49—51.

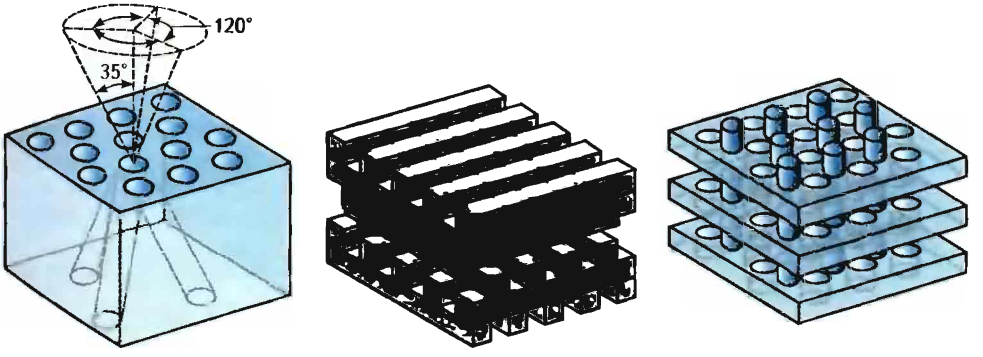


Рис. 7.30. Фотонный кристалл «яблоновит» изготавливается сверлением цилиндрических отверстий в диэлектрическом слое (*а*). В каждой точке двумерной треугольной решетки на поверхности сверлится три отверстия в направлениях, образующих угол 35° с нормалью, с азимутальным углом 120° между ними. Фотонный кристалл «поленница» состоит из чередующихся слоев параллельных стержней, направления которых в соседних слоях повернуты на 90° (*б*). Структура «отверстия и стержни» состоит из чередующихся слоев двумерно-периодических структур: за каждым слоем диэлектрика с гексагональной решеткой из параллельных цилиндрических отверстий следует слой параллельных стержней, расположенных между отверстиями (*в*)

Отверстия и стержни

Еще один пример — структура из отверстий и стержней, показанная на рис. 7.30, *в*. Здесь используется два типа двумерных фотонно-кристаллических слоев, взаимно дополняющих друг друга: диэлектрические стержни в воздухе и воздушные отверстия в диэлектрике. Изготовленная из кремния, такая структура обладает запрещенной фотонной зоной для всех углов наклона в диапазоне длин волн 1,15—1,6 мкм[8].

Как поленница, так и структура из отверстий и стержней позволяет вносить произвольные точечные дефекты, например пропущенное отверстие или стержень, что позволяет изготавливать такие устройства, как фотонно-кристаллические волноводы (см. разд. 8.5), фотонно-кристаллические нанорезонаторы (см. подразд. 10.4.4) и специально управляемые излучатели света[9] (см. гл. 17). Фактически именно возможность внесения дефектов — наиболее ценное качество двумерных и трехмерных фотонно-кристаллических структур, в то время как одномерные структуры служат непревзойденными всенаправленными отражателями.

---

[8] См. *Qi M. et al.* A Three-Dimensional Optical Photonic Crystal with Designed Point Defects. Nature. Vol. 429, 2004. P. 538—542.

[9] См. *Ogawa S.P., Imada M., Yoshimoto D., Okano M., Noda S.* Control of Light Emission by 3D Photonic Crystals. Science. Vol. 305, 2004. P. 227—229.

Рекомендуемая литература

КНИГИ О МНОГОСЛОЙНЫХ И ПЕРИОДИЧЕСКИХ СРЕДАХ

Višňovský Š. Optics in Magnetic Multilayers and Nanostructures. CRC Press, 2006.

Stenzel O. The Physics of Thin Film Optical Spectra: An Introduction. Springer-Verlag, 2005.

Yeh P. Optical Waves in Layered Media. Wiley, 2005.

Brillouin L. Wave Propagation in Periodic Structures: Electric Filters and Crystal Lattices. Dover, 2nd ed. 1953; reprinted 2003.

Nevie1re M., Popov E. Light Propagation in Periodic Media: Differential Theory and Design. Marcel Dekker, 2003.

Born M., Wolf E. Principles of Optics. Cambridge University Press. 7th expanded and corrected ed., 2002. Sec. 1.6.

Kashyap R. Fiber Bragg Gratings. Academic Press, 1999.

Chew W.C. Waves and Fields in Inhomogeneous Media. Van Nostrand Reinhold, 1990.

Yariv A., Yeh P. Optical Waves in Crystals: Propagation and Control of Laser Radiation. Wiley, 1985.

КНИГИ О ФОТОННЫХ КРИСТАЛЛАХ

Yasumoto K., ed. Electromagnetic Theory and Applications for Photonic Crystals. CRC Press, 2006.

Lourtioz J.-M., Benisty H., Berger V., Gérard J.-M., Maystre D., Tchelnokov A. Photonic Crystals: Towards Nanoscale Photonic Devices. Springer-Verlag, 2005.

Sakoda K. Optical Properties of Photonic Crystals. Springer-Verlag, 2nd ed. 2005.

Busch K., Lolkes S., Wehrspohn R.B., Föll H., eds. Photonic Crystals: Advances in Design, Fabrication, and Characterization. Wiley, 2004.

Inoue K., Ohtaka K., eds. Photonic Crystals: Physics, Fabrication and Applications. Springer-Verlag, 2004.

Noda S., Baba T., eds. Roadmap on Photonic Crystals. Kluwer, 2003.

Kochergin V. Omnidirectional Optical Filters. Kluwer, 2003.

Slusher R.E., Eggleton B.J., eds. Nonlinear Photonic Crystals. Springer-Verlag, 2003.

Johnson S.G., Joannopoulos J.D. Photonic Crystals: The Road from Theory to Practice. Springer-Verlag, 2002.

Soukoulis C.M., ed. Photonic Band Gap Materials. Kluwer, 1996.

Joannopoulos J.D., Meade R.D., Winn J.N. Photonic Crystals: Molding the Flow of Light. Princeton University Press, 1995.

Senechal M. Quasicrystals and Geometry. Cambridge University Press, 1995.

СТАТЬИ

Issue on nanophotonics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 12, № 6, 2006.

Adibi A., Lin S.-Y., Scherer A., eds. Photonic crystal materials and devices III. SPIE Proceedings. Vol. 5733, 2005.

Adibi A., Scherer A., Lin S.-Y., eds., Photonic crystal materials and devices II. SPIE Proceedings. Vol. 5360, 2004.

Ogawa S.P., Imada M., Yoshimoto S., Okano M., Noda S. Control of Light Emission by 3D Photonic Crystals. Science. Vol. 305, 2004. P. 227—229.

Qi M., Lidorikis E., Rakich P.T., Johnson S.G., Joannopoulos J.D., Ippen E.P., Smith H.I. A Three-Dimensional Optical Photonic Crystal with Designed Point Defects. Nature. Vol. 429, 2004. P. 538—542.

Adibi A., Scherer A., Lin S.-Y., eds. Photonic crystal materials and devices I. SPIE Proceedings. Vol. 5000, 2003.

Yablonovitch E. Photonic Crystals: Semiconductors of Light. Scientific American. Vol. 285, № 6, 2001. P. 47—55.

Deopura M., Ullal C.K., Temelkuran B., Fink Y. Dielectric Omnidirectional Visible Reflector. Optics Letters. Vol. 26, 2001. P. 1197—1199.

Focus issue on photonic bandgap calculations. Optics Express. Vol. 8, № 3, 2001.

Johnson S.G., Joannopoulos J.D. Block-Iterative Frequency-Domain Methods for Maxwell's Equations in a Planewave Basis. Optics Express. Vol. 8, 2001. P. 173—190.

Müller M., Zentel R., Maka T., Romanov S.G., Sotomayor C.M. Torres, Photonic Crystal Films with High Refractive Index Contrast. Advanced Materials. Vol. 12, 2000. P. 1499—1503.

Blanco A., Chomski E., Grabtchak S., Ibisate M., John S., Leonard S.W., Lopez C., Meseguer F., Miguez H., Mondia J.P., Ozin G.A., Toader O., van Driel H.M. Large-Scale Synthesis of a Silicon Photonic Crystal with a Complete Three-dimensional Bandgap Near 1,5 Micrometres. Nature. Vol. 405, 2000. P. 437—440.

Fink Y., Winn J.N., Fan S., Chen C., Michel J., Joannopoulos J.D., Thomas E.L. A Dielectric Omnidirectional Reflector. Science. Vol. 282, 1998. P. 1679—1682.

Bendickson J.M., Dowling J.P., Scalora M. Analytic Expressions for the Electromagnetic Mode Density in Finite, One-Dimensional, Photonic Band-Gap Structures. Physical Review E. Vol. 53, 1996. P. 4107—4121.

Villeneuve P.R. Piché M. Photonic Bandgaps in Periodic Dielectric Structures. Progress in Quantum Electronics. Vol. 18, 1994. P. 153—200.

John S. Localization of Light. Physics Today. Vol. 44, № 5, 1991. P. 32—40.

Yablonovitch E., Gmitter T.J., Leung K.M. Photonic Band Structures: The Face-Centered Cubic Case Employing Non-Spherical Atoms. Physical Review Letters. Vol. 67, 1991. P. 2295—2298.

Yablonovitch E., Gmitter T.J. Photonic Band Structure: The Face-Centered-Cubic Case. Journal of the Optical Society of America A. Vol. 7, 1990. P. 1792—1800.

John S. Strong Localization of Photons in Certain Disordered Dielectric Superlattices. Physical Review Letters. Vol. 58, 1987. P. 2486—2489.

Yablonovitch E. Inhibited Spontaneous Emission in Solid-State Physics and Electronics. Physical Review Letters. Vol. 58, 1987. P. 2059—2062.

Задачи

К РАЗДЕЛУ 7.1

1. **Светоделительный слой.** Диэлектрический слой толщиной d с показателем преломления n, ориентированный под углом 45° к направлению падающего света, используется в качестве светоделителя. Выведите выражение для коэффициентов отражения и пропускания и постройте графики их спектральной зависимости для поляризаций TE и TM.

2. **Воздушный зазор в стекле.** Определите коэффициент пропускания тонкого плоского воздушного зазора толщиной $d = \lambda/2$ в стекле с показателем преломления n. Рассмотрите:

а) нормальное падение;

б) падение TE-волны под углом больше критического.

Может ли волна пройти (туннелировать) через зазор?

3. **Многослойная структура в несогласованной среде.** Комплексный амплитудный коэффициент отражения многослойной структуры в среде с показателем преломления, равным показателю преломления первого слоя n_1, равен r_m. Покажите, что в среде с другим показателем преломления n он равен

$$r = \frac{r_b + r_m}{1 + r_b r_m},$$

где r_b — коэффициент отражения новой границы раздела,

$$r_b = \frac{n - n_1}{n + n_1}.$$

Определите r для следующих предельных случаев: $r_b = 0$, $r_b = 1$, $r_m = 0$ и $r_m = 1$.

4. **Четвертьволновая пленка: угловая зависимость коэффициента отражения.** Рассмотрите четвертьволновое антиотражательное покрытие, описанное в упражнении 7.1. Выведите выражение для коэффициента отражения как функции угла падения.

5. **Стопа из четвертьволновых и полуволновых слоев.** Выведите выражения для коэффициента отражения стопы из N двойных слоев диэлектрика одинаковой оптической толщины $n_1 d_1 = n_2 d_2$, равной $\lambda/4$ и $\lambda/2$.

6. **Решетка Брэгга из GaAs/AlAs как отражатель.** Отражатель на основе решетки Брэгга состоит из N пар чередующихся слоев GaAs ($n_1 = 3{,}57$) и AlAs ($n_2 = 2{,}94$) толщиной d_1 и d_2, равной четверти длины волны в каждой из сред. Решетка находится в протяженной среде GaAs. Рассчитайте и постройте коэффициенты пропускания и отражения решетки как функции N для $N = 1, 2, ..., 10$ на частоте, равной частоте Брэгга.

7. **Решетка Брэгга: угловая и спектральная зависимость коэффициента отражения.** Напишите компьютерную программу на основе матричной алгебры для определения волновой матрицы передачи и коэффициента отражения решетки Брэгга из N пар чередующихся диэлектрических слоев. С помощью этой программы проверьте представленные на рис. 7.13 и 7.14 графики спектральной и угловой зависимостей коэффициента отражения.

К РАЗДЕЛУ 7.2

1. **Относительная ширина фотонной запрещенной зоны.** С помощью Фурье-оптического подхода определите частоту Брэгга и относительную ширину для низшей фотонной запрещенной зоны одномерной периодической структуры, состоящей из стопы диэлектрических слоев одинаковой толщины с чередующимися показателями преломления $n_1 = 1{,}5$ и $n_2 = 3{,}5$ и периодом $\Lambda = 2$ мкм. Считайте, что волна движется вдоль оси периодичности. Повторите решения для $n_1 = 3{,}4$ и $n_2 = 3{,}6$. Сравните результаты.

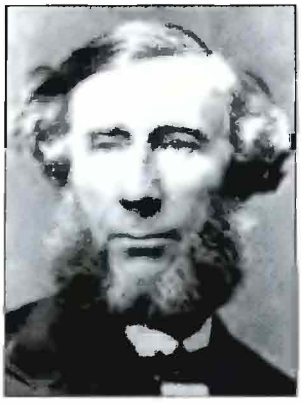
2. **Внеосевая волна в одномерной периодической среде.** Выведите уравнения, аналогичные (7.89)—(7.93) для внеосевой волны, распространяющейся в одномерной периодической среде с поперечной составляющей волнового вектора k_x.

3. **Волна, нормальная к оси, в одномерной периодической среде**. Используйте результаты предыдущей задачи, чтобы показать, что не существует фотонных запрещенных зон для волны, распространяющейся в поперечном направлении к оси одномерной периодической среды, т. е. для $K = 0$.

4. **Всенаправленный отражатель**. Периодическая стопа из двойных слоев диэлектрика с $n_1 d_1 = n_2 d_2$, $n_2 = 2n_1$ и $\Lambda = d_1 + d_2$ используется в качестве всенаправленного отражателя в воздухе. Постройте проекционную дисперсионную диаграмму, показывающую световую линию для воздуха (аналогичную рис. 7.24). Определите полосу частот (в единицах ω_B), в которой происходит всенаправленное отражение.

ГЛАВА 8

ОПТИКА ВОЛНОВОДОВ



Джон Тиндаль (1820—1893) впервые продемонстрировал полное внутреннее отражение — основу оптики волноводов.

В традиционных оптических приборах и системах свет передается в виде пучков, которые коллимируются, передаются, фокусируются и сканируются зеркалами, линзами и призмами. Из-за дифракции пучки расходятся по мере распространения, и чтобы собрать их снова, требуются линзы или зеркала. Так или иначе, объемные оптические элементы, образующие такие системы, часто велики по размеру и громоздки, а объекты, встречаемые пучком на своем пути, могут рассеивать его или препятствовать его прохождению.

Во многих случаях предпочтительно передавать оптические пучки не через свободное пространство, а по диэлектрическому кабелю. Техника, позволяющая достичь этого, называется **оптикой волноводов**. Первоначально ее целью была передача света на большие расстояния без необходимости в дополнительных линзах. Эта техника в настоящее время имеет много важных применений. Среди них передача света на большие расстояния в оптических системах связи, получение изображений в биологии и медицине, где свет должен быть доставлен в труднодоступные области, элементы связи в миниатюрных оптических и оптоэлектронных устройствах и системах.

Основной принцип удержания света прост. Среда с показателем преломления, окруженная средой с меньшим показателем преломления, действует как ловушка для света, в которой лучи удерживаются за счет многократного полно-

го внутреннего отражения на границах. Поскольку этот эффект содействует удержанию света, генерируемого в среде с большим показателем преломления (см. упражнение 1.9), его можно использовать для изготовления световых кабелей — проводников, передающих свет из одного места в другое. **Оптический волновод** — это канал для передачи света, состоящий из диэлектрического слоя, полосы или цилиндра, погруженного в диэлектрический материал с меньшим показателем преломления (рис. 8.1). Свет распространяется во внутренней среде и не излучается во внешнюю среду. Наиболее широко из этих волноводов используется оптическое волокно, состоящее из двух концентрических цилиндров из диэлектрика с низкими потерями, такого, как стекло (см. гл. 9).

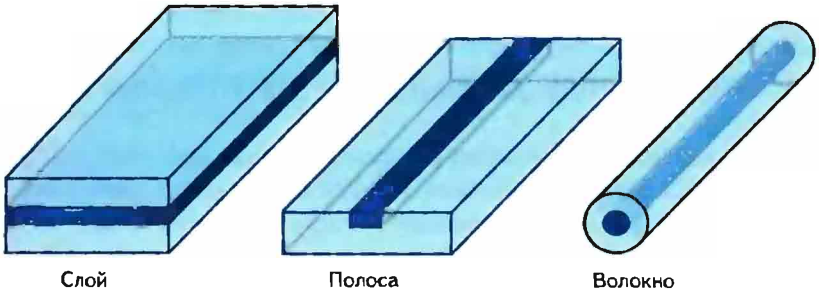


Рис. 8.1. Оптические волноводы

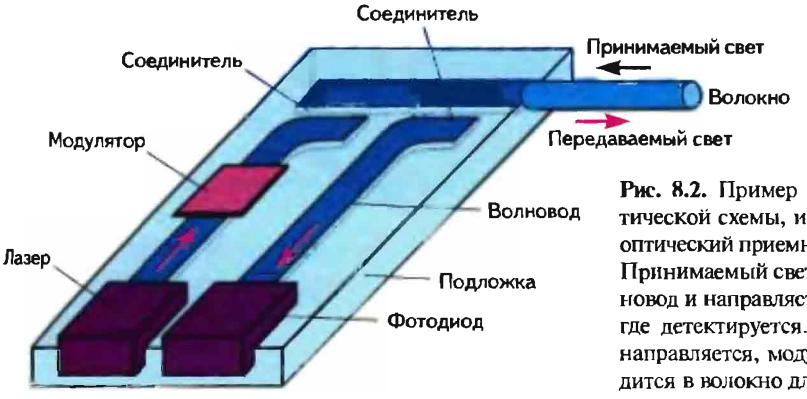


Рис. 8.2. Пример интегрально-оптической схемы, используемой как оптический приемник и передатчик. Принимаемый свет вводится в волновод и направляется на фотодиод, где детектируется. Свет от лазера направляется, модулируется и вводится в волокно для передачи

Интегральная оптика — это технология, в которой на одной подложке («чипе») совмещаются разнообразные оптические устройства и компоненты для генерации, фокусировки, деления, совмещения, изоляции, поляризации, соединения, переключения, модуляции и детектирования света. Оптические волноводы обеспечивают связь между такими элементами. Такие чипы (рис. 8.2) являются оптическим аналогом электронных интегральных схем. Интегральная оптика имеет своей целью миниатюризацию оптических приборов, аналогичную той, которую обеспечило применение интегральных схем в электронике.

О данной главе

Основная теория оптических волноводов представлена в настоящей и следующей главах. В этой главе мы рассматриваем прямоугольные волноводы, которые широко используются в интегральной оптике. Глава 9 посвящена цилиндрическим волноводам, т. е. оптическим волокнам. Если на концах короткого волновода поместить отражатели, получается устройство, удерживающее и накапливающее свет — оптический резонатор. Эти устройства, существенные для лазеров, описаны в гл. 10. Другие компоненты интегральной оптики, такие как полупроводниковые лазеры, детекторы, модуляторы и переключатели, рассматриваются в отдельных главах, специально посвященных этим компонентам и устройствам. Волоконно-оптические системы связи подробно обсуждаются в гл. 24.

8.1. ПЛАНАРНЫЕ ЗЕРКАЛЬНЫЕ ВОЛНОВОДЫ

Начнем с рассмотрения распространения света в волноводе, состоящем из двух параллельных плоских бесконечных *зеркал*, расстояние между которыми равно d (рис. 8.3). Предполагается, что зеркала отражают свет без потерь. Луч света, лежащий в плоскости y—z и образующий угол θ с зеркалами, многократно отражается и распространяется между ними без потери энергии. Таким образом, луч направляется вдоль оси z.

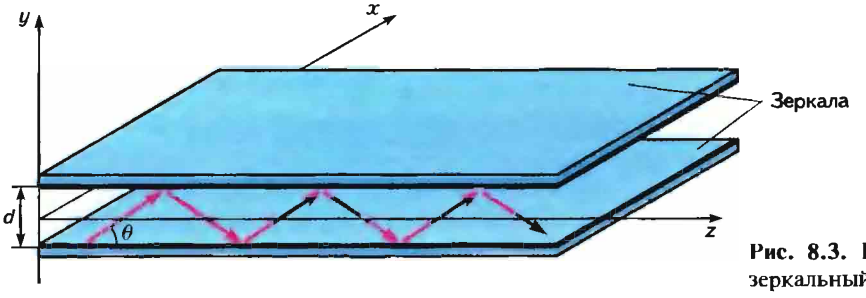


Рис. 8.3. Планарный зеркальный волновод

Такой волновод оказывается идеальным направляющим устройством для световых лучей. Однако он не используется в практических приложениях ввиду дороговизны и сложности изготовления зеркал с низкими потерями. Тем не менее мы исследуем этот простой пример подробно, поскольку он имеет педагогическую ценность для введения в физику диэлектрических волноводов, которые будут рассмотрены в разд. 8.2, и оптических резонаторов, являющихся предметом гл. 10.

Волноводные моды

Картина распространения лучей, направляемых за счет многократных отражений, не может объяснить ряд важных эффектов, требующих использования электромагнитной теории. Простой подход для проведения электромагнитного анализа состоит в том, что каждому лучу света сопоставляется

поперечная электромагнитная (TEM) волна. Полное электромагнитное поле представляет собой сумму этих плоских волн.

Рассмотрим монохроматическую плоскую TEM-волну с длиной волны $\lambda = \lambda_0/n$, волновым числом $k = nk_0$ и фазовой скоростью $c = c_0/n$, где n — показатель преломления среды между зеркалами. Волна поляризована в направлении x, а ее волновой вектор лежит в плоскости y—z и образует угол θ с осью z (см. рис. 8.3). Как и луч света, волна отражается от верхнего зеркала, проходит под углом $-\theta$, отражается от нижнего зеркала, снова распространяется под углом θ и так далее. Поскольку вектор электрического поля параллелен зеркалам, каждое отражение сопровождается сдвигом фазы π для идеального зеркала, однако амплитуда поляризация не меняются. Сдвиг фазы π не гарантирует обращение в нуль суммы падающей и отраженной волн, так что на поверхности зеркал полное поле остается равным нулю. В каждой точке внутри волновода мы имеем TEM-волны, распространяющиеся вверх под углом θ и вниз под углом $-\theta$; все волны поляризованы вдоль оси x.

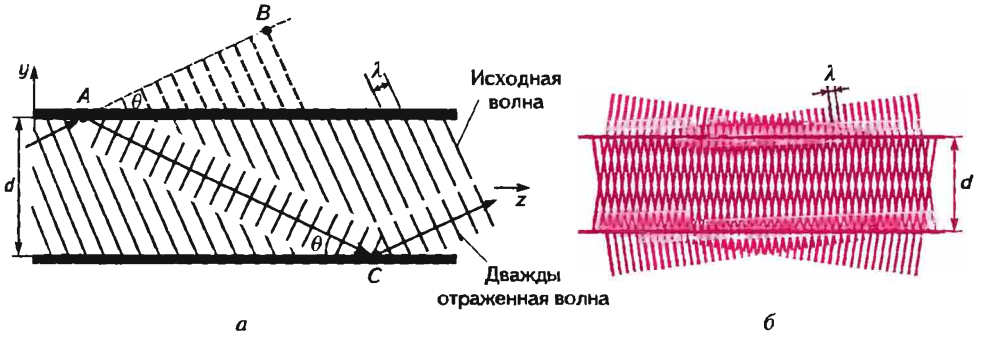


Рис. 8.4. Условие самосогласования: после двух отражений волна воспроизводит сама себя (*а*). При углах, удовлетворяющих условию самосогласования, интерференция двух волн дает картину, которая не меняется вдоль оси z (*б*)

Наложим теперь условие самосогласования, а именно, потребуем, чтобы после двух отражений волна воспроизводила саму себя (рис. 8.4, *а*), т. е. чтобы между зеркалами было всего две разные плоские волны. Поля, удовлетворяющие этому условию, называются модами (или собственными функциями) волновода (см. Приложение В). *Моды — это поля волновода, сохраняющие поперечное распределение и поляризацию во всех точках вдоль оси волновода.* Мы увидим, что условие самосогласования обеспечивает такую инвариантность формы поля. В соответствии с рис. 8.4, фазовый сдвиг $\Delta\varphi$, получаемый исходной волной при распространении от A до B, должен быть равным сдвигу, приобретаемому при отражении в точке A, распространении от A до C и отражении в точке C, или отличаться от него на целое кратное 2π. Учитывая фазовый сдвиг π при каждом отражении, имеем:

$$\Delta\varphi = \frac{2\pi\overline{AC}}{\lambda} - 2\pi - \frac{2\pi\overline{AB}}{\lambda} = 2\pi q,$$

где $q = 0, 1, 2, \ldots$, так что

$$\frac{2\pi(\overline{AC} - \overline{AB})}{\lambda} = 2\pi(q+1).$$

Геометрия, изображенная на рис. 8.4, *а*, вместе с тождеством $\cos(2x) = 1 - 2\sin^2 x$ дает

$$\overline{AC} - \overline{AB} = 2d\sin\theta,$$

где d — расстояние между зеркалами. Таким образом,

$$2\pi\frac{2d\sin\theta}{\lambda} = 2\pi(q+1),$$

так что

$$\frac{2\pi}{\lambda}2d\sin\theta = 2\pi m, \quad m = 1, 2, \tag{8.1}$$

где $m = q + 1$. Таким образом, условие самосогласования выполняется только для определенных углов наклона θ_m, удовлетворяющих условию

$$\boxed{\sin\theta_m = m\frac{\lambda}{2d}, \quad m = 1, 2, \ldots.} \tag{8.2}$$

Углы наклона

Каждому целому m соответствует угол θ_m, а соответствующее поле называется m-й модой. Мода с $m = 1$ имеет наименьший угол

$$\theta_1 = \arcsin\frac{\lambda}{2d};$$

моды с бóльшими m состоят из плоских волн, имеющих больший наклон по отношению к оси z.

При выполнении условия самосогласования фазы плоских волн, идущих вниз и вверх, в точках оси z отличаются на половину фазового сдвига при полном обходе $q\pi$, $q = 0, 1, 2,$, или $(m-1)\pi$, $m = 1, 2,$ так что волны складываются при нечетном m и вычитаются при четном m.

Поскольку y — составляющая постоянной распространения равна $k_y = nk_0\sin\theta$, она приобретает дискретные значения

$$k_{ym} = nk_0\sin\theta_m = \frac{2\pi}{\lambda}\sin\theta_m.$$

Используя (8.2), получаем

$$\boxed{k_{ym} = m\frac{\pi}{d},} \quad m = 1, 2, 3, \tag{8.3}$$

Поперечная компонента волнового вектора

так что k_{ym} отстоят друг от друга на π/d. Уравнение (8.3) утверждает, что фазовый сдвиг при распространении волны на расстояние $2d$ (один полный обход) в направлении y с постоянной распространения k_{ym} должен быть равен целому числу 2π.

Постоянные распространения

Направляемая волна состоит из двух различных плоских волн, распространяющихся в плоскости y—z под углами $\pm\theta$ к оси z. Их волновые векторы имеют компоненты $(0, k_y, k_z)$ и $(0, -k_y, k_z)$. Сумма или разность волн, следовательно, меняется вдоль оси z как $\exp(-jk_z z)$, так что постоянная распространения направляемой волны есть $\beta \equiv k_z = k\cos\theta$. Таким образом, β принимает дискретные значения $\beta_m = k\cos\theta_m$, откуда

$$\beta_m^2 = k^2\left(1 - \sin^2\theta_m\right).$$

С помощью (8.2) получаем

$$\boxed{\beta_m^2 = k^2 - \frac{m^2\pi^2}{d^2}.} \qquad (8.4)$$

Постоянные распространения

Моды более высокого порядка (более наклонные) распространяются с меньшими скоростями. Значения θ_m, k_{ym} и β_m для различных мод показаны на рис. 8.5.

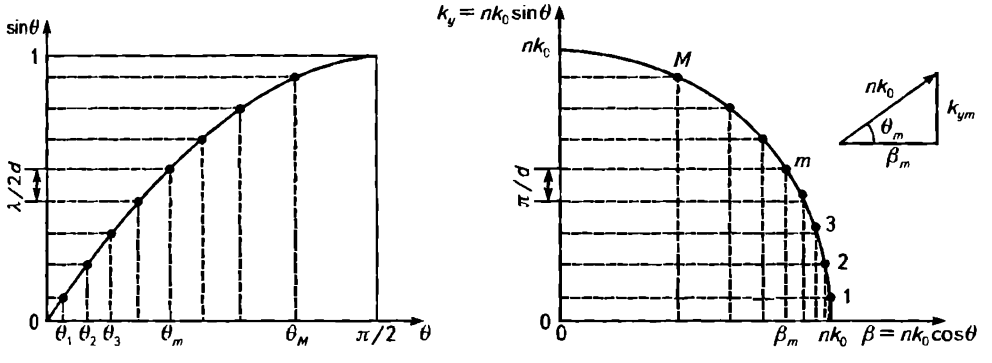


Рис. 8.5. Углы наклона θ_m и компоненты волнового вектора для мод планарного зеркального волновода (показаны точками). Дискретные значения поперечной компоненты $k_{ym} = k\sin\theta_m$ расположены равномерно с интервалом π/d, а углы наклона θ_m и постоянные распространения β_m — неравномерно. Мода $m = 1$ имеет наименьший угол наклона и наибольшую постоянную распространения

Распределения поля

Комплексная амплитуда полного поля в волноводе есть суперпозиция двух многократно отражающихся плоских ТЕМ-волн. Если $A_m \exp(-jk_{ym}y - j\beta_m z)$ — волна, бегущая вверх, то волна, бегущая вниз, будет $e^{j(m-1)\pi}A_m \exp(+jk_{ym}y - j\beta_m z)$

[при $y = 0$ разность фаз этих волн равна $(m - 1)\pi$]. Таким образом, существуют симметричные моды, для которых две плоские волны складываются, и антисимметричные моды, для которых они вычитаются. Полное поле оказывается равным $E_x(y, z) = 2A_m \cos(k_{ym} y) \exp(-j\beta_m z)$ для нечетных мод и $2jA_m \sin(k_{ym} y) \exp(-j\beta_m z)$ для четных мод.

Используя (8.3), запишем комплексную амплитуду электрического поля в виде

$$E_x(y, z) = a_m u_m(y) \exp(-j\beta_m z), \tag{8.5}$$

где

$$u_m(y) = \begin{cases} \sqrt{\dfrac{2}{d}} \cos\left(m\pi \dfrac{y}{d}\right), & m = 1, 3, 5, \\ \sqrt{\dfrac{2}{d}} \sin\left(m\pi \dfrac{y}{d}\right), & m = 2, 4, 6, \end{cases} \tag{8.6}$$

Здесь $a_m = \sqrt{2d} A_m$ — для нечетных и $a_m = j\sqrt{2d} A_m$ — для четных m соответственно. Функции $u_m(y)$ нормированы условием

$$\int_{-d/2}^{d/2} u_m^2(y)\,dy = 1. \tag{8.7}$$

Таким образом, a_m — амплитуда моды m. Можно показать, что функции $u_m(y)$ также удовлетворяют условию

$$\int_{-d/2}^{d/2} u_m(y) u_l(y)\,dy = 0, \quad m \neq l, \tag{8.8}$$

т. е. они ортогональны на интервале $[-d/2, d/2]$.

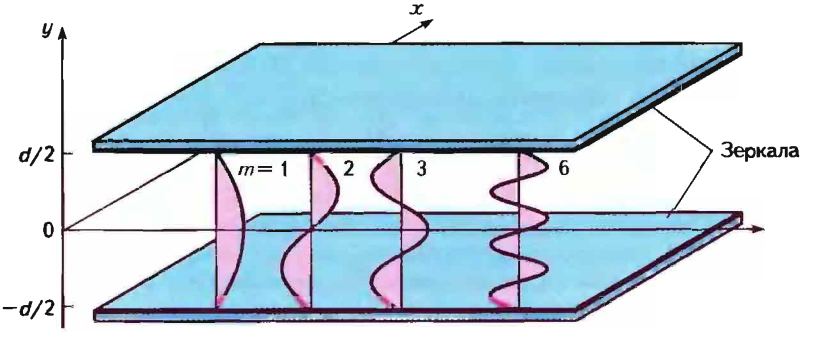


Рис. 8.6. Распределения поля в модах планарного зеркального волновода

Поперечные распределения $u_m(y)$ изображены графически на рис. 8.6. Каждая мода представляет собой волну, стоячую в направлении y и распространяющуюся в направлении z. Моды с большими m меняются в поперечной плос-

кости с большей частотой k_y и распространяются с меньшей постоянной β. Поле обращается в нуль при $y = \pm d/2$ для всех мод, так что граничные условия на поверхности зеркал всегда выполняются.

Поскольку с самого начала мы предположили, что исходная TEM-волна поляризована в направлении x, то полное электрическое поле также направлено вдоль этой оси, и направляемая волна есть поперечная электрическая (TE) волна. Поперечная магнитная (TM) волна анализируется таким же образом, как будет видно из дальнейшего изложения.

Упражнение 8.1

Оптическая энергия

Покажите, что поток энергии света в направлении z, создаваемый модой TE $E_x(y, z) = a_m u_m(y) \exp(-j\beta_m z)$, равен $(|a_m|^2/2\eta)\cos\theta_m$, где $\eta = \eta_0/n$; η_0 — импеданс свободного пространства,

$$\eta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}.$$

Число мод

Поскольку $\sin\theta_m = m\lambda/2d$, $m = 1, 2, \ldots$, то с учетом того, что $\sin\theta_m < 1$, максимальное возможное значение m — это наибольшее целое число, меньшее $1/(\lambda/2d)$, т. е.

$$\boxed{M = \frac{2d}{\lambda}.} \qquad (8.9)$$

Число мод

Символ «$=$» означает, что $2d/\lambda$ уменьшается до ближайшего целого числа. Например, при $2d/\lambda = 0{,}9$; 1 и 1,1 имеем 0; 0 и 1 соответственно. Таким образом, M — число мод волновода. Свет может распространяться в волноводе в виде одной, двух или большего числа мод. Реальное число мод, переносящих энергию, зависит от источника возбуждения, но максимальное число равно M.

Число мод возрастет с ростом отношения расстояния между зеркалами к длине волны. При условии $2d/\lambda \le 1$, т. е. $d \le \lambda/2$, M равно нулю, что означает, что условия самосогласования никогда не выполняются и волновод не поддерживает ни одной моды. Длина волны $\lambda_c = 2d$ называется **длиной волны отсечки** волновода. Это наибольшая длина волны, которую может направлять данная структура. Эта длина волны соответствует **частоте отсечки**

$$\boxed{\nu_c = \frac{c}{2d}} \qquad (8.10)$$

Частота отсечки

или угловой частоте отсечки $\omega_c = 2\pi\nu_c = \pi c/d$ — наименьшей частоте света, который может направляться данным волноводом. Если $1 < 2d/\lambda \le 2$ (т. е. $d \le \lambda < 2d$

или $\nu_c \le \nu < 2\nu_c$), разрешена только одна мода. Такая структура называется **одномодовым волноводом**. Например, при $d = 5$ мкм волновод имеет длину волны отсечки $\lambda_c = 10$ мкм, поддерживает одну моду при $5 \le \lambda < 10$ мкм и большее число мод при $\lambda < 5$ мкм. Уравнение (8.9) можно записать для частоты ν,

$$M \doteq \frac{\nu}{\nu_c} = \frac{\omega}{\omega_c},$$

так что число мод возрастает на единицу, когда угловая частота ω увеличивается на ω_c, как показано на рис. 8.7, *а*.

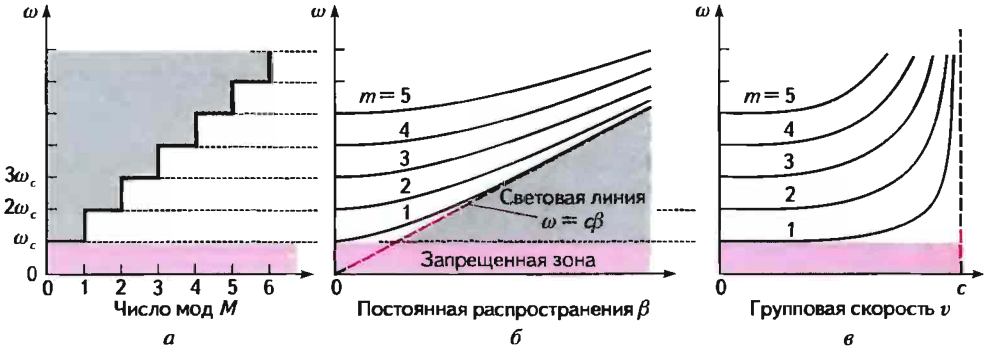


Рис. 8.7. Число мод M как функция угловой частоты ω (*а*). Любые моды запрещены ниже частоты отсечки $\omega_c = \pi c/d$. M увеличивается на единицу, когда ω получает приращение ω_c. Дисперсионное соотношение (*б*). Запрещенная зона существует для угловых частот ниже частоты отсечки ω_c. Групповые скорости мод как функция угловой частоты (*в*)

Дисперсионное соотношение

Связь между постоянной распространения β и угловой частотой ω — важная характеристика волновода, известная как **дисперсионное соотношение**. Для однородной среды дисперсионное соотношение имеет простой вид $\omega = c\beta$. Для моды m планарного зеркального волновода β_m и ω связаны выражением (8.4), поэтому

$$\beta_m^2 = \left(\frac{\omega}{c}\right)^2 - \frac{m^2\pi^2}{d^2}. \tag{8.11}$$

Это соотношение можно записать через угловую частоту отсечки $\omega_c = 2\pi\nu_c = \pi c/d$ как

$$\boxed{\beta_m = \frac{\omega}{c}\sqrt{1 - m^2\frac{\omega_c^2}{\omega^2}}.} \tag{8.12}$$

Дисперсионное соотношение

Как показано на рис. 8.7, *б*, для $m = 1, 2,$ постоянная распространения β для моды m равна нулю при угловой частоте $\omega = m\omega_c$, монотонно возрастает с увеличением частоты и в конечном счете приближается к линейному соотношению $\beta = \omega/c$ для достаточно больших значений β.

Групповые скорости

В среде с заданным дисперсионным соотношением между ω и β импульс (волновой пакет) света с центральной частотой ω распространяется со скоростью $v = \mathrm{d}\omega/\mathrm{d}\beta$, которая называется *групповой скоростью* (см. разд. 5.6). Дифференцируя (8.12) и предполагая, что c не зависит от ω (т. е. пренебрегая дисперсией материала волновода), получаем

$$2\beta_m \frac{\mathrm{d}\beta_m}{\mathrm{d}\omega} = \frac{2\omega}{c^2},$$

так что

$$\frac{\mathrm{d}\omega}{\mathrm{d}\beta_m} = c^2 \frac{\beta_m}{\omega} = c^2 k \frac{\cos\theta_m}{\omega} = c\cos\theta_m,$$

откуда групповая скорость моды равна

$$\boxed{v_m = c\cos\theta_m = c\sqrt{1 - m^2\frac{\omega_c^2}{\omega^2}}.} \qquad (8.13)$$

Групповая скорость

Отсюда следует, что более наклонные моды распространяются с меньшей групповой скоростью, что понятно, поскольку они проходят больший зигзагообразный путь. Зависимость групповой скорости от угловой частоты показана на рис. 8.7, *в*, где видно, что для каждой моды групповая скорость монотонно растет от 0 до c по мере того, как частота увеличивается от частоты отсечки до бесконечности.

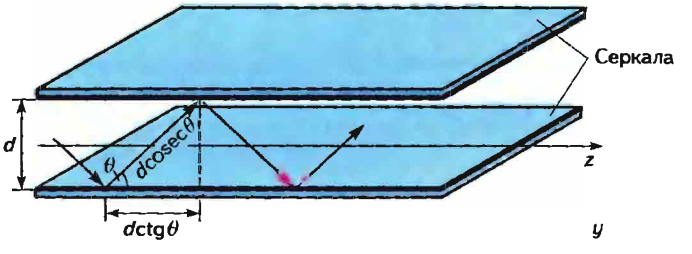


Рис. 8.8. Плоская волна, отражающаяся под углом θ, проходит вдоль z расстояние $d\,\mathrm{ctg}\,\theta$ за время $d\,\mathrm{cosec}\,\theta/c$. Скорость вдоль равна $c\cos\theta$

Равенство (8.13) можно получить геометрическим путем, рассматривая последовательные отражения плоской волны между зеркалами и рассчитывая время перемещения вдоль в ходе ее зигзагообразного распространения. На пути от зеркала до зеркала (рис. 8.8) имеем:

$$v = \frac{\text{расстояние}}{\text{время}} = \frac{d\,\mathrm{ctg}\,\theta}{d\,\mathrm{cosec}\,\theta/c} = c\cos\theta. \qquad (8.14)$$

ТМ-моды

До сих пор мы рассматривали только ТЕ-моды с электрическим полем, направленным по оси x. В зеркальном волноводе могут распространяться и ТМ-моды, у которых вдоль x направлено магнитное поле. Для их исследования рассмотрим плоскую ТЕМ-волну с магнитным вектором вдоль оси x, распространяющуюся под углом θ к поверхности зеркал и отражающуюся от них (рис. 8.9). Тогда электрическое поле имеет составляющие в направлениях y и z. Поскольку z-компонента параллельна зеркалам, она ведет себя в точности так же, как и x компонента в случае моды ТЕ (т. е. приобретает фазовый сдвиг на π при отражении и обращается в нуль на поверхности зеркал). Запись условия самосогласования для этой компоненты приводит к результату, полностью повторяющему случай ТЕ-моды. Углы θ, поперечные компоненты волнового вектора k_y и постоянные распространения β ТМ-мод, связанные с этой компонентой, идентичны найденным для ТЕ-мод. Существует число ТМ-мод [см. (8.9)] (полное число мод $2M$), поддерживаемых волноводом.

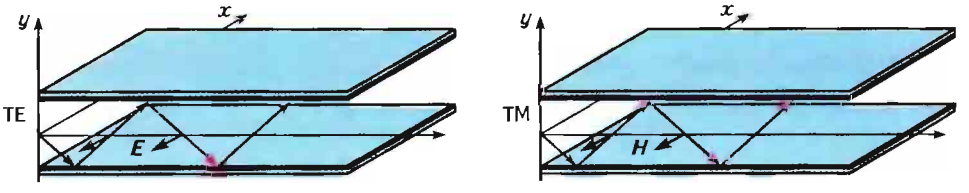


Рис. 8.9. Направляемые моды ТЕ и ТМ

Комплексная амплитуда z-компоненты электрического поля моды m, как и раньше, равна сумме волн, бегущих вверх $A_m \exp(-jk_{ym}y - j\beta_m z)$ и вниз $e^{j(m-1)\pi}A_m \exp(+jk_{ym}y - j\beta_m z)$ с одинаковыми амплитудами и фазовым сдвигом $(m-1)\pi$, так что

$$E_z(y, z) = \begin{cases} a_m\sqrt{\frac{2}{d}}\cos\left(m\pi\frac{y}{d}\right)\exp(-j\beta_m z), & m = 1,\ 3,\ 5, \\ a_m\sqrt{\frac{2}{d}}\sin\left(m\pi\frac{y}{d}\right)\exp(-j\beta_m z), & m = 2,\ 4,\ 6, \end{cases} \tag{8.15}$$

где $a_m = \sqrt{2d}A_m$ — для нечетных и $a_m = j\sqrt{2d}A_m$ — для четных m соответственно. Поскольку вектор электрического поля плоской ТЕМ-волны перпендикулярен направлению распространения, он образует угол $(\pi/2 + \theta_m)$ с осью z для волны, бегущей вверх, и $(\pi/2 - \theta_m)$ для волны, бегущей вниз.

Составляющие электрического поля этих волн равны

$$\begin{gathered} A_m \operatorname{ctg}\theta_m \exp(-jk_{ym}y)\exp(-j\beta_m z); \\ e^{j\pi m}A_m \operatorname{ctg}\theta_m \exp(jk_{ym}y)\exp(-j\beta_m z), \end{gathered} \tag{8.16}$$

так что

$$E_y(y,\ z) = \begin{cases} a_m\sqrt{\dfrac{2}{d}}\ \mathrm{ctg}\,\theta_m \cos\left(m\pi\dfrac{y}{d}\right)\exp(-j\beta_m z), & m = 1,\ 3,\ 5, \\ a_m\sqrt{\dfrac{2}{d}}\ \mathrm{ctg}\,\theta_m \sin\left(m\pi\dfrac{y}{d}\right)\exp(-j\beta_m z), & m = 2,\ 4,\ 6, \end{cases} \tag{8.17}$$

Выполнение граничных условий гарантировано, так как $E_z(y, z)$ обращается в нуль на зеркалах. Составляющую магнитного поля $H_x(y, z)$ можно определить аналогичным образом, если заметить, что отношение электрического и магнитного полей ТЕМ-волны есть импеданс среды η. Естественно, что получающиеся в результате поля $E_y(y, z)$, $E_z(y, z)$ и $H_x(y, z)$ удовлетворяют уравнениям Максвелла.

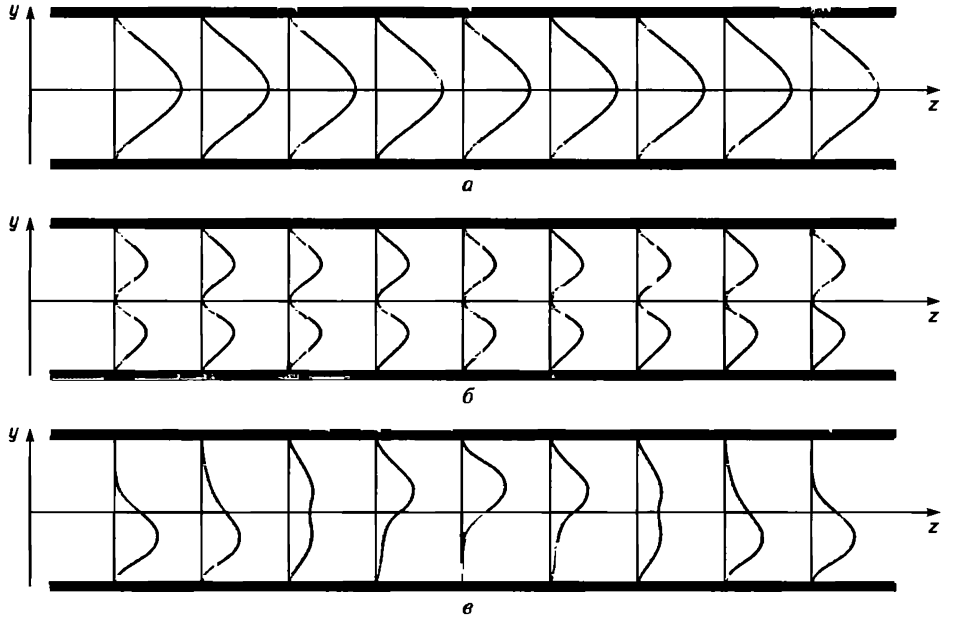


Рис. 8.10. Изменение распределения интенсивности в поперечном направлении при распространении вдоль z:

а — комплексная амплитуда моды 1 равна $E(y, z) = u_1(y)\exp(-j\beta_1 z)$, где $u_1(y) = \sqrt{2/d}\cos(\pi y/d)$. Интенсивность не меняется вдоль оси z; *б* — комплексная амплитуда моды 2 равна $E(y, z) = u_2(y)\exp(-j\beta_2 z)$, где $u_2(y) = \sqrt{2/d}\sin(2\pi y/d)$. Интенсивность не меняется вдоль оси z; *в* — комплексная амплитуда суперпозиции мод 1 и 2 равна $E(y, z) = u_1(y)\exp(-j\beta_1 z) + u_2(y)\exp(-j\beta_2 z)$. Поскольку $\beta_1 \neq \beta_2$, распределение интенсивности меняется вдоль оси z

Многомодовые поля

Чтобы свет распространялся в волноводе, не обязательно, чтобы он имел вид одной из мод. На самом деле поле, удовлетворяющее граничным условиям (исчезающее на зеркалах), но в остальном имеющее совершенно про-

извольное распределение в поперечной плоскости, *может* направляться волноводом. При этом, однако, его энергия делится между модами. Поскольку разные моды имеют разные постоянные распространения и распространяются с разными групповыми скоростями, поперечное распределение поля в процессе распространения по волноводу меняется. На рис. 8.10 показано, как поперечное распределение сохраняется у каждой из мод и меняется у многомодового поля в зависимости от z (показано распределение интенсивности)

Произвольное поле, поляризованное вдоль оси и удовлетворяющее граничным условиям, можно записать в виде взвешенной суперпозиции ТЕ-мод

$$E_x(y,z) = \sum_{m=0}^{M} a_m u_m(y) \exp(-j\beta_m z), \tag{8.18}$$

где a_m — амплитуды различных мод.

Упражнение 8.2

Оптическая мощность в многомодовом поле

Покажите, что мощность, переносимая в направлении z многомодовым полем (8.18), равна сумме мощностей $(|a_m|^2/2\eta)\cos\theta_m$, переносимых отдельными модами.

8.2. ПЛАНАРНЫЕ ДИЭЛЕКТРИЧЕСКИЕ ВОЛНОВОДЫ

Планарный диэлектрический волновод представляет собой слой диэлектрического материала, окруженный средами с более низким показателем преломления. Свет направляется внутри слоя за счет полного внутреннего отражения. В тонкопленочных устройствах внутренний слой называется пленкой, а окружение — подложкой и покрытием. Внутренняя и внешняя части называются также сердцевиной и оболочкой волновода соответственно. В данном разделе мы будем рассматривать распространение света в симметричном планарном диэлектрическом волноводе, состоящем из слоя диэлектрика толщины d с показателем преломления n_1 в оболочке с меньшим показателем преломления n_2, как показано на рис. 8.11. Предполагаем, что все материалы не имеют потерь.

Лучи света, образующие с осью z углы θ в плоскости y—z, испытывают многократное полное внутреннее отражение на границах сердцевины и оболочки при условии, что θ не превышает угла, дополнительного к критическому

$$\bar{\theta}_{\text{кр}} = \frac{\pi}{2} - \arcsin\left(\frac{n_2}{n_1}\right) = \arccos\left(\frac{n_2}{n_1}\right)$$

[см. (1.8) и рис. 6.29 и 6.31]. Они распространяются в направлении z, отражаясь от поверхностей сердцевины без потерь энергии. Лучи с большими углами отклонения частично отражаются, частично преломляются, теряют при каждом отражении часть своей энергии и в конце концов затухают.

Для нахождения мод волновода можно использовать формальный подход, состоящий в поиске решения уравнений Максвелла в сердцевине и оболочке с последующим наложением граничных условий на поверхности раздела (см. задачу 1 к разд. 8.2). Вместо этого мы запишем решение в виде суммы плоских ТЕМ-волн, отражающихся от границ раздела и распространяющихся в сердцевине. Накладывая условия самосогласования, мы определим соответствующие углы наклона, затем постоянные распространения, пространственные распределения полей и групповые скорости. Такой анализ аналогичен проведенному в предыдущем разделе для планарного зеркального волновода.

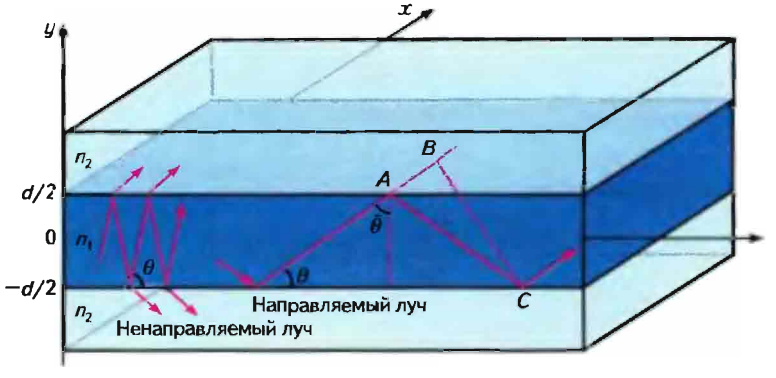


Рис. 8.11. Планарный диэлектрический волновод. Лучи с углами отклонения $\theta < \bar{\theta}_{кр} = \arccos(n_2/n_1)$ от оси z не выходят за границы сердцевины по причине полного внутреннего отражения

8.2.1. Волноводные моды

Предположим, что поле в сердцевине имеет вид монохроматической плоской ТЕМ-волны с длиной волны $\lambda = \lambda_0/n_1$, многократно отражающейся от границ под углом θ, меньшим, чем угол $\bar{\theta}_{кр}$, дополнительный к критическому. Волна распространяется с фазовой скоростью $c_1 = c_0/n_1$, имеет волновое число n_1k_0 и составляющие волнового вектора $k_x = 0$, $k_y = n_1k_0 \sin\theta$ и $k_z = n_1k_0 \cos\theta$. Для нахождения собственных мод наложим условие самосогласования, т. е. потребуем, чтобы волна воспроизводила саму себя после каждого полного обхода.

За один обход дважды отраженная волна отстает от исходной волны на расстояние $\overline{AC} - \overline{AB} = 2d\sin\theta$, как показано на рис. 8.4. Каждое полное внутреннее отражение на границе раздела диэлектриков вносит дополнительный сдвиг фазы φ_r (см. разд. 6.2). Для самосогласования полная разность фаз двух волн должна быть равна нулю или целому кратному 2π

$$\frac{2\pi}{\lambda} 2d\sin\theta - 2\varphi_r = 2\pi m, \quad m = 0,\ 1,\ 2, \tag{8.19}$$

или

$$2k_y d - 2\varphi_r = 2\pi m. \tag{8.20}$$

Единственное отличие от соответствующих условий (8.1) или (8.3) для зеркального волновода состоит в том, что фазовый сдвиг π при отражении от идеального зеркала нужно заменить фазовым сдвигом φ_r, который вносится при отражении на границе раздела диэлектриков.

Фазовый сдвиг при отражении φ_r является функцией угла θ. Он зависит также от поляризации падающей волны, TE или TM. В случае TE-поляризации (электрическое поле вдоль оси x), подставляя $\theta_1 = \pi/2 - \theta$ и $\theta_{кр} = \pi/2 - \bar{\theta}_{кр}$ в (6.37), находим

$$\operatorname{tg}\frac{\varphi_r}{2} = \sqrt{\frac{\sin^2\bar{\theta}_{кр}}{\sin^2\theta} - 1}, \tag{8.21}$$

так что φ_r меняется от π до 0, когда θ меняется от 0 до $\bar{\theta}_{кр}$. Переписывая (8.19) в виде

$$\operatorname{tg}\left(\frac{\pi d \sin\theta}{\lambda} - \frac{m\pi}{2}\right) = \operatorname{tg}\left(\frac{\varphi_r}{2}\right)$$

и используя (8.21), получаем

$$\boxed{\operatorname{tg}\left(\pi\frac{d}{\lambda}\sin\theta - m\frac{\pi}{2}\right) = \sqrt{\frac{\sin^2\bar{\theta}_{кр}}{\sin^2\theta} - 1}.} \tag{8.22}$$

Условие самосогласования (TE-мода)

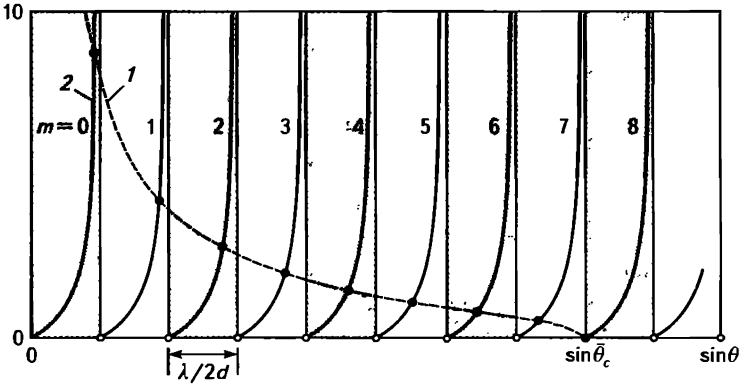


Рис. 8.12. Графическое решение уравнения (8.22) для определения углов наклона θ_m мод планарного диэлектрического волновода. Правая (*1*) и левая (*2*) части уравнения (8.22) показаны графически как функции $\sin\theta$. Точки пересечения, отмеченные черными кружками, определяют значения $\sin\theta_m$. Каждая ветвь функций tg или ctg в левой части соответствует моде. На этом графике $\sin\bar{\theta}_{кр} = 8(\lambda/2d)$, а число мод равно $M = 9$. Светлые кружки соответствуют $\sin\theta_m = m\lambda/2d$, что дает углы наклона мод планарного зеркального волновода тех же размеров

Трансцендентное уравнение (8.22) относительно одного неизвестного $\sin\theta$. Его решения дают углы наклона мод θ_m. Поучительно провести графическое решение. На рис. 8.12 приведены графики зависимости правой и левой частей

уравнения (8.22) от $\sin\theta$. Решения даются точками пересечения графиков. Правая часть (*1*): $\operatorname{tg}(\varphi_r/2)$ — монотонно убывающая функция $\sin\theta$, достигающая 0 при $\sin\theta = \sin\bar{\theta}_{кр}$. Левая часть (*2*) порождает два семейства кривых: $\operatorname{tg}[(\pi d/\lambda)\sin\theta]$ при четном и $\operatorname{ctg}[(\pi d/\lambda)\sin\theta]$ при нечетном m, соответственно. Точки пересечения дают значения θ_m для мод. Углы наклона мод *зеркального* волновода с расстоянием d между зеркалами можно получить из этих же графиков при $\varphi_r = \pi$ или, что равносильно, $\operatorname{tg}(\varphi_r/2) = \infty$. Соответствующие точки показаны на рис. 8.12 светлыми кружками для сравнения.

Углы θ_m лежат между 0 и $\bar{\theta}_{кр}$. Они соответствуют волновым векторам с компонентами $(0, n_1k_0\sin\theta_m, n_1k_0\cos\theta_m)$. Компоненты z являются постоянными распространения

$$\boxed{\beta_m = n_1k_0\cos\theta_m.} \qquad (8.23)$$

Постоянные распространения

Поскольку $\cos\theta_m$ лежит между 1 и $\cos\bar{\theta}_{кр} = n_2/n_1$, β_m заключено между n_2k_0 и n_1k_0, как показано на рис. 8.13.

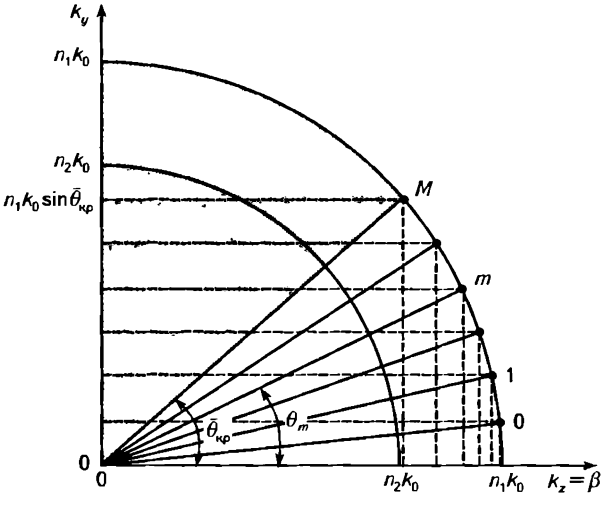


Рис. 8.13. Углы наклона θ_m и соответствующие компоненты k_z и k_y волнового вектора направляемых мод показаны точками. Углы θ_m лежат между 0 и $\bar{\theta}_{кр}$, а постоянные распространения β_m — между n_2k_0 и n_1k_0. Эти результаты нужно сравнить с показанными на рис. 8.5 для планарного зеркального волновода

Для ТМ-мод углы наклона θ_m и постоянные распространения β_m можно найти из того же уравнения (8.19), но с фазовым сдвигом φ_r вида (6.39). Получаются аналогичные результаты.

Число мод

Для определения числа ТЕ-мод, поддерживаемых диэлектрическим волноводом, изучим графики на рис. 8.12. Ось абсцисс разделена на равные отрезки длиной $\lambda/2d$, каждый из которых содержит моду, отмеченную черным кружком. Это верно для углов, удовлетворяющих условию $\sin\theta \le \sin\bar{\theta}_{кр}$. Следовательно, число ТЕ-мод — это наименьшее целое, большее $\sin\bar{\theta}_{кр}/(\lambda/2d)$:

$$M \doteq \frac{\sin\bar{\theta}_{кр}}{\lambda/2d}. \qquad (8.24)$$

Символ «≐» означает, что $\sin \bar{\theta}_{кр}/(\lambda/2d)$ увеличивается до ближайшего целого числа. Например, если эта величина равна 0,9; 1 или 1,1, то $M = 1$, 2 и 2 соответственно. Подставляя в (8.24), получаем

$$\boxed{M \doteq \frac{2d}{\lambda_0}\text{NA},} \qquad (8.25)$$

Число ТЕ-мод

где

$$\boxed{\text{NA} = \sqrt{n_1^2 - n_2^2}} \qquad (8.26)$$

Числовая апертура

есть числовая апертура волновода (синус угла приема лучей из воздуха в слой, см. упражнение 1.8). Аналогичное выражение можно получить для ТМ-мод. Если, например, $d/\lambda_0 = 10$, $n_1 = 1{,}47$ и $n_2 = 1{,}46$, то $\bar{\theta}_{кр} = 6{,}7°$, NA = 0,171 и $M = 4$ (число ТЕ-мод).

Если $\lambda/2d > \sin \bar{\theta}_{кр}$ или $(2d/\lambda_0)\text{NA} < 1$, то существует всего одна мода. Волновод в этом случае является **одномодовым**. Это достигается, когда либо сердцевина достаточно тонкая, либо длина волны достаточно большая. В отличие от зеркального волновода диэлектрический волновод не имеет абсолютной частоты (или длины волны) отсечки. В диэлектрическом волноводе всегда существует, по крайней мере, одна мода ТЕ, поскольку фундаментальная мода $m = 0$ всегда разрешена. Однако каждая из мод $m = 1$, 2, имеет свою длину волны отсечки.

В частотных переменных условие одномодовости состоит в том, что $\nu > \nu_c$ или $\omega > \omega_c$, где частота отсечки моды равна

$$\boxed{\nu_c = \frac{\omega_c}{2\pi} = \frac{1}{\text{NA}}\frac{c_0}{2d}.} \qquad (8.27)$$

Частота отсечки моды

Число мод тогда равно $M \doteq \nu/\nu_c = \omega/\omega_c$ (рис. 8.14). Число M увеличивается на единицу при увеличении частоты на ω_c. Такие же результаты получаются для мод ТМ.

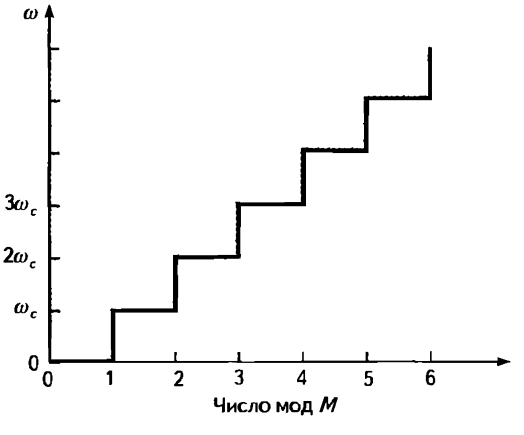


Рис. 8.14. Число ТЕ-мод как функция частоты. Сравните с рис. 8.7, *а* для планарного зеркального волновода. В диэлектрическом волноводе запрещенной зоны нет

Пример 8.1

Моды в волноводе из AlGaAs

Волновод образован слоем $Al_xGa_{1-x}As$ между двумя слоями $Al_yGa_{1-y}As$. Показателями преломления структуры можно управлять, меняя концентрации x и y Al. Если эти концентрации выбрать так, чтобы на рабочей длине волны $\lambda_0 = 0{,}9$ мкм $n_1 = 3{,}5$ и $n_1 - n_2 = 0{,}05$, то при толщине $d = 10$ мкм имеется $M = 14$ ТЕ-мод. Для $d < 0{,}76$ мкм разрешена только одна мода.

8.2.2. Распределения полей

Определим теперь распределения полей для ТЕ-мод.

Внутреннее поле

Поле внутри слоя состоит из двух плоских ТЕМ-волн, бегущих под углами θ_m и $-\theta_m$ к оси z с компонентами волновых векторов $(0, \pm n_1 k_0 \sin\theta_m, n_1 k_0 \cos\theta_m)$. Они имеют одинаковые амплитуды и фазовый сдвиг $m\pi$ (половину сдвига за полный обход) в центре волновода. Следовательно, комплексная амплитуда электрического поля равна

$$E_x(y, z) = a_m u_m(y) \exp(-j\beta_m z),$$

где β_m — постоянная распространения; a_m — постоянная;

$$u_m(y) \propto \begin{cases} \cos\left(2\pi \dfrac{\sin\theta_m}{\lambda} y\right), & m = 0,\ 2,\ 4, \\ \sin\left(2\pi \dfrac{\sin\theta_m}{\lambda} y\right), & m = 1,\ 3,\ 5, \end{cases} \qquad -\frac{d}{2} \le y \le \frac{d}{2}; \tag{8.28}$$

$$\lambda = \frac{\lambda_0}{n_1}.$$

Заметим, что хотя поле гармоническое, оно не исчезает на границах слоя. По мере роста m растет и $\sin\theta_m$, поэтому моды высших порядков быстрее меняются вдоль y.

Внешнее поле

Внешнее поле должно быть согласовано с внутренним во всех точках границы $y = \pm d/2$. Значит, вдоль z оно должно меняться как $\exp(-j\beta_m z)$. Подставляя $E_x(y, z) = a_m u_m(y) \exp(-j\beta_m z)$ в уравнение Гельмгольца

$$\left(\nabla^2 + n_2^2 k_0^2\right) E_x(y, z) = 0,$$

получаем

$$\frac{d^2u_m}{dy^2} - \gamma_m^2 u_m = 0, \tag{8.29}$$

где

$$\gamma_m^2 = \beta_m^2 - n_2^2 k_0^2. \tag{8.30}$$

Поскольку для направляемых мод $\beta_m > n_2 k_0$ (см. рис. 8.13), $\gamma_m^2 > 0$, так что уравнению (8.29) удовлетворяют экспоненциальные функции $\exp(-\gamma_m y)$ и $\exp(\gamma_m y)$. Поскольку поле должно затухать вне сердцевины, выберем $\exp(-\gamma_m y)$ в верхней среде и $\exp(\gamma_m y)$ в нижней среде

$$u_m(y) \propto \begin{cases} \exp(-\gamma_m y), & y > \dfrac{d}{2}; \\ \exp(\gamma_m y), & y < -\dfrac{d}{2}. \end{cases} \tag{8.31}$$

Скорость затухания γ_m называется **коэффициентом экстинкции**, а волна называется **нераспространяющейся**. Подставляя

$$\beta_m = n_1 k_0 \cos\theta_m; \quad \cos\bar{\theta}_c = \frac{n_2}{n_1}$$

в (8.30), получаем

$$\boxed{\gamma_m = n_2 k_0 \sqrt{\frac{\cos^2\theta_m}{\cos^2\bar{\theta}_c} - 1.}} \tag{8.32}$$

Коэффициент экстинкции

С ростом номера моды m величина θ_m растет, а γ_m убывает. Поэтому моды более высокого порядка глубже проникают в подложку и покрытие.

Для определения коэффициента пропорциональности в (8.28) и (8.31) приравняем внутреннее и внешнее поля при $y = d/2$ и используем нормировку

$$\int_{-\infty}^{\infty} u_m^2(y)\,dy = 1. \tag{8.33}$$

Это дает выражение для $u_m(y)$, верное для всех y. Такие функции показаны на рис. 8.15. Как и в зеркальном волноводе, все $u_m(y)$ ортогональны, т. е.

$$\int_{-\infty}^{\infty} u_m(y)u_l(y)\,dy = 0, \quad l \neq m. \tag{8.34}$$

Произвольное ТЕ-поле в диэлектрическом волноводе можно записать как суперпозицию этих мод

$$E_x(y, z) = \sum_m a_m u_m(y)\exp(-j\beta_m z), \tag{8.35}$$

где a_m — амплитуда моды m.

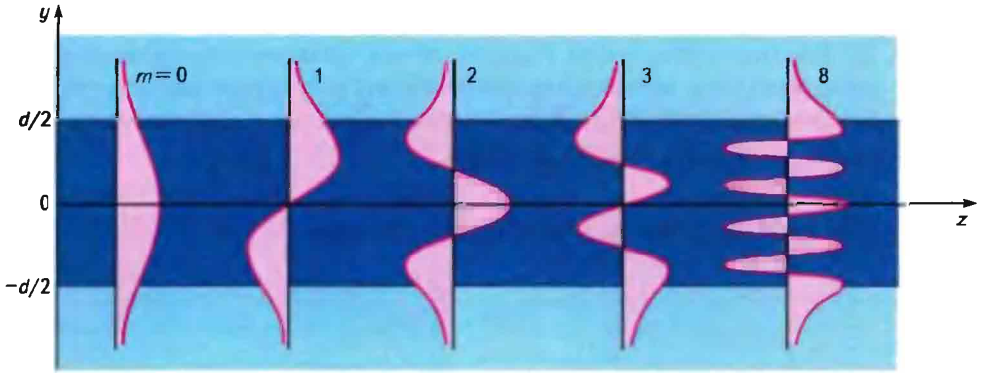


Рис. 8.15. Распределения поля для направляемых ТЕ-мод в диэлектрическом волноводе. Эти результаты следует сравнить с рис. 8.6 для планарного зеркального волновода

Упражнение 8.3

Фактор оптического ограничения

Фактором оптического ограничения называется отношение мощности внутри слоя к полной мощности

$$\Gamma_m = \frac{\int_0^{d/2} u_m^2(y)\,\mathrm{d}y}{\int_0^{\infty} u_m^2(y)\,\mathrm{d}y}. \tag{8.36}$$

Выведите выражение для Γ_m как функции угла θ_m и отношения d/λ. Покажите, что мода наинизшего порядка (с наименьшим θ_m) имеет наибольший фактор оптического ограничения.

Распределения полей в модах ТМ можно найти аналогичным образом (рис. 8.16). Поскольку магнитное поле параллельно границе слоя, компонента z электрического поля ведет себя так же, как компонента x электрического поля у ТЕ-волны. Анализ можно начать с определения $E_z(y, z)$. Используя свойства составляющих ТЕМ-волн, легко найти другие компоненты $E_y(y, z)$ и $H_x(y, z)$, как это было сделано для зеркального волновода. Альтернативным способом определения этих полей является использование уравнений Максвелла.

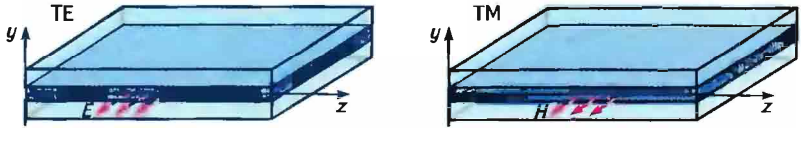
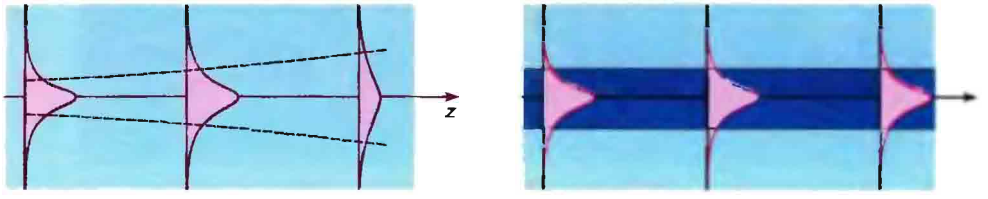


Рис. 8.16. ТЕ- и ТМ-моды в диэлектрическом планарном волноводе

Распределение поля в ТЕ-моде с $m = 0$ по форме напоминает гауссов пучок (см. гл. 3). Однако, в отличие от гауссова пучка, направляемый свет не испытывает расходимости в поперечном направлении в процессе распространения вдоль оси волновода (рис. 8.17). В волноводе тенденция к дифракции компенсируется направляющим действием среды.



а *б*

Рис. 8.17. Гауссов пучок в однородной среде (*а*): направляемая мода в диэлектрическом волноводе (*б*)

8.2.3. Дисперсионные соотношения и групповые скорости

Дисперсионное соотношение (зависимость ω от β) получается, если выразить через ω и β условие самосогласования (8.20). Поскольку

$$k_y^2 = \left(\frac{\omega}{c_1}\right)^2 - \beta^2,$$

из (8.20) получается

$$2d\sqrt{\frac{\omega^2}{c_1^2} - \beta^2} = 2\varphi_r + 2\pi m. \tag{8.37}$$

Так как

$$\cos\theta = \frac{\beta}{\omega/c_1}; \quad \cos\theta_{\text{кр}} = \frac{n_2}{n_1} = \frac{c_1}{c_2},$$

формула (8.21) принимает вид

$$\operatorname{tg}^2\frac{\varphi_r}{2} = \frac{\beta^2 - \omega^2/c_2^2}{\omega^2/c_1^2 - \beta^2}. \tag{8.38}$$

Подставляя (8.38) в (8.37), получаем

$$\boxed{\operatorname{tg}^2\left(\frac{d}{2}\sqrt{\frac{\omega^2}{c_1^2} - \beta^2} - m\frac{\pi}{2}\right) = \frac{\beta^2 - \omega^2/c_2^2}{\omega^2/c_1^2 - \beta^2}.} \tag{8.39}$$

Дисперсионное соотношение

Это соотношение можно изобразить графически, если выразить его в параметрическом виде

$$\frac{\omega}{\omega_c} = \frac{\sqrt{n_1^2 - n_2^2}}{\sqrt{n_1^2 - n^2}}\left(m + \frac{2}{\pi}\operatorname{arctg}\sqrt{\frac{n^2 - n_2^2}{n_1^2 - n^2}}\right); \quad \beta = \frac{n\omega}{c_0} \tag{8.40}$$

через **эффективный показатель преломления** n, определенный в (8.40), где ω_c — частота отсечки моды,

$$\omega_c = \frac{2\pi c_0}{2d\,\mathrm{NA}}.$$

Как видно из рис. 8.18, *а*, графики дисперсионных соотношений для различных мод лежат между линиями $\omega = c_2\beta$ и $\omega = c_1\beta$ — **световыми линиями**, представляющими распространение света в однородных средах с показателями преломления окружающей среды и сердцевинного слоя соответственно. По мере того, как частота все более превышает частоту отсечки каждой моды, график дисперсионной зависимости смещается от световой линии оболочки к световой линии сердцевины, т. е. эффективный показатель преломления растет от n_2 до n_1. Проявление этого эффекта состоит в том, что более короткие волны в большей степени локализованы в пределах слоя с большим показателем преломления.

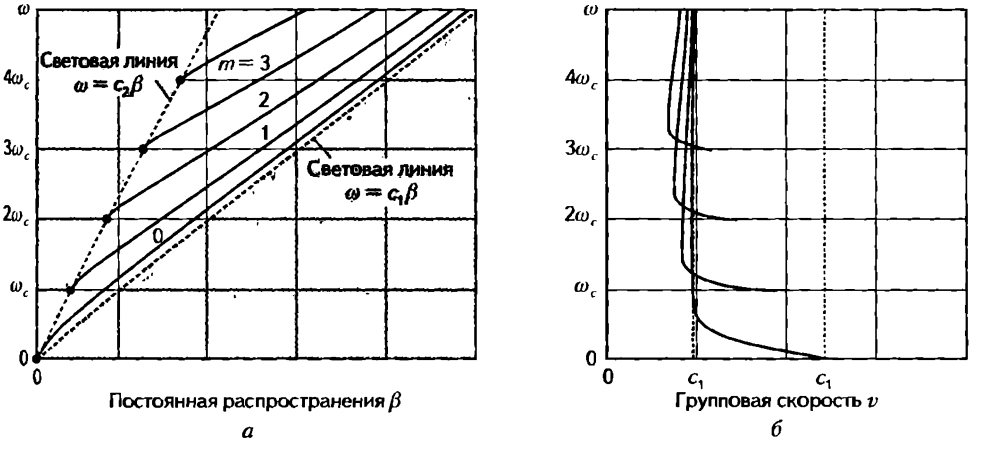


Рис. 8.18. Схематическое представление дисперсионного соотношения для различных ТЕ-мод $m = 0, 1, 2,$ *(а)* и частотной зависимости групповой скорости, которая представляет собой производную дисперсионного соотношения $v = \mathrm{d}\omega/\mathrm{d}\beta$ *(б)*

Групповая скорость получается из дисперсионного соотношения путем определения наклона $v = \mathrm{d}\omega/\mathrm{d}\beta$ для каждой направляемой моды. Зависимость групповой скорости от угловой частоты схематически показана на рис. 8.18, *б*. Для каждой моды по мере роста угловой частоты выше частоты отсечки моды

групповая скорость убывает от своего максимального значения c_2, имеет минимум немного ниже c_1, а затем стремится к c_1 снизу. Таким образом, групповые скорости мод лежат в интервале от максимального значения c_2 до минимального значения немного меньше c_1.

При распространении в многомодовом волноводе оптические импульсы расплываются во времени, так как моды имеют различные скорости. Этот эффект называется **модовой дисперсией**. В одномодовом волноводе оптический импульс расплывается из-за зависимости групповой скорости от частоты. Этот эффект называется **дисперсией групповой скорости** (ДГС). Как показано в разд. 5.6, ДГС имеет место в однородных материалах, она обусловлена зависимостью показателя преломления материала от частоты. Более того, ДГС происходит в волноводах даже в отсутствие дисперсии материала. В этом случае она обязана своим происхождением частотной зависимости коэффициентов распространения, которые определяются зависимостью оптического ограничения волны от ее длины. Как показано на рис. 8.18, *б*, у каждой моды есть такая частота, вблизи которой групповая скорость слабо зависит от частоты (точка, в которой скорость v имеет минимум, так что ее производная по частоте равна нулю).

Приближенное выражение для групповой скорости можно получить, беря полную производную по β от (8.37):

$$\frac{2d}{2k_y}\left(\frac{2\omega}{c_1^2}\frac{\mathrm{d}\omega}{\mathrm{d}\beta}-2\beta\right)=2\frac{\partial\varphi_r}{\partial\beta}+2\frac{\partial\varphi_r}{\partial\omega}\frac{\mathrm{d}\omega}{\mathrm{d}\beta}. \tag{8.41}$$

Подставляя

$$\frac{\mathrm{d}\omega}{\mathrm{d}\beta}=v;\quad \frac{k_y}{\omega/c_1}=\sin\theta;\quad \frac{k_y}{\beta}=\operatorname{tg}\theta$$

и вводя новые параметры

$$\Delta z=\frac{\partial\varphi_r}{\partial\beta};\quad \Delta\tau=\frac{\partial\varphi_r}{\partial\omega}, \tag{8.42}$$

получим

$$v=\frac{d\operatorname{ctg}\theta+\Delta z}{d\operatorname{cosec}\theta/c_1+\Delta\tau}. \tag{8.43}$$

Напомним [см. формулу (8.14) и рис. 8.6], что для планарного зеркального волновода $d\operatorname{ctg}\theta$ есть расстояние, проходимое лучом в направлении z за один полный цикл отражений между границами. Такой цикл занимает время $d\operatorname{csc}\theta/c_1$. Отношение $d\operatorname{ctg}\theta/(d\operatorname{cosec}\theta/c_1)$ дает групповую скорость для зеркального волновода. Выражение (8.43) для групповой скорости в диэлектрическом волноводе показывает, что луч проходит дополнительное расстояние $\Delta z=\partial\varphi_r/\partial\beta$, что занимает дополнительное время $\Delta\tau=-\partial\varphi_r/\partial\omega$. Это можно представлять как эффективное проникновение луча в оболочку или как эффективный латераль-

ный сдвиг луча, как показано на рис. 8.19. Проникновение луча за границу раздела при полном внутреннем отражении называется **сдвигом Гуса—Хенхена** (см. задачу 5 к разд. 6.2). С помощью (8.22) можно показать, что

$$\frac{\Delta z}{\Delta \tau} = \frac{\omega}{\beta} = \frac{c_1}{\cos\theta}.$$

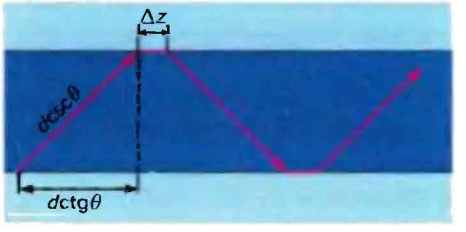


Рис. 8.19. Лучевая модель, в которой сдвиг фазы при отражении заменяется дополнительным расстоянием Δz, проходимым со скоростью $c_1/\cos\theta$

Упражнение 8.4

Асимметричный планарный волновод

Исследуйте поле TE-волны в асимметричном планарном волноводе, состоящем из диэлектрического слоя толщиной d с показателем преломления n_1, расположенного на подложке с меньшим показателем преломления n_2 и покрытого сверху диэлектриком с показателем преломления $n_3 < n_2 < n_1$, как показано на рис. 8.20.

а. Найдите выражение для максимального угла наклона θ плоских волн, испытывающих полное внутреннее отражение, и соответствующую числовую апертуру NA волновода.

б. Запишите выражение для условия самосогласования, аналогичное (8.22).

в. Найдите приближенное выражение для числа мод M, справедливое, когда это число очень велико.

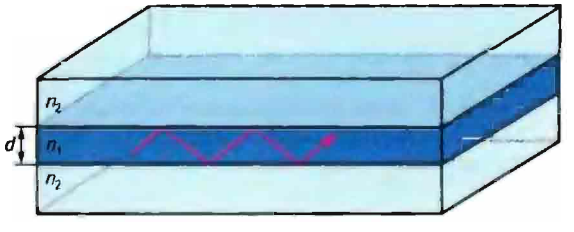


Рис. 8.20. Асимметричный планарный волновод

8.3. ДВУМЕРНЫЕ ВОЛНОВОДЫ

Планарные зеркальные и диэлектрические волноводы, рассмотренные в двух предыдущих разделах, ограничивают свет в одном поперечном направлении (направлении y) и направляют его вдоль оси z. Двумерный волновод

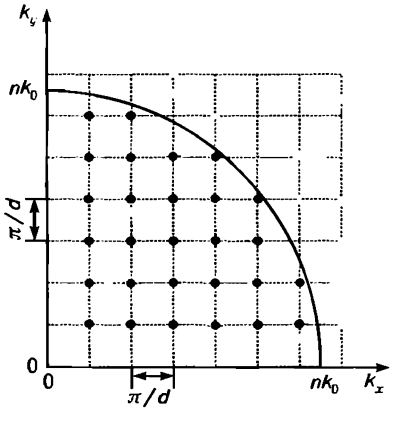
ограничивает свет в двух поперечных направлениях (x и y). Принцип действия и структура мод у двумерных волноводов, в основном, такие же, как у планарных волноводов, только математическое описание несколько длиннее. В этом разделе дается краткое описание природы мод в двумерных волноводах. Детали можно найти в специальной литературе. Глава 9 посвящена важному примеру двумерных волноводов — цилиндрическим волноводам, используемым в волоконной оптике.

Прямоугольный зеркальный волновод

Простейшим обобщением планарного волновода является прямоугольный волновод (рис. 8.21). Если стенки волновода — зеркала, то свет направляется благодаря многократным отражениям между ними под любым углом, как в планарном случае. Предположим для простоты, что поперечное сечение волновода представляет собой квадрат со стороной d. Чтобы плоская волна с волновым вектором (k_x, k_y, k_z) и ее многократные отражения могли самосогласованно существовать в волноводе, они должны удовлетворять условиям

$$\begin{aligned} 2k_x d &= 2\pi m_x, \quad m_x = 1, 2, \\ 2k_y d &= 2\pi m_y, \quad m_y = 1, 2, \end{aligned} \tag{8.44}$$

которые являются очевидным обобщением условий (8.44).



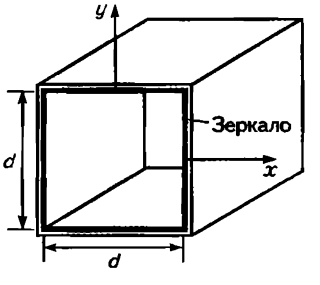


Рис. 8.21. Моды прямоугольного зеркального волновода характеризуются конечным числом дискретных значений k_x и k_y, изображенных точками

Постоянная распространения $\beta = k_z$ может быть определена по k_x и k_y с помощью соотношения

$$k_x^2 + k_y^2 + \beta^2 = n^2 k_0^2.$$

Следовательно, три компоненты волнового вектора имеют дискретные значения, определяющие конечное число мод. Каждая мода полностью характеризуется двумя индексами m_x и m_y (вместо индекса m у планарного волновода).

Эти индексы могут принимать любые целые положительные значения до тех пор, пока

$$k_x^2 + k_y^2 \le n^2 k_0^2,$$

как показано на рис. 8.21.

Число мод легко определить путем подсчета числа точек внутри четверти круга радиуса на плоскости (k_x, k_y) (см. рис. 8.21). Если это число велико, его можно аппроксимировать отношением площади $\pi(nk_0)^2/4$ к площади элементарной ячейки $(\pi/d)^2$

$$M \approx \frac{\pi}{4}\left(\frac{2d}{\lambda}\right)^2 \tag{8.45}$$

Поскольку на каждую моду приходится две поляризации, полное число мод в действительности равно $2M$. Сравнивая это с числом мод одномерного зеркального волновода $M \approx 2d/\lambda$, мы видим, что увеличение размерности приводит к примерно квадратичному росту числа мод. Число мод является мерой числа степеней свободы. При добавлении второго измерения числа степеней свободы просто перемножаются.

Распределения поля, связанные с каждой модой, являются обобщением распределений, найденных выше для планарного волновода. Зависимости, показанные на рис. 8.6, теперь имеют место для каждого из направлений x, y и определяются модовыми индексами m_x и m_y.

Прямоугольный диэлектрический волновод

Рассмотрим диэлектрический брус с показателем преломления n_1 и квадратным поперечным сечением со стороной d, погруженный в среду с немного меньшим показателем преломления n_2. Моды такого волновода можно найти с помощью аналогичной теории. Компоненты волнового вектора (k_x, k_y, k_z) должны удовлетворять условию

$$k_x^2 + k_y^2 \le n_1^2 k_0^2 \sin^2 \bar{\theta}_{\text{кр}},$$

где

$$\bar{\theta}_{\text{кр}} = \arccos\left(\frac{n_2}{n_1}\right),$$

так что k_x, k_y лежат в области, показанной на рис. 8.22. Значения k_x и k_y для каждой моды можно получить из условий самосогласования, в которых необходимо учесть сдвиг фазы на границе раздела диэлектриков, как это делалось в планарном случае.

В отличие от зеркального волновода значения k_x и k_y для мод расположены неравномерно, однако *среднее* расстояние между их соседними значениями равно π/d, как у зеркального волновода. Число мод можно приближенно

найти как число точек внутри малой четверти круга на рис. 8.22, считая, что среднее расстояние между точками равно π/d. В результате получается

$$M \approx \left(\frac{\pi}{4}\right)\frac{n_1 k_0 \sin\bar{\theta}_{кр}}{(\pi/d)^2},$$

откуда

$$\boxed{M \approx \frac{\pi}{4}\left(\frac{2d}{\lambda_0}\right)^2 (\mathrm{NA})^2,} \qquad (8.46)$$

Число ТЕ-мод

где $\mathrm{NA} = \sqrt{n_1^2 - n_2^2}$ — числовая апертура. Приближение удовлетворительно при большом числе мод. Существует такое же число ТМ-мод. Грубо число мод равно квадрату числа мод планарного диэлектрического волновода (8.25).

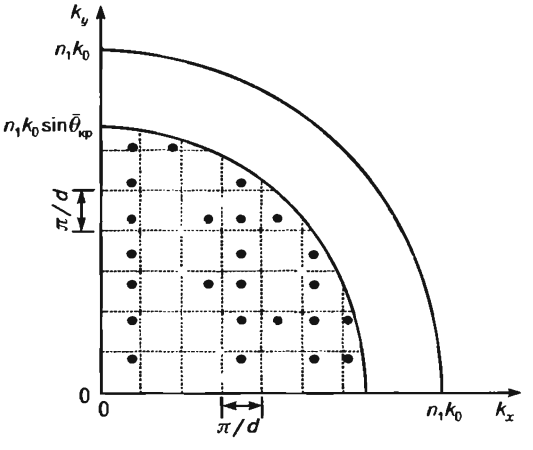


Рис. 8.22. Геометрия прямоугольного диэлектрического волновода. Значения k_x и k_y для мод волновода показаны точками

Геометрия канальных волноводов

Используются следующие типы геометрии волноводов: полосковый, погружной полоской, гребенчатый, нагруженный полоской (рис. 8.23). Точный анализ некоторых из этих геометрий достаточно сложен, поэтому обычно используются те или иные приближения. Для получения более подробной информации по данному вопросу читатель отсылается к специальной литературе.

Волноводы с погружной полоской могут изготавливаться в различных конфигурациях, показанных на рис. 8.24. S-образные колена используются для сдвига оси распространения. Y-образный разветвитель используется для деления или совмещения пучков. Два Y-образных разветвителя могут образовать интерферометр Маха—Цендера. Два волновода на близком расстоянии или при пересечении могут обмениваться энергией и использоваться как направленные разветвители, как мы увидим в следующем разделе.

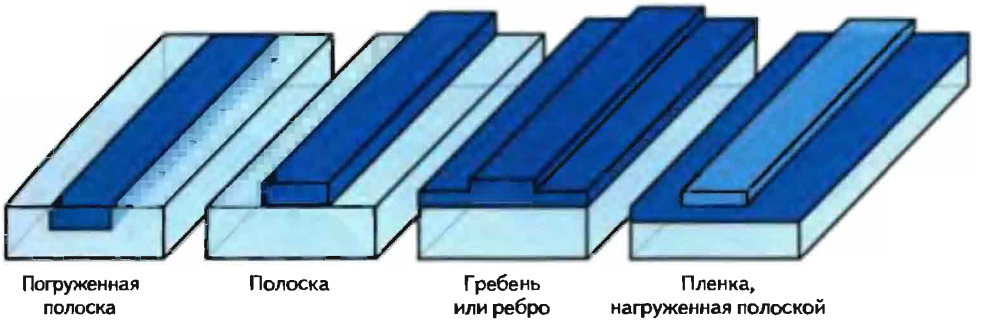


Рис. 8.23. Различные геометрии волноводов. Чем темнее оттенок цвета, тем больше показатель преломления

Рис. 8.24. Различные конфигурации волноводов

Материалы

Наиболее развита технология изготовления волноводов на Ti : $LiNbO_3$. Волновод с погруженной полоской получается путем диффузии титана в подложку из ниобата лития, в результате чего возникает полоска с повышенным показателем преломления. Полосковые GaAs волноводы делаются с применением слоев GaAs и AlGaAs с более низким показателем преломления. Другой полупроводниковый материал, который недавно стал важным в производстве волноводов, это InP. Стеклянные волноводы изготавливают при помощи ионного обмена. Полимерные волноводы все громче заявляют о себе как о жизнеспособной технологии.

Волноводы можно изготавливать также на основе структуры **кремний на диэлектрике** (КНД, $Si—SiO_2$) с использованием стандартных промышленных методов травления. Эта технология называется также **кварц на кремнии** (silica-on-silicon). Поскольку показатель преломления кремния ≈3,5, а кварца — менее 1,5, эта комбинация материалов обладает большой разностью показателей преломления Δn. Типичный КНД-волновод представляет собой кремниевый гребенчатый волновод (рис. 8.25) поверх слоя кварца, служащего оболочкой, под которым находится кремниевая подложка. Технология обработки кремния и изготовления структур на его основе хорошо разработана в микроэлектронной промышленности. Важным достоинством этой технологии является ее со-

вместимость с технологией изготовления комплементарных металлооксидных полупроводников (КМОП).

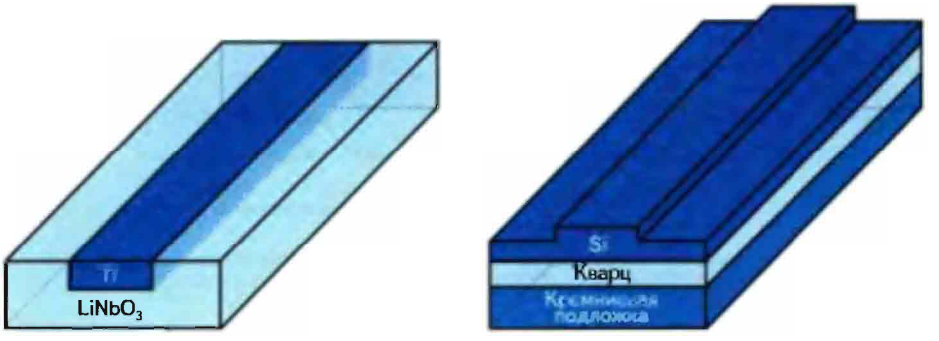


Рис. 8.25. $LiNbO_3$- и КНД-волноводы

Возможность модулировать показатель преломления — важное требование к материалам, используемым в таких устройствах интегральной оптики, как модуляторы и переключатели света, как мы увидим в гл. 20 и 23.

8.4. ФОТОННО-КРИСТАЛЛИЧЕСКИЕ ВОЛНОВОДЫ

Волновод с решетками Брэгга

До сих пор мы видели, что свет можно направлять путем многократного отражения от двух параллельных отражателей — плоских зеркал, как описано в разд. 8.1, или плоских границ раздела диэлектриков, на которых свет испытывает полное внутреннее отражение, как описано в разд. 8.2. Вместо этого в качестве отражателей можно использовать решетки Брэгга (см. разд. 7.1), как показано на рис. 8.26. Решетка Брэгга (РБ) представляет собой пакет чередующихся слоев диэлектрика и имеет специфическую зависимость коэффициента отражения от угла падения и частоты света. Для данного угла коэффициент отражения близок к единице на частотах внутри полосы заграждения. Аналогично, для данной частоты коэффициент отражения близок к единице в некотором диапазоне углов. Возможно также всенаправленное отражение. Таким образом, волна заданной частоты может направляться по волноводу посредством многократных отражений в некотором диапазоне углов. Внутри этого диапазона выполняются условия самосогласования при определенных дискретных значениях угла наклона, соответствующих распространяющимся собственным модам волновода. Распределение поля в распространяющейся моде главным образом сосредоточено в слое диэлектрика между отражателями, при этом затухающие «хвосты» распределения проникают внутрь соседних слоев решетки, как показано на рис. 8.26.

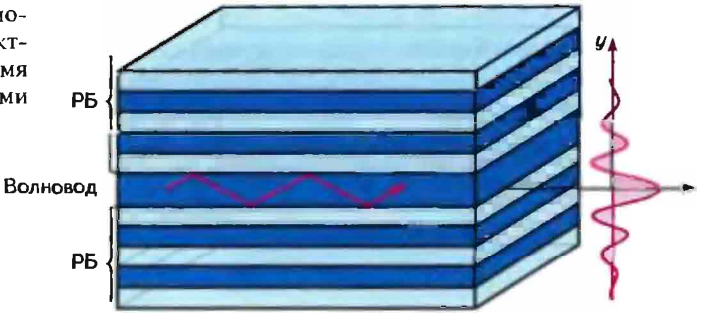


Рис. 8.26. Планарный волновод, состоящий из диэлектрического слоя между двумя отражательными решетками Брэгга

Волновод с решетками Брэгга как фотонный кристалл с дефектным слоем

Если верхняя и нижняя решетки волновода с РБ одинаковы, а толщина центрального слоя диэлектрика сравнима с толщиной периодических слоев, образующих решетку, то вся среда в целом может рассматриваться как одномерный фотонный кристалл с **дефектом**. Например, структура, показанная на рис. 8.26, периодична всюду, кроме центрального слоя, отличающегося показателем преломления и толщиной от остальных слоев. Этот слой можно рассматривать как «дефектный». Как описано в разд. 7.2, идеальный фотонный кристалл имеет дисперсионную зависимость с запрещенными зонами, т. е. полосами частот, на которых не существует распространяющихся мод. При наличии дефекта в фотонной запрещенной зоне может появиться распространяющаяся мода, локализованная преимущественно внутри дефектного слоя. Такая мода соответствует дискретному значению частоты, лежащему внутри фотонной запрещенной зоны на дисперсионной диаграмме, как показано на рис. 8.27. Такая частота является аналогом дефектного (примесного) уровня внутри запрещенной зоны полупроводникового кристалла.

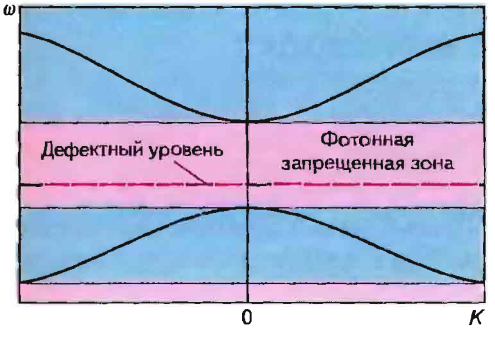
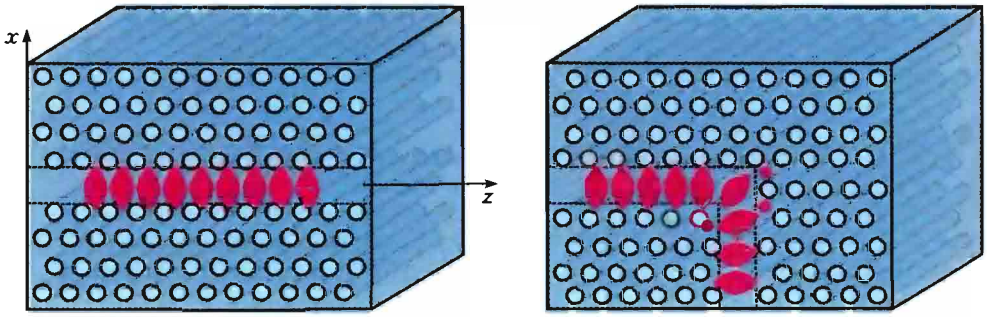


Рис. 8.27. Дисперсионная диаграмма фотонного кристалла с дефектным слоем

Двумерные фотонно-кристаллические волноводы

Волновод можно создать также в виде дорожки дефектов в двумерном фотонном кристалле. На рис. 8.28 двумерный фотонный кристалл, образованный периодически расположенными в виде треугольной решетки цилиндрическими отверстиями в диэлектрике, имеет полностью запрещенную фотон-

ную зону для волн, распространяющихся параллельно плоскости периодичности (по нормали к осям отверстий). Волновод формируется в виде линии пропущенных отверстий. На частотах внутри фотонной запрещенной зоны волна, входящая в волновод, не может вытекать в окружающую периодическую среду и поэтому направляется по волноводу. Типичный профиль распределения поля представлен на рис. 8.28, *а*.



а *б*

Рис. 8.28. Распространяющаяся мода в фотонно-кристаллическом волноводе (*а*); L-образный фотонно-кристаллический волновод (*б*)

Более того, ввиду всенаправленной природы фотонной запрещенной зоны свет можно направлять по фотонно-кристаллическому волноводу с углами и резкими поворотами без утечки энергии в окружающую среду, как иллюстрирует пример L-образного волновода на рис. 8.28, *б*. Такое поведение невозможно для обычных диэлектрических волноводов на основе полного внутреннего отражения.

8.5. ОПТИЧЕСКАЯ СВЯЗЬ В ВОЛНОВОДАХ

8.5.1. Устройства ввода

Возбуждение мод

Как было показано в предыдущих разделах, свет распространяется в волноводе в виде мод. Комплексная амплитуда светового поля в общем случае есть суперпозиция этих мод,

$$E(x,\ y) = \sum_m a_m u_m(y) \exp(-j\beta_m z), \tag{8.47}$$

где a_m — амплитуда; $u_m(y)$ — поперечное распределение (предполагаемое действительным); β_m — постоянная распространения моды с номером m.

Амплитуды отдельных мод зависят от природы источника света, используемого для возбуждения волновода. Если источник имеет распределение поля, точно соответствующее одной из мод, то именно эта мода и будет возбуждаться. В общем

случае источник с произвольным распределением поля $s(y)$ возбуждает разные моды на разном уровне. Доля энергии, передаваемой от источника каждой моде, зависит от степени сходства между $s(y)$ и $u_m(y)$. Для доказательства этого запишем $s(y)$ в виде разложения (взвешенной суперпозиции) ортогональных функций $u_m(y)$

$$s(y) = \sum_m a_m u_m(y), \tag{8.48}$$

где коэффициент a_l, представляющий амплитуду возбуждаемой моды l, равен

$$a_l = \int_{-\infty}^{\infty} s(y) u_l(y) \mathrm{d}y. \tag{8.49}$$

Это выражение можно получить, умножая обе части равенства (8.48) на $u_l(y)$, интегрируя по y и используя соотношение ортогональности

$$\int_{-\infty}^{\infty} u_l(y) u_m(y) \mathrm{d}y = 0$$

для $m \neq l$ вместе с условиями нормировки. Коэффициент a_l количественно определяет степень сходства (или корреляции) между распределением источника $s(y)$ и распределением моды $u_m(y)$.

Устройства ввода

Свет можно ввести в волновод путем непосредственной фокусировки на один его конец (рис. 8.29). Для возбуждения заданной моды поперечное распределение поля у падающего света $s(y)$ должно совпадать с таковым для моды. Поляризация падающего света также должна совпадать с поляризацией желаемой моды. Из-за малых размеров волноводного слоя фокусировка и юстировка обычно сложны, а возбуждение таким способом неэффективно.

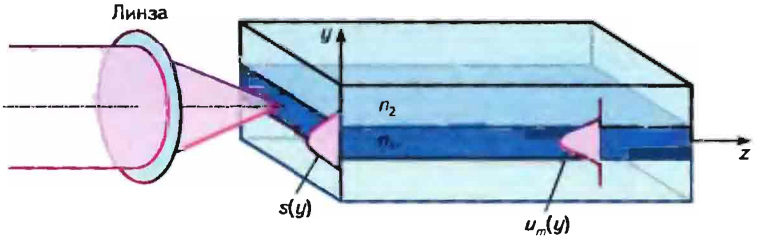


Рис. 8.29. Ввод пучка света в оптический волновод

В многомодовом волноводе оценку величины связи можно произвести с использованием лучевого подхода (рис. 8.30). Направления распространяющихся лучей в волноводе ограничены углом

$$\bar{\theta}_{\text{кр}} = \arccos\left(\frac{n_2}{n_1}\right).$$

Из-за преломления на торце волновода это соответствует внешнему углу θ_a, который удовлетворяет соотношениям

$$\mathrm{NA} = \sin\theta_a = n_1 \sin\bar{\theta}_{\text{кр}} = n_1\sqrt{1-\left(\frac{n_2}{n_1}\right)^2} = \sqrt{n_1^2 - n_2^2},$$

где NA — числовая апертура волновода (см. упражнение 1.6). Для максимальной эффективности возбуждения падающий свет нужно фокусировать под углом θ_a.

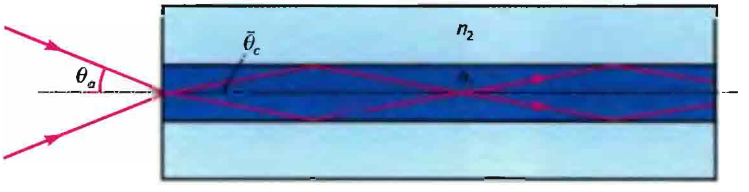
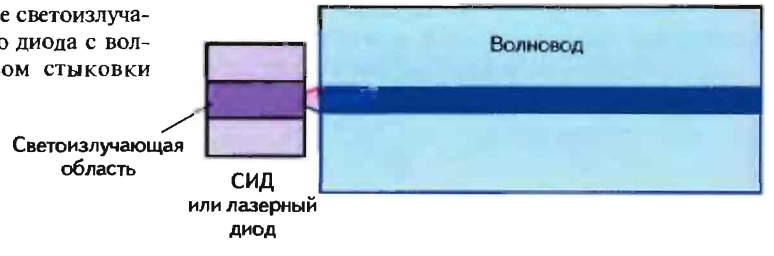


Рис. 8.30. Фокусировка лучей в многомодовый волновод

От полупроводникового источника (светодиода или лазерного диода) свет можно ввести в волновод путем непосредственного совмещения торцов источника и волновода с небольшим зазором между ними, который подбирается так, чтобы связь была максимальной (рис. 8.31). В светоизлучающих диодах свет возникает в области полупроводникового перехода и испускается во всех направлениях. В лазерных диодах испускаемый свет ограничен в пределах их собственного волновода. Светоизлучающие и лазерные диоды описаны в гл. 17. Другие методы ввода излучения в волновод включают использование призм, дифракционных решеток и других волноводов, как обсуждается ниже.

Рис. 8.31. Соединение светоизлучающего или лазерного диода с волноводом посредством стыковки торцов



Боковые устройства ввода на призмах и решетках

Можно ли передать энергию в направляемую моду волновода от волны, входящей сбоку под некоторым углом θ_i в оболочке, как показано на рис. 8.32, *а*? Необходимым для этого условием является равенство аксиальной компоненты волнового вектора падающей волны $n_2 k_0 \cos\theta_i$ постоянной распространения направляемой моды β_m. Поскольку $\beta_m > n_2 k_0$ (рис. 8.32), достижение требуемого условия фазового синхронизма $\beta_m = n_2 k_0 \cos\theta_i$ невозможно. Аксиальная компонента волнового вектора падающей волны попросту слишком мала. Однако данную проблему можно решить с помощью призмы или решетки.

Как показано на рис. 8.32, *б*, призма с показателем преломления $n_p > n_2$ помещается на малом расстоянии d_p от волноводного слоя. Падающая волна, преломляясь, проходит внутрь призмы, где она испытывает полное внутреннее отражение под углом θ_p. Падающая и отраженная волны формируют волну, распространяющуюся в направлении z с постоянной распространения $\beta_p = n_p k_0 \cos\theta_p$. Экспоненциально спадающий «хвост» поперечного распределения поля при этом покрывает зазор между призмой и волноводным слоем. Если расстояние d_p достаточно мало, эта волна возбуждает моду волновода с соответствующей постоянной распространения $\beta_m \approx \beta_p = n_p k_0 \cos\theta_p$. Поскольку $n_p > n_2$ достижение фазового синхронизма возможно, и при выборе надлежащей величины зазора в волновод можно передать значительную энергию. Такое же устройство можно использовать для решения обратной задачи вывода энергии из планарного волновода в свободное пространство.

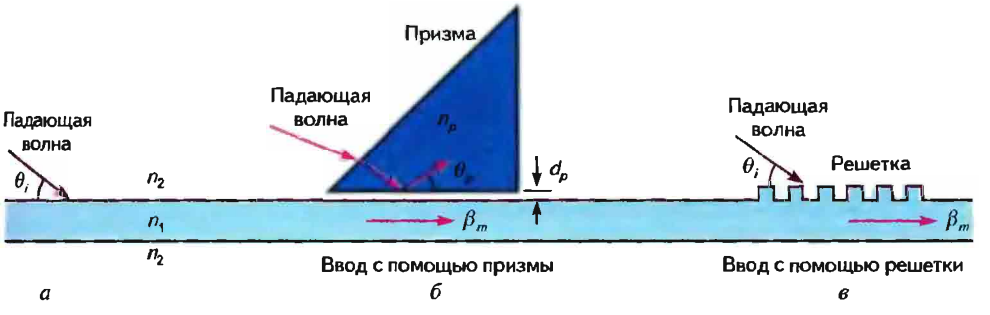


Рис. 8.32. Боковой ввод излучения (*а*) с помощью призмы (*б*) и решетки (*в*)

С помощью решетки (рис. 8.32, *в*) проблема фазового синхронизма решается путем изменения волнового вектора падающей волны. Решетка с периодом Λ модулирует падающую волну фазовыми множителями с фазой $2\pi q/\Lambda z$, где $q = \pm 1, \pm 2,$ Они эквивалентны изменению аксиальной компоненты волнового вектора на величину $2\pi q/\Lambda$. Теперь фазового синхронизма можно достичь, если

$$n_2 k_0 \cos\theta_i + \frac{2\pi q}{\Lambda} = \beta_m,$$

например, при $q = 1$. Решетка может быть спроектирована таким образом, чтобы дополнительно усиливать компоненту с $q = 1$.

8.5.2. Связанные волноводы

Если два волновода расположены достаточно близко, чтобы их поля перекрывались, то свет может передаваться от одного волновода к другому. Оптическая энергия может переноситься из одного волновода в другой, и этот эффект можно использовать для создания оптических соединительных устройств и переключателей. Основные принципы связи волноводов изложены в данном разделе, а соединительные устройства и переключатели рассматриваются в гл. 23 и 24.

Рассмотрим два параллельных планарных волновода в виде двух слоев с показателями преломления n_1 и n_2 толщиной d каждый, расположенных на расстоянии $2a$ друг от друга в среде с показателем преломления n, несколько меньшим, чем n_1 и n_2 (рис. 8.33). Предположим, что каждый из волноводов — одномодовый. Расстояние между слоями таково, что внешнее поле каждого волновода (в отсутствие другого) частично перекрывает сердцевину другого волновода.

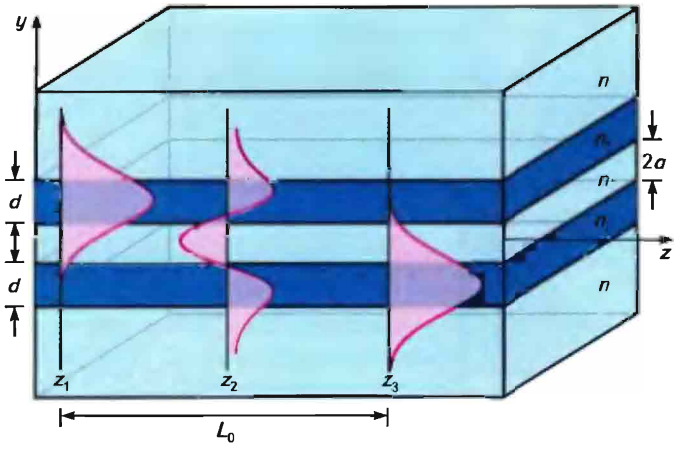


Рис. 8.33. Связь между двумя параллельными планарными диэлектрическими волноводами. В точке $z = z_1$ свет находится преимущественно в волноводе 1, при $z = z_2$ — делится примерно поровну между волноводами, при $z = z_3$ — находится преимущественно в волноводе 2

Формальный подход к изучению распространения света в такой структуре состоит в решении уравнений Максвелла в каждой области с использованием граничных условий для определения собственных мод всей системы. Эти моды отличны от мод каждого из волноводов в отсутствие взаимодействия. Точный анализ сложен и выходит за рамки настоящей книги. Однако для случая слабой связи можно построить приближенную теорию, известную как теория связанных мод, которая часто дает удовлетворительные результаты.

В этой теории моды каждого волновода вначале определяются так, как будто другого волновода нет. При наличии обоих волноводов предполагается, что сами моды остаются приблизительно неизмененными, скажем, $u_1(y) \exp(-j\beta_1 z)$ и $u_2(y) \exp(-j\beta_2 z)$. Считается, что взаимодействие меняет только *амплитуды* связанных мод, оставляя без изменения их поперечные распределения поля и постоянные распространения. Таким образом, амплитуды мод в волноводах 1 и 2 становятся функциями z, $a_1(z)$ и $a_2(z)$. Теория направлена на определение зависимостей $a_1(z)$ и $a_2(z)$ при соответствующих граничных условиях.

Взаимодействие можно рассматривать как эффект рассеяния. Поле волновода 1 рассеивается на волноводе 2, создавая источник света, меняющий амплитуду поля в волноводе 2. Поле волновода 2 оказывает аналогичное воздействие на волновод 1. Анализ такого взаимодействия приводит к двум связанным дифференциальным уравнениям, определяющим изменение $a_1(z)$ и $a_2(z)$.

Можно показать (см. вывод в конце данного раздела), что амплитуды $a_1(z)$ и $a_2(z)$ удовлетворяют двум связанным дифференциальным уравнениям первого порядка

$$\frac{da_1}{dz} = -jC_{21}\exp(j\Delta\beta z)a_2(z); \tag{8.50а}$$

$$\frac{da_2}{dz} = -jC_{12}\exp(j\Delta\beta z)a_1(z), \tag{8.50б}$$

Уравнения связанных мод

где $\Delta\beta$ — фазовая расстройка на единицу длины,

$$\Delta\beta = \beta_1 - \beta_2; \tag{8.51}$$

C_{21}, C_{12} — коэффициенты связи,

$$C_{21} = \frac{1}{2}\left(n_2^2 - n^2\right)\frac{k_0^2}{\beta_1}\int_a^{a+d} u_1(y)u_2(y)dy;$$

$$C_{12} = \frac{1}{2}\left(n_1^2 - n^2\right)\frac{k_0^2}{\beta_2}\int_{-a-d}^{-a} u_2(y)u_1(y)dy. \tag{8.52}$$

Из (8.50) видно, что скорость изменения a_1 пропорциональна a_2 и наоборот. Коэффициенты пропорциональности представляют собой произведения коэффициентов связи и множителя $\exp(j\Delta\beta z)$, зависящего от фазовой расстройки.

Уравнения связанных мод можно решить путем умножения обеих частей (8.50а) на $\exp(-j\Delta\beta z)$, дифференцирования по z, подстановки (8.50б) и решения получившегося дифференциального уравнения второго порядка относительно $a_1(z)$. В результате получается

$$a_1(z) = A(z)a_1(0) + B(z)a_2(0); \tag{8.53а}$$

$$a_2(z) = C(z)a_1(0) + D(z)a_2(0), \tag{8.53б}$$

где $A(z)$, $B(z)$, $C(z)$ — элементы матрицы передачи **T**, связывающей входное и выходное поля,

$$A(z) = D^*(z) = \exp\left(\frac{j\Delta\beta z}{2}\right)\left(\cos\gamma z - j\frac{\Delta\beta}{2\gamma}\sin\gamma z\right); \tag{8.54а}$$

$$B(z) = \frac{C_{21}}{j\gamma}\exp\left(\frac{j\Delta\beta z}{2}\right)\sin\gamma z; \tag{8.54б}$$

$$C(z) = \frac{C_{12}}{j\gamma}\exp\left(-\frac{j\Delta\beta z}{2}\right)\sin\gamma z; \tag{8.54в}$$

$$\gamma^2 = \left(\frac{\Delta\beta}{2}\right)^2 + C^2; \quad C = \sqrt{C_{12}C_{21}}. \tag{8.55}$$

Если предположить, что в волновод 2 свет не вводится, так что $a_2(0) = 0$, то оптические мощности $P_1(z) \propto |a_1(z)|^2$ и $P_2(z) \propto |a_2(z)|^2$ равны

$$P_1(z) = P_1(0)\left[\cos^2 \gamma z + \left(\frac{\Delta\beta}{2\gamma}\right)^2 \sin^2 \gamma z\right]; \tag{8.56а}$$

$$P_2(z) = P_2(0)\frac{|C_{21}|^2}{\gamma^2}\sin^2 \gamma z. \tag{8.56б}$$

Таким образом, происходит периодический обмен энергией между волноводами, как показано на рис. 8.34, *а*. Период этого обмена равен π/γ.

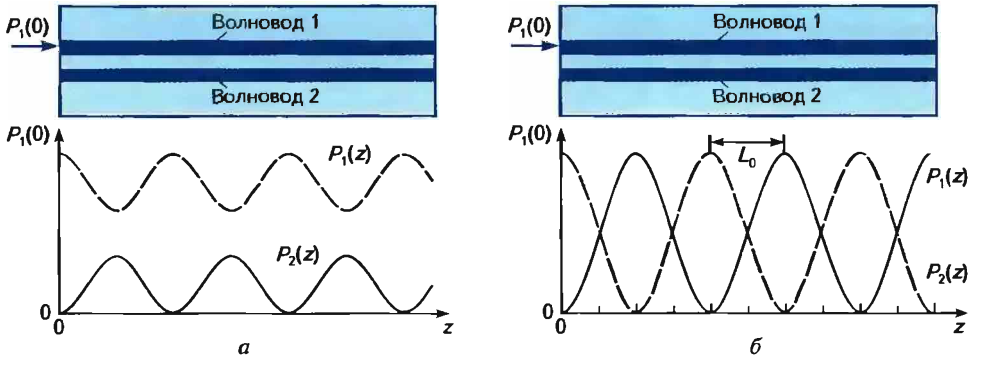


Рис. 8.34. Периодический обмен энергией между волноводами 1 и 2:
а — фазовый синхронизм отсутствует; *б* — случай фазового синхронизма

Если волноводы одинаковы, т. е. $n_1 = n_2$, $\beta_1 = \beta_2$ и $\Delta\beta = 0$, то говорят, что между обеими направляемыми волнами существует фазовый синхронизм. В этом случае $\gamma = C$, $C_{12} = C_{21} = C$ и матрица передачи сводится к упрощенному виду

$$\mathbf{T} = \begin{bmatrix} A(z) & B(z) \\ C(z) & D(z) \end{bmatrix} = \begin{bmatrix} \cos Cz & -j\sin Cz \\ -j\sin Cz & \cos Cz \end{bmatrix}. \tag{8.57}$$

Тогда уравнения (8.56) также упрощаются:

$$P_1(z) = P_1(0)\cos^2 Cz; \tag{8.58а}$$

$$P_2(z) = P_1(0)\sin^2 Cz. \tag{8.58б}$$

В этом случае перекачка энергии между волноводами может быть полной, как показано на рис. 8.34, *б*.

Таким образом, мы получаем устройство, с помощью которого можно передать нужную часть оптической мощности из одного волновода в другой. На

расстоянии $z = L_0 = \pi/2C$, называемом **длиной участка связи**, или **длиной передачи**, мощность полностью переходит из одного волновода в другой. На расстоянии $L_0/2$ передается половина мощности, так что устройство действует как светоделитель 50/50 (рис. 8.35, *б*).

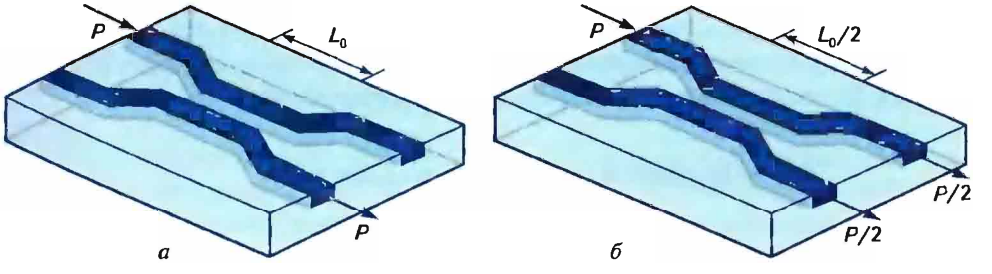


Рис. 8.35. Оптические устройства связи:
а — переключение энергии из одного волновода в другой; *б* — светоделитель 50/50

Переключение путем управления фазовой расстройкой

Устройство связи волноводов фиксированной длины, например $L_0 = \pi/2C$, меняет отношение передачи энергии при введении небольшой фазовой расстройки $\Delta\beta$. С помощью (8.56б) и (8.55) коэффициент передачи энергии $\mathcal{T} = P_2(L_0)/P_1(0)$ можно записать как функцию $\Delta\beta$

$$\mathcal{T} = \frac{\pi^2}{4}\operatorname{sinc}^2\left[\frac{1}{2}\sqrt{1+\left(\frac{\Delta\beta L_0}{\pi}\right)^2}\right], \qquad (8.59)$$

Коэффициент передачи энергии

где

$$\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}.$$

На рис. 8.36 проиллюстрирована зависимость коэффициента передачи энергии $\mathcal{T}$ от параметра расстройки $\Delta\beta L_0$. Коэффициент передачи достигает максимального значения, равного единице, при $\Delta\beta L_0 = 0$, убывает с ростом $\Delta\beta L_0$ и обращается в нуль при $\Delta\beta L_0 = \sqrt{3}\pi$.

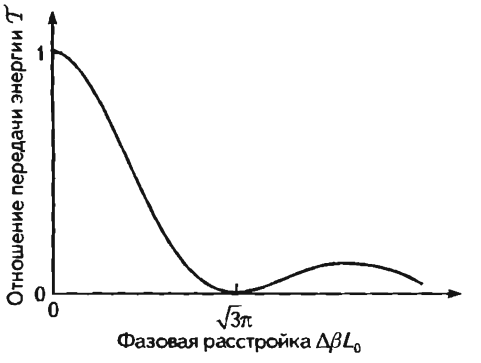


Рис. 8.36. Зависимость коэффициента передачи энергии $\mathcal{T} = P_2(L_0)/P_1(0)$ от параметра фазовой расстройки $\Delta\beta L_0$. Длина волны выбрана такой, чтобы при $\Delta\beta = 0$ (случай фазового синхронизма) максимальная мощность передавалась в волновод 2, т. е. $\mathcal{T} = 1$

Зависимость переданной мощности от фазовой расстройки можно использовать при изготовлении электрически управляемых направленных ответвителей. Если параметр расстройки $\Delta\beta L_0$ переключается между 0 и $\sqrt{3}\pi$, то свет переключается между волноводом 2 и волноводом 1. Электрическое управление величиной можно осуществить, если материал волноводов обладает электрооптическими свойствами (т. е. его показатель преломления меняется при наложении внешнего электрического поля). Такие устройства будут изучены в гл. 20 и 23, посвященных электрооптическим переключателям.

*Вывод уравнений для связанных волн

Перейдем к выводу дифференциальных уравнений (8.50), которым подчиняются амплитуды $a_1(z)$ и $a_2(z)$ связанных мод. Когда между волноводами отсутствует взаимодействие, оптические поля мод имеют комплексные амплитуды вида

$$E_1(y,\ z) = a_1 u_1(y)\exp(-j\beta_1 z); \tag{8.60а}$$

$$E_2(y,\ z) = a_2 u_2(y)\exp(-j\beta_2 z). \tag{8.60б}$$

При этом амплитуды мод a_1 и a_2 постоянны. При наличии взаимодействия предположим, что амплитуды a_1 и a_2 становятся функциями z, однако поперечные распределения полей $u_1(y)$ и $u_2(y)$ и постоянные распространения β_1 и β_2 остаются неизменными. Предполагается, что зависимость a_1 и a_2 от z является медленной по сравнению с обратными постоянными распространения $\beta_{1,2}^{-1}$, имеющими порядок длины волны.

Присутствие волновода 2 рассматривается как возмущение среды вне волновода 1 в виде слоя с показателем преломления $(n_2 - n)$ и шириной d, расположенного на расстоянии $2a$. Избыточный показатель преломления $(n_2 - n)$ и поле E_2 соответствуют избыточной плотности поляризации

$$P = (\varepsilon_2 - \varepsilon)E_2 = \varepsilon_0\left(n_2^2 - n^2\right)E_2,$$

которая создает источник, испускающий оптическое излучение в волновод 1 [см. (5.40)]

$$S_1 = -\frac{\mu_0 \partial^2 P}{\partial t^2}$$

с комплексной амплитудой

$$S_1 = \mu_0\omega^2 P = \mu_0\omega^2\varepsilon_0\left(n_2^2 - n^2\right)E_2 = \left(n_2^2 - n^2\right)k_0^2 E_2 = \left(k_2^2 - k^2\right)E_2. \tag{8.61}$$

Здесь ε_2 и ε — диэлектрические проницаемости, связанные с показателями преломления n_2 и n соответственно, а $k_2 = n_2 k_0$. Этот источник присутствует только в слое волновода 2.

Для определения действия такого источника на поле в волноводе 1 запишем уравнение Гельмгольца в присутствии источника как

$$\nabla^2 E_1 + k_1^2 E_1 = -S_1 = -\left(k_2^2 - k^2\right) E_2. \tag{8.62а}$$

Аналогично напишем уравнение Гельмгольца в волноводе 2 с источником, генерируемым в волноводе 1:

$$\nabla^2 E_2 + k_2^2 E_2 = -S_2 = -\left(k_1^2 - k^2\right) E_1, \tag{8.62б}$$

где $k_1 = n_1 k_0$. Уравнения (8.62) представляют собой связанные уравнения в частных производных, которые мы решаем, чтобы определить E_1 и E_2. Используемый при этом метод возмущений применим только в случае слабой связи волноводов.

Теперь выразим комплексные амплитуды как $E_1(y, z) = a_1(z)e_1(y, z)$ и $E_2(y, z) = a_2(z)e_2(y,z)$, где $e_1(y, z) = u_1(y)\exp(-j\beta_1 z)$ и $e_2(y, z) = u_2(y)\exp(-j\beta_2 z)$, и заметим, что e_1 и e_2 должны удовлетворять уравнениям Гельмгольца

$$\nabla^2 e_1 + k_1^2 e_1 = 0; \tag{8.63а}$$

$$\nabla^2 e_2 + k_2^2 e_2 = 0, \tag{8.63б}$$

где $k_1 = n_1 k_0$ и $k_2 = n_2 k_0$ для точек внутри слоев волноводов 1 и 2 соответственно, и $k_1 = k_2 = nk_0$ во всех остальных точках. Подставляя $E_1 = a_1 e_1$ в (8.62а), получаем

$$\frac{d^2 a_1}{dz^2} e_1 + 2\frac{da_1}{dz}\frac{de_1}{dz} = -\left(k_2^2 - k^2\right) a_2 e_2. \tag{8.64}$$

Замечая, что a_1 меняется медленно, в то время как e_1 меняется быстро вдоль z, пренебрежем первым членом в (8.64) по сравнению со вторым. Отношение этих членов:

$$\frac{(d\Psi/dz)e_1}{2\Psi(de_1/dz)} = \frac{(d\Psi/dz)e_1}{2\Psi(-j\beta_1 e_1)} = \frac{j(d\Psi/\Psi)}{2\beta_1 dz},$$

где $\Psi = da_1/dz$. Наше приближение верно при условии $d\Psi/\Psi \ll \beta_1$, т. е. если изменение $a_1(z)$ происходит медленно по сравнению с длиной β_1^{-1}.

Далее, подставим $e_1 = u_1 \exp(-j\beta_1 z)$ и $e_2 = u_2 \exp(-j\beta_2 z)$ в (8.64). Пренебрегая первым членом, находим

$$2\frac{da_1}{dz}(-j\beta_1)u_1(y)e^{-j\beta_1 z} = -\left(k_2^2 - k^2\right) a_2 u_2(y) e^{-j\beta_2 z}. \tag{8.65}$$

Умножим обе части (8.65) на $u_1(y)$, проинтегрируем по y и используем тот факт, что функция $u_1^2(y)$ нормирована, так что интеграл от нее равен единице. Тогда окончательно получаем

$$\frac{da_1}{dz} e^{-j\beta_1 z} = -jC_{21} a_2(z) e^{-j\beta_2 z}, \tag{8.66}$$

где C_{21} дается формулой (8.52). Такое же уравнение получается, если повторить процедуру для волновода 2. Эти уравнения дают систему связанных дифференциальных уравнений (8.50).

8.5.3. Периодические волноводы

Анализ распространения света в двух связанных параллельных планарных волноводах можно в принципе обобщить на множество волноводов, хотя получающиеся связанные уравнения трудно решить. В пределе большого числа параллельных одинаковых слоев на одинаковом расстоянии друг от друга легко применить теорию распространения света в периодических структурах, представленную в разд. 7.2. Поучительно сравнить дисперсионные диаграммы для распространения света в обычном пластинчатом диэлектрическом волноводе (см. рис. 8.18, *а*) и в периодической диэлектрической среде, состоящей из параллельных диэлектрических слоев, показанной на рис. 7.31. Эти диаграммы вновь приведены на рис. 8.37 для удобства сравнения.

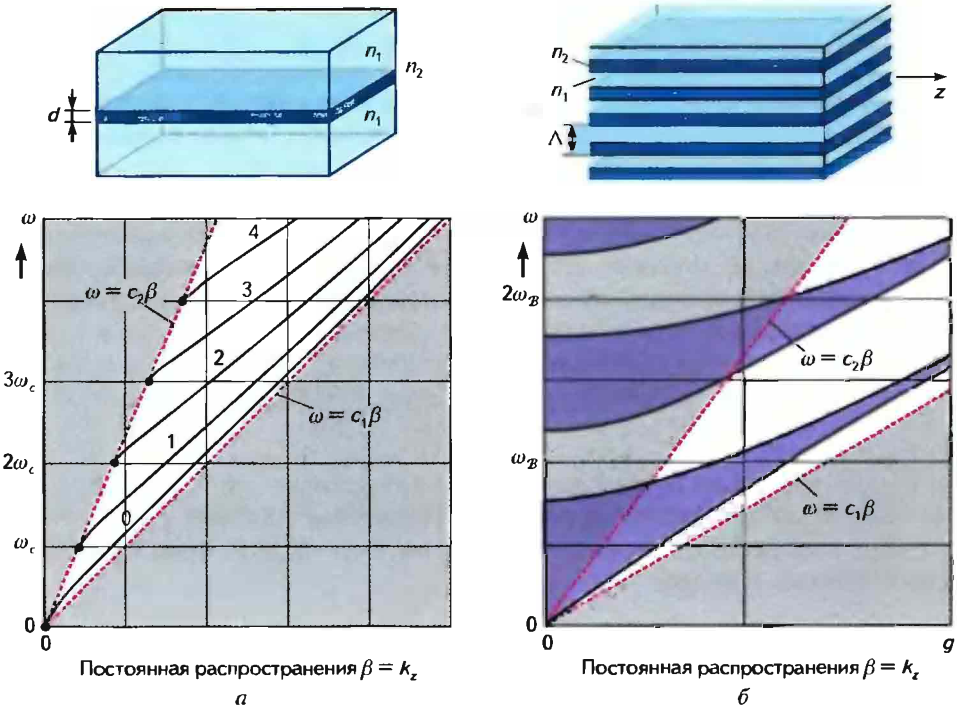


Рис. 8.37. Дисперсионные диаграммы:

а — для пластинчатого волновода с угловой частотой отсечки $\omega_c = (\pi/d)(c_0/N\Lambda)$; *б* — для периодического волновода с брэгговской угловой частотой $\omega_B = (\pi/\Lambda)(c_0/\bar{n})$

В однослойном волноводе свет распространяется в виде мод, дисперсионная кривая каждой из которых лежит между линиями $\omega = c_1\beta$ и $\omega = c_2\beta$. На

каждой частоте существует хотя бы одна мода. В периодическом волноводе дисперсионные кривые сгущаются в разрешенные зоны, разделенные фотонными запрещенными зонами. Здесь мы предполагаем, что моды распространяются в направлении, параллельном слоям (направление z на рис. 8.37, которое соответствует направлению x на рис. 7.31), так что зоны тоже лежат в области между световыми линиями.

8.6. МЕТАЛЛИЧЕСКИЕ ВОЛНОВОДЫ С РАЗМЕРАМИ МЕНЬШЕ ДЛИНЫ ВОЛНЫ (ПЛАЗМОНИКА)

Как показано в предыдущих разделах, оптическую волну трудно ограничить размерами много меньше длины волны (см. также подразд. 4.4.4). В описанном в разд. 8.1 зеркальном волноводе волна с длиной λ не может распространяться, если расстояние между зеркалами d меньше $\lambda/2$ (так как частоты волны в этом случае окажется меньше частоты отсечки $c/2d$). В пластинчатом диэлектрическом волноводе, описанном в разд. 8.2, при уменьшении толщины слоя d до значений меньше $\lambda/2$ может поддерживаться лишь одна мода, причем при дальнейшем уменьшении d происходит существенная утечка направляемой волны в оболочку. Однако свет можно ограничивать и направлять в субволновых масштабах с помощью металлических структур с размерами меньше длины волны, таких как тонкие пленки и металлические частицы, внедренные в диэлектрическую среду. Такой подход стал возможен в последние годы благодаря достижениям нанотехнологии (технологии наноструктур и наночастиц) и прогрессу в области, носящей название **плазмоника**.

Распространение света в объеме металла описано в подразд. 5.5.4. Напомним, что на частотах ниже плазменной частоты оптическая волна затухает, причем коэффициент ее ослабления убывает с ростом частоты и обращается в нуль на плазменной частоте; свободные электроны тогда испытывают продольные колебания, связанные с **плазмонами**. Ясно, что объемные металлы не годятся для ограничения и направления оптических волн. Однако на границе между металлом и диэлектриком уравнения Максвелла допускают решения в виде волн плотности заряда, связанных с оптическими волнами. Такие решения обычно называют **поверхностными плазмонными поляритонами** (ППП). Электроны проводимости осциллируют в продольном направлении, а электромагнитное поле ограничено слоем субволновых размеров вблизи поверхности металла. Эти связанные волны можно возбуждать на частотах ниже плазменной частоты; их наибольшая локализация достигается на плазменной частоте. ППП позволяют управлять и манипулировать светом в нанометровых пространственных масштабах, сохраняя высокую временную частоту, присущую оптическим волнам.

Реализовать идею волновода на основе ППП можно, например, с использованием диэлектрического слоя, покрытого с обеих сторон металлической оболочкой. Толщина диэлектрического слоя должна быть достаточно мала, чтобы ППП-волны, локализованные вблизи поверхностей «металл—диэлектрик» могли перекрываться, давая возможность связанным ППП-волнам направленно рас-

пространяться. Дисперсионные соотношения для такой структуры можно получить с помощью граничных условий на поверхностях раздела «металл—диэлектрик», описывая металл, например, моделью Друде (см. подразд. 5.5.4). При достаточно малых значениях толщины слоя можно достичь высоких постоянных распространения даже на частотах намного ниже плазменной частоты объемного металла. Эти плазмонные волноводы делаются из гетероструктур **металл—диэлектрик—металл (МДМ)** субмикронных размеров. Моды ближнего инфракрасного диапазона длин волн можно локализовать до нанометровых размеров, однако при этом длина распространения ограничена.

В другом классе плазмонных волноводов с субволновыми размерами мод используются матрицы из металлических наночастиц, расположенных достаточно близко, так чтобы их локализованные плазмонные поля перекрывались. Такие **метаматериалы** (см. разд. 5.7) допускают существование направляемых мод субмикронного размера на плазмонных частотах индивидуальных частиц или на резонансной частоте зазора между частицами.

Плазмоника стремится соединить микроэлектронику высокой степени интеграции (размеры менее 100 нм) с фотоникой оптических частот (полоса частот более 100 ТГц). Ожидается, что она найдет множество ценных применений в нанооптике, включая внутренние соединения микросхем, пропускание света через объекты, которые при обычных условиях непрозрачны (за счет плазмонных возбуждений наноразмерных отверстий в материале), создание распределенных точечных источников света, генерируемого на поверхностях наноразмерных объектов с металлическим покрытием, а также таких устройств, как наноантенны, нанорезонаторы и нановолноводы, которые аналогичны элементам электрических схем, но работают в видимом диапазоне частот. Биосенсорные применения, основанные на высокой чувствительности плазмонных возбуждений к свойствам диэлектрической среды, окружающей металлическую наноструктуру, включают измерение толщины коллоидных пленок, а также скрининг и подсчет событий связывания белков.

Рекомендуемая литература

книги

Chen C.-L. Foundations for Guided Wave Optics, Wiley, 2006.

Iga K., Kokobun Y., eds. Encyclopedic Handbook of Integrated Optics. CRC Press, 2006.

Sharma A., ed. Guided Wave Optics. Anshan, 2006.

Okamoto K. Fundamentals of Optical Waveguides. Elsevier, 2nd ed. 2005.

Pal B.P., ed. Guided Wave Optical Components and Devices: Basics, Technology, and Applications. Academic Press, 2005.

Reed G.T., Knights A.P. Silicon Photonics. Wiley, 2004.

Lifante G. Integrated Photonics: Fundamentals. Wiley, 2003.

Pollock C., Lipson M. Integrated Photonics. Kluwer, 2003.

Barybin A.A., Dmitriev V.A. Modern Electrodynamics and Coupled-Mode Theory: Application to Guided- Wave Optics. Rinton Press, 2002.

Hunsperger R.G. Integrated Optics: Theory and Technology. Springer-Verlag, 1982; 5th ed. 2002.

Iizuka K. Elements of Photonics. Vol. 2: For Fiber and Integrated Optics. Wiley, 2002.

Waynant R.W., Lowell J.K. Electronic and Photonic Circuits and Devices. IEEE Press, 1998.

Leigh W.B. Devices for Optoelectronics. Marcel Dekker, 1996.

Coldren L.A., Corzine S.W. Diode Lasers and Photonic Integrated Circuits. Wiley, 1995.

Zappe H.P. Introduction to Semiconductor Integrated Optics. Artech, 1995.

Hunsperger R.G., ed. Photonic Devices and Systems. Marcel Dekker, 1994.

Suematsu Y., Adams A.R., eds. Handbook of Semiconductor Lasers and Photonic Integrated Circuits. Chapman & Hall. English ed., 1994.

Wada O., ed. Optoelectronic Integration: Physics, Technology, and Applications. Kluwer, 1994.

Ebeling K.J. Integrated Optoelectronics: Waveguide Optics, Photonics, Semiconductors. Springer-Verlag, 1993.

Mickelson A.R. Guided Wave Optics. Springer-Verlag, 1993.

Hornak L.A., ed. Polymers for Lightwave and Integrated Optics: Technology and Applications. Marcel Dekker, 1992.

Midwinter J.E., Guo Y.L. Optoelectronics and Lightwave Technology. Wiley, 1992.

Syms R., Cozens J. Optical Guided Waves and Devices. McGraw-Hill, 1992.

Young M. Optics and Lasers: Including Fibers and Optical Waveguides. Springer-Verlag, 4th revised ed., 1992.

Marcuse D. Theory of Dielectric Optical Waveguides. Academic Press, 1974; 2nd ed., 1991.

Tamir T., ed. Guided- Wave Optoelectronics. Springer-Verlag, 2nd ed., 1990.

Marcuse D. Light Transmission Optics. Van Nostrand Reinhold, 1972; 2nd ed., 1982; Krieger reissued, 1989.

Nishihara H., Haruna M., Suhara T. Optical Integrated Circuits. McGraw-Hill, 1989.

Solimeno S., Crosignani B., Di Porto P. Guiding, Diffraction, and Confinement of Optical Radiation. Academic Press, 1986.

СТАТЬИ

Issue on nanophotonics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 12, № 6, 2006.

Paniccia M., Koehl S. The Silicon Solution. IEEE Spectrum. Vol. 42, № 10, 2005. P. 38—43.

Issue on integrated optics and optoelectronics. Part II. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 11, № 2, 2005.

Issue on integrated optics and optoelectronics. Part I. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 11, № 1, 2005.

Special issue on integrated optics. Applied Physics B: Lasers and Optics. Vol. 73, № 5—6, 2001.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Issue on integrated optics and optoelectronics. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 1, 2000.

Hall D.G., ed. Selected Papers on Coupled Mode Theory in Guided- Wave Optics. SPIE Optical Engineering Press (Milestone Series Volume 84), 1993.

Задачи

К РАЗДЕЛУ 8.1

1. **Распределение поля.**

a. Покажите, что одна плоская ТЕМ-волна $E_x(y, z) = A\exp(-jk_y y)\exp(-j\beta z)$ не может удовлетворить граничным условиям $E_x(\pm d/2, z) = 0$ при любом z в зеркальном волноводе, изображенном на рис. 8.3.

б. Покажите, что сумма двух плоских ТЕМ-волн, записанная в виде

$$E_x(y, z) = A_1 \exp(-jk_{y1}y) \exp(-j\beta_1 z) + A_2 \exp(-jk_{y2}y) \exp(-j\beta_2 z),$$

удовлетворяет граничным условиям, если $A_1 = \pm A_2$, $\beta_1 = \beta_2$ и $k_{y1} = -k_{y2} = m\pi/d$, где $m = 1, 2,$

2. **Модовая дисперсия.** Свет с длиной волны $\lambda_0 = 0{,}633$ мкм проходит через зеркальный волновод с расстоянием между зеркалами $d = 10$ мкм и $n = 1$. Определите число ТЕ- и ТМ-мод. Определите групповые скорости для наиболее быстрой и медленной моды. Если короткий импульс света переносится всеми модами волновода на расстояние 1 м, то насколько импульс уширяется из-за различия групповых скоростей?

К РАЗДЕЛУ 8.2

1. **Параметры диэлектрического волновода.** Свет, имеющий в свободном пространстве длину волны $\lambda = 0{,}87$ мкм, направляется тонкой плоской пленкой толщины $d = 2$ мкм с показателем преломления $n_1 = 1{,}6$, которую окружает среда с показателем преломления $n_2 = 1{,}4$.

а. Определите критический угол $\theta_{кр}$ и дополнительный к нему $\bar{\theta}_{кр}$, числовую апертуру NA и максимальный угол приема для света, падающего из воздуха ($n = 1$).

б. Определите число ТЕ-мод.

в. Определите угол наклона θ и групповую скорость v для ТЕ-моды с $m = 0$.

2. **Влияние оболочки.** Повторите расчет из предыдущей задачи, считая, что пленка находится в воздухе ($n = 1$). Сравните результаты.

3. **Распределение поля.** Поперечное распределение $u_m(y)$ комплексной амплитуды электрического поля ТЕ-моды в пластинчатом волноводе дается формулами (8.28) и (8.31). Выведите выражение для отношения констант пропорциональности. Постройте график ТЕ-моды с $m = 0$ для пластинчатого волновода с параметрами $n_1 = 1{,}48$, $n_2 = 1{,}46$, $d = 0{,}5$ мкм, $\lambda_0 = 0{,}85$ мкм и определите фактор ограничения (процент энергии, локализованной внутри сердцевины).

4. **Вывод распределения поля с использованием уравнений Максвелла.** Предполагая, что электрическое поле в симметричном диэлектрическом волноводе ведет себя гармонически внутри слоя, экспоненциально — вне слоя и в обеих средах имеет постоянную распространения β, можно записать $E_x(y, z) = u(y)e^{-j\beta z}$, где

$$u(y) = \begin{cases} A\cos\left(k_y y + \varphi\right), & -\dfrac{d}{2} \le y \le \dfrac{d}{2}; \\ B\exp(-\gamma y), & y > \dfrac{d}{2}; \\ B\exp(\gamma y), & y < \dfrac{d}{2}. \end{cases}$$

Чтобы удовлетворить уравнению Гельмгольца, необходимо

$$k_y^2 + \beta^2 = n_1^2 k_0^2; \quad -\gamma^2 + \beta^2 = n_2^2 k_0^2.$$

Используйте уравнения Максвелла, чтобы вывести выражения для $H_y(y, z)$ и $H_z(y, z)$. Покажите, что граничные условия удовлетворяются, если β, γ и k_y принимают значения β_m, γ_m и k_{ym}, выведенные в тексте, и проверьте выполнение условия самосогласования (8.22).

5. **Одномодовый волновод.** Какова наибольшая толщина d планарного симметричного диэлектрического волновода с показателями преломления $n_1 = 1{,}50$ и $n_2 = 1{,}46$, для которого существует только одна ТЕ-мода на длине волны $\lambda_0 = 1{,}3$ мкм? Каково число мод, если при той же толщине слоя заменить длину волны на $\lambda_0 = 0{,}85$ мкм?

6. **Отсечка моды.** Покажите, что условие отсечки для ТЕ-моды $m > 0$ в симметричном пластинчатом волноводе с $n_1 \approx n_2$ приближенно записывается как

$$\lambda_0^2 \approx 8n_1 \frac{\Delta n d^2}{m^2},$$

где $\Delta n = n_1 - n_2$.

7. **ТМ-моды.** Выведите выражение для углов наклона ТМ-мод, аналогичное (8.22). С помощью компьютера постройте график, аналогичный рис. 8.12, для ТМ-мод в волноводе с $\sin \bar{\theta}_{кр} = 0{,}3$ и $\lambda/2d = 0{,}1$. Каково число ТМ-мод?

К РАЗДЕЛУ 8.3

1. **Моды прямоугольного диэлектрического волновода.** Прямоугольный диэлектрический волновод имеет квадратное сечение площадью 10^{-2} мм^2 и числовую апертуру NA = 0,1. Используйте (8.46) для построения графика числа ТЕ-мод как функции частоты ν. Сравните ваши результаты с рис. 8.14.

К РАЗДЕЛУ 8.4

1. **Коэффициенты связи между двумя слоями.**

а. Используйте (8.52) для определения коэффициентов связи между двумя идентичными пластинчатыми волноводами толщиной $d = 0{,}5$ мкм, расстояние между которыми $2a = 1$ мкм, показатели преломления $n_1 = n_2 = 1{,}48$, в среде с показателем преломления $n = 1{,}46$ при $\lambda_0 = 0{,}85$ мкм. Считайте, что оба волновода работают на ТЕ-моде с $m = 0$, и используйте результаты задачи 3 к разд. 8.2 для нахождения поперечного распределения поля.

б. Определите длину волновода, при которой он действует как 50 %-делитель.

ГЛАВА 9 ВОЛОКОННАЯ ОПТИКА

Чарльз Као (род. 1933) пропагандировал концепцию использования оптоволокна с низкими потерями в практических телекоммуникационных системах.

Филипп Джон Рассел (род. 1953) в 1991 г. изобрел фотонно-кристаллическое волокно, нашедшее многочисленные применения.

Оптическое волокно — это цилиндрический диэлектрический волновод, изготовленный из материала с малыми потерями, такого как кварцевое стекло. Оно имеет центральную **сердцевину**, по которой направляется свет, окруженную снаружи **оболочкой** с немного меньшим показателем преломления (рис. 9.1). Лучи света, падающие на границу раздела между сердцевиной и оболочкой под углами больше критического, испытывают полное внутреннее отражение и направляются по сердцевине, не проходя в оболочку. Лучи с большим отклонением от оси волокна теряют часть своей энергии при каждом отражении и не канализируются в сердцевине.

Рис. 9.1. Оптическое волокно представляет собой цилиндрический диэлектрический волновод с сердцевиной внутри и оболочкой снаружи

Замечательные достижения в производстве оптических волокон в течение последних двух десятилетий позволили достичь передачи света на 1 км по стеклянному оптоволокну с потерями всего ≈0,15 дБ (≈3,4 %) на длине волны максимума прозрачности. Благодаря таким низким потерям оптические волокна давно вытеснили медные коаксиальные кабели в качестве предпочтительного средства передачи в наземных и трансокеанских системах голосовой и цифровой связи.

В данной главе рассмотрены основные принципы передачи света в оптических волокнах. Эти принципы по существу те же, что и для планарных диэлектрических волноводов (см. гл. 8); наиболее заметное различие состоит в том, что оптические волокна имеют цилиндрическую геометрию. В обоих типах волноводов свет распространяется в виде мод. Каждая мода распространяется вдоль оси волновода с определенной постоянной распространения и групповой скоростью, сохраняя при этом поперечное распределение поля и поляризацию. В планарном диэлектрическом волноводе каждая мода представляется суммой многократно отражающихся внутри центрального слоя ТЕМ-волн с определенным углом наклона лучей к оси волновода. Этот подход приближенно применим и к описанию цилиндрического волновода. Если диаметр сердцевины мал, то поддерживается лишь одна мода, и волокно называется **одномодовым**.

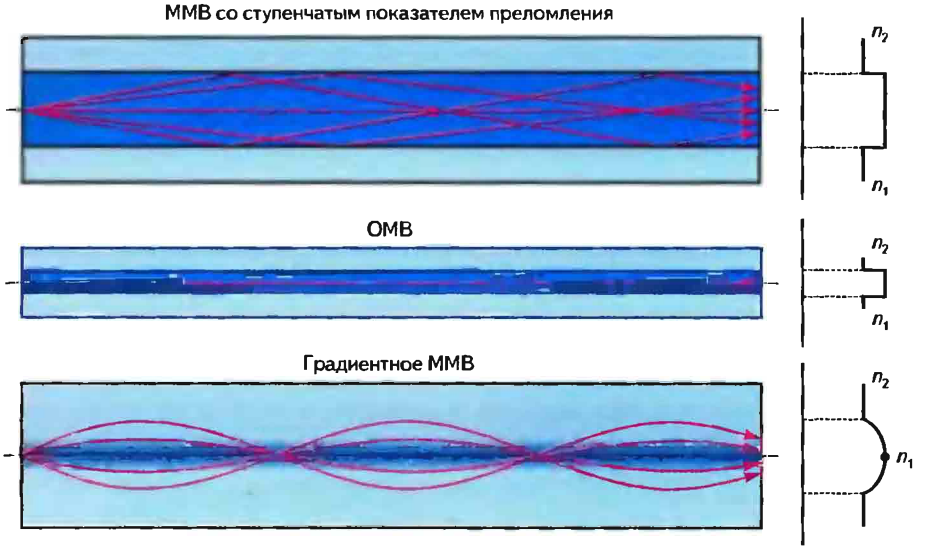


Рис. 9.2. Геометрия, профиль показателя преломления и типичный ход лучей в многомодовом волокне (ММВ) со ступенчатым профилем показателя преломления, одномодовом (ОМВ) и градиентном (ММВ) волокне

Волокна с большим диаметром сердцевины являются **многомодовыми**. Одна из трудностей передачи оптических сигналов по многомодовым волокнам связана с тем, что групповые скорости мод различны. Это приводит к разбросу времен прохождения и имеет своим результатом уширение импульса по мере

его прохождения через оптоволокно. Этот эффект, называемый модовой дисперсией, ограничивает частоту следования входных импульсов: интервал между запускаемыми импульсами должен быть достаточно большим, чтобы они не перекрывались на дальнем (выходном) конце волокна. Следовательно, модовая дисперсия ограничивает скорость действия коммуникационных систем на основе многомодовых оптических волокон.

Модовую дисперсию можно уменьшить за счет плавного уменьшения показателя преломления сердцевины от максимального значения в центре до минимального — на границе между сердцевиной и оболочкой. В этом случае волокно называется **градиентным** (GRIN), тогда как обычное волокно называется **волокном со ступенчатым профилем показателя преломления**. В градиентных волокнах скорость распространения растет с удалением от оси сердцевины, поскольку уменьшается показатель преломления. Хотя лучи с большим отклонением от оси проходят большее расстояние, они распространяются быстрее. Это позволяет выровнять времена распространения различных мод.

Итак, оптические волокна можно разделить на волокна со ступенчатым показателем преломления и градиентные волокна, а также на одно- и многомодовые (рис. 9.2).

О данной главе

Эта глава начинается с лучевого описания волокон со ступенчатым и градиентным профилем показателя преломления (см. разд. 9.1). Электродинамический подход, выявляющий природу оптических мод и одномодового режима, описан далее в разд. 9.2. Оптические свойства материалов для волокон (обычно это плавленый кварц), включая поглощение и материальную дисперсию, обсуждаются в разд. 9.3 вместе с модовой, волноводной и поляризационно-модовой дисперсией. Поскольку волокна обычно используются для передачи информации в виде оптических импульсов, в разд. 9.3 дается также краткое введение в физику распространения импульсов в волокнах. Микроструктурные (дырчатые) и фотонно-кристаллические волокна, имеющие более сложный профиль показателя преломления и необычные дисперсионные характеристики, вводятся в разд. 9.4. Мы вернемся к этой теме в гл. 22 и 24, посвященных, соответственно, оптике ультракоротких импульсов и волоконно-оптическим коммуникационным системам.

9.1. НАПРАВЛЯЕМЫЕ ЛУЧИ

9.1.1. Волокна со ступенчатым профилем показателя преломления

Волокно со ступенчатым показателем преломления представляет собой цилиндрический диэлектрический волновод, свойства которого определяются показателями преломления n_1 и n_2 и радиусами a и b сердцевины

и оболочки соответственно (см. рис. 9.1). Примеры стандартных отношений диаметра сердцевины к диаметру оболочки (в единицах мкм/мкм) таковы: $2a/2b$ = 8/125; 50/125; 62,5/125; 85/125 и 100/140. Показатели преломления сердцевины и оболочки отличаются лишь незначительно, поэтому их относительная разность мала:

$$\Delta \equiv \frac{n_1^2 - n_2^2}{2n_1^2} \approx \frac{n_1 - n_2}{n_1} \ll 1. \tag{9.1}$$

Большинство волокон, используемых в современных оптических системах связи, изготовлено из плавленого кварцевого стекла (SiO_2) высокой степени химической чистоты. Слабые изменения показателя преломления вызываются добавлением малых концентраций легирующих материалов, например титана, германия, бора. Показатель преломления выбирается от 1,44 до 1,46 в зависимости от длины волны, а Δ в типичных случаях лежит между 0,001 и 0,02.

Луч света в волокне со ступенчатым профилем показателя преломления направляется посредством полного отражения внутри сердцевины, если угол его падения на границу между сердцевиной и оболочкой больше критического угла $\theta_{кр} = \arcsin(n_2/n_1)$ и остается таковым при многократном отражении.

Меридиональные лучи

Меридиональными называются лучи, лежащие в плоскости, проходящей через ось волокна (рис. 9.3). Для них условие волноводного распространения особенно просто. Эти лучи пересекают ось волокна и отражаются в одной плоскости без изменения угла падения, т. е. ведут себя так же, как в планарном волноводе. Меридиональные лучи не выходят из сердцевины при условии, что угол θ, образуемый ими с осью волокна, меньше угла, дополнительного к критическому, т. е. если

$$\theta < \bar{\theta}_{кр} = \frac{\pi}{2} - \theta_{кр} = \arccos\left(\frac{n_2}{n_1}\right).$$

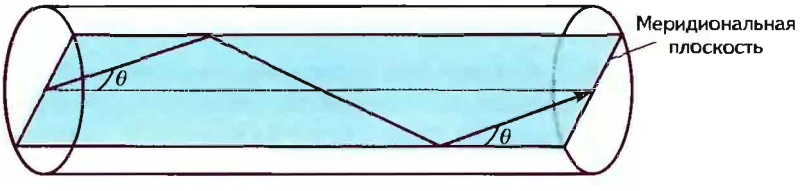


Рис. 9.3. Траектория меридионального луча лежит в плоскости, проходящей через ось волокна. Луч направляется волноводом, если $\theta < \bar{\theta}_{кр} = \arccos(n_2/n_1)$

Поскольку $n_1 \approx n_2$, $\sin\bar{\theta}_{кр}$ обычно мал, и направляемые лучи приблизительно параксиальны.

Косые лучи

Произвольный луч характеризуется плоскостью, в которой он находится и которая параллельна оси волокна, а также углом отклонения от направления оси в этой плоскости (рис. 9.4). Эта плоскость пересекает границу раздела между сердцевиной и оболочкой, образуя угол ϕ с нормалью к ней. Расстояние от оси до указанной плоскости обозначим через R. Таким образом, луч задается углом отклонения от направления оси θ и углом наклона его плоскости ϕ. Когда $\phi \neq 0$ ($R \neq 0$), луч называется косым. Для меридиональных лучей $\phi = 0$ и $R = 0$.

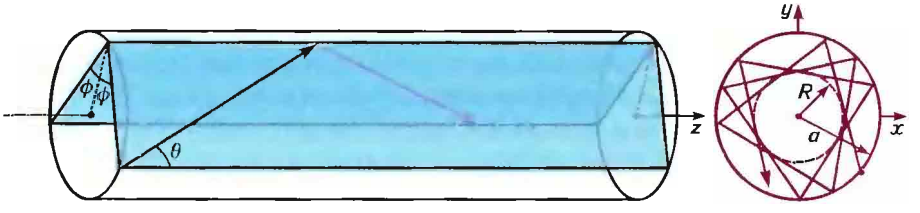


Рис. 9.4. Косой луч лежит в плоскости, отстоящей от оси на расстояние R. Луч характеризуется углами θ и ϕ. Его траектория — ломаная винтовая линия, лежащая внутри цилиндрического слоя с внутренним и внешним радиусами R и a соответственно. Проекция луча на поперечную плоскость — правильный многоугольник, не обязательно замкнутый

Косой луч многократно отражается, переходя в плоскости, образующие такой же угол ϕ с границей раздела сердцевины и оболочки; его траектория представляет собой ломаную винтовую линию, заключенную внутри цилиндрического слоя с внешним радиусом a и внутренним R, как показано на рис. 9.4. Проекция траектории на поперечную плоскость x—y представляет собой правильный многоугольник, не обязательно замкнутый. Условие полного внутреннего отражения косого луча заключается в том, что угол его отклонения θ от направления оси z должен быть меньше угла, дополнительного к критическому, т. е. $\theta < \bar{\theta}_{кр}$.

Числовая апертура

Луч, падающий из воздуха на торец волокна, становится направляемым лучом, если после преломления на торце сердцевины он образует с осью волокна угол θ, который меньше $\bar{\theta}_{кр}$. Как следует из рис. 9.5, *а*, если применить к границе между воздухом и сердцевиной закон Снелла, то угол θ_a в воздухе, соответствующий углу $\bar{\theta}_{кр}$ в сердцевине, получается из условия $1 \cdot \sin \theta_a = n_1 \sin \bar{\theta}_{кр}$, которое приводит к

$$\sin \theta_a = n_1 \sqrt{1 - \cos^2 \bar{\theta}_c} = n_1 \sqrt{1 - \left(\frac{n_2}{n_1}\right)^2} = \sqrt{n_1^2 - n_2^2}$$

(см. упражнение 1.6). Угол приема (входная угловая апертура) волокна равен

$$\theta_a = \arcsin \text{NA}, \tag{9.2}$$

где NA — числовая апертура волокна

$$\boxed{\text{NA} = \sqrt{n_1^2 - n_2^2} \approx n_1\sqrt{2\Delta},} \tag{9.3}$$

Числовая апертура

поскольку $n_1 - n_2 = n_1\Delta$ и $n_1 + n_2 \approx 2n_1$.

Входная угловая апертура волокна θ_a определяет конус падающих лучей, которые могут направляться волокном. Лучи, падающие под большими углами, проникают в волокно, но проходят вдоль него лишь небольшое расстояние, поскольку для них внутреннее отражение не является полным. Таким образом, числовая апертура описывает способность волокна собирать свет, что иллюстрируется рис. 9.5, *б*.

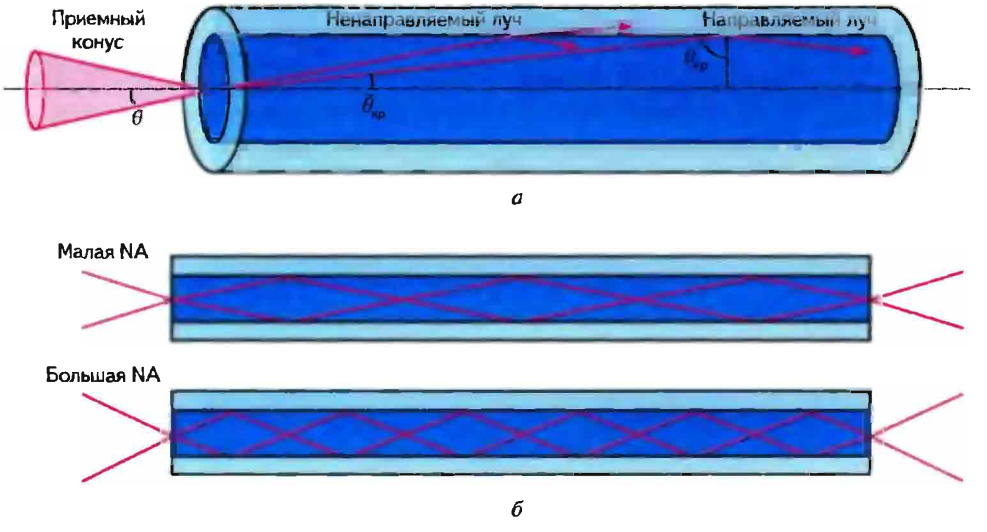


Рис. 9.5. Угол приема (входная угловая апертура) волокна θ_a (*а*). Лучи внутри приемного конуса становятся направляемыми за счет полного внутреннего отражения. Числовая апертура NA = sin θ_a. Углы θ_a и $\bar{\theta}_{кр}$ в типичном случае весьма малы, на рисунке они для наглядности преувеличены. Способность волокна собирать свет тем больше, чем больше NA (*б*)

Когда направляемые лучи достигают выходного торца волокна, они преломляются и выходят снова внутри конуса с углом θ_a. Таким образом, входная угловая апертура волокна представляет собой ключевой проектный параметр для ввода—вывода излучения.

Пример 9.1

Волокно с оболочкой и без нее

У волокна из кварцевого стекла с $n_1 = 1{,}46$ и $\Delta = (n_1 - n_2)/n_1 = 0{,}01$ угол, дополнительный к критическому, $\bar{\theta}_{кр} = \arccos(n_2/n_1) = 8{,}1°$ и входная угловая апертура $\theta_a = 11{,}9°$, что соответствует числовой апертуре NA = 0,206. Для сравнения, у волокна с сердцевиной из такого же кварцевого стекла с $n_1 = 1{,}46$ и оболочкой с намного меньшим показателем преломления $n_2 = 1{,}064$: $\bar{\theta}_{кр} = 43{,}2°$, $\theta_a = 90°$ и NA = 1. Лучи, падающие *со всех направлений*, становятся направляемыми, поскольку внутри сердцевины после преломления на ее торце они лежат в пределах конуса с образующим углом $\theta_a = 43{,}2°$. Аналогично, для волокна без оболочки ($n_2 = 1$) $\bar{\theta}_{кр} = 46{,}8°$, и лучи, падающие из воздуха под любым углом, становятся направляемыми. Хотя способность собирать свет у волокна без оболочки максимальная, такое волокно, как правило, не подходит для использования в качестве оптического волновода из-за большого числа мод, как будет пояснено ниже.

9.1.2. Градиентные волокна

Создание плавного профиля показателя преломления — остроумный способ уменьшить расплывание импульса из-за различия в групповых скоростях мод в многомодовом волокне. Сердцевина градиентного волокна имеет плавно меняющийся показатель преломления, который максимален в центре и уменьшается к оболочке, достигая наименьшего значения на границе. Поэтому фазовая скорость света минимальна в центре волокна и постепенно растет в направлении к границе раздела сердцевины и оболочки. Лучи, наиболее близкие к оси, проходят меньшее расстояние, но с меньшей скоростью. Лучи с большим наклоном, идущие более зигзагообразно, проходят большее расстояние, но большей частью в той области, где фазовая скорость выше. Таким образом, разница в длинах пути компенсируется противоположной разницей в фазовых скоростях. Как следствие, разброс времени прохождения между разными составляющими импульса уменьшается. В данном разделе мы рассмотрим распространение света в градиентных волокнах.

Показатель преломления сердцевины градиентного волокна $n(r)$ является функцией радиальной координаты r. Как показано на рис. 9.6, наибольшее значение $n(r)$ достигается на оси волокна, $n(0) = n_1$, а наименьшее — на расстоянии, равном радиусу сердцевины, $n(a) = n_2$. Показатель преломления оболочки постоянен и равен n_2.

Изменяемый профиль показателя преломления, обладающий этими основными свойствами, можно описать степенным законом

$$n^2(r) = n_1^2\left[1 - 2\left(\frac{r}{a}\right)^p \Delta\right], \quad r \le a, \tag{9.4}$$

где

$$\Delta = \frac{n_1^2 - n_2^2}{2n_1^2} \approx \frac{n_1 - n_2}{n_1}. \qquad (9.5)$$

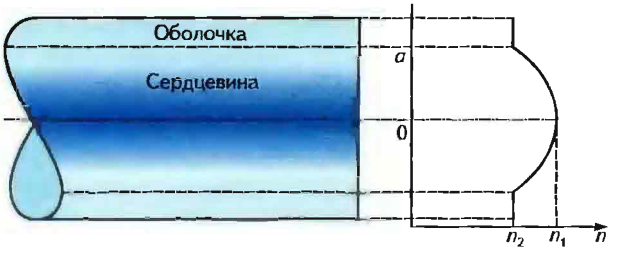


Рис. 9.6. Геометрия и профиль показателя преломления градиентного оптического волокна

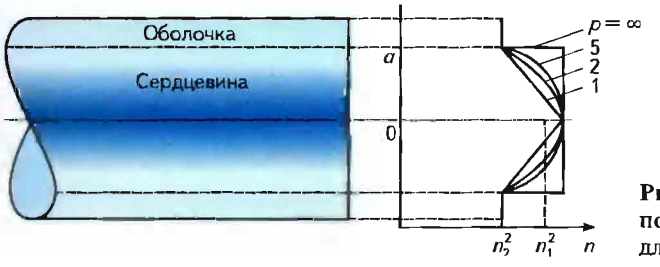


Рис. 9.7. Степенной профиль показателя преломления $n^2(r)$ для различных значений p

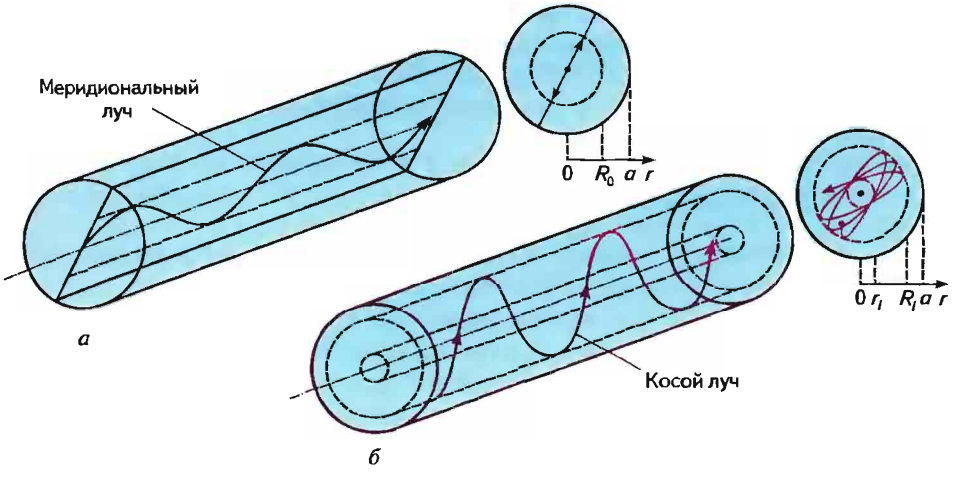


Рис. 9.8. Направляемые лучи в сердцевине градиентного волокна:

а — меридиональный луч ограничен меридиональной плоскостью в пределах цилиндра радиуса R_0; *б* — косой луч идет по винтовой траектории между двумя цилиндрическими поверхностями с радиусами r_l и R_l

Профильный параметр p определяет степень крутизны профиля. Как показано на рис. 9.7, $n(r)$ является линейной функцией при $p = 1$ и квадратичной — при

$p = 2$. Края профиля становятся круче с ростом p, и в пределе при $p \to \infty$ профиль становится ступенчатым. Таким образом, волокно со ступенчатым профилем показателя преломления является частным случаем градиентного волокна.

Прохождение световых лучей по градиентному волноводу с параболическим профилем показателя преломления обсуждается в разд. 1.3. В меридиональной плоскости лучи следуют плоским волнистым траекториям, тогда как косые лучи распространяются вдоль винтовых траекторий, точки поворота которых образуют цилиндрические каустические поверхности, как показано на рис. 9.8. Направляемые лучи локализованы в сердцевине и не достигают оболочки.

Числовую апертуру градиентного оптоволокна можно найти, определив наибольший угол наклона падающего луча, при котором он останется в пределах сердцевины и не достигнет оболочки. Для меридиональных лучей в градиентном волокне с параболическим профилем числовая апертура дается формулой (9.3) (см. упражнение 1.9).

9.2. НАПРАВЛЯЕМЫЕ ВОЛНЫ

Перейдем к построению электромагнитной теории распространения света в оптическом волокне. Мы будем искать электрическое и магнитное поля направляемых волн исходя из уравнений Максвелла и граничных условий, накладываемых цилиндрической диэлектрической сердцевиной и оболочкой. Как и для всех волноводов, существуют специальные решения этой задачи, называемые модами (см. приложение В), каждое из которых имеет определенную постоянную распространения, характеристическое распределение поля по поперечному сечению и два независимых поляризационных состояния. Поскольку точное решение представляет значительные трудности, будет использован ряд приближений.

Уравнение Гельмгольца

Оптическое волокно представляет собой диэлектрическую среду с показателем преломления $n(r)$. У волокна со ступенчатым профилем $n(r) = n_1$ в сердцевине ($r < a$) и $n(r) = n_2$ в оболочке ($r > a$). У градиентного волокна $n(r)$ — непрерывная функция в сердцевине и постоянная величина $n(r) = n_2$ в оболочке. Во всех случаях будем считать, что внешний радиус оболочки b достаточно велик, чтобы можно было считать толщину оболочки бесконечной при рассмотрении направляемого света в сердцевине и окрестности ее границы с оболочкой.

Каждая из компонент монохроматического электрического и магнитного полей подчиняется уравнению Гельмгольца

$$\nabla^2 U + n^2(r) k_0^2 U = 0,$$

где $k_0 = 2\pi/\lambda_0$. Это уравнение является точным для обеих областей волокна со ступенчатым профилем показателя преломления и приближенным — в сердцевине градиентного волокна, при условии, что $n(r)$ меняется медленно по сравне-

нию с длиной волны (см. разд. 5.3). В цилиндрических координатах (рис. 9.9) уравнение Гельмгольца записывается как

$$\frac{\partial^2 U}{\partial r^2} + \frac{1}{r}\frac{\partial U}{\partial r} + \frac{1}{r^2}\frac{\partial^2 U}{\partial \phi^2} + \frac{\partial^2 U}{\partial z^2} + n^2 k_0^2 U = 0, \tag{9.6}$$

где $U = U(r, \phi, z)$. Направляемые моды — это волны, распространяющиеся в направлении z с постоянной распространения β, поэтому зависимость U от z имеет вид $e^{-j\beta z}$. Они периодичны по углу ϕ с периодом 2π, поэтому зависимость от ϕ представим в виде гармонической функции $e^{-jl\phi}$, где l — целое число. Подставляя

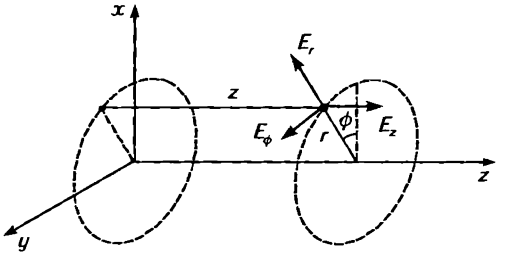


Рис. 9.9. Цилиндрическая система координат

$$U(r, \phi, z) = u(r)e^{-jl\phi}e^{-j\beta z}, \quad l = 0, \pm 1, \pm 2, \tag{9.7}$$

в (9.6), получаем обыкновенное дифференциальное уравнение для радиального профиля:

$$\frac{\mathrm{d}^2 u}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}u}{\mathrm{d}r} + \left(n^2(r)k_0^2 - \beta^2 - \frac{l^2}{r^2}\right)u = 0. \tag{9.8}$$

9.2.1. Волокна со ступенчатым профилем показателя преломления

Как мы убедились в подразд. 8.2.2, волна направляется (или связана) в волноводе, если постоянная ее распространения меньше волнового числа в сердцевине ($\beta < n_1 k_0$) и больше волнового числа в оболочке ($\beta > n_2 k_0$). Удобно ввести величины

$$k_T^2 = n_1^2 k_0^2 - \beta^2; \tag{9.9а}$$

$$\gamma^2 = \beta^2 - n_2^2 k_0^2 \tag{9.9б}$$

так, что у направляемых волн k_T^2 и γ^2 положительны, а k_T и γ действительны. Уравнение (9.8) запишем отдельно в сердцевине и в оболочке:

$$\frac{\mathrm{d}^2 u}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}u}{\mathrm{d}r} + \left(k_T^2 - \frac{l^2}{r^2}\right)u = 0, \quad r < a \quad \text{(сердцевина)}; \tag{9.10а}$$

$$\frac{\mathrm{d}^2 u}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}u}{\mathrm{d}r} - \left(\gamma^2 + \frac{l^2}{r^2}\right)u = 0, \quad r > a \quad \text{(оболочка)}. \tag{9.10б}$$

Уравнения (9.10) — хорошо известные дифференциальные уравнения, решением которых является семейство функций Бесселя. Исключая функции, стремящиеся к ∞ при $r \to 0$ в сердцевине или при $r \to \infty$ в оболочке, получаем ограниченные решения

$$u(r) \propto \begin{cases} J_l(k_T r), & r < a \quad \text{(сердцевина)}; \\ K_l(\gamma r), & r > a \quad \text{(оболочка)}, \end{cases} \tag{9.11}$$

где $J_l(x)$ — функция Бесселя первого рода порядка l, а $K_l(x)$ — модифицированная функция Бесселя второго рода порядка l. Функция $J_l(x)$ осциллирует подобно синусу или косинусу, но с убывающей амплитудой. Функция $K_l(x)$ при больших x экспоненциально затухает. Два примера радиального распределения $u(r)$ показаны на рис. 9.10.

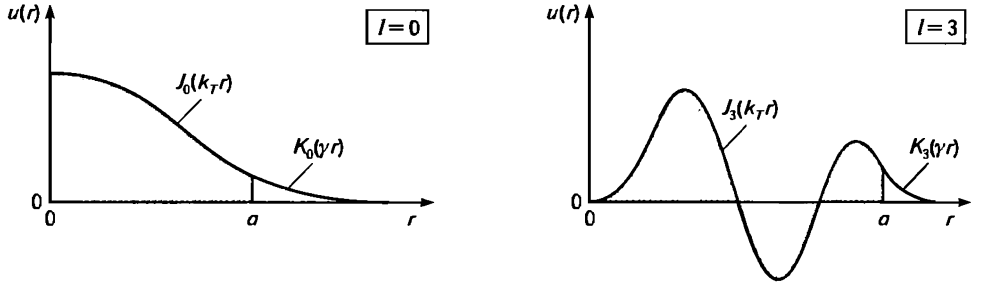


Рис. 9.10. Примеры радиальных распределений $u(r)$, даваемых формулами (9.11) при $l = 0$ и $l = 3$. Затененная часть находится внутри сердцевины, а незатененная — в оболочке. Параметры k_T и γ, а также коэффициенты пропорциональности в (9.11) выбраны так, чтобы функция $u(r)$ была непрерывной и имела непрерывную производную при $r = a$. Увеличение k_T и γ приводит к увеличению числа осцилляций $u(r)$

Параметры k_T и γ определяют скорость изменения $u(r)$ в сердцевине и оболочке соответственно. Большая величина k_T означает большее число осцилляций радиального распределения в оболочке. Большее значение γ означает более быстрый спад и меньшее проникновение волны в оболочку. Как видно из (9.9), сумма квадратов k_T и γ постоянна

$$k_T^2 + \gamma^2 = \left(n_1^2 - n_2^2\right) k_0^2 = (\mathrm{NA})^2 k_0^2, \tag{9.12}$$

поэтому если k_T растет, то γ убывает, и поле проникает глубже в оболочку. Для значений k_T, превосходящих NA k_0, величина γ становится мнимой, и волна перестает быть связанной в оболочке.

Параметр волокна V

Удобно нормировать k_T и γ путем введения величин

$$X = k_T a; \quad Y = \gamma a. \tag{9.13}$$

В соответствии с (9.12)

$$X^2 + Y^2 = V^2, \tag{9.14}$$

где $V = \mathrm{NA} \cdot k_0 a$, откуда

$$\boxed{V = 2\pi \frac{a}{\lambda_0} \mathrm{NA}.} \tag{9.15}$$

Параметр V

Важно помнить, что для направляемых волн X должно быть меньше V.

Как мы вскоре убедимся, V является важным параметром, который определяет число мод волокна и их постоянные распространения. Он называется параметром волокна, или V-параметром. Он прямо пропорционален отношению радиуса сердцевины к длине волны a/λ_0 и числовой апертуре NA. Выражение (9.15) не лишено сходства с формулой (8.25) для числа ТЕ-мод в планарном диэлектрическом волноводе.

Моды

Рассмотрим теперь граничные условия. Начнем с записи аксиальных компонент комплексных амплитуд электрического и магнитного полей E_z и H_z в виде (9.7). Условие их непрерывности на границе раздела сердцевины и оболочки $r = a$ даст соотношение между коэффициентами пропорциональности в (9.11), так что мы имеем лишь одно неизвестное для E_z и одно — для H_z. С помощью уравнений Максвелла

$$j\omega\varepsilon_0 n^2 \boldsymbol{E} = \nabla \times \boldsymbol{H}; \quad j\omega\mu_0 \boldsymbol{H} = \nabla \times \boldsymbol{E}$$

[см. (5.53) и (5.54)] оставшиеся четыре компоненты E_ϕ, H_ϕ, E_r и H_r выражаются через E_z и H_z. Непрерывность E_ϕ и H_ϕ даст еще два уравнения. Одно из них связывает два неизвестных коэффициента пропорциональности, входящих в E_z и H_z, а другое дает условие, которому должна удовлетворять постоянная распространения β. Это условие, называемое **характеристическим уравнением**, или **дисперсионным соотношением**, является уравнением относительно β, в которое в качестве известных параметров входят отношение a/λ_0, а также показатели преломления n_1 и n_2.

Для каждого азимутального индекса l характеристическое уравнение имеет множество решений, т. е. дискретных значений постоянной распространения β_{lm}, каждое из которых соответствует моде. Соответствующие значения k_T и γ, определяющие пространственное распределение поля в сердцевине и оболочке, находятся с помощью (9.9) и обозначаются k_{Tlm} и γ_{lm}. Таким образом, мода задается индексами l и m, характеризующими ее азимутальное и радиальное распределения соответственно. Функция $u(r)$ зависит как от l, так и от m; $l = 0$ соответствует меридиональным лучам. Более того, для каждой моды существует две независимые конфигурации векторов $\boldsymbol{E}$ и $\boldsymbol{H}$, соответствующие двум состояниям поляризации. Классификация и обозначения этих конфигураций, вообще говоря, довольно сложны (подробности приводятся в специальных книгах, см. список литературы в конце главы).

Характеристическое уравнение (слабонаправляющее волокно)

Большинство волокон являются слабонаправляюшими ($n_1 \approx n_2$ или $\Delta \ll 1$), так что направляемые лучи параксиальны, т. е. почти параллельны оси волокна. Продольные компоненты электрического и магнитного полей в этом случае намного меньше поперечных, и направляемые волны по своей природе приблизительно являются поперечными ТЕМ-волнами. В этом случае поляризация в поперечной плоскости характеризуется двумя ортогональными поляризационными состояниями. Линейно поляризованная мода (l, m) обычно обозначается как мода LP_{lm}. Обе ортогональных поляризационных составляющих моды (l, m) имеют одну и ту же постоянную распространения и одинаковое пространственное распределение.

Для слабонаправляющих волокон характеристическое уравнение, полученное согласно описанной выше процедуре, приближенно эквивалентно условиям непрерывности скалярной функции $u(r)$ (9.11) и ее первой производной на границе раздела $r = a$. Эти два условия удовлетворяются, если

$$\frac{(k_T a)J_l'(k_T a)}{J_l(k_T a)} = \frac{(\gamma a)K_l'(\gamma a)}{K_l(\gamma a)}. \tag{9.16}$$

Производные J' и K' от функций Бесселя удовлетворяют тождествам

$$J_l'(x) = \pm J_{l\mp 1}(x) \mp l\frac{J_l(x)}{x}; \tag{9.17}$$

$$K_l'(x) = -K_{l\mp 1}(x) \mp l\frac{K_l(x)}{x}. \tag{9.18}$$

Поставляя эти тождества в (9.16) и используя нормированные параметры $X = k_T a$ и $Y = \gamma a$, получаем характеристическое уравнение

$$\boxed{X\frac{J_{l\pm 1}(X)}{J_l(X)} = \pm Y\frac{K_{l\pm 1}(Y)}{K_l(Y)}, \quad Y = \sqrt{V^2 - X^2}.} \tag{9.19}$$

Характеристическое уравнение

При данных V и l характеристическое уравнение содержит единственную неизвестную переменную X. Заметим, что $J_{-l}(x) = (-1)^l J_l(x)$ и $K_{-l}(x) = K_l(x)$, так что уравнение остается неименным при замене l на $-l$.

Характеристическое уравнение можно решить графически, если построить графики его правой и левой частей как функций X и найти точки пересечения. Как показано на рис. 9.11 для $l = 0$, левая часть имеет множество ветвей, в то время как правая часть монотонно убывает с ростом X и обращается в нуль при $X = V (Y = 0)$. Следовательно, имеется множество точек пересечения в интервале $0 < X \leq V$. Каждая точка пересечения соответствует моде с определенным значением X. Эти значения обозначим X_{lm} в порядке возрастания X. После того,

как найдены X_{lm}, формулы (9.13), (9.9) и (9.11) позволяют определить соответствующие поперечные постоянные распространения k_{zlm}, постоянные затухания γ_{lm}, постоянные распространения β_{lm} и функции радиального распределения $u_{lm}(r)$. График на рис. 9.11 по своему характеру сходен с графиком на рис. 8.12 для нахождения мод планарного диэлектрического волновода.

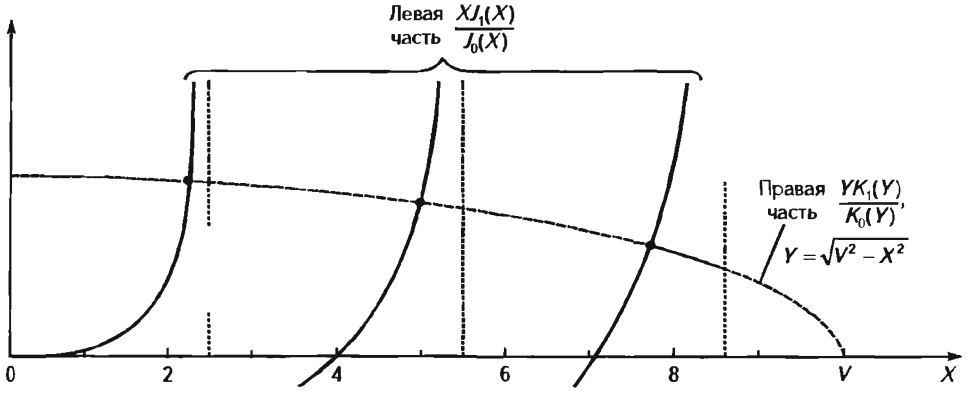
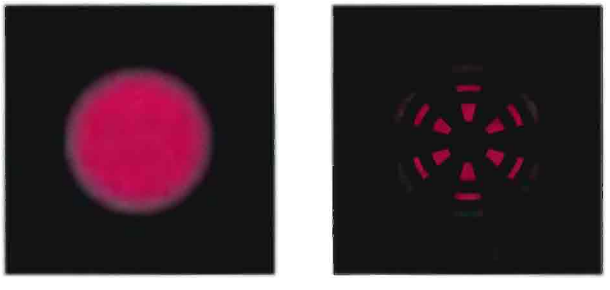


Рис. 9.11. Графическое построение для решения характеристического уравнения (9.19). Левая и правая части отложены как функции X. Решения даются точками пересечения. Левая часть имеет множество ветвей, пересекающих ось абсцисс при значениях корней $J_{l\pm1}(X)$. Правая часть пересекает каждую ветвь один раз и достигает оси абсцисс при $X = V$. Число мод равно числу корней $J_{l\pm1}(X)$, которое меньше, чем V. На данном графике $l = 0$, $V = 10$. а в формуле (9.19) можно пользоваться как знаком «+», так и знаком «−»



а *б*

Рис. 9.12. Распределение интенсивности для мод LP_{01} (*а*) и LP_{34} (*б*) в поперечной плоскости. Предполагается, что азимутальное распределение имеет вид $\cos(l\phi)$. Для основной моды LP_{01} распределение напоминает гауссов пучок, описанный в гл. 3

Каждая мода имеет свое радиальное распределение. Например, два радиальных распределения $u(r)$, показанных на рис. 9.10, соответствуют моде LP_{01} ($l = 0$, $m = 1$) в волокне с $V = 5$ и моде LP_{34} ($l = 3$, $m = 4$) в волокне с $V = 25$. Поскольку моды с (l, m) и ($-l$, m) имеют одинаковые постоянные распространения, интересно исследовать пространственное распределение их суперпозиции с равными весами. Комплексная амплитуда суммы этих мод пропорциональна $u_{lm}(r)\cos(l\phi)\exp(-j\beta_{lm}z)$. Интенсивность, пропорциональная $u_{lm}^2(r)\cos^2(l\phi)$, показана на рис. 9.12 для случая мод LP_{01} и LP_{34} (тех же самых, для которых распределение $u(r)$ показано на рис. 9.10).

Модовая отсечка

Из графического построения на рис. 9.11 очевидно, что по мере роста V число пересечений (мод) растет, так как левая часть характеристического уравнения (9.19) не зависит от V, в то время как график правой части с ростом V сдвигается вправо. При знаке «минус» в характеристическом уравнении ветви левой части пересекают ось абсцисс, когда $J_{l-1}(X) = 0$. Эти корни обозначим через x_{lm}, $m = 1, 2, \ldots$. Число мод равно числу корней, которые меньше V. Мода (l, m) может существовать, если $V > x_{lm}$. Мода достигает точки отсечки, когда $V = x_{lm}$. По мере убывания V мода $(l, m-1)$ также достигает своей точки отсечки, когда достигается новый корень, и т. д. Наименьший корень $J_{l-1}(X)$ есть $x_{01} = 0$ для $l = 0$, а следующий по возрастанию $x_{11} = 2{,}405$ для $l = 1$. Численные значения некоторых корней приведены в табл. 9.1.

Таблица 9.1. **Значения параметра V, соответствующие отсечке мод LP_{lm} низших порядков***

l/m	1	2	3
0	0	3,832	7,016
1	2,405	5,520	8,654

* Отсечка мод с $l = 0$ происходит в точках, где $J_{-1}(X) = -J_1(X)$ обращается в нуль. Отсечка мод с $l = 1$ соответствует корням $J_0(X)$ и т. д.

Когда $V < 2{,}405$, все моды, кроме фундаментальной моды LP_{01}, отсекаются. Волокно в этом случае действует как одномодовое. Условие одномодовости волокна, следовательно, имеет вид

$$\boxed{V < 2{,}405.} \qquad (9.20)$$

Условие одномодовости

Так как V пропорционально оптической частоте [см. (9.15)], условие отсечки основной моды (9.20) дает соответствующую частоту отсечки

$$\boxed{\nu_c = \frac{\omega_c}{2\pi} = \frac{1}{\mathrm{NA}}\frac{c_0}{2{,}61a}.} \qquad (9.21)$$

Частота отсечки

Для сравнения, в соответствии с (8.27) частота отсечки основной моды в диэлектрическом планарном волноводе толщины d равна

$$\nu_c = \frac{1}{\mathrm{NA}}\frac{c_0}{2d}.$$

Число мод

График числа мод M_l как функции V имеет вид лестницы, поскольку это число увеличивается на единицу при достижении каждого корня x_{lm} функции Бесселя $J_{l-1}(X)$. Подсчет полного числа мод M (для всех значений l)

в зависимости от V представлен на рис. 9.13. Каждый корень нужно учитывать дважды, поскольку для каждой моды с азимутальным индексом $l > 0$ существует соответствующая мода с $-l$, идентичная первой во всем, кроме противоположной полярности угла ϕ (такие моды соответствуют правым и левым винтовым траекториям лучей). В характеристическом уравнении это соответствует замене знака «минус» на «плюс». Кроме того, каждая мода имеет два состояния поляризации и должна учитываться дважды.

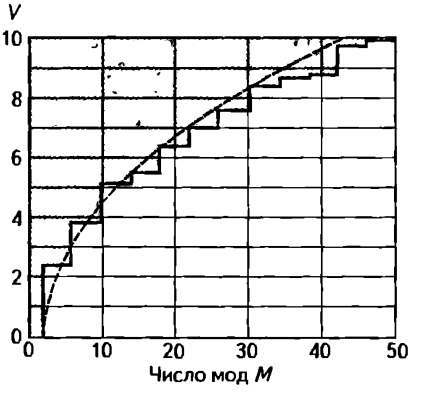


Рис. 9.13. Полное число мод M в зависимости от параметра волокна $V = 2\pi(a/\lambda_0)\mathrm{NA}$. Учитывается по две винтовые полярности для мод с $l > 0$ и по два состояния поляризации, приходящиеся на каждую моду. Для $V < 2{,}405$ существует единственная мода LP_{01} с двумя поляризациями. Штриховая линия — график функции $M = 4V^2/\pi^2 + 2$, дающей приближенно число мод при $V \gg 1$

Число мод (волокно с большим параметром V)

Для волокон с большим параметром V существует много корней функции $J_l(X)$ в интервале $0 < X < V$. Поскольку

$$J_l(X) \approx \left(\frac{2}{\pi X}\right)^{1/2} \cos\left[X - \left(l + \frac{1}{2}\right)\frac{\pi}{2}\right],$$

при $X \gg 1$ ее корни x_{lm} приближенно равны

$$x_{lm} = \left(l + \frac{1}{2}\right)\frac{\pi}{2} + (2m - 1)\frac{\pi}{2}.$$

Таким образом,

$$x_{lm} = \left(l + 2m - \frac{1}{2}\right)\frac{\pi}{2},$$

поэтому когда m велико, точки отсечки мод (l, m), являющиеся корнями $J_{l\pm1}(X)$, приближенно даются выражением

$$x_{lm} \approx \left(l + 2m - \frac{1}{2} \pm 1\right)\frac{\pi}{2} \approx (l + 2m)\frac{\pi}{2}, \quad l = 0,\ 1, \qquad m \gg 1. \tag{9.22}$$

Для фиксированного l эти корни расположены равномерно на расстоянии π друг от друга, поэтому число корней M_l удовлетворяет равенству

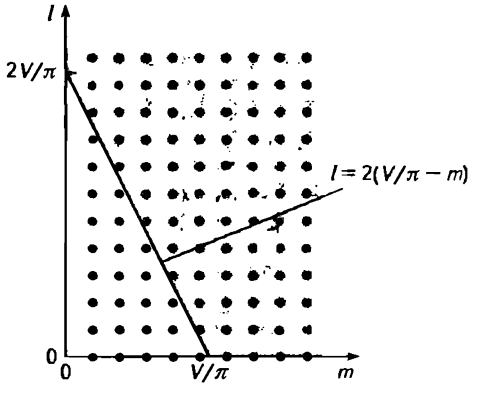
$$(l + 2M_l)\frac{\pi}{2} = V,$$

откуда

$$M_l \approx \frac{V}{\pi} - \frac{l}{2}.$$

Таким образом, M_l убывает линейно с ростом l, начиная с $M_l \approx V/\pi$ при $l = 0$ и кончая $M_l = 0$ при $l = l_{max}$, где $l_{max} = 2V/\pi$, как показано на рис. 9.14. Итак, полное число мод

$$M = \sum_{l=0}^{l_{max}} M_l = \sum_{l=0}^{l_{max}} \left(\frac{V}{\pi} - \frac{l}{2} \right).$$



Поскольку число членов в этой сумме предполагается большим, ее можно аппроксимировать площадью светлого треугольника на рис. 9.14:

$$M \approx \frac{1}{2} \frac{2V}{\pi} \frac{V}{\pi} = \frac{V^2}{\pi^2}.$$

Рис. 9.14. Индексы направляемых мод лежат в интервалах от $m = 1$ до $m \approx V/\pi - l/2$ и от $l = 0$ до $l \approx 2V/\pi$, чему соответствует светлый треугольник на рисунке

Учитывая две степени свободы, соответствующие положительным и отрицательным l, а также два состояния поляризации для каждой пары индексов (l, m), окончательно получаем

$$\boxed{M \approx \frac{4}{\pi^2} V^2.} \qquad (9.23)$$

Число мод ($V \gg 1$)

Заметим, что (9.23) справедливо только для больших V. Это приближенное число мод сравнивается с точным на рис. 9.13.

Выражение для числа мод круглого волновода

$$M \approx \left(\frac{4a}{\lambda_0} \right)^2 (\mathrm{NA})^2$$

вытекающее из (9.23), аналогично выражению для числа мод волновода прямоугольного сечения

$$M \approx \frac{\pi}{4} \left(\frac{2d}{\lambda_0} \right)^2 (\mathrm{NA})^2$$

[формула (8.46)].

Пример 9.2

Приблизительное число мод

Кварцевое волокно с $n_1 = 1{,}452$ и $\Delta = 0{,}01$ имеет числовую апертуру

$$\mathrm{NA} = \sqrt{n_1^2 - n_2^2} \approx n_1\sqrt{2\Delta} \approx 0{,}205.$$

Если $\lambda_0 = 0{,}85$ мкм и радиус сердцевины $a = 25$ мкм, то $V = 37{,}9$. Следовательно, существует примерно

$$M \approx \frac{4V^2}{\pi^2} \approx 585 \text{ мод.}$$

Если удалить оболочку, чтобы сердцевина находилась в непосредственном контакте с воздухом, $n_2 = 1$ и $\mathrm{NA} = 1$, тогда $V = 184{,}8$ и может существовать более 13 800 мод.

Постоянные распространения (волокна с большим параметром V)

Как отмечалось выше, постоянные распространения определяются путем решения дисперсионного уравнения (9.19) относительно X_{lm} с последующим использованием формул (9.9а) и (9.13) для получения

$$\beta_{lm} = \left(n_1^2 k_0^2 - \frac{X_{lm}^2}{a^2}\right)^{1/2}$$

В литературе приводится ряд приближенных формул для X_{lm}, которые можно применять в определенных пределах, но точных явных формул не существует.

Если $V \gg 1$, то наиболее грубым приближением будет считать, что X_{lm} равны значениям отсечки x_{lm}. Это эквивалентно предположению, что ветви на рис. 9.11 примерно являются вертикальными линиями, так что $X_{lm} \approx x_{lm}$. Поскольку $V \gg 1$, большинство корней тогда будут большими, поэтому можно применить приближение (9.22) и получить

$$\beta_{lm} \approx \sqrt{n_1^2 k_0^2 - (l + 2m)^2 \frac{\pi^2}{4a^2}}. \tag{9.24}$$

Так как

$$M \approx \frac{4}{\pi^2} V^2 = \frac{4}{\pi^2}(\mathrm{NA})^2 a^2 k_0^2 \approx \frac{4}{\pi^2}\left(2n_1^2\Delta\right) k_0^2 a^2, \tag{9.25}$$

формулы (9.24) и (9.25) дают

$$\beta_{lm} \approx n_1 k_0 \sqrt{1 - 2\frac{(l + 2m)^2}{M}\Delta}. \tag{9.26}$$

Поскольку Δ мало́, можно использовать приближение

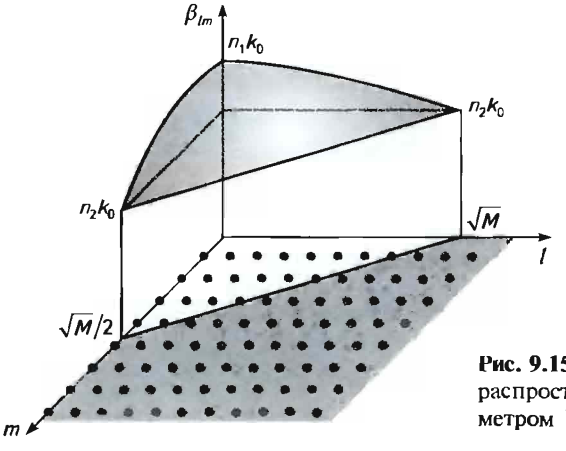
$$\sqrt{1+\delta} \approx 1 + \frac{\delta}{2}$$

для $|\delta| \ll 1$ и получить

$$\boxed{\beta_{lm} \approx n_1 k_0 \left[1 - \frac{(l+2m)^2}{M}\Delta\right]}. \qquad (9.27)$$

Постоянные распространения ($V \gg 1$)

$l = 0, 1, \ldots, \sqrt{M}$; $m = 1, 2, \ldots, \frac{1}{2}(\sqrt{M} - l)$



Так как $l + 2m$ меняется между 2 и $\approx 2V/\pi = \sqrt{M}$ (см. рис. 9.14), то β_{lm} меняется примерно между n_1k_0 и $n_1k_0(1 - \Delta) \approx n_2k_0$, как показано на рис. 9.15.

Рис. 9.15. Приближенные значения постоянных распространения β_{lm} волокна с большим параметром V как функции модовых индексов l и m

Групповые скорости (волокна с большим параметром V)

Для определения групповой скорости

$$v_{lm} = \frac{d\omega}{d\beta_{lm}}$$

моды (l, m) выразим β_{lm} как явную функцию ω путем подстановки

$$n_1 k_0 = \frac{\omega}{c}; \quad M = \frac{4}{\pi^2}\left(2n_1^2\Delta\right)k_0^2a^2 = \frac{8}{\pi^2}\frac{a^2m^2\Delta}{c_1^2}$$

в формулу (9.27) и предположим, что c_1 и Δ не зависят от ω. Дифференцирование $d\omega/d\beta_{lm}$ дает

$$v_{lm} \approx c_1\left[1 + \frac{(l+2m)^2}{M}\Delta\right]^1 \qquad (9.28)$$

Поскольку $\Delta \ll 1$, приближенное разложение $(1 + \delta)^{-1} \approx 1 - \delta$ при $|\delta| \ll 1$ приводит к выражению

$$\boxed{v_{lm} \approx c_1\left[1 - \frac{(l+2m)^2}{M}\Delta\right]}. \qquad (9.29)$$

Групповые скорости ($V \gg 1$)

Так как минимальное и максимальное значения $(l + 2m)$ равны 2 и $\sqrt{M}$, соответственно, и $M \gg 1$, групповая скорость меняется приблизительно между c_1 и $c_1(1 - \Delta) = c_1(n_1/n_2)$. Таким образом, групповые скорости мод низкого порядка приблизительно равны фазовой скорости в материале сердцевины, тогда как для мод высших порядков они меньше.

Относительная разница максимальной и минимальной групповых скоростей в грубом приближении равна Δ — относительной разнице показателей преломления сердцевины и оболочки волокна. Волокна с большим Δ хотя и обладают большой числовой апертурой и, следовательно, высокой способностью собирать свет, они обладают также большим количеством мод, большой модовой дисперсией и, следовательно, большой скоростью расплывания импульсов. Эти эффекты особенно остро проявляются при удалении оболочки.

9.2.2. Одномодовые волокна

Как обсуждалось ранее, волокно с радиусом сердцевины a и числовой апертурой NA действует как одномодовое с основной модой LP_{01}, если

$$V = 2\pi \frac{a}{\lambda_0} \text{NA} < 2{,}405.$$

Таким образом, одномодовый режим достигается малостью диаметра сердцевины и числовой апертуры (для чего показатель преломления n_2 должен быть близок к n_1) либо за счет достаточно низкой оптической частоты [ниже частоты отсечки $\nu_c = (1/\text{NA})(c_0/2{,}61a)$].

Основная мода имеет колоколообразное пространственное распределение, аналогичное гауссову (см. рис. 9.10 для $l = 0$ и 9.12, *а*). Она обеспечивает наибольшую локализацию световой энергии внутри сердцевины.

Пример 9.3

Одномодовый режим

Волокно из кварцевого стекла с $n_1 = 1{,}447$ и $\Delta = 0{,}01$ (NA = 0,205) работает на длине волны $\lambda_0 = 1{,}3$ мкм как одномодовое волокно, если $V = 2\pi(a/\lambda_0)\text{NA} < 2{,}405$, т. е. если диаметр сердцевины $2a < 4{,}86$ мкм. Если понизить Δ до 0,0025, то одномодовый режим будет поддерживаться для диаметра $2a < 9{,}72$ мкм.

Зависимость эффективного показателя преломления $n = \beta/k_0$ от параметра V для основной моды показана на рис. 9.16, *а*, а соответствующее дисперсионное соотношение (зависимость ω от β) — на рис. 9.16, *б*. По мере увеличения параметра V, т. е. когда увеличивается либо частота, либо диаметр сердцевины, эффективный показатель преломления n растет от n_2 до n_1. Этого можно было ожидать, поскольку мода в большей степени локализована в сердцевине при меньшей длине волны.

Использование одномодовых волокон в оптических системах связи имеет множество преимуществ. Как уже объяснялось ранее, моды в многомодовом волокне распространяются с различными групповыми скоростями, из-за чего короткий импульс многомодового света расплывается во времени, так его составляющие тратят различное время на прохождение одного и того же расстояния. Количественные измерения модовой дисперсии изучаются в подразд. 9.3.2. В одномодовом волокне, напротив, существует лишь одна мода с одной групповой скоростью, поэтому короткий импульс проходит по волокну без искажений, связанных с задержкой. Как объясняется в подразд. 9.3.2, расплывание импульса в одномодовых волокнах является следствием других механизмов дисперсии, которые гораздо слабее модовой дисперсии.

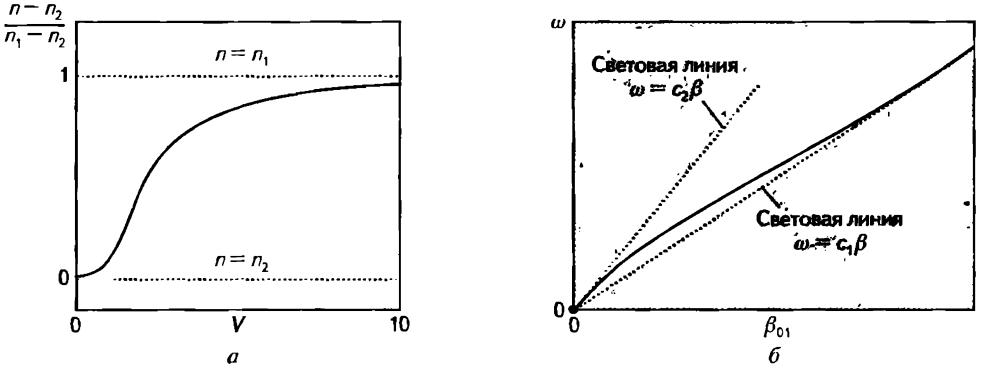


Рис. 9.16. Схематическая иллюстрация характеристик распространения основной моды LP_{01}: *а* — эффективный показатель преломления $n = \beta/k_0$ как функция параметра V; *б* — дисперсионное соотношение (зависимость ω от β)

Далее, как показано в подразд. 9.3.1, скорость затухания энергии в одномодовом волокне меньше, чем в многомодовом. Это, вместе с меньшей скоростью расплывания импульса, дает возможность существенно увеличить скорость передачи данных за счет замены многомодовых волокон одномодовыми. Эта тема рассматривается далее в гл. 22 и 24.

Другой трудностью при использовании многомодовых волокон является случайная интерференция мод. В результате неконтролируемых дефектов, механических напряжений и температурных флуктуаций каждая мода приобретает случайный фазовый сдвиг, вследствие чего сумма комплексных амплитуд дает интенсивность, случайно меняющуюся во времени и пространстве. Такая случайная зависимость известна как **модовый шум**, или **спекл-шум**. Эффект похож на замирание радиосигнала из-за прохождения по множеству путей. В одномодовом волокне путь всего один и поэтому модового шума нет.

Волокна, сохраняющие поляризацию света

В волокне с круговым поперечным сечением каждая мода имеет два независимых состояния поляризации с одной и той же постоянной распро-

странения. Так, основная мода LP_{01} в одномодовом слабонаправляющем волокне может быть поляризована в направлении x или y, причем обе ортогонально поляризованные волны имеют одинаковую постоянную распространения и групповую скорость.

В принципе между двумя поляризационными компонентами не должно быть обмена энергией. Если энергия от источника света сообщается исключительно одной поляризационной компоненте, то энергия и останется в ней. На практике, однако, слабые случайные дефекты и неконтролируемые механические напряжения приводят к случайной перекачке энергии между двумя поляризациями. Такая перекачка облегчается тем обстоятельством, что обе поляризационные компоненты имеют одну и ту же постоянную распространения и поэтому для них выполняется условие фазового синхронизма. Поэтому в общем случае свет, поляризованный линейно на входе в волокно, приобретает эллиптическую поляризацию на выходе. Хотя полная оптическая мощность остается постоянной (рис. 9.17), эллиптичность принимаемого света флуктуирует случайным образом во времени из-за флуктуаций механических напряжений в материале, температуры и длины волны источника света. Рандомизация деления мощности между двумя поляризационными компонентами не создает трудностей, если целью является только передача энергии света, при условии что на выходе собирается вся мощность.

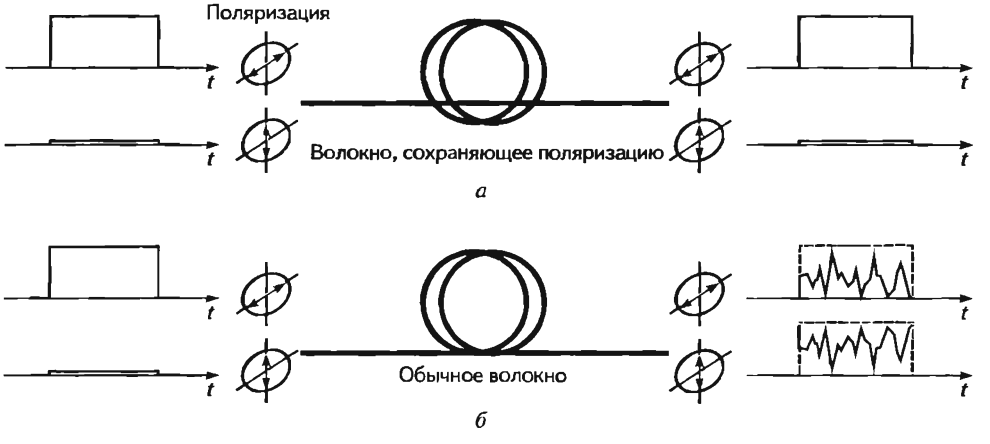


Рис. 9.17. Идеальное волокно, сохраняющее поляризацию (*а*); случайная перекачка энергии между двумя поляризационными компонентами (*б*)

Однако во многих областях применения волоконной оптики, например, в устройствах интегральной оптики, оптических сенсорах на основе интерферометрии и когерентной оптической связи, от волокна требуется передача комплексной амплитуды (величины и фазы) сигнала с определенной поляризацией. Для таких приложений необходимы волокна, сохраняющие поляризацию. Для конструирования такого волокна необходимо нарушить аксиальную симмет-

рию обычного волокна, например, придавая волокну эллиптический профиль или создавая анизотропию показателя преломления за счет механического напряжения. Тогда поляризационное вырождение снимается, и постоянные распространения поляризационных компонент становятся различными. Нарушение фазового синхронизма приводит к уменьшению эффективности взаимодействия компонент.

9.2.3. Квазиплоские волны в волокнах со ступенчатым и градиентным профилем показателя преломления

Для определения мод градиентного волокна нужно записать уравнение Гельмгольца (9.6) с $n = n(r)$, решить его, определив пространственное распределение компонент поля, а затем использовать уравнения Максвелла и граничные условия для получения характеристического уравнения, как это было сделано в случае волокна со ступенчатым профилем показателя преломления. Вообще говоря, это сложная процедура.

В данном разделе вместо этого мы используем приближенный подход, основанный на рассмотрении поля как квазиплоской волны, распространяющейся в сердцевине приблизительно по траектории оптического луча. Квазиплоская волна — это волна, локальные свойства которой такие же, как у плоской волны, но направление и амплитуда медленно меняются в процессе распространения. Такой подход позволяет сохранить простоту лучевого описания, но не только фазу, присущую волновой оптике, так что можно использовать условие самосогласования для нахождения постоянных распространения направляемых мод (как это делалось для планарного диэлектрического волновода в разд. 8.2). Этот приближенный метод, называемый методом Вентцеля—Крамерса—Бриллюэна (ВКБ), применим только к волокнам с большим числом мод (большим параметром).

Квазиплоские волны

Рассмотрим решение уравнения Гельмгольца (9.6) в виде квазиплоской волны (см. разд. 2.3)

$$U(\boldsymbol{r}) = a(\boldsymbol{r})\exp\left[-jk_0 S(\boldsymbol{r})\right], \tag{9.30}$$

где $a(\boldsymbol{r})$ и $S(\boldsymbol{r})$ — действительные функции координат, меняющиеся медленно по сравнению с длиной волны $\lambda_0 = 2\pi/k_0$. Из (2.1) известно, что $S(\boldsymbol{r})$ приближенно удовлетворяет уравнению эйконала $|\nabla S|^2 \approx n^2$ и что лучи распространяются в направлении градиента ∇S. Если положить $k_0 S(\boldsymbol{r}) = k_0 s(r) + l\phi + \beta z$, где $s(r)$ — медленно меняюшаяся функция r, то уравнение эйконала дает

$$\left(k_0 \frac{\mathrm{d}s}{\mathrm{d}r}\right)^2 + \beta^2 + \frac{l^2}{r^2} = n^2(r)k_0^2. \tag{9.31}$$

Локальная пространственная частота волны в радиальном направлении есть частная производная фазы $k_0 S(r)$ по r:

$$k_r = k_0 \frac{ds}{dr}, \tag{9.32}$$

поэтому (9.30) принимает вид

$$\boxed{U(r) = a(r) \exp\left(-j \int_0^r k_r dr\right) e^{-jl\phi} e^{-j\beta z}}, \tag{9.33}$$

Квазиплоская волна

а (9.31) дает

$$\boxed{k_r^2 = n^2(r) k_0^2 - \beta^2 - \frac{l^2}{r^2}}. \tag{9.34}$$

Определяя $k_\phi = l/r$ так, что $\exp(-jl\phi) = \exp(-jk_\phi r\phi)$ и $k_z = \beta$, получаем из (9.34)

$$k_r^2 + k_\phi^2 + k_z^2 = n^2(r) k_0^2.$$

Следовательно, квазиплоская волна имеет локальный волновой вектор $\boldsymbol{k}$ длиной $n(r)k_0$ с цилиндрическими компонентами (k_r, k_ϕ, k_z). Поскольку $n(r)$ и k_ϕ являются функциями r, k_r также зависит от координат. Направление $\boldsymbol{k}$ медленно меняется в зависимости от r (рис. 9.18), следуя винтовой линии, такой же как траектория косого луча, показанная на рис. 9.8, *б*.

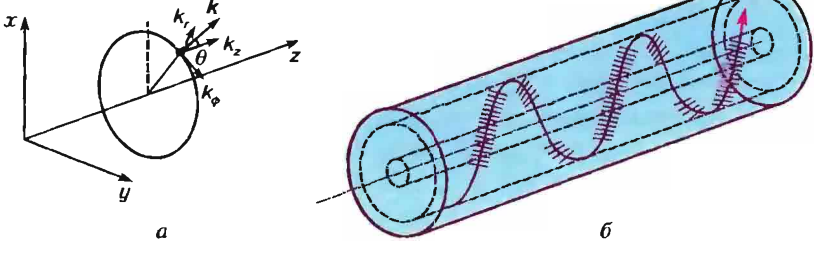


Рис. 9.18. Волновой вектор $\boldsymbol{k} = (k_r, k_\phi, k_z)$ в цилиндрической системе координат (*а*); квазиплоская волна, следующая по направлению луча (*б*)

Для определения области локализации волны в сердцевине найдем значения r, при которых k_r действительно, или $k_r^2 > 0$. Для данных l и β построим график

$$k_r^2 = n^2(r) k_0^2 - \frac{l^2}{r^2} - \beta^2$$

как функции от r. Сначала построим график зависимости $n^2(r)k_0^2$ от r (толстая сплошная кривая на рис. 9.19, *а*). Далее вычтем член l^2/r^2 (штриховая линия).

Значение β^2 отмечено тонкой сплошной вертикальной линией. Величина k_r^2 представляется отклонением штриховой линии от этой вертикали и на рисунке выделена темным. Области, в которых k_r^2 положительно и отрицательно, отмечены знаками «+» и «—» соответственно.

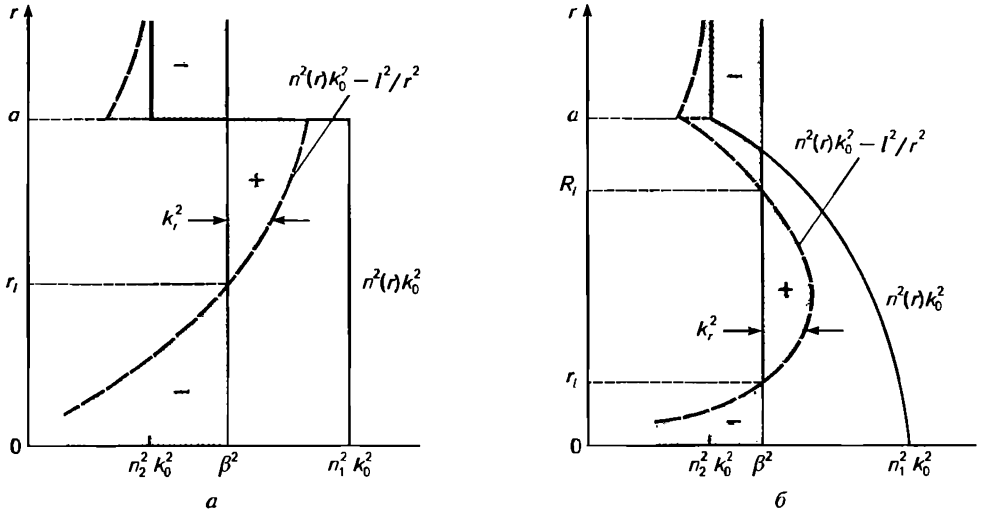


Рис. 9.19. Зависимость $n^2(r)k_0^2$, $n^2(r)k_0^2 - l^2/r^2$ и $k_r^2 = n^2(r)k_0^2 - l^2/r^2 - \beta^2$ от координаты r. При любом r k_r^2 есть ширина темной области, знаки «+» и «—» соответствуют положительным и отрицательным k_r^2:

а — волокно со ступенчатым профилем показателя преломления: k_r^2 положительно в области $r_l < r < a$; *б* — градиентное волокно: k_r^2 положительно в области $r_l < r < R_l$

Для волокна со ступенчатым профилем показателя преломления имеем $n(r) = n_1$ при $r < a$ и $n(r) = n_2$ при $r > a$. В этом случае квазиплоская волна направляется по сердцевине посредством отражений от границы между сердцевиной и оболочкой при $r = a$. Как видно из рис. 9.19, *а*, в этом случае область ограничения поля $r_l < r < a$, где

$$n_1^2 k_0^2 - \frac{l^2}{r^2} - \beta^2 = 0. \tag{9.35}$$

Волна идет по винтовой линии, как косой луч на рис. 9.4. В оболочке ($r > a$) и вблизи центра сердцевины ($r < r_l$) k_r^2 отрицательно и, значит, k_r — мнимое; волна в этом случае экспоненциально затухает. Заметим, что r_l зависит от β. Для больших β (или больших l) r_l так велико, что волна ограничена тонким цилиндрическим слоем вблизи границы сердцевины.

Для градиентного волокна, как показано на рис. 9.19, *б*, величина k_r действительна в области $r_l < r < R_l$, где r_l, R_l — корни уравнения

$$n^2(r)k_0^2 - \frac{l^2}{r^2} - \beta^2 = 0. \tag{9.36}$$

Получается, что волна в основном удерживается внутри цилиндрического слоя между поверхностями радиусов r_l и R_l, так же как траектория винтового луча, показанная на рис. 9.8, *б*.

Моды

Моды волокна определяются наложением условия самосогласования, согласно которому волна воспроизводит себя через один период винтового распространения от r_l до R_l и обратно. Азимутальная длина пути, соответствующая углу 2π, должна вносить фазовый сдвиг, равный целому числу 2π, т. е. $k_\varphi 2\pi r = 2\pi l$; $l = 0, \pm 1, \pm 2,$ Это условие, очевидно, выполняется, поскольку $k_\varphi = l/r$. Далее, радиальная длина полного обхода должна соответствовать фазовому сдвигу, кратному 2π, т. е.

$$2\int_{r_l}^{R_l} k_r \mathrm{d}r = 2\pi m, \quad m = 1, 2, \qquad M_l, \tag{9.37}$$

где $R_l = a$ для волокна со ступенчатым профилем показателя преломления. Это условие, аналогичное условию самосогласования (8.20) для планарных волноводов, приводит к характеристическому уравнению, из которого определяются значения постоянной распространения мод β_{lm}. Эти значения схематически показаны на рис. 9.20; мода с $m = 1$ имеет наибольшее значение β (приблизительно $n_1 k_0$), а мода с $m = M_l$ — наименьшее значение (приблизительно $n_2 k_0$).

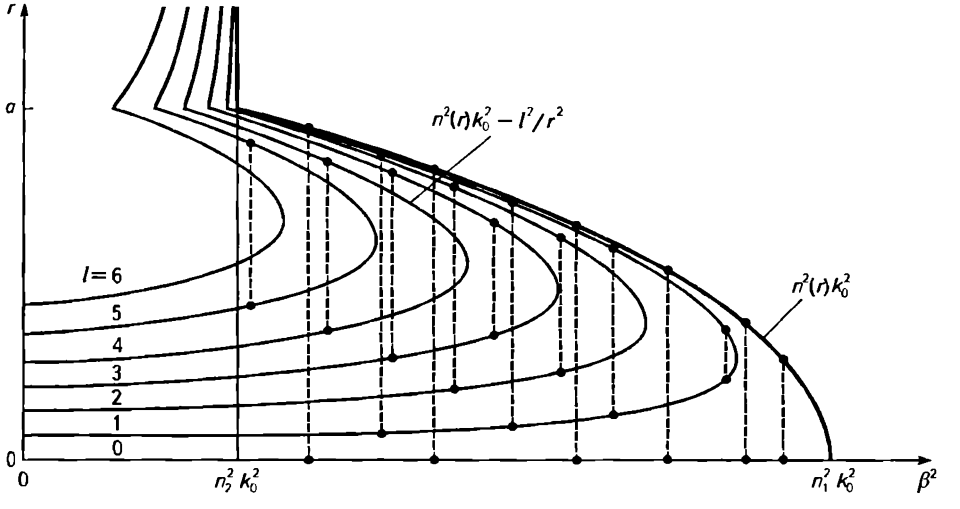


Рис. 9.20. Постоянные распространения и области ограничения мод волокна. Каждая кривая соответствует индексу l, принимающему на данном графике значения от 0 до 6. Каждая мода (соответствующая некоторому значению m) схематически отмечена двумя точками, соединенными вертикальной штриховой линией. Координаты точек отмечают радиальные границы моды r_l и R_l. По оси абсцисс отложены значения квадрата постоянной распространения β

Число мод

Полное число мод можно определить путем сложения чисел мод M_l для $l = 0, 1, ..., l_{max}$. Однако мы используем для этого другой метод. Сначала определим число q_β мод с постоянными распространения, превышающими данное значение β. Для каждого l число мод $M_l(\beta)$ с постоянными распространения, большими чем β, представляет собой число отрезков длиной 2π, укладывающихся на величине интеграла (9.37), т. е.

$$M_l(\beta) = \frac{1}{\pi}\int_{r_l}^{R_l} k_r \mathrm{d}r = \frac{1}{\pi}\int_{r_l}^{R_l} \sqrt{n^2(r)k_0^2 - \frac{l^2}{r^2} - \beta^2}\,\mathrm{d}r, \tag{9.38}$$

где r_l и R_l — радиальные границы моды, соответствующие постоянной распространения β и определяемые уравнением (9.36). Очевидно, что r_l и R_l зависят от β, и $R_l = a$ для волокна со ступенчатым профилем показателя преломления.

Полное число мод с постоянными распространения, превышающими β, тогда равно

$$q_\beta = 4\sum_{l=0}^{l_{max}(\beta)} M_l(\beta) \tag{9.39}$$

где $l_{max}(\beta)$ — максимальное значение l, дающее связанную моду с постоянной распространения больше β, т. е. значение l, для которого максимальное значение функции $n^2(r)k_0^2 - l^2/r^2$ превышает β^2. Итоговое суммарное число мод M есть q_β при $\beta = n_2k_0$. Множитель 4 в (9.39) учитывает две поляризации и две полярности угла ϕ, соответствующие правым и левым винтовым траекториям для каждого (l, m). Если число мод достаточно велико, можно заменить суммирование в (9.39) интегрированием, после чего

$$q_\beta \approx 4\int_0^{l_{max}(\beta)} M_l(\beta)\mathrm{d}l. \tag{9.40}$$

Для волокон со степенным профилем показателя преломления подставим (9.4) в (9.38), а затем в (9.40). Вычисление интеграла приводит к выражению

$$q_\beta \approx M\left[\frac{1-(\beta/n_1k_0)^2}{2\Delta}\right]^{p+2/p} \tag{9.41}$$

с обозначением

$$M \approx \frac{p}{p+2}n_1^2k_0^2a^2\Delta = \frac{p}{p+2}\frac{V^2}{2}, \tag{9.42}$$

где $\Delta = (n_1 - n_2)/n_1$ и $V = 2\pi(a/\lambda_0)\mathrm{NA}$ — параметр волокна. Так как $q_\beta \approx M$ при $\beta = n_2k_0$, M — действительно полное число мод.

Для волокна со ступенчатым профилем показателя преломления ($p \to \infty$) формулы (9.41) и (9.42) приобретают вид

$$q_\beta \approx M\left[\frac{1-(\beta/n_1k_0)^2}{2\Delta}\right]; \tag{9.43}$$

$$\boxed{M \approx \frac{V^2}{2}} \tag{9.44}$$

Число мод (ступенчатый показатель преломления)

соответственно. Это выражение для M почти такое же, как (9.23), $M \approx 4V^2/\pi^2 \approx 0{,}41V^2$, которое было получено в разд. 9.2 с использованием другого приближения.

Постоянные распространения

Постоянная распространения β_q для моды q получается обращением к (9.41):

$$\beta_q \approx n_1k_0\sqrt{1-2\left(\frac{q}{M}\right)^{p/(p+2)}\Delta}, \quad q = 1, 2, \ldots, M, \tag{9.45}$$

где индекс q_β заменен на q, а β — на β_q. Поскольку $\Delta \ll 1$, к формуле (9.45) можно применить приближение

$$\sqrt{1+\delta} \approx 1+\frac{\delta}{2},$$

справедливое при малых $|\delta| \ll 1$, что дает

$$\boxed{\beta_q \approx n_1k_0\left[1-\left(\frac{q}{M}\right)^{p/(p+2)}\Delta\right].} \tag{9.46}$$

Постоянные распространения

Следовательно, постоянная распространения β_q убывает от $\approx n_1k_0$ при $q = 1$ до n_2k_0 при $q = M$, как видно на рис. 9.21.

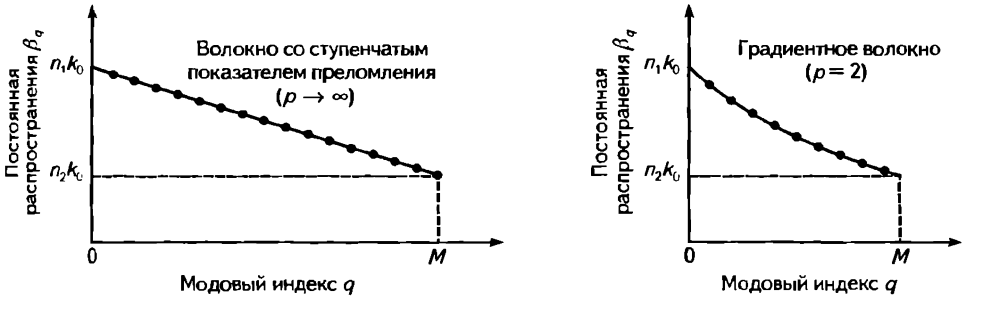


Рис. 9.21. Зависимость постоянной распространения β_q от модового индекса $q = 1, 2, \ldots M$ для волокна со ступенчатым профилем показателя преломления ($p \to \infty$) и для оптимального градиентного волокна ($p = 2$)

Для волокна со ступенчатым профилем показателя преломления ($p \to \infty$) выражение (9.45) сводится к виду

$$\beta_q \approx n_1 k_0 \left(1 - \frac{q}{M}\Delta\right), \tag{9.47}$$

Постоянные распространения (ступенчатый профиль)

Это выражение совпадает с (9.27), если индекс $q = 1, 2, ..., M$ заменить на $(l + 2m)^2$, с $l = 0, 1, ..., \sqrt{M}$; $m = 1, 2, ..., \frac{1}{2}(\sqrt{M} - 1)$.

Групповые скорости

Для определения групповой скорости $v_q = \mathrm{d}\omega/\mathrm{d}\beta_q$ запишем β_q как функцию ω путем подстановки (9.42) в (9.46), в результат которой подставим $n_1 k_0 = \omega/c_1$ и вычислим производную $v_q = (\mathrm{d}\beta_q/\mathrm{d}\omega)^{-1}$. С помощью приближения $(1 + \delta)^{-1} \approx 1 - \delta$, справедливого при $|\delta| \ll 1$, и предполагая, что c_1 и Δ не зависят от ω (т. е. игнорируя материальную дисперсию), получим

$$v_q \approx c_1 \left[1 - \frac{p-2}{p+2}\left(\frac{q}{M}\right)^{p/(p+2)}\Delta\right]. \tag{9.48}$$

Групповые скорости

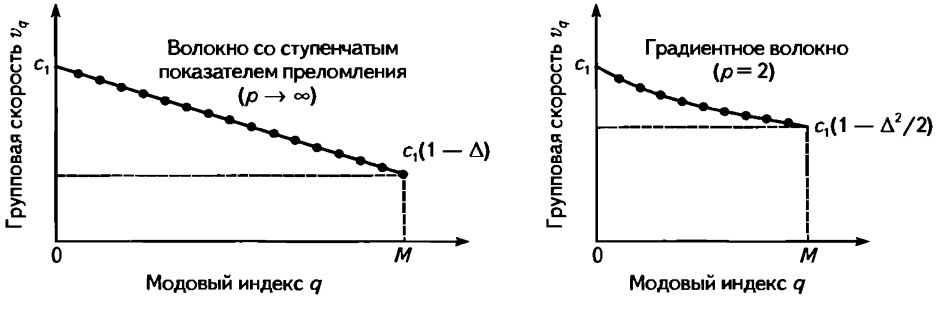


Рис. 9.22. Групповые скорости мод для волокна со ступенчатым профилем показателя преломления ($p \to \infty$) и оптимального градиентного волокна ($p = 2$)

Для волокна со ступенчатым профилем показателя преломления ($p \to \infty$) формула (9.48) дает

$$v_q \approx c_1 \left(1 - \frac{q}{M}\Delta\right), \tag{9.49}$$

что воспроизводит выражение (9.29). Таким образом, групповая скорость меняется примерно от c_1 до $c_1(1 - \Delta)$, как показано на рис. 9.22, *а*.

Оптимальный профиль показателя преломления

Выражение (9.48) показывает, что градиентный параметр профиля $p = 2$ дает групповую скорость $v_q \approx c_1$ для всех q, так что все моды распространяются с примерно одинаковой скоростью c_1. В этом состоит большое преимущество использования такого градиентного волокна в многомодовых волоконно-оптических системах связи.

Для улучшения точности определения групповой скорости вернемся к выводу v_q из (9.45) при $p = 2$. Производя разложение в ряд Тейлора с точностью не до двух, а до трех членов, т. е.

$$\sqrt{1+\delta} \approx 1 + \frac{1}{2}\delta - \frac{1}{8}\delta^2,$$

получаем

$$\boxed{v_q \approx c_1\left(1 - \frac{q}{M}\frac{\Delta^2}{2}\right).} \qquad (9.50)$$

Групповые скорости (градиентное волокно $p = 2$)

Таким образом, групповые скорости меняются примерно от c_1 при $q = 1$ до $c_1(1 - \Delta^2/2)$ при $q = M$. Сравнение с результатами для волокна со ступенчатым профилем показателя преломления показано на рис. 9.22. Разница в групповых скоростях у параболического градиентного волокна составляет $\Delta^2/2$, что значительно меньше, чем Δ у волокна со ступенчатым профилем показателя преломления. Таким образом, при идеальных условиях градиентное волокно снижает разницу групповых скоростей на множитель $\Delta/2$, что обеспечивает достижение поставленной цели выровнять скорости мод. Однако, поскольку анализ, приводящий к формуле (9.50), основан на ряде приближений, этот результат является лишь грубой оценкой и лишь частично достигается на практике.

Число мод M у градиентного волокна с параметром профиля p дается выражением (9.42). Для $p = 2$ оно превращается в

$$\boxed{M \approx \frac{1}{4}V^2.} \qquad (9.51)$$

Число мод (градиентное волокно $p = 2$)

Сравнивая этот результат с ранее полученным $M \approx V^2/2$ для волокна со ступенчатым профилем [формула (9.44)], видим, что у оптимального градиентного волокна число мод примерно вдвое меньше, чем у волокна со ступенчатым профилем при тех же параметрах n_1, n_2 и a.

9.3. ЗАТУХАНИЕ И ДИСПЕРСИЯ

Затухание и дисперсия ограничивают возможности использования оптических волокон в качестве каналов для передачи данных. Затухание, связанное с потерями различного рода, ограничивает величину передаваемой оп-

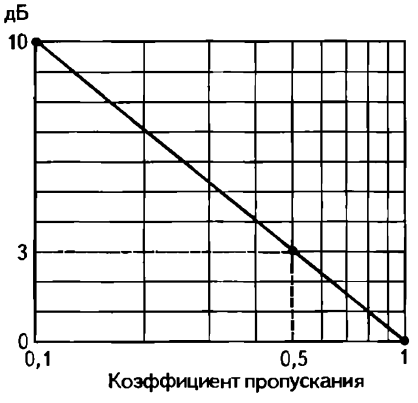
тической мощности. Дисперсия, ответственная за временное расплывание оптических импульсов, ограничивает частоту следования несущих информацию импульсов и, следовательно, скорость передачи информации.

9.3.1. Затухание

Коэффициент затухания

Мощность света, распространяющегося по оптическому волокну, экспоненциально убывает с расстоянием в результате поглощения и рассеяния. Соответствующий коэффициент затухания обычно определяется в единицах децибел на километр (дБ/км) и обозначается символом α

$$\alpha = \frac{1}{L} 10 \lg_{10} \frac{1}{\mathcal{T}}, \tag{9.52}$$



где $\mathcal{T} = P(L)/P(0)$ — коэффициент пропускания (отношение мощности прошедшего света к мощности падающего) для волокна длиной L км. Преобразование $\mathcal{T}$ в децибелы иллюстрирует рис. 9.23. Например, затухание в 3 дБ/км соответствует коэффициенту пропускания $\mathcal{T} = 0{,}5$ у волокна длиной $L = 1$ км.

Рис. 9.23. Значение отношения в дБ. Например, 3 дБ эквивалентно отношению 0,5; 10 дБ соответствует $\mathcal{T} = 0{,}1$; 20 дБ соответствует $\mathcal{T} = 0{,}01$

При прохождении света через каскад систем с потерями коэффициенты пропускания *перемножаются*. Поскольку в выражение (9.52) входит логарифм коэффициента пропускания, полное затухание в децибелах есть *сумма* затуханий всех элементов каскада в децибелах. При передаче на расстояние z км потери составляют αz дБ. Соответствующий коэффициент пропускания получается обращением (9.52) и равен

$$\frac{P(Z)}{P(0)} = 10^{-\alpha z/10} \approx e^{-0{,}23z} \tag{9.53}$$

Выражение (9.53) применяется, когда α выражено в единицах дБ/км. Если же коэффициент затухания выражен в 1/км, то

$$\frac{P(Z)}{P(0)} = e^{-\alpha z}, \tag{9.54}$$

где $\alpha \approx 0{,}23\alpha$. Для других не волоконных оптических системах коэффициент затухания α обычно выражается в 1/см, тогда ослабление мощности определяется формулой (9.54), где z выражено в см.

Поглощение

Коэффициент затухания плавленого кварца (SiO_2) сильно зависит от длины волны, как показано на рис. 9.24. Этот материал имеет две полосы сильного поглощения: в среднем инфракрасном диапазоне, связанном с колебательными переходами, и в ультрафиолетовой области, связанной с электронными и молекулярными переходами. «Хвосты» этих полос образуют окно в ближнем инфракрасном диапазоне спектра, в котором собственное поглощение мало.

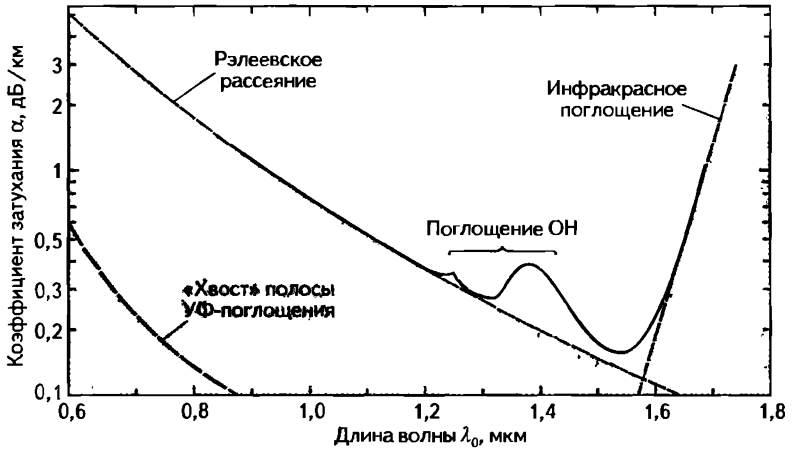


Рис. 9.24. Коэффициент затухания для кварцевого стекла в зависимости от длины волны λ_0. При 1,3 мкм имеется локальный минимум ($\alpha \approx 0{,}3$ дБ/км) и при 1,55 мкм — абсолютный минимум ($\alpha \approx 0{,}15$ дБ/км)

Рассеяние

Рэлеевское рассеяние — еще один собственный эффект, вносящий вклад в ослабление света в стекле. Случайно расположенные изменения положения молекул в самом стекле создают случайные неоднородности показателя преломления, которые действуют как крошечные рассеивающие центры. Амплитуда рассеянного света пропорциональна, где — угловая частота света[1]. Интенсивность рассеянного света тогда пропорциональна ω^2 или ω, так что короткие волны рассеиваются сильнее, чем длинные. Поэтому синий свет больше рассеивается, чем красный (аналогичный эффект — молекулярное рассеяние солнечного света в атмосфере — является причиной того, что небо выглядит голубым).

Функциональная зависимость интенсивности рэлеевского рассеяния, которая убывает как $1/\lambda_0^4$, известна как **закон обратной четвертой степени Рэлея**. В видимом диапазоне рэлеевское рассеяние света является более значитель-

[1] Рассеивающая среда создает плотность поляризации $\mathcal{P}$, которая соответствует источнику излучения, пропорционального $d^2\mathcal{P}/dt^2 = -\omega^2\mathcal{P}$ [см. (5.40)].

ным источником потерь, чем «хвост» полосы УФ-поглощения, как видно из рис. 9.24. Однако для длин волн больше 1,6 мкм рэлеевское рассеяние становится пренебрежимо малым по сравнению с инфракрасным поглощением.

Мы заключаем, что окно прозрачности кварцевого стекла ограничено рэлеевским рассеянием с коротковолновой стороны и инфракрасным поглощением с длинноволновой стороны (см. рис. 9.24). Оптические системы связи специально проектируются для работы в этом окне.

Примесные эффекты

Кроме указанных собственных эффектов существуют примесные полосы поглощения, обусловленные наличием примесей, главным образом, металлов ионов и радикалов ОН, связанных с растворенными в стекле парами воды. Большинство металлических примесей легко удаляется, однако примесь ОН удалить сложнее. Лишь совсем недавно стали доступны специальные волокна с пониженным поглощением ОН. В общем случае рабочие длины волн для волокна выбираются с таким расчетом, чтобы не попасть в полосы поглощения ОН.

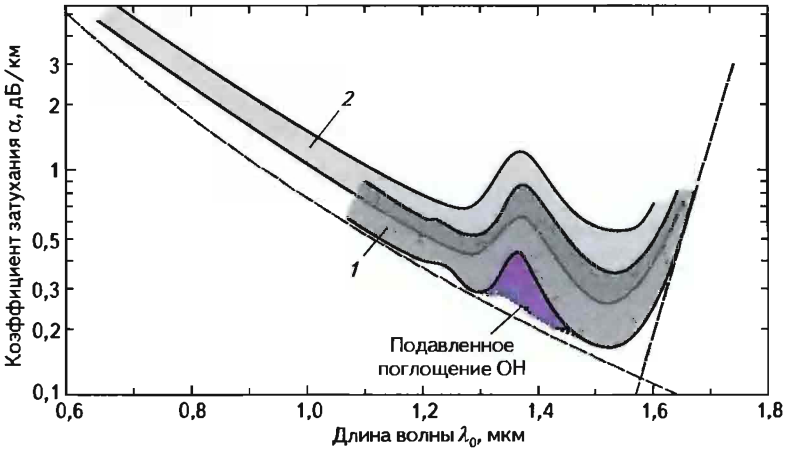


Рис. 9.25. Области коэффициентов затухания для одномодовых (*1*) и многомодовых (*2*) волокон из плавленого кварца

Потери из-за рассеяния света также увеличиваются при добавлении легирующих примесей, которые обычно применяются для получения градиента показателя преломления. Коэффициент затухания света, передаваемого по стеклянному волокну, зависит от поглощения и рассеяния в материале как сердцевины, так и оболочки. Каждая мода имеет свою глубину проникновения в оболочку, из-за чего лучи проходят разные эффективные расстояния, а коэффициент затухания становится зависящим от моды. В общем случае затухание выше для мод более высокого порядка. Поэтому одномодовые волокна, как правило, имеют меньший коэффициент затухания, чем многомодовые волокна (рис. 9.25). Потери возникают также из-за малых случайных изменений геометрии волокна, а также его изгибов.

Альтернативы кварцевому стеклу

Помимо кварцевого стекла, ряд материалов исследуется с точки зрения потенциальных применений в оптоволоконных системах среднего инфракрасного диапазона. Сюда относятся легированные тяжелыми металлами фторидные, германатные и халькогенидные стекла. У этих материалом полосы инфракрасного поглощения расположены дальше в глубь ИК-диапазона, чем у кварцевого стекла, поэтому появляется возможность работать на более длинных волнах, где слабее рэлеевское рассеяние (пропорциональное $1/\lambda_0^4$). В частности, у волокон из фторидного стекла с тяжелыми металлами предсказываемое оптическое затухание примерно в 10 раз меньше, чем у волокон из кварцевого стекла, и достигает минимального значения ≈0,01 дБ/км при $\lambda_0 \approx 2{,}5$ мкм. Однако примесные механизмы потерь, как правило, преобладают над собственными, а новые материалы не так прочны, как кварцевое стекло. Хотя квантово-каскадные лазеры обеспечивают работу в среднем ИК-диапазоне при комнатных температурах, в этом диапазоне, вообще говоря, нет высокоэффективных фотоприемников.

9.3.2. Дисперсия

Когда короткий импульс распространяется по оптическому волокну, его энергия «рассеивается» во времени, так что импульс приобретает большую длительность. Существует пять главных источников дисперсии в оптических волокнах:

- модовая дисперсия;
- материальная дисперсия;
- волноводная дисперсия;
- поляризационно-модовая дисперсия;
- нелинейная дисперсия.

Как будет показано далее, совместный вклад этих эффектов в расплывание импульсов не обязательно аддитивен.

Модовая дисперсия

Модовая дисперсия возникает в многомодовых волокнах из-за различия групповых скоростей различных мод. Одиночный импульс света, входящий в M-модовое волокно при $z = 0$, распадается на M импульсов, у которых относительная задержка растет как функция z. Для волокна длиной L временные задержки, порожденные различными скоростями, составляют $\tau_q = L/v_q$, $q = 1, \ldots, M$, где v_q — групповая скорость моды q. Если $v_{\min}$ и $v_{\max}$ — наименьшая и наибольшая групповые скорости соответственно, то принимаемый импульс расплывается по временному интервалу $L/v_{\min} - L/v_{\max}$. Поскольку моды обычно возбуждаются неодинаково, форма принимаемого импульса в общем случае имеет плавную огибающую (рис. 9.26). Оценка полной длительности импульса

в предположении треугольной огибающей и при определении ширины по полувысоте составляет

$$\sigma_\tau = \frac{1}{2}\left(\frac{L}{v_{\min}} - \frac{L}{v_{\max}}\right),$$

что представляет собой время отклика волокна, обусловленное модовой дисперсией.

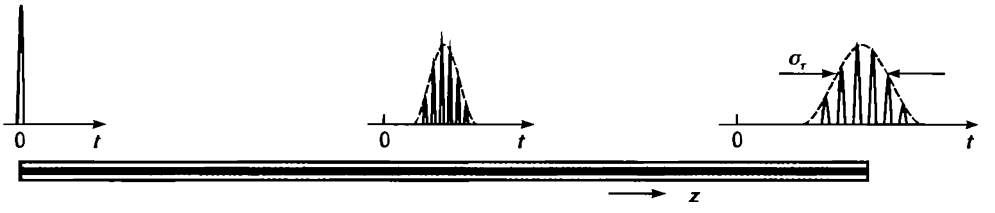


Рис. 9.26. Расплывание импульса, вызванное модовой дисперсией

В волокне со ступенчатым профилем показателя преломления и большим числом мод, $v_{\min} \approx c_1(1 - \Delta)$ и $v_{\max} \approx c_1$ (см. подразд. 9.2.3 и рис. 9.22, *а*). Поскольку $(1 - \Delta)^{-1} \approx 1 + \Delta$ при $\Delta \ll 1$, время отклика превращается в долю $\Delta/2$ времени задержки L/c_1:

$$\boxed{\sigma_\tau \approx \frac{L}{c_1}\frac{\Delta}{2}.} \qquad (9.55)$$

Время отклика (многомодовое ступенчатое волокно)

Модовая дисперсия значительно меньше у градиентных волокон, чем у волокон со ступенчатым профилем показателя преломления, поскольку групповые скорости у градиентных волокон меньше различаются и разница времен задержки различных мод $\tau_q = L/v_q$ меньше. В подразд. 9.2.3 и на рис. 9.22, *б* показано, что у градиентного волокна с оптимальным профилем показателя преломления и большим числом мод $v_{\max} \approx c_1$ и $v_{\min} \approx c_1(1 - \Delta^2/2)$. В этом случае время отклика на множитель $\Delta/2$ меньше, чем у волокна со ступенчатым профилем:

$$\boxed{\sigma_\tau \approx \frac{L}{c_1}\frac{\Delta^2}{4}.} \qquad (9.56)$$

Время отклика (градиентное волокно)

Пример 9.4

Скорость расплывания многомодового импульса

В волокне со ступенчатым профилем показателя преломления при $\Delta = 0{,}01$ и $n_1 = 1{,}46$ импульсы расплываются со скоростью примерно

$$\frac{\sigma_\tau}{L} = \frac{\Delta}{2c_1} = n_1\frac{\Delta}{2c_0} \approx 24 \text{ нс/км}.$$

В 100-километровом волокне импульс расплывается до ≈2,4 мкс. Если в этом волокне подобрать оптимальный градиентный профиль показателя преломления, то скорость расплывания импульса будет примерно $n_1\Delta^2/4c_0 \approx 122$ пс/км, т. е. значительно ниже.

Уширение импульса из-за модовой дисперсии пропорционально длине волокна L как у ступенчатых, так и у градиентных волокон. Однако из-за взаимодействия мод эта зависимость может меняться, если длина волокна больше некоторой критической величины. Взаимодействие происходит между модами, имеющими почти одинаковые постоянные распространения, вследствие малых дефектов волокна, таких как случайные неровности поверхности или неоднородности показателя преломления. За счет этого энергия перекачивается из одной моды в другую. При определенных условиях время отклика σ_τ волокна с взаимодействующими модами пропорционально L для малых длин волокна и $\sqrt{L}$, когда превышается критическая длина, после чего импульсы расплываются с уменьшенной скоростью[2].

Материальная дисперсия

Стекло — диспергирующий материал, т. е. его показатель преломления является функцией длины волны. Как обсуждалось в разд. 5.6, оптический импульс распространяется в диспергирующей среде, имеющей показатель преломления n, с групповой скоростью $v = c_0/N$, где

$$N = n - \lambda_0 \frac{\mathrm{d}n}{\mathrm{d}\lambda_0}.$$

Поскольку импульс представляет собой волновой пакет, состоящий из набора компонент с различными длинами волн, каждая из которых распространяется со своей групповой скоростью, его ширина увеличивается. Временная длительность оптического импульса со спектральной шириной σ_λ (нм) после прохождения расстояния L в диспергирующем материале представляет собой

$$\sigma_\tau = \left|\frac{\mathrm{d}}{\mathrm{d}\lambda_0}\frac{L}{v}\right|\sigma_\lambda = \left|\frac{\mathrm{d}}{\mathrm{d}\lambda_0}\frac{LN}{c_0}\right|\sigma_\lambda.$$

Это приводит к времени отклика [см. (5.125), (5.130) и (5.131)]

$$\boxed{\sigma_\tau = |D_\lambda|\sigma_\lambda L,} \qquad (9.57)$$

Время отклика (материальная дисперсия)

где коэффициент материальной дисперсии D_λ

$$D_\lambda = -\frac{\lambda_0}{c_0}\frac{\mathrm{d}^2 n}{\mathrm{d}\lambda_0^2}. \qquad (9.58)$$

---

[2] См., например, *Midwinter J.E.* Optical Fibers for Transmission. Wiley, 1979; Krieger, переиздание, 1992.

Время отклика линейно растет с расстоянием L. Обычно L измеряется в км, σ_τ — в пс, а σ_λ — в нм, так что имеет размерность D_λ пс/(км · нм). Такой тип дисперсии называется **материальной дисперсией**.

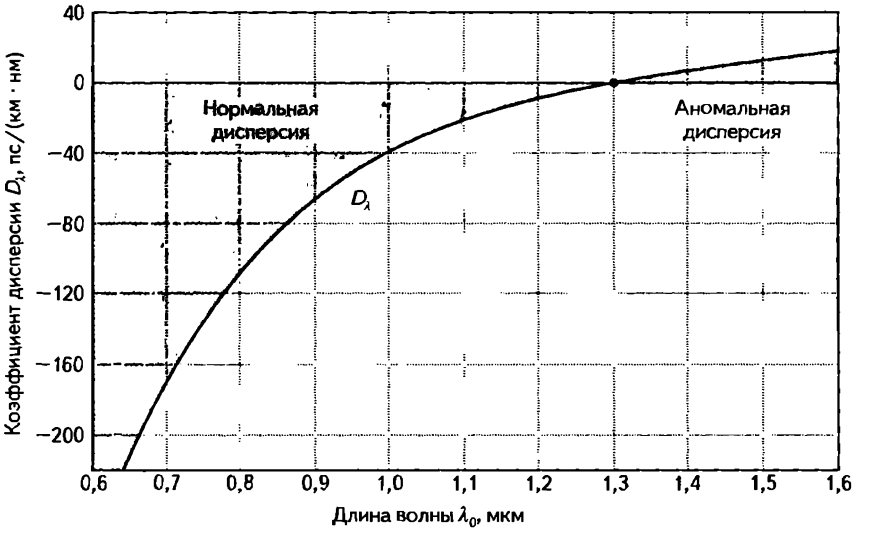


Рис. 9.27. Коэффициент дисперсии D_λ для волокна из кварцевого стекла как функция длины волны λ_0. Результат похож, но отличается от результата для плавленого кварца (см. рис. 5.28)

Зависимость коэффициента дисперсии D_λ от длины волны для волокна из кварцевого стекла показана на рис. 9.27. На длинах волн менее 1,3 мкм коэффициент дисперсии отрицателен, так что волновые пакеты из длинных волн распространяются быстрее, чем коротковолновые. Например, на длине волны $\lambda_0 = 0{,}87$ мкм коэффициент дисперсии D_λ составляет примерно −80 пс/км нм. С другой стороны, при $\lambda_0 = 1{,}55$ мкм $D_\lambda \approx +17$ пс/км нм. При $\lambda_0 \approx 1{,}312$ мкм коэффициент дисперсии исчезает, так что σ_τ в формуле (9.57) также исчезает. Более точное выражение для σ_τ, учитывающее расплывание спектральной ширины σ_λ вблизи $\lambda_0 = 1{,}312$ мкм, дает небольшую, но отличную от нуля ширину.

Пример 9.5

Уширение импульса, связанное с материальной дисперсией

Коэффициент дисперсии D_λ для кварцевого стекла примерно равен −80 пс/(км нм) при $\lambda_0 = 0{,}87$ мкм. Для источника со спектральной шириной $\sigma_\lambda = 50$ нм (например, светодиода) скорость расплывания импульса в одномодовом волокне без других источников дисперсии равна $|D_\lambda|\sigma_\lambda = 4$ нс/км. Следовательно, импульс света, проходящий по волокну расстояние $L = 100$ км, приобретает ширину $\sigma_\tau = |D_\lambda|\sigma_\lambda L = 0{,}4$ мкс. Таким образом, время отклика волокна составляет $\sigma_\tau = 0{,}4$ мкс. Другой пример: импульс с более узкой спектральной полосой $\sigma_\lambda = 2$ нм (например, генерируемый лазерным диодом) на 1,3 мкм, где

коэффициент дисперсии равен 1 пс/км нм, расплывается со скоростью всего 2 пс/км. В этом случае 100-километровое волокно имеет существенно меньшее время отклика $\sigma_\tau = 0{,}2$ нс.

Сочетание материальной и модовой дисперсий

Влияние материальной дисперсии на уширение импульсов в многомодовых волокнах можно определить, если вернуться к исходным уравнениям для постоянных распространения β_q и найти групповые скорости $v_q = (\mathrm{d}\beta_q/\mathrm{d}\omega)^{-1}$ с n_1 и n_2, заданными как функции ω. Рассмотрим, например, постоянные распространения градиентного волокна с большим числом мод, определяемые соотношениями (9.46) и (9.42). Хотя n_1 и n_2 зависят от ω, разумно считать, что отношение $\Delta = (n_1 - n_2)/n_1$ почти не зависит от ω. Используя это приближение и вычисляя $v_q = (\mathrm{d}\beta_q/\mathrm{d}\omega)^{-1}$, получаем

$$v_q \approx \frac{c_0}{N_1}\left[1 - \frac{p-2}{p+2}\left(\frac{q}{M}\right)^{p/(p+2)}\Delta\right], \tag{9.59}$$

где N_1 — групповой показатель преломления материала сердцевины,

$$N_1 = \frac{\mathrm{d}}{\mathrm{d}\omega}\omega n_1 = n_1 - \lambda_0\frac{\mathrm{d}n_1}{\mathrm{d}\lambda_0}.$$

В этом приближении выражение (9.48) для скорости v_q остается неизменным, за исключением того, что показатель преломления n_1 заменяется групповым показателем преломления N_1. Для волокна со ступенчатым профилем показателя преломления ($p \to \infty$) групповые скорости мод меняются от c_0/N_1 до $(c_0/N_1)(1 - \Delta)$, так что время отклика равно

$$\boxed{\sigma_\tau \approx \frac{L}{(c_0/N_1)}\frac{\Delta}{2}}. \tag{9.60}$$

Время отклика (многомодовое волокно, ступенчатый профиль, материальная дисперсия)

Это выражение нужно сравнить с формулой (9.55), которую можно использовать в отсутствие материальной дисперсии.

Упражнение 9.1

Оптимальный параметр градиентного профиля

Используйте (9.46) и (9.42), чтобы вывести следующее выражение для групповой скорости v_q, когда и n_1, и Δ зависят от длины волны:

$$v_q \approx \frac{c_0}{N_1}\left[1 - \frac{p-2-p_s}{p+2}\left(\frac{q}{M}\right)^{p/(p+2)}\Delta\right], \quad q = 1,\ 2,\ \ldots\ M, \tag{9.61}$$

где

$$p_s = 2\frac{n_1}{N_1}\frac{\omega}{\Delta}\frac{\mathrm{d}\Delta}{\mathrm{d}\omega}.$$

Каково оптимальное значение параметра градиентного профиля p, обеспечивающее минимальность модовой дисперсии?

Волноводная дисперсия

Групповые скорости мод в волноводе зависят от длины волны, даже если материальная дисперсия пренебрежимо мала. Этот эффект, известный как **волноводная диснерсия**, является результатом зависимости распределения поля в волокне от отношения радиуса сердцевины к длине волны (a/λ_0). Таким образом, распределение энергии света между сердцевиной и оболочкой зависит от λ_0. Поскольку фазовые скорости в сердцевине и оболочке различны, групповая скорость моды меняется. Волноводная дисперсия особенно существенна в одномодовых волокнах, где модовая дисперсия отсутствует, и на длинах волн, где материальная дисперсия мала (в кварцевом стекле — вблизи $\lambda_0 = 1{,}3$ мкм), поскольку здесь она преобладает.

Как обсуждалось в подразд. 9.2.1, групповая скорость $v = (\mathrm{d}\beta/\mathrm{d}\omega)^{-1}$ и постоянная распространения β определяются из характеристического уравнения, которое зависит от параметра

$$V = 2\pi\frac{a}{\lambda_0}\mathrm{NA} = \frac{a\cdot\mathrm{NA}}{c_0}\omega.$$

В отсутствие материальной дисперсии (т. е. когда NA не зависит от ω) V прямо пропорционально ω, так что

$$\frac{1}{v} = \frac{\mathrm{d}\beta}{\mathrm{d}\omega} = \frac{\mathrm{d}\beta}{\mathrm{d}V}\frac{\mathrm{d}V}{\mathrm{d}\omega} = \frac{a\cdot\mathrm{NA}}{c_0}\frac{\mathrm{d}\beta}{\mathrm{d}V}. \tag{9.62}$$

Уширение импульса от источника со спектральной шириной σ_λ связано с временной задержкой L/v посредством формулы

$$\sigma_\tau = \left|\frac{\mathrm{d}}{\mathrm{d}\lambda_0}\frac{L}{v}\right|\sigma_\lambda.$$

Таким образом,

$$\sigma_\tau = |D_w|\sigma_\lambda L, \tag{9.63}$$

где коэффициент волноводной дисперсии D_w дается формулой

$$D_w = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{1}{v}\right) = -\frac{\omega}{\lambda_0}\frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{1}{v}\right). \tag{9.64}$$

Подстановка (9.62) в (9.64) приводит к выражению

$$D_w = -\left(\frac{1}{2\pi c_0}\right) V^2 \frac{d^2\beta}{dV^2}. \qquad (9.65)$$

Таким образом, групповая скорость обратно пропорциональна $d\beta/dV$, а коэффициент волноводной дисперсии волокна пропорционален $V^2 d^2\beta/dV^2$. Зависимость β от V показана на рис. 9.16, *а* для основной моды LP_{01}. Поскольку β меняется нелинейно в зависимости от V, коэффициент волноводной дисперсии D_w сам является функцией V и, следовательно, функцией длины волны[3]. Зависимостью D_w от λ_0 можно управлять путем изменения радиуса сердцевины или, для градиентных волокон, формы профиля показателя преломления.

Сочетание материальной и волноводной дисперсии

Совместные проявления материальной и волноводной дисперсии (которые мы будем называть **хроматической дисперсией**) можно описать путем учета зависимости показателей преломления n_1, n_2 и, следовательно, NA от длины волны при нахождении $d\beta/d\omega$ из характеристического уравнения. Хотя волноводная дисперсия обычно меньше материальной, она смещает длину волны, на которой полная хроматическая дисперсия минимальна.

Поскольку хроматическая дисперсия ограничивает возможности одномодовых волокон, разработки более совершенных типов волокон направлены на уменьшение этого эффекта. Для этого профиль показателя преломления сердцевины нужно выбрать так, чтобы длина волны, на которой волноводная дисперсия компенсирует материальную, смещалась в сторону рабочей длины волны волокна. **Волокна со смещенной дисперсией** были успешно изготовлены с использованием линейно спадающего профиля показателя преломления сердцевины и уменьшенного ее радиуса, как показано на рис. 9.28, *а*. Эта методика позволяет сместить длину волны, на которой хроматическая дисперсия равна нулю, с 1,3 до 1,55 мкм, где волокно имеет наименьший коэффициент затухания. Были разработаны другие градиентные профили, для которых хроматическая дисперсия исчезает на двух длинах волн и снижена между ними. Эти волокна, называемые **волокнами со сглаженной дисперсией**, были изготовлены на основе четырехслойного распределения показателя преломления, показанного на рис. 9.28, *б*. Заметим, что сам процесс изготовления такого профиля вносит потери из-за использования легирующих добавок.

Волокна с другими профилями показателя преломления могут быть спроектированы так, чтобы комбинированный коэффициент материальной и волноводной дисперсии был пропорционален такому коэффициенту для обычного волокна со ступенчатым показателем преломления, но имел противоположный знак. Этого можно достичь в широком диапазоне длин волн, как показано на рис. 9.28, *в*. Тогда расплывание импульса, вносимое обычным волокном, мож-

[3] Подробности по данному вопросу — в литературе для чтения, особенно в статьях Gloge.

но обратить путем объединения двух профилей в один. Волокна с обращенным коэффициентом дисперсии известны как **волокна с компенсирующей дисперсией**. Короткая вставка из такого волокна может компенсировать дисперсию, вносимую длинным отрезком обычного волокна.

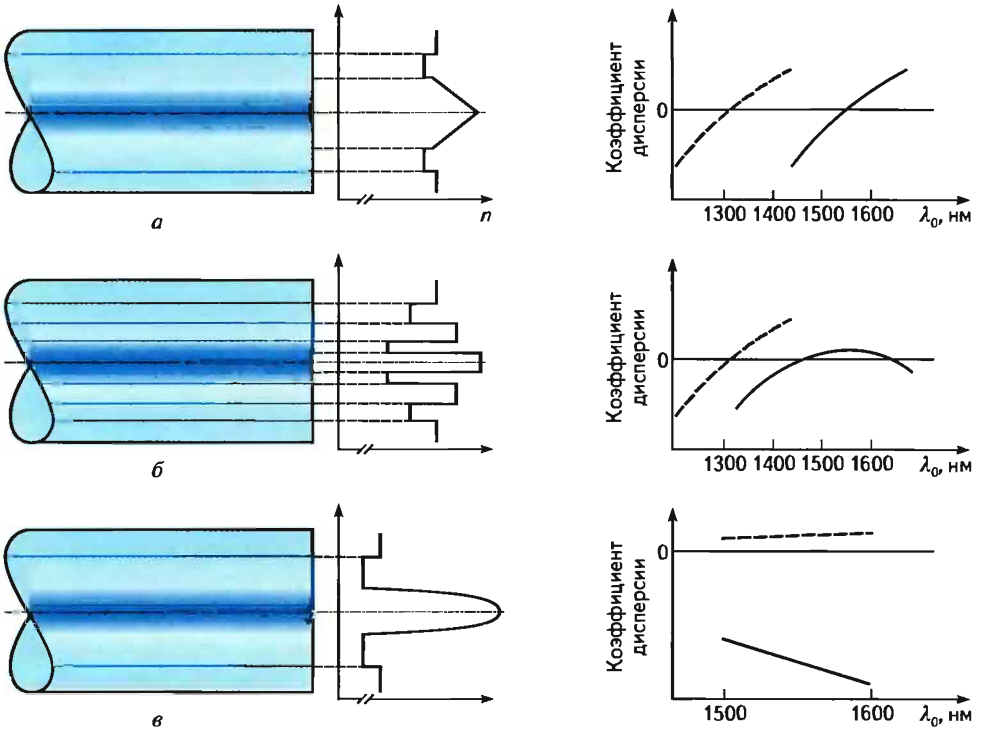


Рис. 9.28. Профили показателя преломления и схематическая зависимость от длины волны для коэффициента материальной дисперсии (штриховые линии), а также для полного коэффициента хроматической дисперсии (материальной и волноводной) (сплошные линии):

а — волокно со смещенной дисперсией (DSF); *б* — волокно со сглаженной дисперсией (DFF); *в* — волокно с компенсирующей дисперсией (DCF)

Поляризационно-модовая дисперсия

Как отмечалось выше, основная пространственная мода (LP_{01}) оптического волокна имеет две поляризационные моды, например линейно-поляризованные в направлениях *x* и *y*. Если волокно имеет идеальную осевую симметрию, а его материал полностью изотропен, эти две поляризационные моды вырождены, т. е. распространяются с одинаковой скоростью. Однако волокна, находящиеся в реальных окружающих условиях, всегда обладают небольшим двулучепреломлением, которое случайным образом меняется вдоль волокна. Оно обусловлено слабыми вариациями показателя преломления и эллиптичности поперечного сече-

ния волокна. Поскольку влияние таких неоднородностей и анизотропии на поляризационные моды и дисперсию оптических импульсов в общем случае трудно оценить, мы рассмотрим такие эффекты в рамках упрощенных моделей.

Рассмотрим вначале волокно как однородную анизотропную среду с главными осями x и y и главными показателями преломления n_x и n_y. Третья главная ось z, конечно, направлена вдоль оси волокна. Материал волокна предполагается диспергирующим, так что n_x и n_y зависят от частоты, однако направления главных осей считаются не зависящими от частоты в пределах рассматриваемого спектрального диапазона. Если импульс на входе линейно поляризован вдоль оси x, то на отрезке волокна длиной L он приобретет групповую задержку $\tau_x = N_x L/c_0$; если же он поляризован по оси y, то $\tau_y = N_y L/c_0$. Здесь N_x и N_y — групповые показатели преломления, связанные с n_x и n_y (см. разд. 5.6). Импульс с поляризационными компонентами вдоль обеих осей приобретет дифференциальную групповую задержку (ДГЗ) $\delta\tau = |\tau_y - \tau_x|$, даваемую выражением

$$\boxed{\delta\tau = \frac{\Delta N L}{c_0},} \qquad (9.66)$$

Дифференциальная групповая задержка

где $\Delta N = |N_y - N_x|$. В процессе распространения импульс расщепится на два ортогонально-поляризованных импульса, центры которых будут все больше разделяться по времени (рис. 9.29). ДГЗ соответствует поляризационно-модовой дисперсии (ПМД), которая линейно растет по длине волокна со скоростью $\Delta N/c_0$, обычно выражаемой в пс/км.

Рис. 9.29. Дифференциальная групповая задержка (ДГЗ), связанная с поляризационно-модовой дисперсией (ПМД)

Поскольку длинное волокно, как правило, подвержено воздействию окружающей среды и структурных факторов, меняющихся вдоль его оси, простая модель, рассмотренная выше, часто не является адекватной. При таких условиях более реалистическая модель состоит из последовательности коротких однородных отрезков волокна, каждый со своими главными осями и главными показателями преломления. Считается, что главные оси слегка поворачиваются при переходе от отрезка к отрезку. Такая каскадная система в общем случае описывается 2 × 2 матрицей Джонса **T**, которая получается как произведение матриц Джонса отдельных отрезков (см. подразд. 6.1.2). Поляризационные моды комбинированной системы представляют собой собственные векторы **T** и не обязательно линейно поляризованы. Если волокно не вносит потерь, матрица **T** унитарна. Тогда ее собственные значения представляют собой фазовые

множители $\exp(j\varphi_1)$ и $\exp(j\varphi_2)$, которые можно записать в виде $\exp(jn_1k_0L)$ и $\exp(jn_2k_0L)$, где и n_1 и n_2 — эффективные показатели преломления двух поляризационных мод, а L — длина волокна. Прохождение светом волокна такой длины можно описать путем разложения входной волны на составляющие вдоль поляризационных мод; эти составляющие будут распространяться с эффективными показателями преломления n_1 и n_2.

Поскольку волокно обладает дисперсией, **T** зависит от частоты, следовательно, от частоты зависят n_1 и n_2, а также соответствующие групповые показатели преломления N_1 и N_2. Входной импульс, поляризационное состояние которого совпадает с первой поляризационной модой волокна, распространяется с эффективным групповым показателем преломления N_1. Если поляризационное состояние входного импульса — вторая мода волокна, то эффективный групповой показатель преломления будет N_2. Импульс, включающий обе моды, будет испытывать ДГЗ в соответствии с (9.66), где $\Delta N = |N_1 - N_2|$.

Статистическая модель, описывающая случайные изменения величины и ориентации двулучепреломления по длине волокна, приводит к выражению для среднеквадратичной величины уширения импульса из-за ДГЗ. Она оказывается пропорциональной $\sqrt{L}$, а не L:

$$\boxed{\sigma_{\text{ПМД}} = D_{\text{ПМД}}\sqrt{L},} \qquad (9.67)$$

Поляризационно-модовая дисперсия

где $D_{\text{ПМД}}$ — дисперсионный параметр, обычно лежащий в интервале от 0,1 до 1 пс/$\sqrt{\text{км}}$.

Кроме ДГЗ присутствуют также эффекты дисперсии высших порядков. Каждая из поляризационных мод подвержена дисперсии групповой скорости с коэффициентом дисперсии, пропорциональным второй производной ее показателя преломления (см. разд. 5.6).

Другой эффект высшего порядка имеет отношение к взаимосвязанной природе поляризационных и спектральных свойств системы. Поскольку матрица **T** зависит от частоты, не только ее собственные значения (т. е. показателя преломления n_1 и n_2), но и собственные векторы (т. е. поляризационные моды) зависят от частоты. Если импульс имеет достаточно малую спектральную ширину (т. е. импульс не слишком короткий), можно приближенно использовать поляризационные моды на центральной частоте. Однако для ультракоротких импульсов требуется более подробный анализ, включающий комбинированное спектральное и поляризационное описание системы. Можно найти такие поляризационные состояния, у которых групповые задержки не чувствительны к частоте, так что связанная с ними дисперсия групповой скорости минимальна. Однако эти состояния не являются собственными векторами матрицы Джонса, поэтому поляризационные состояния на входе и на выходе не совпадают[4].

[4] Более подробные сведения по данному вопросу см. в *Poole C.D., Wagner R.E.* Phenomenological Aproach to Poarization Dispersion in Long Single-Mode Fibers. Electronics Letters. Vol. 22, 1986. P. 1029—1030.

Упражнение 9.2

Дифференциальная групповая задержка в волокне из двух частей

Рассмотрим распространение оптического импульса по волокну длиной 1 км, состоящему из двух отрезков равной длины. Каждый из отрезков представляет собой одномодовое анизотропное волокно с главными групповыми показателями преломления $N_x = 1{,}462$ и $N_y = 1{,}463$. Соответствующие коэффициенты дисперсии групповой скорости равны $D_x = D_y = 20$ пс/км нм. Главные оси одного отрезка расположены под углом 45° к главным осям другого, как показано на рис. 9.30.

а. Пусть входной импульс имеет длительность 100 пс и линейно поляризован под углом 45° по отношению к направлениям *x* и *y* волокна. Изобразите временной профиль импульса на выходном конце волокна. Спектральную ширину источника импульса примите равной 50 нм.

б. Определите поляризационные моды всего волокна и временной профиль выходного импульса, если входной импульс совпадает с одной из поляризационных мод.

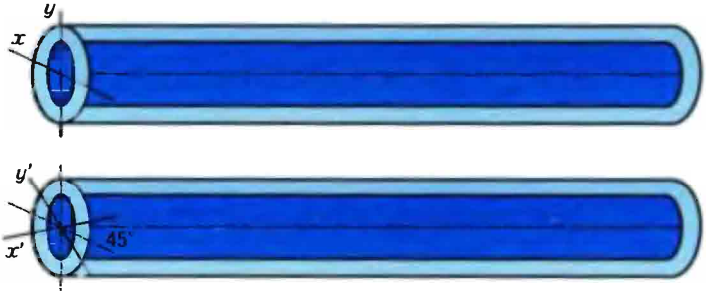


Рис. 9.30. Двулучепреломляющее волокно из двух отрезков

Нелинейная дисперсия

Еще один дисперсионный эффект возникает, когда интенсивность света в сердцевине волокна достаточно велика, поскольку в этом случае показатель преломления становится зависящим от интенсивности, и материал демонстрирует нелинейное поведение. Поскольку фаза пропорциональна показателю преломления, части импульса с высокой интенсивностью сдвигаются иначе, чем менее интенсивные части, что приводит к различной величине сдвига мгновенной частоты. Этот нелинейный эффект, называемый **самомодуляцией фазы** (СМФ), вносит вклад в дисперсию импульса. При определенных условиях СМФ может компенсировать дисперсию групповой скорости, связанную с материальной дисперсией, и тогда импульс может распространяться без изменения своего временного профиля. Такая направляемая волна называется **солитоном**. Нелинейная оптика вводится в рассмотрение в гл. 21, а оптические солитоны обсуждаются в гл. 22.

Резюме

Распространение импульсов в оптических волокнах регулируется затуханием и несколькими видами дисперсии. На рис. 9.31 дана схематическая иллюстрация для сравнения профилей импульсов, распространяющихся по волокнам различных типов.

- В многомодовом волокне в формировании ширины импульса, принимаемого на дальнем конце волокна, преобладает модовая дисперсия. Она определяется разбросом групповых задержек отдельных мод.
- В одномодовом волокне модовой дисперсии нет, и уширение импульса определяется комбинацией материальной и волноводной дисперсии (хроматической дисперсией). Ширина выходного импульса зависит от дисперсии групповой скорости.
- Материальная дисперсия обычно намного сильнее волноводной. Однако на длинах волн, где материальная дисперсия мала, волноводная дисперсия становится важной. Волокна со специальными профилями показателя преломления можно использовать для изменения характеристик хроматической дисперсии, создания волокон со сглаженной, смещенной и компенсирующей дисперсией.
- При распространении импульсов по длинным одномодовым волокнам, в которых хроматическая дисперсия пренебрежимо мала, преобладает поляризационно-модовая дисперсия. Малые изменения анизотропии волокна, вызванные, например, окружающими условиями, меняют поляризационные моды, так что входной импульс распространяется в виде двух поляризационных мод с разными групповыми показателями преломления. Эта дифференциальная групповая задержка проявляется в небольшом расплывании импульса.
- При определенных условиях интенсивный импульс, называемый солитоном, делает волокно нелинейным и распространяется по нему без уширения. Это является результатом баланса между материальной дисперсией и самомодуляцией фазы (зависимостью показателя преломления от интенсивности света), как обсуждается в гл. 22.

9.4. МИКРОСТРУКТУРНЫЕ И ФОТОННО-КРИСТАЛЛИЧЕСКИЕ ВОЛОКНА

Микроструктурное (дырчатое) волокно — это волокно из чистого кварцевого стекла, содержащее многочисленные цилиндрические воздушные отверстия, параллельные оси по всей длине. В поперечном сечении эти отверстия образуют периодическую структуру. Как показано на рис. 9.31, сердцевина волокна определяется наличием в этой структуре дефекта, например пропущенного отверстия или отверстия другого размера, или лишнего отверстия.

Отверстия характеризуются расстоянием между центрами Λ и диаметрами *d*. Величина Λ, которая также называется **шагом**, в типичном случае лежит в интервале 1—10 мкм. Стекло не обязательно легировать какими-либо добавками.

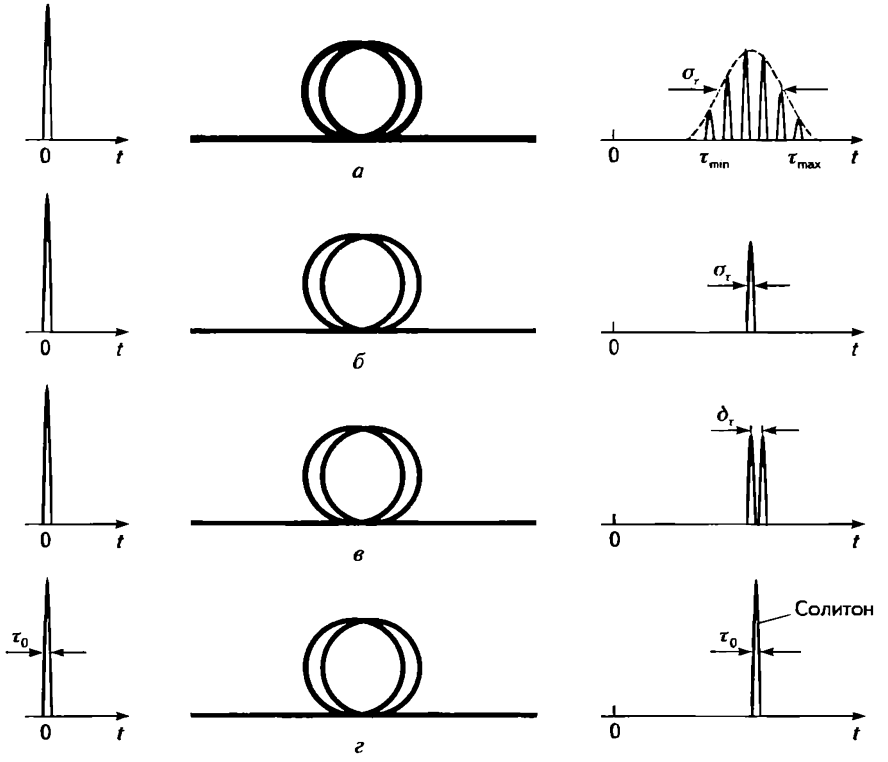


Рис. 9.31. Уширение короткого оптического импульса после прохождения различных типов волокон:

а — модовая дисперсия в многомодовом волокне (ММВ); *б* — материальная и волноводная дисперсия в одномодовом волокне (ОМВ); *в* — поляризационно-модовая дисперсия (ПМД) в ОМВ; *г* — распространение солитона в нелинейном ОМВ

Дырчатые волокна направляют оптические волны посредством двух механизмов: за счет **эффективного показателя преломления** и **фотонной запрещенной зоны**.

Волноводный эффект за счет эффективного показателя преломления

Если диаметр отверстий много меньше длины волны ($d \ll \lambda$), то периодическая оболочка ведет себя приблизительно как сплошная среда, **эффективный показатель преломления** которой n_2 равен среднему показателю преломления дырчатого материала (рис. 9.32, *а*). Волноводный эффект достигается за счет сплошной сердцевины с показателем $n_1 > n_2$. Свет направляется за

счет полного внутреннего отражения, как в обычных волокнах. Отверстия служат в качестве распределенной «отрицательной легирующей добавки», снижающей показатель преломления оболочки по сравнению с сердцевиной. Поэтому отверстия могут располагаться не периодически, а случайным образом и не обязательно должны продолжаться по всей длине волокна.

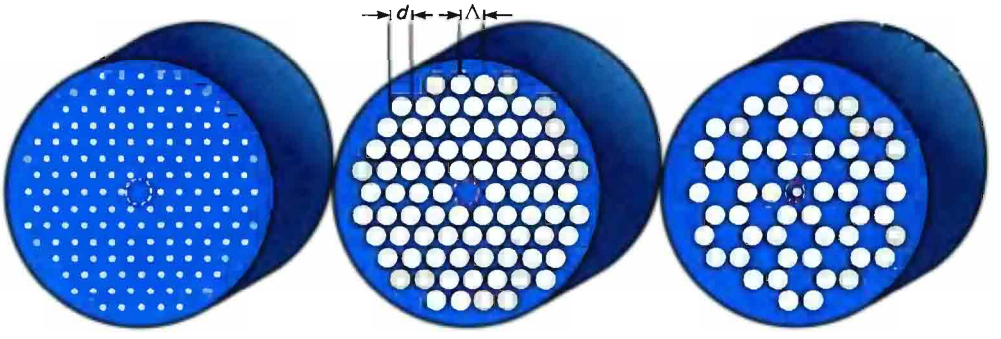


Рис. 9.32. Различные типы дырчатых волокон:

а — сплошная сердцевина (пунктирная окружность), окруженная оболочкой из того же материала, но пронизанного периодическим набором воздушных отверстий, диаметры которых много меньше длины волны. Средний показатель преломления у оболочки ниже, чем у сердцевины; *б* — фотонно-кристаллическое дырчатое волокно с оболочкой, содержащей периодический набор больших воздушных отверстий, и сплошной сердцевиной (пунктирная окружность); *в* — фотонно-кристаллическое дырчатое волокно, оболочка которого содержит периодический набор больших воздушных отверстий, а сердцевина представляет собой воздушное отверстие другого размера (пунктирная окружность)

Если размер отверстий не мал по сравнению с длиной волны, то оболочка должна рассматриваться как двумерная периодическая среда. В этом случае эффективный показатель преломления равен среднему показателю преломления, взвешенному с распределением оптической интенсивности в среде, и поэтому сильно зависящему от длины волны, размера и геометрии отверстий. Поскольку более короткие волны сильнее локализуются в среде с более высоким показателем преломления, эффективный показатель преломления оболочки $n_2(\lambda)$ является убывающей функцией длины волны. Аналогичный эффект наблюдается в одномерных фотонных кристаллах, где эффективный показатель преломления является возрастающей функцией длины волны на частотах, лежащих в низшей разрешенной фотонной зоне (см. рис. 7.20). Вследствие этого дырчатое волокно обладает сильной волноводной дисперсией, которая может быть очень полезным свойством.

Одно из следствий этой волноводной дисперсии состоит в том, что дырчатое волокно может работать в одномодовом режиме в широком диапазоне длин волн от инфракрасного до ультрафиолетового[5]. Это свойство, называемое **бес-**

[5] См. *Birks T.A., Knight J.C., Russel P.St.J.* Endlessly Single-Mode Photonic Crystal Fibre. Optics Letters. Vol. 22, 1997. P. 961—963.

конечной одномодовой волноводностью (endless sigle-mode guidance), проявляется, когда параметр волокна

$$V = \frac{2\pi a}{\lambda}\sqrt{n_1^2 - n_2^2(\lambda)}$$

почти не зависит от λ. Условия для этого обеспечиваются убыванием эффективного показателя преломления $n_2(\lambda)$ с ростом длины волны по такому закону, чтобы выполнялось соотношение

$$\sqrt{n_1^2 - n_2^2(\lambda)} \propto \lambda.$$

Для обычного волокна, напротив, V обратно пропорционально λ, так что одномодовый режим, имеющий место для некоторой длины волны (2,405), превращается в многомодовый при ее уменьшении, достаточном, чтобы V превысило 2,405.

Другим интересным свойством является возможность одномодового режима с **большой площадью сечения** моды. Оптические волокна с большой площадью сечения моды полезны для приложений, требующих ввода в волокно большой мощности. В обычном волокне выполнение условия одномодового режима ($V = 2\pi(a/\lambda_0)\mathrm{NA} < 2{,}405$) при большом диаметре сердцевины $2a$ можно обеспечить за счет малой числовой апертуры. Аналогично, у дырчатого волокна диаметр направляемой моды можно увеличить за счет большего расстояния между отверстиями Λ (что увеличивает диаметр сердцевины) и меньшего диаметра отверстий d (что снижает числовую апертуру и способствует большему проникновению поля в оболочку). Получено существенное увеличение площади сечения моды при сравнительно малых изменениях размера отверстий. В литературе имеются сведения об увеличении площади сечения моды на несколько порядков по сравнению с обычным волокном.

Волноводный эффект за счет фотонной запрещенной зоны

Оболочка дырчатого волокна может рассматриваться как двумерный **фотонный кристалл**. Например, треугольная микроструктура из отверстий, показанная на рис. 9.32, *б*, имеет дисперсионную диаграмму с фотонными запрещенными зонами, изображенную на рис. 7.27 и обсуждавшуюся в подразд. 7.3.1. Если оптическая частота лежит в фотонной запрещенной зоне, то распространение волны в оболочке запрещено, и волокно действует как фотонно-кристаллический волновод (см. разд. 8.5).

Фотонно-кристаллическое волокно (ФКВ) может иметь сплошную или полую сердцевину, как показано на рис. 9.32, *б* и *в* соответственно. Волокна с полой сердцевиной являются принципиально новыми, так как они не могут работать по принципу эффективного показателя преломления (см. выше), поскольку в них невозможно полное внутреннее отражение. У волны, направляемой по ФКВ с воздушной сердцевиной, снижаются потери и нелинейные эффекты, поэтому она может переносить большую мощность. Такие волокна могут иметь множество уникальных приложений. В них можно добиться как

сглаженной дисперсии в широком диапазоне длин волн, так и сдвига в область ниже длины волны, на которой материальная дисперсия равна нулю. Можно построить мощные волоконные лазеры, работающие в широком диапазоне длин волн. Возможно также огромное количество других приложений, например анализ газов путем введения их в сердцевину волокна.

Рекомендуемая литература

книги

См. также книги по оптике волноводов к гл. 8.

DeCusatis C., DeCusatis C.J. Sher. Fiber Optic Essentials. Elsevier, 2005.

Hecht J. Understanding Fiber Optics. Prentice Hall, 5th ed. 2005.

Zolla F., Renversez G., Nicolet A., Kuhlmey B., Guenneau S., Felbacq D. Foundations of Photonic Crystal Fibres. Imperial College Press (London), 2005.

Galtarossa A., Menyuk C.R., eds. Polarization Mode Dispersion. Springer-Verlag, 2005.

Khare R.P. Fiber Optics and Optoelectronics. Oxford University Press, 2004.

Buck J.A. Fundamentals of Optical Fibers. Wiley, 1995; 2nd ed. 2004.

Hecht J. City of Light: The Story of Fiber Optics. Oxford University Press, 2004.

Bjarklev A., Broeng J., Bjarklev A.S. Photonic Crystal Fibers. Springer-Verlag, 2003.

Petersen J.K. Fiber Optics Illustrated Dictionary. CRC Press, 2003.

Kaminow I., Li T., eds. Optical Fiber Telecommunications IVA: Components. Academic Press, 2002.

Manolatou C., Haus H.A. Passive Components for Dense Optical Integration. Kluwer, 2002.

Goff D.R., Hansen K.S., eds. Fiber Optic Reference Guide. A Practical Guide to the Technology, Focal. 3rd ed. 2002.

Tricker R. Optoelectronics and Fiber Optic Technology. Newnes, 2002.

Bates R.J. Basic Fiberoptics Technologies. McGraw-Hill, 2001.

Crisp J. Introduction to Fiber Optics. Newnes, 2nd ed. 2001.

Othonos A., Kalli K. Fiber Bragg Gratings: Fundamentals and Applications in Telecommunications and Sensing. Artech, 1999.

Ghatak A., Thyagarajan K. An Introduction to Fiber Optics. Cambridge University Press, 1998.

Sanghera J.S., Aggarwal I.D., eds. Infrared Fiber Optics. CRC Press, 1998.

Powers J.P. An Introduction to Fiber Optic Systems. Irwin, 2nd ed. 1997.

Weik M.H. Fiber Optics Standard Dictionary. Chapman & Hall, 3rd ed. 1997.

Kumar A. Antenna Design with Fiber Optics. Artech House, 1996.

Miller J.L., Friedman E. Photonics Rules of Thumb: Optics, Electro-Optics, Fiber Optics, and Lasers. McGraw-Hill, 1996.

Chen C.-L. Elements of Optoelectronics and Fiber Optics. Irwin, 1995.

Dyott R.B. Elliptical Fiber Waveguides. Artech House, 1995.

Kashima N. Passive Optical Components for Optical Fiber Transmission. Artech, 1995.

Krivoshlykov S.G. Quantum-Theoretical Formalism for Inhomogeneous Graded-Index Waveguides, Akademie- Verlag, 1994.

Cancellieri G. Single-Mode Optical Fiber Measurement: Characterization and Sensing. Artech House, 1993.

Midwinter J.E. Optical Fibers for Transmission. Wiley, 1979; Krieger, reissued 1992.

Chang K., ed. Fiber and Electro-Optical Components. Wiley, 1991.

Cheo P.K. Fiber Optics and Optoelectronics. Prentice Hall, 1985; 2nd ed. 1990.

Diament P. Wave Transmission and Fiber Optics. Macmillan, 1989.

Marcuse D. Light Transmission Optics. Van Nostrand Reinhold, 1972; 2nd ed. 1982; Krieger, reissued 1989.

Kao C.K. Optical Fiber Systems: Technology, Design, and Applications. McGraw-Hill, 1982; reprinted 1986.

Marcuse D. Principles of Optical Fiber Measurements. Academic Press, 1981.

СТАТЬИ

Bayindir M., Abouraddy A.F., Shapira O., Viens J., Saygin-Hinczewski D., Sorin F., Arnold J., Joannopoulos J.D., Fink Y. Kilometer-Long Ordered Nanophotonic Devices by Preform-to-Fiber Fabrication. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 12, 2006. P. 1202—1213.

Russell P. Photonic Crystal Fibers. Science. Vol. 299, 2003. P. 358—362.

Issue on novel and specialty fibers. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 7, № 3, 2001.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Gordon J.P., Kogelnik H. PMD Fundamentals: Polarization Mode Dispersion in Optical Fibers. Proceedings of the National Academy of Sciences (USA). Vol. 97, 2000. P 4541—4550.

Issue on fiber-optic passive components. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 5, № 5, 1999.

Birks T.A., Knight J.C., Russell P.St.J. Endlessly Single-Mode Photonic Crystal Fibre. Optics Letters. Vol. 22, 1997. P. 961—963.

Clarricoats P.J.B. Optical Fibre Waveguides-A Review. In Progress in Optics. Vol. 14. E. Wolf, ed. North-Holland, 1977.

Gloge D. Weakly Guiding Fibers. Applied Optics. Vol. 10, 1971. P. 2252—2258.

Gloge D. Dispersion in Weakly Guiding Fibers. Applied Optics. Vol. 10, 1971. P. 2442—2445.

Задачи

К РАЗДЕЛУ 9.1

1. **Коэффициент передачи оптической мощности.**

а. Источник излучает свет мощностью P_0 с распределением $I(\theta) = (1/\pi)P_0 \cos\theta$, где $I(\theta)$ — мощность, приходящаяся на единицу телесного угла в направлении, образующем угол θ с осью волокна. Покажите, что мощность, собираемая волокном, равна $P = (\mathrm{NA})^2 P_0$, так что коэффициент передачи составляет $(\mathrm{NA})^2$, где NA — числовая апертура волокна.

б. Пусть источником является плоский светодиод с показателем преломления n_s, присоединенный к волокну, причем площадь сечения волокна больше, чем площадь излучающей поверхности светодиода. Рассчитайте числовую апертуру волокна и коэффициент передачи, если $n_1 = 1{,}46$, $n_2 = 1{,}455$ и $n_s = 3{,}5$.

2. **Числовая апертура градиентного волокна.** Сравните числовые апертуры волокна со ступенчатым профилем показателя преломления с $n_1 = 1{,}45$ и $\Delta = 0{,}01$ и градиентного волокна с $n_1 = 1{,}45$, $\Delta = 0{,}01$ и параболическим профилем показателя преломления ($p = 2$) (см. упражнение 1.9).

К РАЗДЕЛУ 9.2

1. **Моды.** Волокно со ступенчатым профилем показателя преломления имеет радиус $a = 5$ мкм, показатель преломления сердцевины $n_1 = 1{,}45$ и относительную разность показателей преломления $\Delta = 0{,}002$. Определите наименьшую длину волны λ_c, при которой волокно является одномодовым. На длине волны $\lambda_c/2$ определите индексы (l, m) всех направляемых мод.

2. **Модовая дисперсия.** Волокно со ступенчатым профилем показателя преломления имеет числовую апертуру NA = 0,16, радиус сердцевины $a = 45$ мкм и показатель преломления сердцевины $n_1 = 1{,}45$. На используемой длине волны $\lambda_0 = 1{,}3$ мкм дисперсия материала пренебрежимо мала. Очень короткий импульс входит в волокно при $t = 0$ и проходит расстояние 1 км. Нарисуйте форму принимаемого импульса:

а) используя лучевую оптику и рассматривая только меридиональные лучи;

б) используя волновую оптику и рассматривая только меридиональные ($l = 0$) моды.

3. **Постоянные распространения и групповые скорости.** Ступенчатое волокно с показателями преломления $n_1 = 1{,}444$ и $n_2 = 1{,}443$ работает на длине волны $\lambda_0 = 1{,}55$ мкм. Определите радиус сердцевины, при котором параметр волокна V равен 10. Используйте рис. 9.11 для оценки постоянных распространения всех направляемых мод с $l = 0$. Пусть теперь радиус сердцевины изменился так, что $V = 4$. Используйте рис. 9.16, *а* для определения фазовой скорости, постоянной распространения и групповой скорости моды LP_{01}. Дисперсией материала пренебречь.

4. **Постоянные распространения и волновой вектор (ступенчатое волокно).** Волокно со ступенчатым профилем показателя преломления радиуса $a = 20$ мкм с показателями преломления $n_1 = 1{,}47$ и $n_{21} = 1{,}46$ работает на длине волны $\lambda_0 = 1{,}55$ мкм. Применяя теорию квазиплоских волн и рассматривая только направляемые моды с азимутальным индексом $l = 1$:

а) определите наименьшую и наибольшую постоянные распространения;

б) для моды с наименьшей постоянной распространения определите внешний и внутренний радиусы цилиндрического слоя, в котором заключена волна, а также компоненты волнового вектора $\boldsymbol{k}$ при $r = 5$ мкм.

5. **Постоянные распространения и волновой вектор (градиентное волокно).** Проведите те же самые расчеты, что и в предыдущей задаче, но для градиентного волокна с параболическим профилем показателя преломления ($p = 2$).

К РАЗДЕЛУ 9.3

1. **Потери из-за рассеяния.** На длине волны $\lambda_0 = 820$ нм потери из-за поглощения в волокне составляют 0,25 дБ/км, а потери из-за рассеяния — 2,25 дБ/км. На длине волны $\lambda_0 = 600$ нм калориметрические измерения нагрева волокна из-за поглощения света дают потери 2 дБ/км. Найдите полный коэффициент затухания на $\lambda_0 = 600$ нм.

2. **Модовая дисперсия в ступенчатом волокне.** Определите радиус сердцевины многомодового волокна со ступенчатым профилем показателя преломления,

имеющего числовую апертуру NA = 0,1, если на длине волны $\lambda_0 = 0{,}87$ мкм число мод составляет $M = 5000$. При показателе преломления сердцевины $n_1 = 1{,}445$, групповом показателе преломления $N_1 = 1{,}456$ и Δ, приблизительно не зависящем от длины волны, определите время отклика из-за модовой дисперсии σ_τ для волокна длиной 2 км.

3. **Модовая дисперсия в градиентном волокне**. Рассмотрим градиентное волокно с $a/\lambda_0 = 10$, $n_1 = 1{,}45$, $\Delta = 0{,}01$ и степенным профилем с показателем p. Определите число мод M и скорость расплывания импульса из-за модовой дисперсии σ_τ/L при $p = 1{,}9$; 2; 2,1 и ∞.

4. **Распространение импульса**. Импульс с начальной длительностью τ_0 передается по градиентному волокну длиной L км со степенным (p) профилем показателя преломления. Наибольший показатель преломления n_1 зависит от длины волны, причем

$$D_\lambda = -\frac{\lambda_0}{c_0}\frac{\mathrm{d}^2 n_1}{\mathrm{d}\lambda_0^2},$$

Δ почти не зависит от длины волны, спектральная ширина источника света составляет σ_λ, а рабочая длина волны равна λ_0. Обсудите влияние увеличения каждого из параметров L, τ_0, p, $|D_\lambda|$, σ_λ и λ_0 на длительность принимаемого импульса.

ГЛАВА 10 ОПТИКА РЕЗОНАТОРОВ

Шарль Фабри (1867—1945)

Альфред Перо (1863—1925)

Фабри и Перо сконструировали оптический резонатор для использования в качестве интерферометра. В настоящее время известный как эталон Фабри—Перо, он широко используется в лазерах.

Оптический резонатор является оптическим аналогом колебательного контура в электротехнике. Он локализует и накапливает свет на резонансных частотах, определяемых его конфигурацией. Его можно рассматривать как оптическую передающую систему с обратной связью: свет циркулирует или многократно отражается внутри резонатора. Различные типы оптических резонаторов показаны на рис. 10.1. Простейший из них **резонатор Фабри—Перо** состоит из двух параллельных плоских зеркал; свет многократно отражается между ними при условии достаточно малых потерь. Другие зеркальные конфигурации включают сферические зеркала, кольцевые расположения зеркал, а также прямоугольные двух- и трехмерные полости. Широко используются также **волоконные и интегрально-оптические кольцевые** резонаторы.

В **диэлектрических резонаторах** вместо зеркал используется полное внутреннее отражение на границе между двумя диэлектриками с малыми потерями. Удерживаемые внутри диэлектрика лучи испытывают скользящее отражение

от его замкнутой границы, при этом угол падения всегда больше критического, что не дает лучам выходить наружу. В **микродисках**, **микроторах** и **микросферах** свет циркулирует за счет отражения при почти скользящем падении на границу. Соответствующие моды известны как **моды шепчущей галереи**. Периодические диэлектрические структуры, такие как **распределенные брэгговские отражатели** (РБО), играют роль зеркал в обычных резонаторах Фабри—Перо, обеспечивая обратную связь в таких структурах, как резонаторы типа **микростолбиков**. Двухмерные **фотонные кристаллы** с дефектом также используются для изготовления микрорезонаторов.

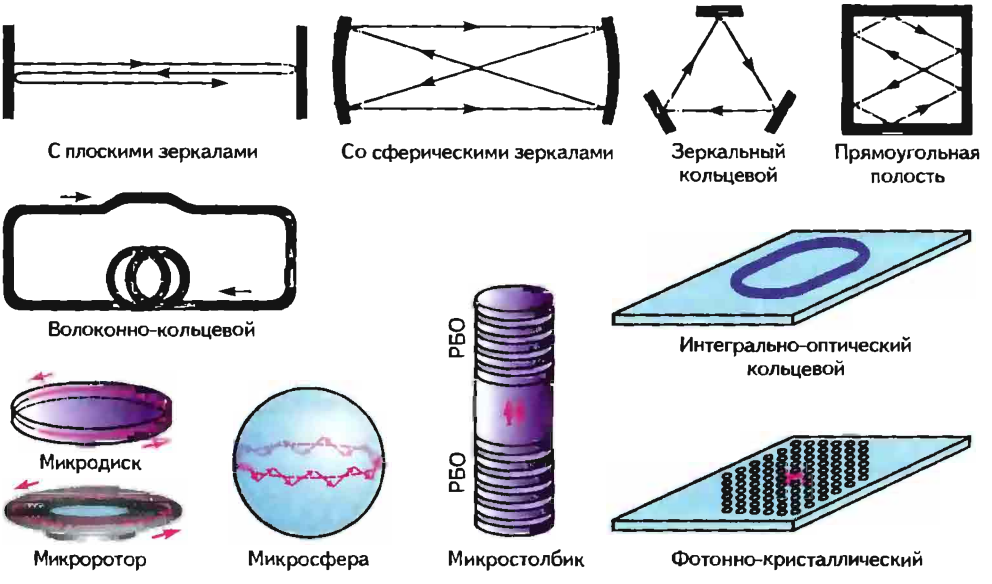


Рис. 10.1. Удержание света в оптических резонаторах за счет: многократного отражения от зеркал; распространения по замкнутым оптическим волокнам и интегрально-оптическим волноводам; отражения вблизи поверхности дисков, торов и сфер типа мод шепчущей галереи; отражений от периодических структур типа брэгговских решеток; дефектов фотонного кристалла

Оптический резонатор характеризуется двумя ключевыми параметрами:

- объемом моды V, т. е. объемом, который занимает удерживаемая оптическая мода;
- добротностью Q, которая пропорциональна времени жизни поля в резонаторе, отнесенному к оптическому периоду.

Эти параметры характеризуют способность резонатора удерживать свет в пространстве и во времени. Степень пространственной локализации удалось повысить за счет разработки микрорезонаторов различной геометрии, а увеличить время удержания — за счет использования материалов с низкими потерями и конфигураций с низкой утечкой.

В силу частотной избирательности оптические резонаторы служат оптическими фильтрами или анализаторами спектров, как обсуждалось в гл. 7. Однако их самая важная роль — роль «контейнера», внутри которого лазерный свет генерируется и формируется. Лазер включает в себя среду, которая усиливает свет внутри оптического резонатора. Резонатор частично определяет частоту и пространственное распределение генерируемого лазером света. Поскольку резонаторы обладают способностью хранить энергию, они могут быть также использованы для генерации импульсов лазерной энергии. Лазеры обсуждаются в гл. 15 и 17, и материал данной главы является существенным для их понимания.

О данной главе

Для описания действия оптических резонаторов полезны несколько теоретических подходов, уже рассмотренных в предыдущих главах.

- Простейший подход основан на *оптике лучей* (см. гл. 1). Рассматриваются лучи, многократно отражающиеся внутри резонатора, и устанавливаются геометрические условия, обеспечивающие их удержание.
- *Оптика волн* (см. гл. 2) используется для определения мод резонатора, т. е. резонансных частот и волновых функций оптических волн, которые могут самосогласованно существовать в резонаторе.
- Изучение *оптики пучков* (см. гл. 3) полезно для понимания поведения поля в резонаторах со сферическими зеркалами; моды такого резонатора — пучки Гаусса и Гаусса—Эрмита.
- *Фурье-оптика* и теория распространения и дифракции света (см. гл. 4) определяет как конечные размеры зеркал резонатора влияют на его потери и пространственные характеристики мод.
- *Оптика фотонных кристаллов* и многослойных сред (см. гл. 7) важна для изучения оптических резонаторов, поскольку в них часто применяются многослойные диэлектрические структуры и периодические среды (например, распределенные брэгговские отражатели и фотонные кристаллы) в качестве зеркал.
- Анализ мод резонатора аналогичен анализу мод в *оптике волноводов* (см. гл. 8), проведенному для планарных зеркальных и диэлектрических волноводов, поскольку оптический резонатор можно рассматривать как оптический волновод с отражателями на обоих концах, что обеспечивает многократное отражение распространяющегося света и его удержание внутри резонатора с малыми потерями.

Очевидно, что оптический резонатор — превосходный объект для приложения различных теорий света, представленных в предыдущих главах. Мы начнем с изучения резонатора с плоскими (см. разд. 10.1) и сферическими (см. разд. 10.2) зеркалами. Затем мы познакомим читателя с двух- и трехмерными резонаторами (см. разд. 10.4) и рассмотрим микрорезонаторы (см. разд. 10.4).

10.1. РЕЗОНАТОРЫ С ПЛОСКИМИ ЗЕРКАЛАМИ

10.1.1. Моды резонатора

В данном разделе мы изучим моды оптического резонатора, образованного двумя параллельными хорошо отражающими плоскими зеркалами, расположенными на расстоянии друг от друга (рис. 10.2). Этот простой одномерный резонатор известен как эталон Фабри—Перо. Сначала мы рассмотрим идеализированный вариант, при котором зеркала не имеют потерь, влияние которых будет затем учтено.

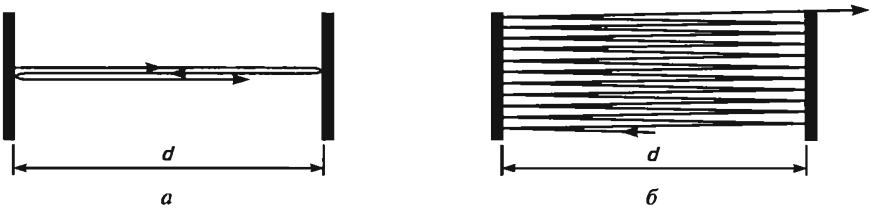


Рис. 10.2. Двухзеркальный плоский резонатор (эталон Фабри—Перо):

а — лучи света, перпендикулярные зеркалам, отражаются туда и обратно, не выходя из резонатора; *б* — лучи с небольшим отклонением в конце концов покидают резонатор. То же самое происходит при не вполне параллельных зеркалах

Моды резонатора как стоячие волны

Как обсуждалось в разд. 2.2, 5.3 и 5.4, монохроматическая волна с частотой ν характеризуется волновой функцией

$$u(\boldsymbol{r},\ t) = \mathrm{Re}\{U(\boldsymbol{r})\exp(j2\pi\nu t)\}, \tag{10.1}$$

представляющей поперечную компоненту электрического поля. Комплексная амплитуда $U(\boldsymbol{r})$ удовлетворяет уравнению Гельмгольца

$$\nabla^2 U(\boldsymbol{r}) + k^2 U(\boldsymbol{r}) = 0,$$

где $k = 2\pi\nu/c$ — волновое число; c — скорость света в среде. Моды резонатора представляют собой решения уравнения Гельмгольца при соответствующих граничных условиях. Для плоскозеркального резонатора без потерь поперечная компонента электрического поля исчезает на поверхности зеркал (см. разд. 5.1), так что $U(\boldsymbol{r}) = 0$ в плоскостях $z = 0$ и $z = d$ на рис. 10.3.

Стоячая волна $U(\boldsymbol{r}) = A\sin kz$, где A — постоянная, удовлетворяет уравнению Гельмгольца и исчезает при $z = 0$ и $z = d$, если k удовлетворяет условию $kd = q\pi$, где q — целое. Это определяет возможные значения k:

$$k_q = q\frac{\pi}{d}, \quad q = 1,\ 2, \tag{10.2}$$

так что моды имеют комплексные амплитуды

$$U(\boldsymbol{r}) = A_q \sin k_q z, \tag{10.3}$$

где A_q — постоянные. Отрицательные значения q не соответствуют независимым модам, поскольку $\sin k_{-q} z = -\sin k_q z$. Более того, $q = 0$ связано с модой, которая не несет энергии, поскольку $k_0 = 0$ и $\sin k_0 z = 0$. Таким образом, моды резонатора — это стоячие волны $A_q \sin k_q z$, где положительное целое число $q = 1, 2,$ называется **индексом моды**. Произвольная волна в резонаторе может быть записана в виде суперпозиции мод резонатора

$$U(\boldsymbol{r}) = \sum_q A_q \sin k_q z. \tag{10.4}$$

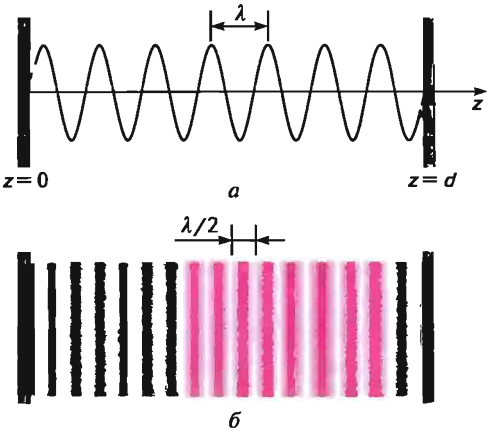


Рис. 10.3. Комплексная амплитуда моды идеального плоскозеркального резонатора (*а*). Поскольку в данном примере на длине резонатора укладывается 14 полуволн, индекс моды $q = d/(\lambda/2) = 14$. Распределение интенсивности (*б*)

Из (10.2) следует, что соответствующие модам частоты $\nu = ck/2\pi$ принимают дискретные значения

$$\nu_q = q\frac{c}{2d}, \quad q = 1, 2, \tag{10.5}$$

представляющие собой резонансные частоты резонатора. Как показано на рис. 10.4, соседние резонансные частоты разделены одинаковым интервалом

$$\boxed{\nu_F = \frac{c}{2d}.} \tag{10.6}$$

Межмодовое расстояние

Соответствующие резонансные длины волн

$$\lambda_q = \frac{c}{\nu_q} = \frac{2d}{q}.$$

Таким образом, длина полного обхода резонатора должна равняться целому числу волн

$$2d = q\lambda_q, \quad q = 1, 2, \tag{10.7}$$

Важно помнить, что c есть скорость света в среде между зеркалами и что λ_q — длина волны в этой среде.

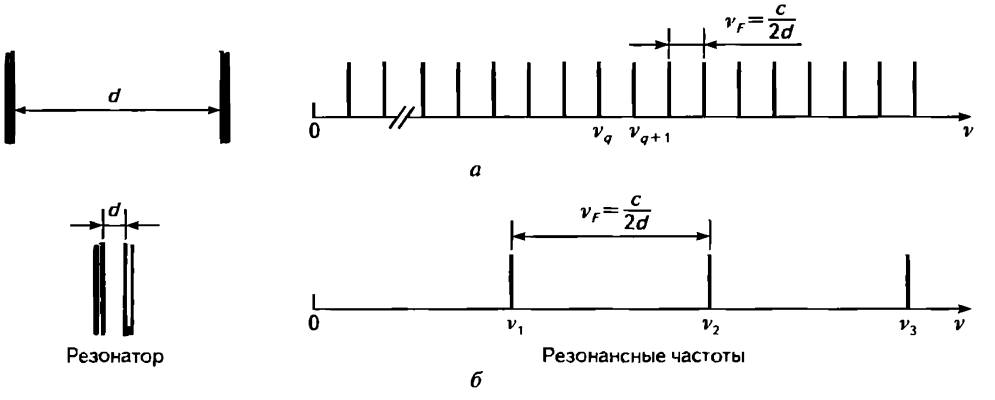


Рис. 10.4. Соседние резонансные частоты плоскозеркального резонатора разделены расстоянием $\nu_F = c/2d = c_0/2dn$, что иллюстрируется двумя примерами:

а — резонатор длиной d = 30 см, заполненный воздухом (n = 1), имеет межмодовое расстояние ν_F = 500 МГц; *б* — намного более короткий резонатор длиной d = 3 мкм имеет ν_F = 50 ТГц, так что первая мода имеет частоту, соответствующую длине волны 6 мкм, и оптическую область 700—900 нм, соответствующую частотной полосе шириной 95 ТГц, попадают всего две моды

Моды резонатора как бегущие волны

В качестве альтернативы можно определить моды, рассматривая волну, бегущую вперед и назад между двумя зеркалами (рис. 10.5, *а*). Модой является волна, воспроизводящая себя после полного обхода резонатора (см. приложение В). Набег фазы за один полный обход (расстояние $2d$), равный

$$\varphi = k2d = \frac{4\pi\nu d}{2c},$$

должен составлять целое число 2π:

$$\varphi = k2d = q2\pi, \quad q = 1, 2, \tag{10.8}$$

Этот результат не меняется при учете дополнительного фазового сдвига 2π, который может вноситься за счет отражения от двух зеркал (см. разд. 6.2). Как и ожидалось, мы снова получаем $kd = q\pi$, как и в (10.2), с тем же самым набором резонансных частот (10.5). Равенство (10.8) можно рассматривать как условие положительной обратной связи в системе, изображенной на рис. 10.5, *б*; оно означает, что выходной сигнал должен быть в фазе с входным сигналом.

Покажем теперь, что в стационарном режиме в резонаторе могут существовать только самовоспроизводящиеся волны или их комбинации. Рассмотрим монохроматическую волну с комплексной амплитудой U_0 в точке P, распространяющуюся вправо вдоль оси резонатора (см. рис. 10.5, *а*). Волна отражается зеркалом 2 и распространяется обратно к зеркалу 1, где снова отражается. Ее комплексная амплитуда в точке P становится равной U_1. Следующий полный

обход дает амплитуду U_2 и так далее. Поскольку исходная волна монохроматическая, она существует вечно. Все парциальные волны U_0, U_1, U_2, монохроматические и одновременно существуют всегда. Далее, их амплитуды одинаковы, поскольку предполагается, что их отражение и распространение не сопровождаются потерями. Полная волна представляет собой сумму бесконечного числа векторов одинаковой длины на комплексной плоскости

$$U = U_0 + U_1 + U_2 + \qquad (10.9)$$

как показано на рис. 10.5, *в*.

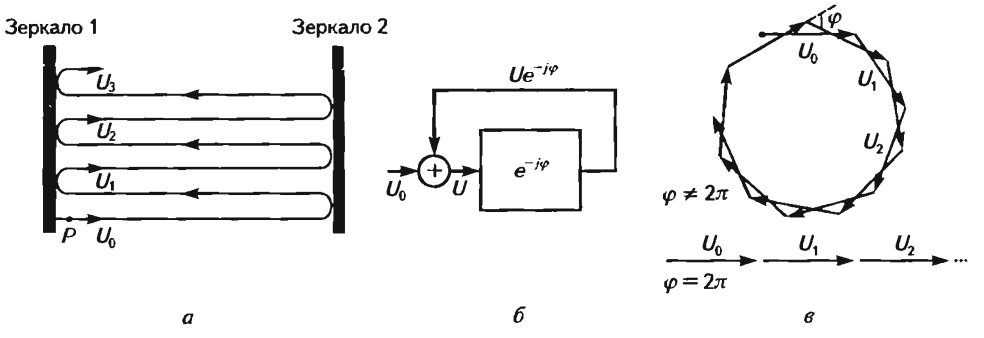


Рис. 10.5. Волна отражается туда и обратно между двумя зеркалами, приобретая набег фазы φ за один полный обход (*а*). Блок-схема оптической системы с обратной связью с фазовой задержкой φ (*б*). Векторная диаграмма на комплексной плоскости, представляющая сумму $U = U_0 + U_1 + ...$ для $\varphi \neq q^2\pi$ и для $\varphi = q^2\pi$ (*в*)

Угол между соседними векторами на комплексной плоскости равен набегу фазы $\varphi = k2d$ за один полный обход. Если длина начального вектора бесконечно мала, то бесконечно малой будет и длина всех остальных векторов. Длина суммы бесконечно малых векторов также бесконечно мала, если они не выстроены в одном направлении, как показано в нижней части рисунка 10.5, *в*, а в этом случае $\varphi = q2\pi$. Таким образом, бесконечно малая исходная волна может создать в резонаторе конечный запас энергии только в случае $\varphi = q2\pi$.

Резонаторы бегущей волны

В резонаторе бегущей волны оптическая мода распространяется в одном направлении по замкнутому пути, представляющему полный обход, и воспроизводит себя без обращения направления. Примерами являются **кольцевой резонатор и резонатор типа «галстук-бабочка»**, схематически показанные на рис. 10.6. Резонансные частоты мод можно получить путем приравнивания набега фазы за один полный обход к целому числу 2π. Каждой моде, распространяющейся по часовой стрелке, соответствует мода, распространяющаяся против часовой стрелки, и в случае совпадения такие моды называются вырожденными.

Упражнение 10.1

Резонансные частоты резонатора бегущей волны

Выведите выражения для резонансных частот и межмодового расстояния для трехзеркального кольцевого резонатора и четырехзеркального резонатора типа «галстук-бабочка», показанных на рис. 10.6. Считать, что на каждом зеркале приобретается фазовый сдвиг.

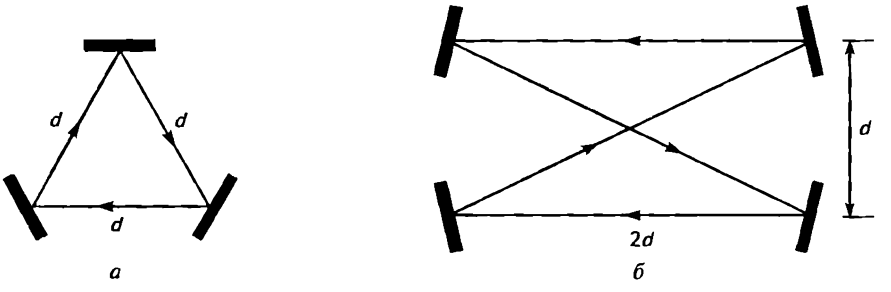


Рис. 10.6. Резонаторы бегущей волны:

а — трехзеркальный кольцевой резонатор; *б* — четырехзеркальный резонатор типа «галстук-бабочка»

Плотность мод

Число мод на единицу частоты есть обратная величина по отношению к межмодовому расстоянию, т. е. $1/\nu_F = 2d/c$ для каждой из двух ортогональных поляризаций. Тогда плотность мод $M(\nu)$, представляющая собой число мод на единицу частоты и на единицу длины резонатора, равна

$$\boxed{M(\nu) = \frac{4}{c}.} \qquad (10.10)$$

Плотность мод (одномерный резонатор)

Число мод в резонаторе длиной d в полосе частот $\Delta\nu$ равно $(4/c)d\Delta\nu$. Эта величина характеризует число степеней свободы оптических волн, существующих в резонаторе, т. е. число независимых способов организации этих волн.

Потери и спектральная ширина резонансов

Строгое условие, накладываемое на частоты оптических волн, могущих существовать в резонаторе, смягчается при наличии у резонатора потерь. Рассмотрим снова рис. 10.5, *а* и проследим за исходной волной U_0 в резонаторе в процессе ее многократного прохождения туда и обратно между зеркалами. Как обсуждалось выше, в результате получается бесконечная сумма

векторов на комплексной плоскости, показанной на рис. 10.5, *в*, где разность фаз, приобретаемая за один полный обход, составляет

$$\varphi = 2kd = \frac{4\pi\nu d}{c}. \tag{10.11}$$

Отражение от двух зеркал вносит дополнительный фазовый сдвиг, обычно 2π.

Однако при наличии потерь длина векторов на комплексной плоскости перестает быть одинаковой. Два последовательных вектора связаны комплексным коэффициентом ослабления за один обход $h = |r|e^{-j\varphi}$, возникающим из-за потерь за счет отражения на зеркалах и поглощения в среде, заполняющей резонатор. Соответствующий коэффициент ослабления интенсивности есть $|r|^2$, где $|r| < 1$. Таким образом, $U_1 = hU_0$, $U_2 = hU_1$, и т. д. В конце концов, получается суперпозиция бесконечного числа волн, каждая из которых отличается от предыдущей постоянным фазовым сдвигом и амплитудой, убывающей по закону геометрической прогрессии. Легко видеть, что

$$U = U_0 + U_1 + U_2 + \ldots = U_0 + hU_0 + h^2U_0 + \ldots = \frac{U_0}{1-h}.$$

Окончательный результат $U = U_0/(1 - h)$ легко понять с помощью простой схемы обратной связи, показанной на рис. 10.5, *б*.

Теперь интенсивность света в резонаторе получается равной

$$I = |U|^2 = \frac{|U_0|^2}{\left|1 - |r|e^{-j\varphi}\right|^2} = \frac{I_0}{1 + |r|^2 - 2|r|\cos\varphi}, \tag{10.12}$$

что можно переписать как

$$I = \frac{I_{\max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2(\varphi/2)}; \quad I_{\max} = \frac{I_0}{(1 - |r|)^2}. \tag{10.13}$$

Здесь $I_0 = |U_0|^2$ — интенсивность начальной волны; $\mathcal{F}$ — **параметр резкости** для резонатора,

$$\boxed{\mathcal{F} = \frac{\pi\sqrt{|r|}}{1 - |r|}.} \tag{10.14}$$

Параметр резкости

Предложенный выше подход аналогичен ранее развитому в подразд. 2.5.2, где комплексный коэффициент ослабления за один полный обход был выбран равным $h = |h|e^{+j\varphi}$. В рассматриваемом здесь случае этот коэффициент выбран равным $h = |r|e^{-jk\,2d} = |r|e^{-j\varphi}$, поскольку последовательность комплексных векторов определяется *задержкой* волны при ее распространении туда и обратно между зеркалами. Это различие носит несущественный характер и не влияет на результаты.

Действительно, (10.13) совпадает с (2.63), графически представленным на рис. 2.30, *б*. Интенсивность $I(\varphi)$ — периодическая функция φ с периодом 2π. При больших $\mathcal{F}$ $I(\varphi)$ имеет острые пики с центрами в точках $\varphi = q2\pi$, которые соответствуют выстраиванию всех векторов на комплексной плоскости в одну прямую линию. Пики имеют ширину по полувысоте $\Delta\varphi \approx 2\pi/\mathcal{F}$, что соответствует (2.65).

Интенсивность излучения внутри резонатора (10.13) можно иначе выразить как функцию оптической частоты монохроматической волны внутри резонатора $I(\nu)$ с использованием соотношения (10.11), из которого $\varphi = 4\pi\nu d/c$. Эта функция имеет вид

$$\boxed{I = \frac{I_{\max}}{1 + (2\mathcal{F}/\pi)^2 \sin^2(\pi\nu/\nu_F)}}; \quad I_{\max} = \frac{I_0}{(1-|r|)^2}, \tag{10.15}$$

где $\nu_F = c/2d$. Этот результат проиллюстрирован на рис. 10.7 и, очевидно, совпадает с показанным на рис. 2.30. Максимальная внутренняя интенсивность $I_{\max}$ достигается, когда второй член в знаменателе обращается в нуль, т. е. на резонансных частотах

$$\nu = \nu_q = q\nu_F; \quad q = 1, 2, \tag{10.16}$$

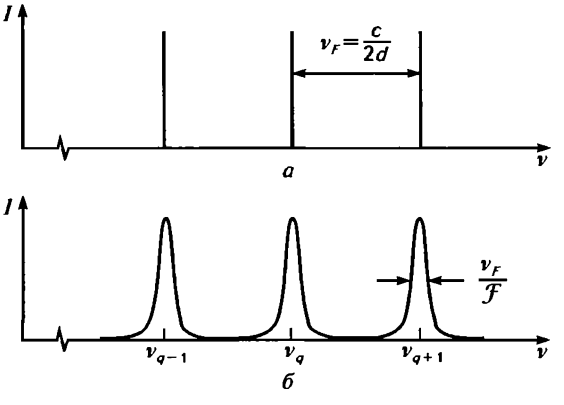


Рис. 10.7. В стационарном состоянии резонатор без потерь ($\mathcal{F} = \infty$) поддерживает световые волны только на резонансных частотах ν_q (*а*). В резонаторе с потерями наилучшим образом поддерживаются волны с частотами вблизи резонансов, однако могут существовать и волны с другими частотами (*б*)

Минимальная интенсивность, достигаемая между резонансами, составляет

$$I_{\min} = \frac{I_{\max}}{1 + (2\mathcal{F}/\pi)^2}. \tag{10.17}$$

Когда параметр резкости велик ($\mathcal{F} \gg 1$), спектральный отклик резонатора, очевидно, локализован в виде узких пиков вблизи резонансных частот, и $I_{\min}/I_{\max}$ мало. В этом случае ширина резонансных пиков по полувысоте $\delta\nu \approx \nu_F/\mathcal{F}$, поскольку $\delta\nu = (c/4\pi d)\Delta\varphi$ и $\Delta\varphi \approx \pi/\mathcal{F}$ в соответствии с (2.65). Этот простой результат дает основание для определения параметра резкости с помощью формулы (10.14).

Короче, спектральный отклик оптического резонатора Фабри—Перо характеризуется двумя параметрами:

- межмодовым расстоянием ν_F между частотами соседних резонансов

$$\boxed{\nu_F = \frac{c}{2d};} \qquad (10.18)$$

Межмодовое расстояние

- спектральной шириной отдельных мод резонатора

$$\boxed{\delta\nu \approx \frac{\nu_F}{\mathcal{F}}.} \qquad (10.19)$$

Спектральная ширина

Равенство (10.19) справедливо в обычном случае, когда $\mathcal{F} \gg 1$. Спектральная ширина $\delta\nu$ обратно пропорциональна параметру резкости $\mathcal{F}$. С ростом потерь $\mathcal{F}$ понижается и, следовательно, $\delta\nu$ растет.

Источники потерь резонатора

Существует два основных источника потерь в оптических резонаторах:

- Потери, возникающие вследствие неидеального отражения на зеркалах. Коэффициент отражения зеркал может быть понижен по двум причинам. Во-первых, небольшое его снижение необходимо, чтобы выводить наружу лазерное излучение, генерируемое в резонаторе. Во-вторых, конечный размер зеркал приводит к тому, что часть излучения попадает мимо них и теряется. Последний эффект меняет также пространственное распределение отраженной волны, обрезая ее по размеру зеркала. Отраженный свет создает на противоположном зеркале дифракционную картину, которая снова обрезается. Такие дифракционные потери можно рассматривать как эффективное снижение коэффициента отражения зеркала. Дальнейшие подробности, касающиеся дифракционных потерь, будут рассмотрены в подразд. 10.2.5.
- Потери, которые можно приписать поглощению и рассеянию, происходящему в среде между зеркалами. Фактор ослабления за один обход резонатора, связанный с этими эффектами, равен $\exp(-2\alpha_s d)$, где α_s — локальный коэффициент потерь среды, связанный с поглощением и рассеянием.

Для зеркал с коэффициентами отражения $\mathcal{R}_1 = |r_1|^2$ и $\mathcal{R}_2 = |r_2|^2$ интенсивность волны в результате двух отражений за один полный обход уменьшается на множитель $\mathcal{R}_1\mathcal{R}_2$. Такие потери называются «сосредоточенными», поскольку они происходят только в дискретных точках, где расположены зеркала. Если учесть также «распределенные» потери в толще среды, заполняющей резонатор, то фактор ослабления за один полный обход получается равным

$$|r|^2 = \mathcal{R}_1\mathcal{R}_2 \exp(-2\alpha_s d), \qquad (10.20)$$

что обычно записывается в виде

$$|r|^2 = \exp(-2\alpha_r d), \qquad (10.21)$$

где α_r — эффективный полный коэффициент распределенных потерь. Приравнивая (10.20) и (10.21) и беря натуральный логарифм от обеих частей, можно выразить α_r через параметры распределенных и сосредоточенных потерь α_s и $\mathcal{R}_1\mathcal{R}_2$ соответственно:

$$\boxed{\alpha_r = \alpha_s + \frac{1}{2d}\ln\frac{1}{\mathcal{R}_1\mathcal{R}_2}.} \qquad (10.22)$$

Коэффициент потерь

Последнему выражению можно придать вид

$$\alpha_r = \alpha_s + \alpha_{m_1} + \alpha_{m_2}, \qquad (10.23)$$

где величины

$$\alpha_{m_1} = \frac{1}{2d}\ln\frac{1}{\mathcal{R}_1}; \quad \alpha_{m_2} = \frac{1}{2d}\ln\frac{1}{\mathcal{R}_2} \qquad (10.24)$$

представляют эффективные коэффициенты распределенных потерь, связанных с зеркалами 1 и 2 соответственно.

Эти коэффициенты потерь можно представить в более простом виде для зеркал с высоким коэффициентом отражения. Если $\mathcal{R}_1 \approx 1$, то

$$\ln\left(\frac{1}{\mathcal{R}_1}\right) = -\ln\mathcal{R}_1 = -\ln\left[1-(1-\mathcal{R}_1)\right] \approx 1-\mathcal{R}_1,$$

где использовано разложение в ряд Тейлора $\ln(1-\Delta) \approx -\Delta$, верное при $|\Delta| \ll 1$. Это позволяет записать

$$\alpha_{m_1} \approx \frac{1-\mathcal{R}_1}{2d}. \qquad (10.25)$$

Аналогично, если $\mathcal{R}_2 \approx 1$, имеем

$$\alpha_{m_2} \approx \frac{1-\mathcal{R}_2}{2d}.$$

Далее, если $\mathcal{R}_1 = \mathcal{R}_2 = \mathcal{R} \approx 1$, то

$$\alpha_r \approx \alpha_s + \frac{1-\mathcal{R}}{d}. \qquad (10.26)$$

Параметр резкости $\mathcal{F}$ можно выразить как функцию эффективного коэффициента потерь α_r путем подстановки (10.21) в (10.14). Результат имеет вид

$$\mathcal{F} = \frac{\pi\exp(-\alpha_r d/2)}{1-\exp(-\alpha_r d)}, \qquad (10.27)$$

что графически показано на рис. 10.8. Ясно, что параметр резкости убывает с ростом потерь. Если $\alpha_r d \ll 1$, то $\exp(-\alpha_r d) \approx 1 - \alpha_r d$, тогда

$$\boxed{\mathcal{F} \approx \frac{\pi}{\alpha_r d}}. \qquad (10.28)$$

Параметр резкости и коэффициент потерь

Это показывает, что в данном предельном случае параметр резкости обратно пропорционален фактору потерь $\alpha_r d$.

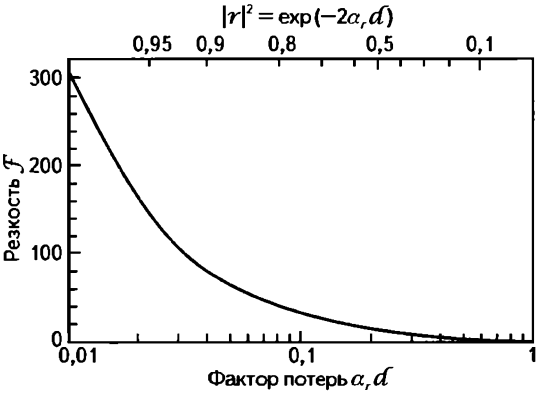


Рис. 10.8. Параметр резкости для оптического резонатора в зависимости от фактора нотерь $\alpha_r d$, где α_r — эффективный полный коэффициент распределенных потерь. За один обход интенсивность ослабляется на множитель $|r|^2 = \exp(-2\alpha_r d)$

Упражнение 10.2

Моды резонатора и спектральная ширина

Определите межмодовое расстояние и спектральную ширину мод резонатора Фабри—Перо с коэффициентами отражения зеркал 0,98 и 0,99, расположенных на расстоянии $d = 100$ см. Считайте, что среда имеет показатель преломления $n = 1$ и пренебрежимо малые потери. Справедливо в этом случае приближение, использованное при выводе (10.28)?

Время жизни фотона

Связь между шириной линии резонанса и потерями в резонаторе можно рассматривать как проявление соотношения неопределенности между частотой и временем, что мы сейчас покажем. Поставляя (10.18) и (10.28) в (10.19), получаем

$$\delta\nu \approx \frac{c/2d}{\pi/\alpha_r d} = \frac{c\alpha_r}{2\pi}. \qquad (10.29)$$

Поскольку α_r — потери на единицу длины, то $c\alpha_r$ — потери в единицу времени. Определяя характерное время затухания

$$\tau_p = \frac{1}{c\alpha_r} \qquad (10.30)$$

как **время жизни излучения** в резонаторе или **время жизни фотона**, получаем

$$\delta\nu = \frac{1}{2\pi\tau_p}. \tag{10.31}$$

Произведение неопределенностей времени и частоты тогда составляет $\delta\nu\tau_p = 1/2\pi$. Таким образом, уширение резонансных линий можно рассматривать как следствие затухания оптической энергии, обусловленного наличием потерь в резонаторе. Электрическое поле, затухающее как $\exp(-t/2\tau_p)$ и соответствующее энергии, затухающей как $\exp(-t/\tau_p)$, имеет Фурье-образ, пропорциональный функции $1/(1 + j4\pi\nu\tau_p)$, которая имеет спектральную ширину на уровне полувысоты, равную $\delta\nu = 1/2\pi\tau_p$.

Добротность Q

Добротность Q часто используется для характеристики электрических цепей и микроволновых резонаторов. Этот параметр определяется как

$$Q = 2\pi \frac{\text{запасенная энергия}}{\text{потери энергии за цикл}}. \tag{10.32}$$

Большие значения Q связаны с малыми потерями в резонаторе. Последовательная RLC-цепь имеет резонансную частоту

$$\nu_0 \approx \frac{1}{2}\pi\sqrt{LC}$$

и добротность

$$Q = 2\pi\nu_0 \frac{L}{R},$$

где R, L и C — сопротивление, индуктивность и емкость резонансной цепи соответственно.

Добротность оптического резонатора определяется, исходя из того факта, что запасенная в нем энергия теряется со скоростью $c\alpha_r$ (в единицу времени), что эквивалентно скорости $c\alpha_r/\nu_0$ (за один обход), так что

$$Q = \frac{2\pi}{c\alpha_r/\nu_0} = \frac{2\pi\nu_0}{c\alpha_r}. \tag{10.33}$$

Поскольку $\delta\nu = c\alpha_r/2\pi$,

$$Q = \frac{\nu_0}{\delta\nu}. \tag{10.34}$$

С учетом (10.33) добротность выражается через время жизни поля в резонаторе (время жизни фотона) $\tau_p = 1/c\alpha_r$ как

$$Q = 2\pi\nu_0\tau_p. \tag{10.35}$$

Окончательно, комбинируя (10.19) и (10.34), получаем связь между Q и параметром резкости $\mathcal{F}$ резонатора

$$Q = \frac{\nu_0}{\nu_F}\mathcal{F}. \tag{10.36}$$

Поскольку частоты оптического резонатора ν_0 в типичном случае намного больше межмодового расстояния ν_F, имеем $Q \gg \mathcal{F}$. Добротность оптического резонатора обычно намного больше, чем у резонаторов микроволнового диапазона.

Резюме

- Для характеристики потерь в оптическом резонаторе удобны два параметра: коэффициент потерь α_r (см$^{-1}$) и время жизни фотона $\tau_p = 1/c\alpha_r$ (с).
- Качество оптического резонатора длиной d, работающего на частоте ν_0, характеризуется двумя безразмерными параметрами: параметром резкости $\mathcal{F} = \pi/\alpha_r d$ и добротностью $Q = 2\pi\nu_0\tau_p$.
- Спектральные характеристики оптического резонатора описываются двумя частотами: межмодовым расстоянием $\nu_F = c/2d$, которое также называется областью дисперсии, и спектральной шириной $\delta\nu = \nu_F/\mathcal{F}$.

10.1.2. Внеосевые моды резонатора

Оптический резонатор с идеально параллельными плоскими зеркалами бесконечных размеров может поддерживать также наклонные, или внеосевые, моды. Плоская волна, распространяющаяся под углом θ к оси резонатора (оси z), отражается туда и обратно между зеркалами (рис. 10.9, *а*) как волна, направляемая вдоль поперечной оси x. Такие направляемые волны были описаны в разд. 8.1.

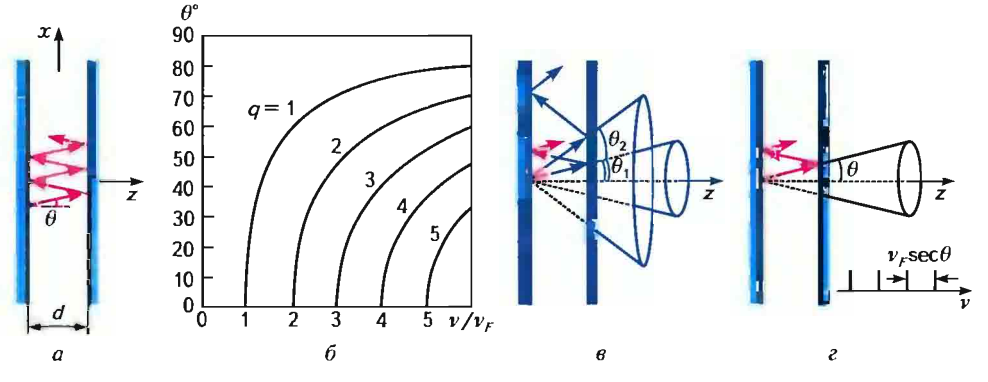


Рис. 10.9. Внеосевая мода плоского зеркального резонатора (*а*). Связь между углами отклонения и резонансными частотами мод (*б*). Внеосевые моды на фиксированной частоте $\nu > \nu_F$ (*в*). Резонансные частоты внеосевой моды с заданным углом θ (*г*)

Граничные условия на зеркалах требуют, чтобы аксиальная компонента постоянной распространения, $k_z = k\cos\theta$, была кратна π/d. Однако такое условие не накладывается на поперечную компоненту k_x, поскольку резонатор открыт в направлении x. Условие $k\cos\theta = q\pi/d$, где q — целое, можно записать в виде

$$\nu = q\nu_F \sec\theta, \quad q = 1, 2, \tag{10.37}$$

где $\nu_F = c/2d$. Это соотношение, графически иллюстрируемое рис. 10.8, *б*, эквивалентно условию самосогласования для направляемых мод плоского зеркального волновода (см. разд. 8.1). Оно идентично также условию (7.43) максимальности коэффициента пропускания наклонной волны через эталон Фабри—Перо. Как показано на рис. 10.9, *в*, на данной частоте ν существуют моды с дискретными значениями угла θ_q, которые удовлетворяют условию $\cos\theta_q = q\nu_F/\nu$. Это углы наклона направляемых мод волновода. С другой стороны, для каждого заданного угла θ частоты мод равны $\nu_q = q\nu_F/\cos\theta$, как показано на рис. 10.9, *г*. Чем больше угол отклонения, тем больше межмодовое частотное расстояние.

10.2. РЕЗОНАТОРЫ СО СФЕРИЧЕСКИМИ ЗЕРКАЛАМИ

Конфигурация резонатора с плоскими зеркалами, обсуждавшаяся в предыдущем разделе, весьма неустойчива к разъюстировке. Если зеркала не идеально параллельны или лучи не идеально перпендикулярны поверхности зеркал, то лучи претерпевают последовательность поперечных смещений, в результате которых, в конце концов, выходят из резонатора конечных размеров (см. рис. 10.2). Более устойчивую конфигурацию для удержания лучей имеют резонаторы со сферическими зеркалами, что делает их менее чувствительными к разъюстировке, при выполнении определенных геометрических условий.

Резонатор со сферическими зеркалами состоит из двух сферических зеркал с радиусами R_1 и R_2, разделенных расстоянием d (рис. 10.10). Линия, соединяющая центры зеркал, определяет оптическую ось (ось z), относительно которой система симметрична. Каждое из зеркал может быть вогнутым ($R < 0$) или выпуклым ($R > 0$). Плоский зеркальный резонатор представляет собой частный случай сферического с $R_1 = R_2 = \infty$. Используя результаты подразд. 1.4.4, вначале изучим условия, необходимые для удержания лучей. Затем, используя результаты гл. 3, определим моды резонатора и их резонансные частоты. В конце мы кратко обсудим следствия конечного размера зеркал.

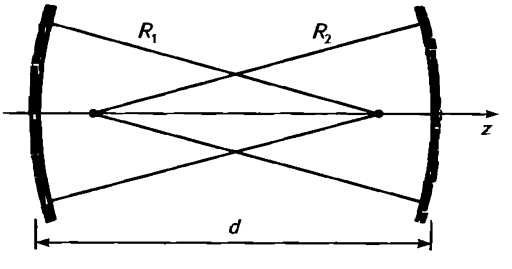


Рис. 10.10. Геометрия резонаторы со сферическими зеркалами. На этой иллюстрации оба зеркала вогнутые (их радиусы кривизны отрицательны)

10.2.1. Удержание лучей

Начнем с лучевой оптики для определения условий удержания лучей света в резонаторе со сферическими зеркалами. Будем рассматривать только меридиональные лучи (лучи, лежащие в плоскости, проходящей через оптическую ось) и ограничимся параксиальными лучами (лучами, образующие с оптической осью малые углы). В этом случае для изучения траекторий лучей в резонаторе можно использовать методы матричной оптики, введенные в разд. 1.4 и применимые только для меридиональных и параксиальных лучей в системах с осевой симметрией.

Резонатор — периодическая оптическая система, поскольку после полного обхода распространение луча повторяется. Поэтому можно использовать результаты анализа периодических оптических систем, представленные в подразд. 1.4.4. Пусть y_m и θ_m — положение и угол наклона луча после m полных обходов (рис. 10.11). По данным y_m и θ_m определим y_{m+1} и θ_{m+1}, отслеживая ход луча в системе.

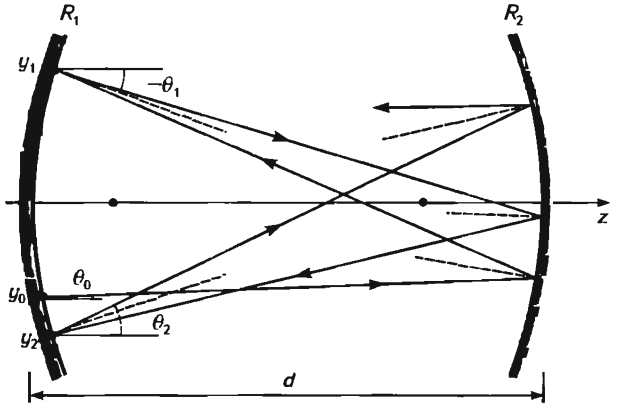


Рис. 10.11. Положение и наклон луча после m обходов обозначены через y_m и θ_m соответственно, где $m = 0, 1, 2,$ На схеме $\theta_1 < 0$, поскольку луч направлен вниз. Углы преувеличены для наглядности, на самом деле все лучи параксиальны, так что $\sin\theta \approx \operatorname{tg}\theta \approx \theta$, а расстояние, проходимое всеми лучами от зеркала до зеркала, $\approx d$

Для параксиальных лучей, когда все углы малы, соотношение между (y_{m+1}, θ_{m+1}) и (y_m, θ_m) линейно и может быть записано в матричном виде как

$$\begin{bmatrix} y_{m+1} \\ \theta_{m+1} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_m \\ \theta_m \end{bmatrix}. \tag{10.38}$$

Матрица передачи луча, ход которого показан на рис 10.11, начиная с его нижней левой части, где заданы y_0 и θ_0, для одного полного обхода

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \dfrac{2}{R_1} & 1 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \dfrac{2}{R_2} & 1 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}. \tag{10.39}$$

Этот каскад матриц передачи представляет справа налево [см. (1.43) и (1.48)]:

- распространение на расстояние d в свободном пространстве;
- отражение от зеркала радиусом R_2;
- отражение от зеркала радиусом R_1.

Как показано в подразд. 1.4.4, решение разностного уравнения (10.38) имеет вид $y_m = y_{\max} F^m \sin(m\varphi + \varphi_0)$, где $F^2 = AD - BC$; $\varphi = \arccos(b/F)$; $b = (A + D)/2$; $y_{\max}$ и φ_0 — постоянные, определяемые из начального положения и отклонения луча. В рассматриваемом случае $F = 1$, так что

$$y_m = y_{\max} \sin(m\varphi + \varphi_0); \tag{10.40}$$

$$\varphi = \arccos b, \quad b = 2\left(1 + \frac{d}{R_1}\right)\left(1 + \frac{d}{R_2}\right) - 1. \tag{10.41}$$

Решение (10.40) — гармоническая и, следовательно, ограниченная функция при условии, что $\varphi = \arccos b$ — действительная величина. Последнее обеспечивается, если $|b| \le 1$, т. е. $-1 \le b \le 1$, так что

$$0 \le \left(1 + \frac{d}{R_1}\right)\left(1 + \frac{d}{R_2}\right) \le 1. \tag{10.42}$$

Удобно записать это условие через величины $g_1 = 1 + d/R_1$ и $g_2 = 1 + d/R_2$, которые называются **g-параметрами**:

$$\boxed{0 \le g_1 g_2 \le 1.} \tag{10.43}$$

Условие устойчивости

Резонатор **устойчив**, если это условие выполнено. Этот же результат получается в волновой оптике, что мы впоследствии продемонстрируем.

Если условие устойчивости (10.43) не выполняется, то φ — мнимая величина, так что y_m в (10.40) превращается в гиперболический синус, который неограниченно растет с увеличением m. В этом случае говорят, что резонатор **неустойчив**. На границе области устойчивости (когда неравенства превращаются в равенства) резонатор называется **условно устойчивым**.

Полезное графическое представление условия устойчивости (рис. 10.12) получается, если каждой паре параметров g_1 и g_2 сопоставить точку на плоскости с координатами (g_1, g_2). Левое неравенство в (10.43) эквивалентно {$g_1 \ge 0$ и $g_2 \ge 0$ или $g_1 \le 0$ и $g_2 \le 0$}, так что все точки устойчивости должны лежать в первом или третьем квадрантах. Правое неравенство в (10.43) означает, что точки устойчивости лежат в области, ограниченной гиперболой $g_1 g_2 = 1$. Светлая часть плоскости на рис. 10.12 — область, где оба условия выполняются одновременно, т. е. область устойчивости.

Симметричные резонаторы по определению имеют одинаковые зеркала ($R_1 = R_2 = R$), поэтому $g_1 = g_2 = g$. Резонаторам такого типа на диаграмме

(рис. 10.12) соответствует прямая $g_1 = g_2$. Условие устойчивости в этом случае приобретает вид $g^2 \le 1$, или $-1 \le 1 + d/R \le 1$, откуда

$$\boxed{0 \le \frac{d}{(-R)} \le 2.} \qquad (10.44)$$

Условие устойчивости (симметричный резонатор)

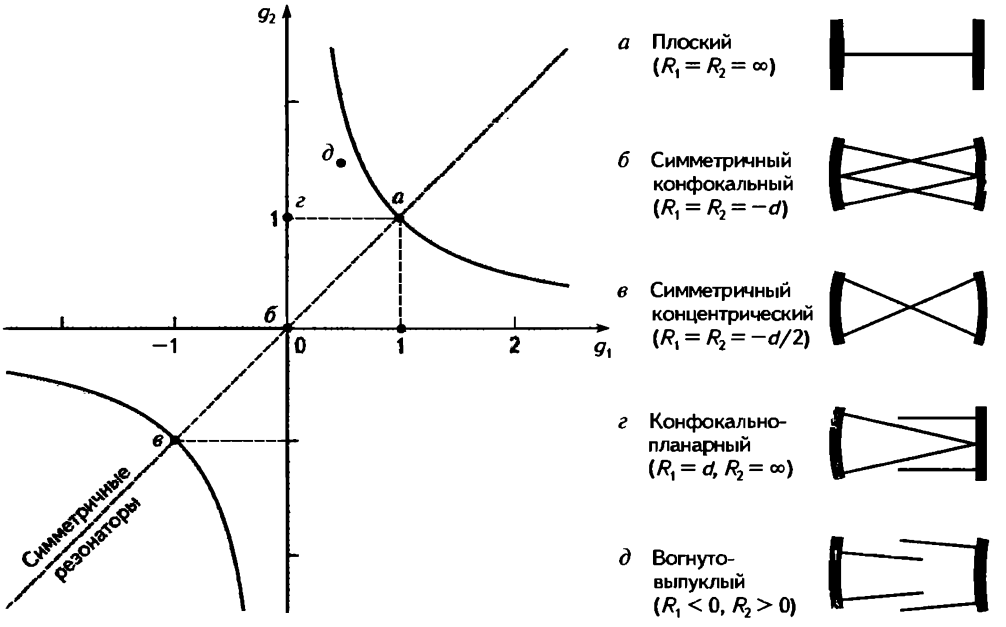


Рис. 10.12. Диаграмма устойчивости резонатора. Резонатор со сферическими зеркалами устойчив, если параметры $g_1 = 1 + d/R_1$ и $g_2 = 1 + d/R_2$ лежат в светлых частях плоскости, которые ограничены линиями $g_1 = 0$, $g_2 = 0$ и гиперболой $g_2 = 1/g_1$. R отрицательно для вогнутого зеркала и положительно для выпуклого. Часто используемые конфигурации резонаторов показаны буквами на диаграмме и схематически изображены справа. Симметричным резонаторам соответствует прямая $g_1 = g_2$

Чтобы удовлетворить (10.44), устойчивый симметричный резонатор должен иметь вогнутые зеркала ($R < 0$), радиусы которых больше половины длины резонатора. Три типа резонаторов этого класса представляют особый интерес: $d/(-R) = 0$, 1 и 2, что соответствует **плоскому**, **конфокальному** и **концентрическому** резонатору.

У симметричного конфокального резонатора $(-R) = d$, так что центры кривизны обоих зеркал совпадают. Тогда $b = -1$ и $\varphi = \pi$, так что положение луча в (10.40) определяется соотношением $y_m = y_{\max} \sin(m\pi + \varphi_0)$, т. е. $y_m = (-1)^m y_0$. Лучи, начинающиеся в положении y_0, при любом начальном отклонении отображаются в точку $y_1 = -y_0$, затем в $y_2 = y_0$ и т. д. Каждый луч повторяет себя

после двух полных обходов (рис. 10.13). Таким образом, все параксиальные лучи удерживаются в резонаторе независимо от их начального положения и отклонения. Это существенное улучшение по сравнению с плоским резонатором, для которого только лучи с нулевым отклонением повторяют себя, схематически показано на рис. 10.2.

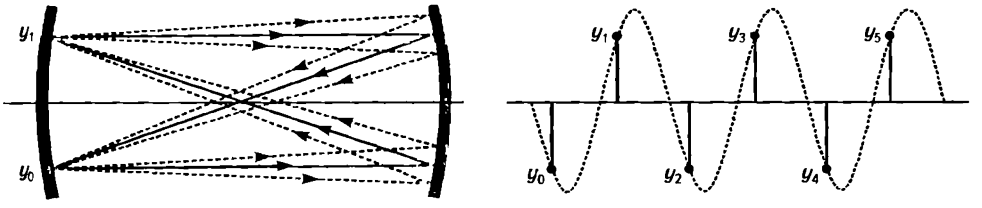


Рис. 10.13. Все параксиальные лучи в симметричном конфокальном резонаторе повторяют себя после двух полных обходов независимо от их начального положения и отклонения. Углы на рисунке преувеличены для наглядности

Резюме

Условие устойчивости параксиальных лучей в резонаторе, состоящем из двух сферических зеркал с радиусами кривизны R_1 и R_2, разделенных расстоянием d, есть $0 \le g_1 g_2 \le 1$, где $g_1 = 1 + d/R_1$ и $g_2 = 1 + d/R_2$. Условие устойчивости симметричного резонатора есть $0 \le d/(-R) \le 2$. Это условие, в частности, определяет устойчивость плоских, симметричных конфокальных и симметричных концентрических резонаторов.

Упражнение 10.3

Максимальная длина устойчивого резонатора

Резонатор образован вогнутыми зеркалами с радиусами 50 см и 100 см. Определите максимальную длину резонатора, удовлетворяющего условию устойчивости.

10.2.2. Гауссовы моды

Хотя подход геометрической оптики, рассмотренный в предыдущем разделе, полезен для определения геометрических условий удержания лучей, он не может дать информации о резонансных частотах и пространственном распределении интенсивности мод резонатора. Для их определения необходимо обратиться к волновой оптике. Покажем теперь, что гауссовы пучки являются решениями параксиального уравнения Гельмгольца при граничных условиях, накладываемых наличием двух сферических зеркал резонатора. Покажем также, что в более общем случае модами сферического резонатора являются пучки Гаусса—Эрмита. В ходе нашего анализа мы получим выражения для резонансных частот и пространственных распределений интенсивности мод резонатора.

Гауссовы пучки

Как обсуждалось в гл. 3, гауссов пучок представляет собой аксиально-симметричную волну, энергия которой сосредоточена вблизи ее оси (оси z), а нормали к волновому фронту — параксиальные лучи (рис. 10.14). В соответствии с (3.12) на продольном расстоянии от перетяжки пучка распределение его интенсивности в плоскости определяется гауссовой функцией

$$I = I_0 \left(\frac{W_0}{W(z)} \right)^2 \exp\left[-\frac{2(x^2 + y^2)}{W^2(z)} \right].$$

Ширина этого распределения дается формулой (3.8):

$$W(z) = W_0 \sqrt{1 + \left(\frac{z}{z_0} \right)^2} \tag{10.45}$$

где z_0 — расстояние, называемое рэлеевской длиной, на котором кривизна волнового достигает максимума. Ширина (радиус) пучка $W(z)$ растет от минимального значения W_0 в обе стороны от перетяжки, расположенной при $z = 0$. Радиус кривизны волнового фронта, даваемый формулой (3.9),

$$R(z) = z \left[1 + \left(\frac{z_0}{z} \right)^2 \right] \tag{10.46}$$

убывает от ∞ при $z = 0$ до минимального значения при $z = z_0$, а затем снова растет, причем при больших z_0 этот рост становится линейным. При $z > 0$ волна расходится и $R(z) > 0$, а при $z < 0$ сходится, $R(z) < 0$. Рэлеевская длина z_0 связана с радиусом перетяжки пучка соотношением (3.11)

$$z_0 = \frac{\pi W_0^2}{\lambda}. \tag{10.47}$$

Глубина резкости равна $2z_0$, т. е. удвоенной рэлеевской длине.

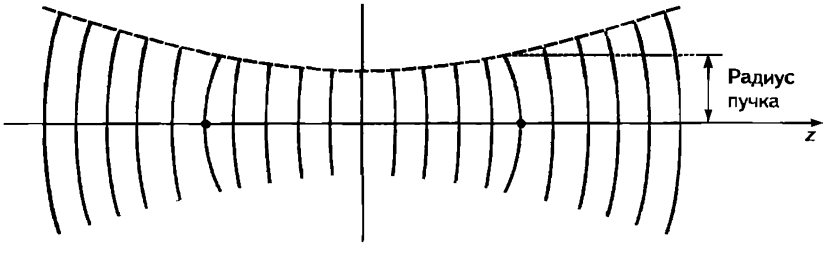


Рис. 10.14. Волновые фронты гауссова пучка (сплошные кривые) и ширина пучка (штриховая линия)

Гауссов пучок — мода резонатора со сферическими зеркалами

Гауссов пучок после отражения от сферического зеркала будет повторять себя, если радиус кривизны волнового фронта равен радиусу кривизны зеркала (см. подразд. 3.2.3). Следовательно, если радиусы кривизны волновых фронтов гауссова пучка в плоскостях, разделенных расстоянием d, согласуются с радиусами кривизны зеркал, находящихся на таком же расстоянии d друг от друга, то пучок, падающий на первое зеркало, отразится и, повторяя себя, достигнет второго зеркала, отразившись от которого повторится и дойдет до первого зеркала, и т. д. Это значит, что пучок может самосогласованно существовать в резонаторе со сферическими зеркалами, удовлетворяя при этом уравнению Гельмгольца и граничным условиям на зеркалах. При условии, что фаза также воспроизводит саму себя, как обсуждалось в подразд. 10.2.3, в этом случае говорят, что гауссов пучок является собственной модой резонатора.

Перейдем к определению гауссова пучка, согласованного с резонатором со сферическими зеркалами, имеющими радиусы кривизны R_1 и R_2 и находящимися на расстоянии d друг от друга. Задача иллюстрируется рис. 10.15 для частного случая, когда оба зеркала вогнутые ($R_1 < 0$ и $R_2 < 0$).

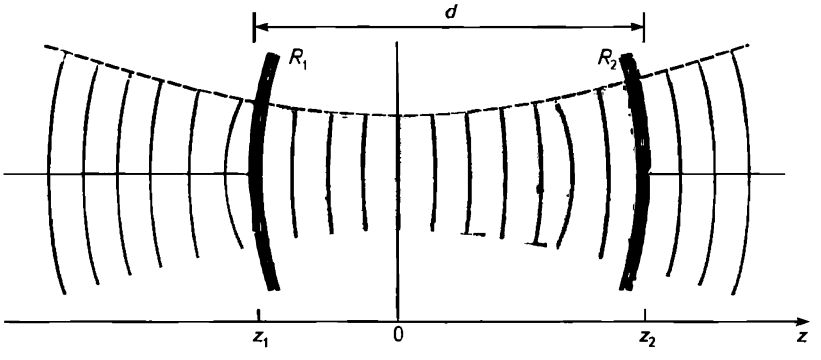


Рис. 10.15. Подгонка гауссова пучка к двум зеркалам, расположенным на расстоянии. Их радиусы кривизны R_1 и R_2. Оба зеркала взяты вогнутыми, так что R_1 и R_2 отрицательны, как и z_1

Ось z задается центрами кривизны зеркал. Центр пучка, который нужно найти, выбран за начало координат $z = 0$, зеркала R_1 и R_2 расположены в точках z_1 и

$$z_2 = z_1 + d \tag{10.48}$$

соответственно. Отрицательное значение z_1 отражает тот факт, что центр пучка находится справа от первого зеркала, положительное значение — что слева. Значения z_1 и z_2 определяются путем приравнивания радиуса кривизны волнового фронта пучка к радиусам R_1 зеркал z_1 при R_2 и при z_2. Особое внимание следует уделить знакам. Если оба зеркала вогнутые, то они имеют отрицательные радиусы. Однако радиус кривизны волнового фронта был определен как по-

ложительный при $z > 0$ (на зеркале 2) и отрицательный при $z < 0$ (на зеркале 1). Поэтому приравниваем $R_1 = R(z_1)$, но $R_2 = -R(z_2)$, и получаем

$$R_1 = z_1 + \frac{z_0^2}{z_1}; \tag{10.49}$$

$$-R_2 = z_2 + \frac{z_0^2}{z_2}. \tag{10.50}$$

Решая (10.48), (10.49) и (10.50) относительно z_1, z_2 и z_0, находим

$$z_1 = \frac{-d(R_2 + d)}{R_1 + R_2 + 2d}; \quad z_2 = z_1 + d; \tag{10.51}$$

$$z_0^2 = \frac{-d(R_1 + d)(R_2 + d)(R_1 + R_2 + d)}{(R_2 + R_1 + 2d)^2}, \tag{10.52}$$

что согласуется с (3.27) и (3.28) (если R_2 заменить на $-R_2$).

После того как определены положение центра пучка и глубина резкости $2z_0$, о пучке известно все (см. подразд. 3.1.2). Радиус перетяжки равен

$$W_0 = \sqrt{\frac{\lambda z_0}{\pi}},$$

а радиусы кривизны волнового фронта на зеркалах

$$W_i = W_0\sqrt{1 + \left(\frac{z_i}{z_0}\right)^2}, \quad i = 1,\ 2. \tag{10.53}$$

Чтобы решения (10.51), (10.52) действительно представляли гауссов пучок, величина z_0 должна быть действительной. Мнимое значение z_0 означало бы, что гауссов пучок представляет собой параболоидальную волну, которая является неограниченным решением параксиального уравнения Гельмгольца (см. подразд. 3.1.1). Используя (10.52), нетрудно показать, что условие $z_0^2 > 0$ эквивалентно

$$0 \le \left(1 + \frac{d}{R_1}\right)\left(1 + \frac{d}{R_2}\right) \le 1. \tag{10.54}$$

Это условие устойчивости резонатора (10.42), полученное выше в рамках оптики лучей.

Упражнение 10.4

Плоско-вогнутый резонатор

При условии, что зеркало 1 плоское ($R_1 = \infty$), определите условия устойчивости и глубину резкости, а также ширину пучка в перетяжке и на каждом зеркале как функцию $d/|R_2|$.

Гауссова мода симметричного резонатора со сферическими зеркалами

Результаты, даваемые формулами (10.48)—(10.52), существенно упрощаются для симметричного резонатора с вогнутыми зеркалами. Поставляя $R_1 = R_2 = -|R|$ в (10.51), находим $z_1 = -d/2$ и $z_2 = d/2$. Центр пучка находится в центре резонатора и

$$z_0 = \frac{d}{2}\sqrt{2\frac{|R|}{d} - 1}; \tag{10.55}$$

$$W_0^2 = \frac{\lambda d}{2\pi}\sqrt{2\frac{|R|}{d} - 1}; \tag{10.56}$$

$$W_1^2 = W_2^2 = \frac{\lambda d/\pi}{\sqrt{(d/|R|)\left[2 - (d/|R|)\right]}}. \tag{10.57}$$

Условие устойчивости принимает вид

$$0 \le \frac{d}{|R|} \le 2. \tag{10.58}$$

Теперь для резонатора с заданным расстоянием d между зеркалами изучим влияние кривизны зеркал на радиус пучка в перетяжке W_0 и на зеркалах $W_1 = W_2$. Увеличение кривизны означает увеличение $d/|R|$, поскольку *радиус кривизны* уменьшается, когда *кривизна* растет. Результаты показаны на рис. 10.16. Для резонатора с плоскими зеркалами $d/|R| = 0$, поэтому W_0 и W_1 бесконечны, что соответствует плоской волне, а не гауссову пучку. По мере роста $d/|R|$ величина W_0 уменьшается и для концентрического резонатора ($d/|R| = 2$) обращается в нуль; при этом $W_1 = W_2 = \infty$. В этом предельном случае резонатор вместо гауссова пучка поддерживает сферическую волну.

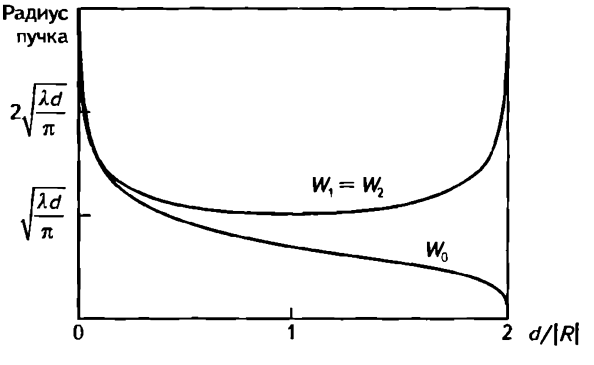


Рис. 10.16. Ширина пучка в перетяжке W_0 и на зеркалах $W_1 = W_2$ для симметричного резонатора с вогнутыми сферическими зеркалами как функция отношения $d/|R|$. Резонатору с плоскими зеркалами соответствует $d/|R| = 0$. У симметричного конфокального и концентрического резонаторов $d/|R| = 1$ и $d/|R| = 2$ соответственно

Ширина пучка на зеркалах достигает своего минимального значения

$$W_1 = W_2 = \sqrt{\frac{\lambda d}{\pi}},$$

когда $d/|R| = 1$, т. е. для симметричного конфокального резонатора. В этом случае

$$z_0 = \frac{d}{2}; \tag{10.59}$$

$$W_0 = \sqrt{\frac{\lambda d}{2\pi}}, \tag{10.60}$$

$$W_1 = W_2 = \sqrt{2}W_0. \tag{10.61}$$

При этом глубина резкости $2z_0$ равна длине резонатора d, как показано на рис. 10.17. По этой причине параметр $2z_0$ иногда называют конфокальным параметром. Длинный резонатор имеет большую глубину резкости. Радиус перетяжки пропорционален квадратному корню из расстояния между зеркалами. Например, при $\lambda_0 = 633$ нм (длина волны Hc—Ne-лазера) гауссов пучок в резонаторе длиной $d = 100$ см имеет радиус перетяжки $W_0 = 0{,}32$ мм, в то время как резонатор длиной 25 см на той же длине волны поддерживает пучок с перетяжкой вдвое меньшего радиуса 0,16 мм. Ширина пучка на каждом из зеркал больше, чем в перетяжке, в $\sqrt{2}$ раз.

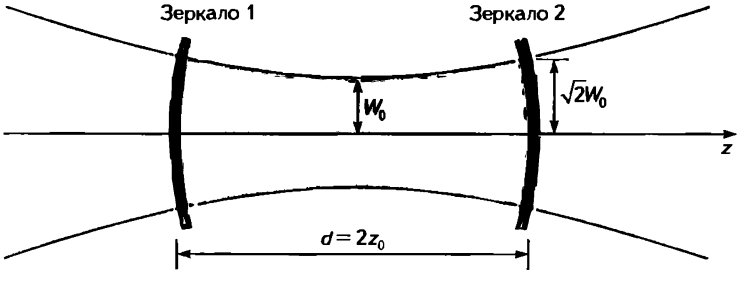


Рис. 10.17. Гауссов пучок в симметричном конфокальном резонаторе с вогнутыми зеркалами. Глубина резкости $2z_0$ равна длине резонатора d. Ширина пучка на зеркалах в $\sqrt{2}$ раз больше, чем в перетяжке

10.2.3. Резонансные частоты

Как отмечалось в подразд. 10.2.2, гауссов пучок является собственной модой резонатора со сферическими зеркалами при условии, что волновые фронты отражаются сами на себя, всегда проходя тот же путь, при этом фаза также воспроизводит саму себя.

В соответствии с (3.23) фаза гауссова пучка

$$\varphi(\rho, z) = kz - \zeta(z) + \frac{k\rho^2}{2R(z)}, \tag{10.62}$$

где $\zeta(z) = \operatorname{arctg}(z/z_0)$ и $\rho^2 = x^2 + y^2$. В точках оптической оси ($\rho = 0$) имеем $\varphi(0, z) = kz - \zeta(z)$, так что задержка фазы по сравнению с плоской волной равна $\zeta(z)$. В местах расположения зеркал, z_1 и z_2, имеем

$$\varphi(0, z_1) = kz_1 - \zeta(z_1); \tag{10.63}$$

$$\varphi(0, z_2) = kz_2 - \zeta(z_2). \tag{10.64}$$

Поскольку поверхности зеркал совпадают с волновыми фронтами, все точки каждого зеркала имеют одинаковую фазу. При распространении пучка от зеркала 1 до зеркала 2 его фаза меняется на

$$\varphi(0, z_2) - \varphi(0, z_1) = k(z_2 - z_1) - [\zeta(z_2) - \zeta(z_1)] = kd - \Delta\zeta, \tag{10.65}$$

где

$$\Delta\zeta = \zeta(z_2) - \zeta(z_1). \tag{10.66}$$

Таким образом, за один полный обход фаза меняется на $2kd - 2\Delta\zeta$.

Чтобы пучок действительно воспроизводил себя, набег фазы за один полный обход должен быть равен нулю или кратен $\pm 2\pi$, т. е. $2kd - 2\Delta\zeta = 2\pi q$, $q = 0, \pm 1, \pm 2,$ Осуществляя подстановку $k = 2\pi\nu/c$ и $\nu_F = c/2d$, получаем частоты, которые удовлетворяют этому условию:

$$\boxed{\nu_q = q\nu_F + \frac{\Delta\zeta}{\pi}\nu_F.} \tag{10.67}$$

Резонансные частоты Гауссовы моды

Частотное расстояние между соседними модами составляет $\nu_F = c/2d$, что совпадает с результатом, полученным в подразд. 10.1.1 для резонатора с плоскими зеркалами. Очевидно, что для резонатора со сферическими зеркалами межмодовое расстояние не зависит от кривизны зеркал. Второй член в (10.67), зависящий от кривизны зеркал, просто представляет сдвиг всех резонансных частот.

Упражнение 10.5

Резонансные частоты конфокального резонатора

Симметричный конфокальный резонатор имеет длину $d = 30$ см и заполнен средой с показателем преломления $n = 1$. Определите межмодовое расстояние ν_F и частотный сдвиг $(\Delta\zeta/\pi)\nu_F$. Определите все резонансные частоты, лежащие в полосе 5 $10^{14} \pm 2 \cdot 10^9$ Гц.

10.2.4. Моды Гаусса—Эрмита

В разд. 3.3 было показано, что гауссов пучок — не единственное пучковое решение параксиального уравнения Гельмгольца. Семейство пучков Гаусса—Эрмита также является множеством решений этого уравнения. Хотя пу-

чок Гаусса—Эрмита порядка (l, m) имеет амплитудное распределение, отличное от такового для гауссова пучка, их волновые фронты одинаковы. В результате конструкция резонатора, согласующаяся с заданным гауссовым пучком, или, наоборот, пучок, подобранный к заданному резонатору, не отличаются от таковых в случае гауссова пучка и одинаковы для всех (l, m). Поэтому все члены семейства пучков Гаусса—Эрмита представляют собой моды резонатора со сферическими зеркалами.

Резонансные частоты моды (l, m), однако, зависят от индексов (l, m). Это происходит из-за зависимости фазового сдвига Гюи от l и m. Как следует из (3.66), фаза моды (l, m) на оси пучка равна

$$\varphi(0,\ z) = kz - (l + m + 1)\zeta(z). \tag{10.68}$$

Фазовый сдвиг, приобретаемый распространяющейся волной за полный обход резонатора длиной d, должен быть целым кратным $\pm 2\pi$, чтобы пучок воспроизводил себя. Таким образом,

$$2kd - 2(l + m + 1)\Delta\zeta = 2\pi q, \quad q = 0,\ \pm 1,\ \pm 2, \tag{10.69}$$

где, как и раньше, $\Delta\zeta = \zeta(z_2) - \zeta(z_1)$, а z_1 и z_2 — положения зеркал. Учитывая, что $k = 2\pi\nu/c$ и $\nu_F = c/2d$, получаем резонансные частоты

$$\boxed{\nu_{l,m,q} = q\nu_F + (l + m + 1)\frac{\Delta\zeta}{\pi}\nu_F.} \tag{10.70}$$

Резонансные частоты
Моды Гаусса—Эрмита

Моды с разными q, но одинаковыми (l, m), имеют одинаковое распределение интенсивности [см. (3.68)]. Они называются **продольными**, или **аксиальными**, модами. Индексы (l, m) нумеруют разные пространственные зависимости от поперечных координат x и y, а соответствующие моды называются **поперечными** (см. рис. 3.13).

Равенство (10.70) определяет следующие свойства резонансных частот мод Гаусса—Эрмита.

- Продольные моды, соответствующие данной поперечной моде, имеют частоты, расстояние между которыми $\nu_F = c/2d$, поскольку $\nu_{l,m,q+1} - \nu_{l,m,q} = \nu_F$. Этот результат совпадает с полученными для гауссовой моды (0, 0) и для резонатора с плоскими зеркалами.
- Все поперечные моды, для которых сумма индексов $l + m$ одинакова, имеют одинаковые резонансные частоты
- Две поперечные моды (l, m), (l', m'), соответствующие одной и той же продольной моде, имеют разность резонансных частот

$$\nu_{l,m,q} - \nu_{l',m',q} = \left[(l + m) - (l' + m')\right]\frac{\Delta\zeta}{\pi}\nu_F. \tag{10.71}$$

Это выражение определяет частотный сдвиг между наборами продольных мод с индексами (l, m) и (l', m').

Упражнение 10.6

Резонансные частоты симметричного конфокального резонатора

Покажите, что для симметричного конфокального резонатора продольные моды, связанные с различными поперечными модами, либо совпадают, либо смещены на $\nu_F/2$, как показано на рис. 10.18.

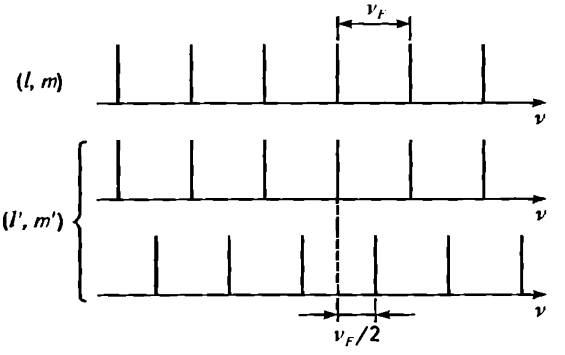


Рис. 10.18. В симметричном конфокальном резонаторе продольные моды, связанные двумя поперечными модами (l, m) и (l', m'), либо совпадают, либо смещены на половину продольного межмодового расстояния

10.2.5. Конечные апертуры и дифракционные потери

Поскольку пучки Гаусса и Гаусса—Эрмита имеют бесконечную поперечную протяженность, а зеркала резонатора — конечную, часть оптической энергии вытекает мимо зеркал из резонатора при каждом проходе. Оценку потерь энергии можно получить путем расчета мощности той части пучка, которая не перекрывается зеркалами. Если, например, пучок гауссов и имеет ширину W, а зеркала круглые и имеют радиус $a = 2W$, то небольшая часть $\exp(-2a^2/W^2) \approx 3{,}35\ \ 10^{-4}$ мощности пучка вытекает при каждом проходе резонатора [см. (3.17)], а остальная часть отражается (или проходит через зеркало). Поперечные моды более высоких порядков теряют большую мощность, так как имеют большую пространственную протяженность в поперечной плоскости.

Если радиусы зеркал меньше $2W$, то потери больше. В этом случае пучки Гаусса и Гаусса—Эрмита уже не являются хорошим приближением для описания мод резонатора. Задача определения мод резонатора со сферическими зеркалами конечных размеров является сложной. Волна является модой, если она воспроизводит свою амплитуду (с точностью до постоянного множителя) и фазу (с точностью до целого числа 2π) после полного обхода резонатора. Один из часто используемых методов расчета мод состоит в численном моделировании амплитуды и фазы бегущей волны, испытывающей последовательность отражений от зеркал резонатора, аналогично тому, как определялись положение и отклонение луча. После множества обходов этот процесс сходится к одной из мод.

Пусть $U_1(x, y)$ — комплексная амплитуда волны у поверхности зеркала 1 (рис. 10.19), а $U_2(x, y)$ — комплексная амплитуда после полного обхода резонатора. Функция $U_1(x, y)$ является модой, если $U_2(x, y) = \mu U_1(x, y)$ и при условии,

что $\arg\{\mu\}$ — целое кратное 2π (т. е. μ — действительное и положительное). После одного обхода интенсивность волны уменьшается на множитель μ^2, а фаза воспроизводится с точностью до целого числа 2π. Для определения $U_2(x, y)$ из $U_1(x, y)$ можно воспользоваться методами Фурье-оптики (см. гл. 4). Эти величины можно рассматривать как входной и выходной сигналы линейной системы (см. приложение Б), характеризуемой функцией отклика на импульсное воздействие $h(x, y; x', y')$, тогда

$$U_2(x, y) = \int_{-\infty}^{\infty}\int h(x, y; x', y')U_1(x', y')\mathrm{d}x'\mathrm{d}y' \tag{10.72}$$

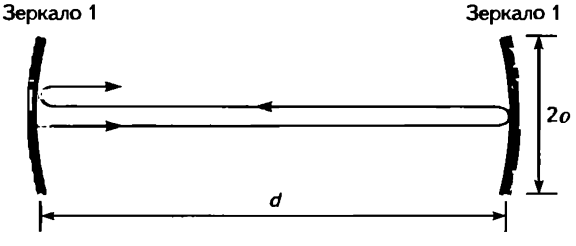


Рис. 10.19. Распространение волны в резонаторе со сферическими зеркалами. Комплексная амплитуда $U_1(x, y)$ соответствует моде, если она воспроизводится после полного обхода, т. е. если $U_2(x, y) = \mu U_1(x, y)$ и $\arg\{\mu\} = q2\pi$

Если функция отклика на импульсное воздействие h известна, то моды можно определить путем решения задачи на собственные значения для интегрального уравнения (см. приложение В):

$$\int_{-\infty}^{\infty}\int h(x, y; x', y')U(x', y')\mathrm{d}x'\mathrm{d}y' = \mu U(x, y). \tag{10.73}$$

Решение этой задачи дает собственные функции $U_{l,m}(x, y)$ и собственные значения $\mu_{l,m}$, нумеруемые индексами (l, m). Собственные функции — это моды, а собственные значения — постоянные множители, приобретаемые за один обход. Квадрат модуля $|\mu_{l,m}|^2$ определяет коэффициент ослабления интенсивности за один обход для моды (l, m). Очевидно, что если размер зеркал стремится к бесконечности и выполняются условия параксиального приближения, то моды сводятся к семейству пучков Гаусса—Эрмита, которые обсуждались выше.

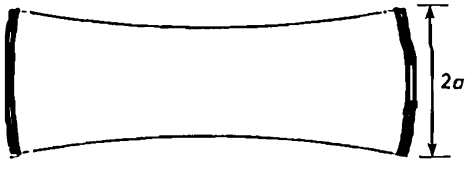
Остается определить $h(x, y; x', y')$ и решить интегральное уравнение (10.73). Один проход резонатора включает распространение на расстояние d, обрезание апертурой зеркала и отражение от него. Второй проход, завершающий полный обход резонатора, аналогичен. Функцию отклика на импульсное воздействие $h(x, y; x', y')$ можно найти по теории дифракции Френеля (см. подразд. 4.3.2). В общем случае, однако, моды и связанные с ними потери удается определить лишь посредством численного решения интегрального уравнения (10.73). Итерационное численное решение начинается с начального приближения U_1, по которому вычисляется U_2 и так далее до достижения сходимости процесса.

Этот метод был использован для определения потерь, связанных с различными модами резонатора со сферическими зеркалами, имеющими круговую апертуру радиусом a. Результаты представлены на рис. 10.20 для симметрично-

го конфокального резонатора. Потери определяются единственным параметром — числом Френеля $N_F = a^2/\lambda d$. Это происходит потому, что число Френеля определяет дифракцию Френеля между двумя зеркалами, как обсуждалось в подразд. 4.3.2. Для симметричного конфокального резонатора, описываемого формулами (10.60) и (10.61), ширина пучка на зеркалах:

$$W = \sqrt{\frac{\lambda d}{\pi}},$$

так что $\lambda d = \pi W^2$, откуда для числа Френеля легко получаем выражение $N_F = a^2/\pi W^2$. Таким образом, N_F пропорционально отношению a^2/W^2; большее число Френеля соответствует меньшим потерям. Из рис. 10.20 находим, что потери на один проход резонатора для моды самого низшего порядка, так же как симметричного конфокального резонатора, составляет примерно 0,1 % при $a/W = 0{,}94$. Это число Френеля соответствует $N_F = 1{,}72$. Если бы пучок был гауссовым с шириной W, то процент мощности вне круга радиусом $a = 1{,}72W$, был бы $\exp(-2a^2/W^2) \approx 0{,}27$ %. Это больше потерь 0,1 % за один проход для действительной моды резонатора. Моды высших порядков претерпевают большие потери в связи с их большей пространственной протяженностью.



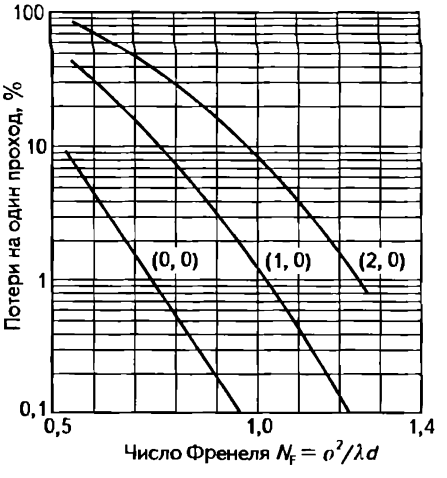


Рис. 10.20. Процент дифракционных потерь за один проход (половину полного обхода) как функция числа Френеля $N_F = a^2/\lambda d$ для мод (0, 0), (1, 0) и (2, 0) в симметричном конфокальном резонаторе. (Заимствовано из *Siegman A.E.* Lasers. University Science, 1986, рис. 19.19, слева)

10.3. ДВУ- И ТРЕХМЕРНЫЕ РЕЗОНАТОРЫ

10.3.1. Двумерные прямоугольные резонаторы

Двумерный (2D) резонатор с плоскими зеркалами образован двумя ортогональными парами параллельных зеркал, т. е. парой, перпендикулярной оси z, и другой парой, перпендикулярной оси x. Свет удерживается в плоскости z—y посредством серии отражений, как показано на рис. 10.21, *а*.

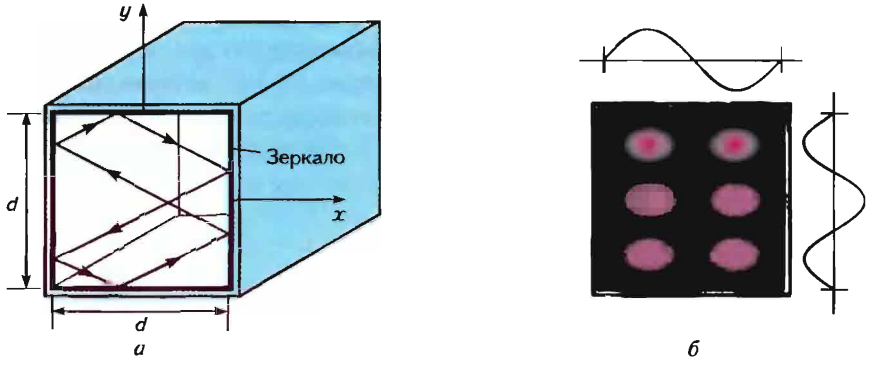


Рис. 10.21. Двумерный резонатор с плоскими зеркалами:
а — ход луча; *б* — структура стоячей волны с модовыми индексами $q_y = 3$ и $q_z = 2$

Граничные условия определяют моды двумерного резонатора в значительной степени так же, как для резонатора Фабри—Перо. Если расстояние между зеркалами равно d, то для стоячих волн компоненты волнового вектора $\boldsymbol{k} = (k_y, k_z)$ принимают дискретные значения

$$k_y = q_y \frac{\pi}{d}, \quad q_y = 1,\ 2, \qquad k_z = q_z \frac{\pi}{d}, \quad q_z = 1,\ 2, \tag{10.74}$$

где q_y и q_z — модовые индексы для направлений y и z соответственно. Эти условия являются обобщением условия (10.2). Каждой паре целых чисел (q_y, q_z) соответствует мода резонатора

$$U(\boldsymbol{r}) \propto \sin\left(\frac{q_y \pi y}{d}\right) \sin\left(\frac{q_z \pi z}{d}\right),$$

как показано на рис. 10.21, *б*. Наименьший порядок моды (1, 1), поскольку моды $(q_y, 0)$ и $(0, q_z)$ имеют нулевую амплитуду, т. е. $U(\boldsymbol{r}) = 0$. Удобно представлять моды точками на плоскости (k_y, k_z), которые образуют периодическую решетку с периодом π/d (рис. 10.22).

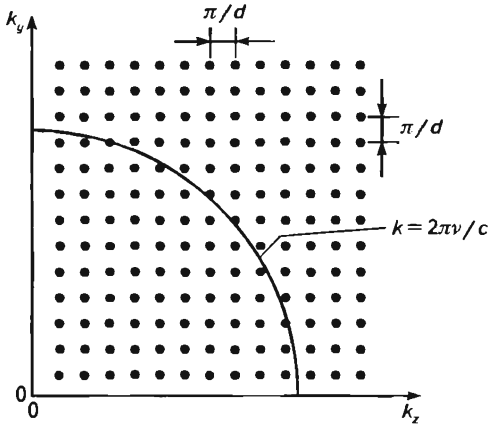


Рис. 10.22. Точки показывают положения концов волновых векторов $\boldsymbol{k} = (k_y, k_z)$ для мод двумерного резонатора

Волновое число моды k есть расстояние от соответствующей точки до начала координат. Связанная с модой частота равна $\nu = ck/2\pi$. Таким образом, частоты мод резонатора определяются из уравнения

$$k^2 = k_y^2 + k_z^2 = \left(\frac{2\pi\nu}{c}\right)^2 \tag{10.75}$$

откуда

$$\boxed{\nu_{\boldsymbol{q}} = \nu_F\sqrt{q_y^2 + q_z^2}, \quad q_y,\ q_z = 1,\ 2,\ \ldots;} \quad \nu_F\frac{c}{2d}, \tag{10.76}$$

Резонансные частоты

где $\boldsymbol{q} = (q_y, q_z)$.

Для определения числа мод в заданной полосе частот $\nu_1 < \nu < \nu_2$ нужно провести две окружности с радиусами $k_1 = 2\pi\nu_1/c$ и $k_2 = 2\pi\nu_2/c$ на k-диаграмме (см. рис. 10.22) и подсчитать число точек внутри получившегося кольца. Эта процедура преобразует разрешенные значения вектора $\boldsymbol{k}$ в разрешенные значения частоты ν.

Упражнение 10.7

Плотность мод двумерного резонатора

а. Получите приближенное выражение для числа мод двумерного резонатора с частотами, лежащими в интервале от 0 до ν, предполагая, что $2\pi\nu/c \gg \pi/d$, т. е. $d \gg 1/2$, с учетом двух ортогональных поляризаций для каждой моды.

б. Покажите, что число мод на единицу площади внутри частотного интервала от ν до $\nu + d\nu$ равно $M(\nu)d\nu$, где плотность числа мод $M(\nu)$ (число мод на единицу площади и на единицу частоты) на частоте ν равна

$$\boxed{M(\nu) = \frac{4\pi\nu}{c^2}.} \tag{10.77}$$

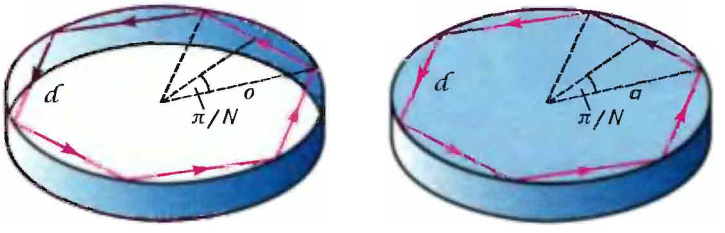
Плотность числа мод (2D-резонатор)

Моды, рассмотренные в данном разделе до сих пор, распространяются в плоскости y—z 2D-резонатора. Неплоские моды имеют постоянную распространения с ненулевой компонентой в направлении x. Это направляемые моды, бегущие вдоль оси двумерного волновода, как описано в разд. 8.3. В то время как компоненты k_y и k_z волнового вектора принимают дискретные значения, диктуемые граничными условиями, компонента k_x принимает непрерывные значения, поскольку 2D-резонатор открыт в направлении x.

10.3.2. Круговые резонаторы и моды шепчущей галереи

Свет можно удерживать в двумерном круговом резонаторе за счет повторяющихся отражений от его круговой границы. Как показано на рис. 10.23, луч самовоспроизводится после N отражений и проходит путь длиной Nd,

где $d = 2a \sin(\pi/N)$; a — радиус круга. Для моды бегущей волны резонансные частоты определяются приравниванием длины пути кругового обхода к целому числу длин волн, как в (10.7). Без учета фазового сдвига при каждом отражении это приводит к $Nd = q\lambda = qc/\nu$, т. е. к резонансным частотам $\nu_q = qc/Nd$, где $q = 1, 2,$ Межмодовое расстояние получается равным $\nu_F = c/Nd$.



Зеркальный резонатор Диэлектрический резонатор

Рис. 10.23. Отражения в круговом резонаторе

Для $N = 2$ имеем

$$\nu_F = \frac{c}{2d} = \frac{c}{4a},$$

что совпадает с (10.5). Аналогично, $N = 3$ дает

$$\nu_F = \frac{c}{3d} = \frac{c}{3\sqrt{3}a},$$

что совпадает с результатом для трехзеркального резонатора (см. упражнение 10.1). В пределе $N \to \infty$ длина пути приближается к длине окружности цилиндра $2\pi a$, а соответствующее межмодовое расстояние становится равным

$$\boxed{\nu_F = \frac{c}{2\pi a}.} \qquad (10.78)$$

Расстояние между резонансными частотами

При этом луч стелется по внутренней поверхности цилиндра, отражаясь от нее при почти скользящем падении, как показано на рис. 10.23. Такие оптические моды называются модами шепчущей галереи (МШГ). Эти оптические моды похожи на акустические моды известных в архитектуре «шепчущих галерей» — помещений, названных так из-за легкости, с которой тихий шепот передается вдоль их вогнутой внутренней стены, но при этом почти не слышен в середине помещения.

Двумерные резонаторы с другими формами сечения также находят применение. Например, круговое сечение может быть сплющено в структуру, напоминающую дорожку стадиона. В такой продолговатой конфигурации поддер-

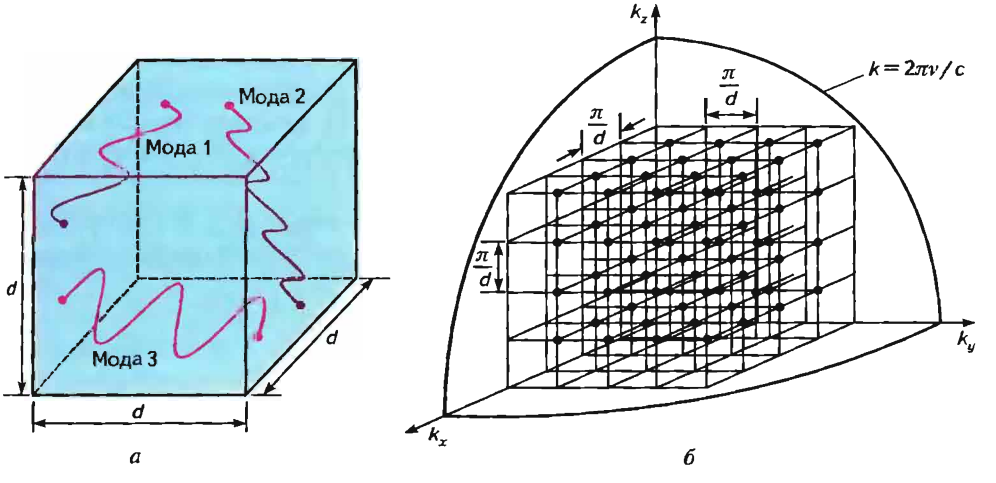
живаются моды типа галстука-бабочки (см. рис. 10.6, *б*), у которых при обходе луч претерпевает четыре локальных отражения в тех местах периметра резонатора, где кривизна такая же, как в обычном конфокальном резонаторе со сферическим зеркалами (см. подразд. 10.2.1).

10.3.3. Трехмерные резонаторы в виде прямоугольной полости

Трехмерный (3D) резонатор может быть построен из трех пар параллельных зеркал, образующих закрытый прямоугольный ящик с размерами d_x, d_y и d_z. Структура прямоугольного резонатора показана на рис. 10.24, *а*. Для существования решения в виде стоячих волн в резонаторе требуется, чтобы компоненты волнового вектора имели дискретные значения

$$k_x = q_x \frac{\pi}{d_x}; \quad k_y = q_y \frac{\pi}{d_y}; \quad k_z = q_z \frac{\pi}{d_z}, \quad q_x, q_y, q_z = 1, 2, \qquad (10.79)$$

где q_x, q_y и q_z — положительные целые числа, нумерующие моду. Каждая мода q, характеризуемая тремя числами (q_x, q_y, q_z), представляется точкой в пространстве с координатами (k_x, k_y, k_z). Расстояния между точками вдоль осей обратно пропорциональны соответственным размерам резонатора. На рис. 10.24, *б* показан пример k-пространства для кубического резонатора с $d_x = d_y = d_z = d$.



а *б*

Рис. 10.24. Волны в трехмерном кубическом резонаторе ($d_x = d_y = d_z = d$) (*а*). Концы волновых векторов (k_x, k_y, k_z) для мод трехмерного резонатора показаны точками (*б*). Волновое число моды k равно расстоянию от начала координат до соответствующей точки. На каждую точку в k-пространстве приходится объем $(\pi/d)^3$. Все моды с частотой меньше ν лежат в положительном октанте внутри сферы радиусом $k = 2\pi\nu/c$

Длины волновых векторов и соответствующие резонансные частоты удовлетворяют соотношению

$$k^2 = k_x^2 + k_y^2 + k_z^2 = \left(\frac{2\pi\nu}{c}\right)^2 \tag{10.80}$$

Поверхность постоянной частоты ν представляет собой сферу радиусом $k = 2\pi\nu/c$. Резонансные частоты определяются из (10.79) и (10.80):

$$\boxed{\nu_q = \sqrt{q_x^2\nu_{F_x}^2 + q_y^2\nu_{F_y}^2 + q_z^2\nu_{F_z}^2}, \quad q_x,\ q_y,\ q_z = 1,\ 2,\ \ldots,} \tag{10.81}$$

Резонансные частоты

где ν_{F_x}, ν_{F_y}, ν_{F_z} — частотные расстояния между соседними модами, обратно пропорциональные размерам резонатора вдоль осей x, y и z соответственно,

$$\nu_{F_x} = \frac{c}{2d_x}; \quad \nu_{F_y} = \frac{c}{2d_y}; \quad \nu_{F_z} = \frac{c}{2d_z}. \tag{10.82}$$

Для резонаторов, размеры которых много больше длины волны, межмодовое расстояние много меньше оптической частоты. Например, при $d = 1$ см и $n = 1$ $\nu_F = 15$ ГГц. Однако для микрорезонаторов это не так, что будет показано в разд. 10.4.

Плотность мод

Если все размеры резонатора много больше длины волны, то межмодовое расстояние $\nu_F = c/2d$ мало, и аналитически трудно пересчитывать дискретные моды. В этом случае удобно обратиться к непрерывному приближению и ввести понятие плотности мод, применимость которого зависит от относительной величины частотной полосы, представляющей интерес, и частотного интервала между соседними модами.

Число мод, лежащих в частотном интервале между 0 и ν, соответствует числу точек в объеме положительного октанта шара радиуса k на k-диаграмме (см. рис. 10.24, *б*). Это число равно

$$2\frac{1}{8}\frac{(3/4)\pi k^3}{(\pi/d)^3} = \frac{k^3}{3\pi^2}d^3.$$

Первый множитель 2 учитывает наличие двух ортогональных поляризаций для каждой моды, а знаменатель $(\pi/d)^3$ определяет объем k-пространства, приходящийся на одну моду. Поскольку $k = 2\pi\nu/c$, число мод, лежащих между 0 и ν, равно

$$\frac{(2\pi\nu/c)^3}{3\pi^2}d^3 = \frac{8\pi\nu^3}{3c^3}d^3.$$

Число мод, лежащих в пределах приращения частоты от ν до $\nu + \Delta\nu$, тогда оказывается равным

$$\frac{\mathrm{d}}{\mathrm{d}\nu}\frac{8\pi\nu^3}{3c^3}d^3\Delta\nu = \frac{8\pi\nu^2}{c^3}d^3\Delta\nu.$$

Плотность мод $M(\nu)$, т. е. число мод на единицу объема резонатора и на единицу частоты в полосе вблизи частоты ν, получается равной

$$\boxed{M(\nu) = \frac{8\pi\nu^2}{c^3}.} \qquad (10.83)$$

Плотность мод (3D-резонатор)

Эта формула была впервые получена Рэлеем и Джинсом в связи со спектром излучения черного тела (см. подразд. 13.4.2). Величина $M(\nu)$ — квадратичная функция частоты, так что число мод внутри фиксированной полосы частот $\Delta\nu$ растет с ростом частоты, как показано на рис. 10.25. При $\nu = 3\ \ 10^{14}$ Гц ($\lambda_0 = 1$ мкм) $M(\nu) = 0{,}08$ мод/(см^3 · Гц). В полосу частот 1 ГГц тогда попадает $\approx 8 \cdot 10^7$ мод/см^3. Число мод на единицу объема в некотором интервале частот $\nu_1 < \nu < \nu_2$ — просто интеграл $\int_{\nu_1}^{\nu_2} M(\nu)\mathrm{d}\nu$.

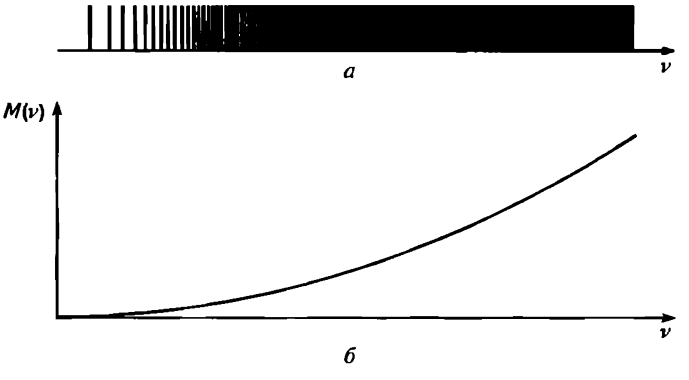


Рис. 10.25. Частотное расстояние между соседними модами уменьшается с ростом частоты (*а*). Плотность мод $M(\nu)$ для трехмерного оптического резонатора есть квадратично возрастающая функция частоты (*б*)

Плотность мод в двух и трех измерениях была получена на основе геометрии квадратного и кубического резонатора. Однако полученные результаты применимы к любой геометрии при условии, что размеры резонатора много больше длины волны.

Здесь, по-видимому, уместно заметить, что выполненный подсчет электромагнитных мод, математически идентичен вычислению разрешенных квантовых состояний электронов, находящихся между идеально отражающими стенками. Последняя важная модель определяет плотность возможных состояний электронов как функцию их энергии в полупроводниковых материалах (см. подразд. 16.1.3).

10.4. МИКРОРЕЗОНАТОРЫ

Микрорезонаторы — это резонаторы, размер которых в одном или нескольких направлениях составляет несколько длин волны света или менее. Термин **резонатор в виде микрополости** или, для краткости, просто **микрополость** чаще относится к микрорезонаторам, размеры которых малы во всех пространственных направлениях, так что частотные межмодовые расстояния велики по всем направлениям в k-пространстве и резонансные частоты расположены редко. Однако часто эти термины используются взаимозаменяемо.

Отсутствие резонансных мод в протяженных спектральных интервалах может подавлять излучение света источниками, расположенными внутри микрополости. В то же время, испускание света в избранные моды высокодобротного микрорезонатора малого объема может быть усилено по сравнению с обычными оптическими модами, как описано в разд. 13.3.5. Такие эффекты могут быть важны для работы лазеров с микрорезонаторами (см. подразд. 17.4.2).

Микрорезонаторы можно изготовить из диэлектрических материалов в различной геометрической конфигурации, например:

1) микростолбики с брэгговскими отражающими решетками;

2) микродиски и микросферы, в которых отраженный свет вблизи внутренней поверхности имеет характер мод шепчущей галереи;

3) микроторы, напоминающие маленькие волоконные кольцевые резонаторы;

4) двумерные фотонные кристаллы с удерживающими свет дефектами, работающие как микрорезонаторы.

Все соответствующие технологии направлены на выполнение двух главных проектных требований:

- уменьшение модового объема V, который определяется как объемный интеграл от плотности оптической энергии моды $\varepsilon \mathcal{E}^2/2$, нормированный на ее максимальное значение;
- увеличение добротности Q.

Типичные значения модового объема и добротности упомянутых структур сведены в табл. 10.1.

Таблица 10.1. **Нормированные модовые объемы V/λ^3 и добротность Q различных микрорезонаторов**

	Микростолбик	Микродиск	Микротор	Микросфера	Фотонный кристалл
V/λ^3	5	5	10^3	10^3	1
Q	10^3	10^4	10^8	10^{10}	10^4

Точный анализ мод диэлектрических микрорезонаторов требует полной электромагнитной теории. Уравнение Гельмгольца решается в системе координат, учитывающей геометрию резонатора; на компоненты электрического и магнитного полей на плоских, цилиндрических или сферических границах накладываются граничные условия. Решение дает резонансные частоты мод и их простран-

ственные распределения, которые можно использовать для определения модового объема для каждой моды. Поскольку этот анализ для всех практически применяемых геометрий сложен, часто необходимы численные решения.

В следующем разделе мы опишем некоторые свойства простого прямоугольного микрорезонатора (ящика), стенки которого сделаны из идеально отражающих зеркал. Простой анализ мод такой структуры дает резонансные частоты и пространственные распределения поля. В высокодобротных микрорезонаторах не применяются зеркала, поскольку реально они имеют довольно большие потери, а конфигурация прямоугольного ящика не относится к числу типичных на практике. Тем не менее этот анализ полезен для выявления связи между резонансными частотами и размерами резонатора, а также для иллюстрации частотной зависимости плотности мод для ящиков с различным соотношением размеров.

10.4.1. Прямоугольные микрорезонаторы

Простейшая структура микрорезонатора — прямоугольный ящик, изготовленный из плоских параллельных зеркал. Моды в этом случае представляют собой синусоидальные стоячие волны во всех трех измерениях, а резонансные частоты даются формулой (10.81). Если размеры ящика малы, то лишь моды низших порядков лежат в оптическом диапазоне. Для кубического резонатора резонансные частоты приведены в табл. 10.2 в единицах $\nu_F = c/2d$. Например, если $d = 1$ мкм, а среда имеет показатель преломления $n = 1{,}5$, мы получаем $\nu_F = 100$ ТГц. Частоты мод низших порядков тогда соответствуют длинам волн $\lambda_0 = 2{,}13$; 1,73; 1,34; 1,22; 1,06; 1,00 и 0,87 мкм, далеко отстоящим друг от друга.

Таблица 10.2. **Резонансные частоты мод низших порядков в кубическом микрорезонаторе**

Мода (q_x, q_y, q_z)*	$(011)^{(3)}$	$(111)^{(1)}$	$(012)^{(6)}$	$(112)^{(3)}$	$(022)^{(3)}$	$(122)^{(3)}$	$(222)^{(1)}$
Частота (в единицах ν_F)	1,41	1,73	2,24	2,45	2,83	3	3,46

* Верхние индексы в скобках обозначают кратность вырождения, т. е. число мод с данной резонансной частотой. Например, три моды имеют одну и ту же резонансную частоту $1{,}41\nu_F$: (011), (101) и (110).

Если геометрия резонатора характеризуется сочетанием малых и больших размеров, как у вытянутого ящика, то моды располагаются в узлах анизотропной решетки в k-пространстве (см. рис. 10.24, *б*). Решетка частая в направлениях больших размеров и редкая в направлении малых. Для подсчета числа мод тогда можно применять непрерывную аппроксимацию только в тех направлениях, в которых решетка частая. Получающаяся плотность мод показана на рис. 10.26 для различных случаев.

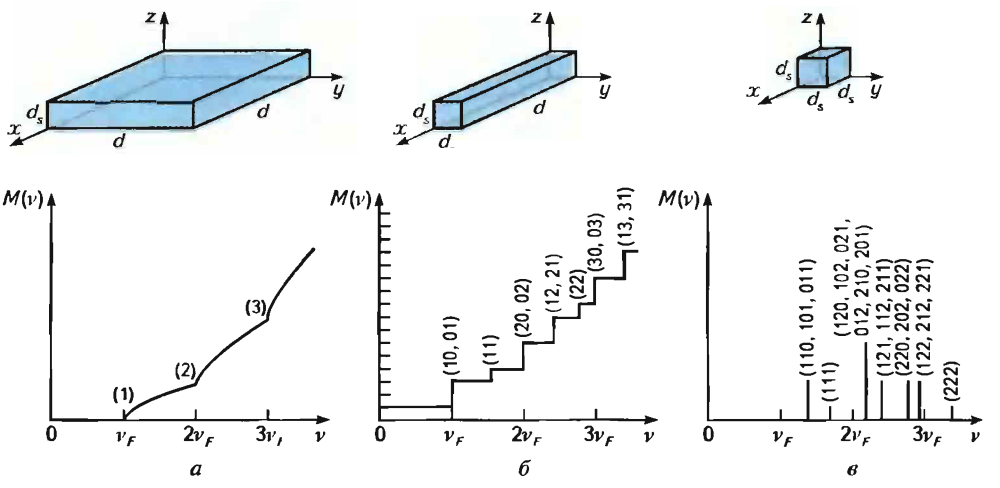


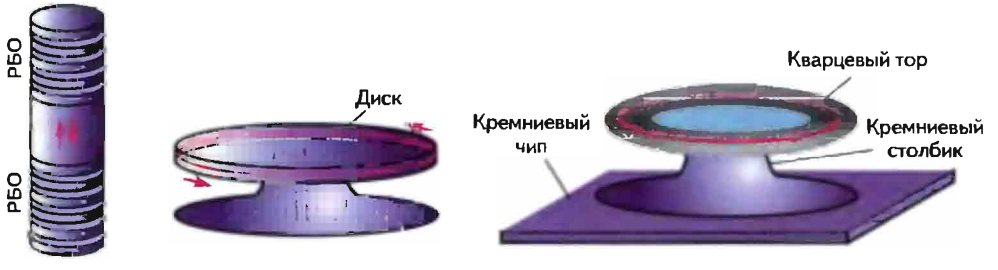
Рис. 10.26. Плотность мод $M(\nu)$ для прямоугольных микрорезонаторов с одной (*а*), двумя (*б*) и тремя (*в*) сторонами малого размера $d_s \ll d$. Расстояние между частотами мод, связанных с малым размером, равно $\nu_F = c/2d_s$. Когда все размеры малы, как в случае (*в*), резонансные частоты дискретны, а их частоты приведены в табл. 10.2 для кубического микрорезонатора. Результат, приведенный на рисунке (*б*), представляет собой комбинацию дискретных мод двумерного микрорезонатора и квазинепрерывных мод одномерного резонатора большой длины, имеющего однородное распределение плотности мод [см. (10.10)]. Результат, показанный на рисунке (*а*), — комбинация дискретных мод одномерного микрорезонатора и квазинепрерывных мод большого двумерного резонатора, у которого плотность мод пропорциональна частоте [см. (10.77)]

10.4.2. Резонаторы в виде микростолбиков, микродисков и микроторов

Диэлектрические резонаторы изготавливаются в множестве конфигураций, к которым относятся микростолбики, микродиски и микроторы, показанные на рис. 10.27. Свет удерживается в этих структурах благодаря полному внутреннему отражению (см. рис. 10.23).

Резонатор в форме микростолбика или микроштыря — цилиндр из материала с высоким показателем преломления, заключенный между диэлектрическими слоями, образующими распределенную отражательную брэгговскую решетку (ОБР), как показано на рис. 10.27, *а*. Свет в аксиальном направлении удерживается за счет отражения от ОБР, как в резонаторе Фабри—Перо. В поперечном направлении свет удерживается за счет полного внутреннего отражения от стенок цилиндра. Микростолбики, как правило, изготавливают из сложных полупроводников путем обычной литографии и травления; слои ОБР часто делают из AlAs/GaAs или AlGaAs/GaAs. Сам столбик может содержать активную область, такую как структура из множества квантовых ям, обеспечивающую усиление света при накачке (см. разд. 17.4).

Резонатор в форме микродиска, показанный на рис. 10.27, *б*, — круговой резонатор, в котором свет распространяется в виде мод шепчущей галереи, полностью отражаясь при почти скользящем падении на внутреннюю поверхность цилиндрической границы диэлектрика (см. подразд. 10.3.2). Размеры микростолбиков и микродисков обычно колеблются от нескольких микромстров до нескольких десятков микрометров, а их добротность Q существенно выше, чем у зеркальных резонаторов, поскольку потери намного ниже (см. табл. 10.1). Тем не менее их функциональные возможности ограничены качеством поверхности материала, поскольку свет распространяется вблизи границы.



а *б* *в*

Рис. 10.27. Резонаторы в виде микростолбиков (*а*), микродисков (*б*) и микроторов (*в*)

Тороидальный диэлектрический микрорезонатор, показанный на рис. 10.27, *в*, сильно напоминает кольцевой волоконный резонатор, в котором моды представляют собой направляемые бегущие волны. Такие микрорезонаторы обычно изготавливаются из кварца и поддерживаются на кремниевом чипе кремниевым столбиком. Тор формируется силами поверхностного натяжения, когда материал находится в жидком состоянии; таким образом, внешняя граница приобретает почти атомный класс чистоты поверхности. В результате потери на рассеяние удается существенно понизить по сравнению с микродисковым резонатором. Кварцевые тороидальные микрорезонаторы на чипах имеют исключительно высокие значения добротности $Q > 10^8$ (см. табл. 10.2).

10.4.3. Микросферические резонаторы

Диэлектрические сферы используются в качестве трехмерных оптических микрорезонаторов. Некоторые волны направляются вдоль траекторий (орбит) локализованных вблизи большого круга сферы, что дает моды шепчущей галереи.

Моды диэлектрической сферы можно определить путем решения уравнения Гельмгольца (5.57) для векторов электрического и магнитного полей при соответствующих граничных условиях. Эти моды похожи на волновые функции электрона в атоме водорода (см. подразд. 13.1.1) по причине сферической симметрии обеих задач, однако между ними имеются различия ввиду векторной природы электромагнитного поля.

Электрическое и магнитное векторные поля непосредственно связаны со скалярной потенциальной функцией U, удовлетворяющей уравнению Гельмгольца[1]. Для сферы радиусом a с показателем преломления n, находящейся в воздухе, метод разделения переменных в сферической системе координат (r, θ, ϕ) дает решение в виде

$$U(r, \theta, \phi) \propto \sqrt{r} J_{\ell+1/2}(nk_0 r) P_{\mathrm{m}}^{\ell}(\cos\theta) \exp(\pm j\mathrm{m}\phi), \quad r \le a; \tag{10.84}$$

$$U(r, \theta, \phi) \propto \sqrt{r} \mathfrak{h}_{\ell+1/2}^{(1)}(nk_0 r) P_{\mathrm{m}}^{\ell}(\cos\theta) \exp(\pm j\mathrm{m}\phi), \quad r > a, \tag{10.85}$$

где $J_\ell(\ )$ — функция Бесселя первого рода порядка ℓ; $\mathfrak{h}_\ell^{(1)}(\cdot)$ — функция Ханкеля первого рода порядка ℓ; P_m^ℓ — присоединенный полином Лежандра; m и ℓ — неотрицательные целые числа. Граничные условия при $r = a$ дают характеристическое уравнение, определяющее дискретный набор значений k_0, соответствующих резонансным частотам. Они нумеруются третьим целым числом ν. Кроме того, имеется две поляризационные моды — E-мода, у которой $H_r = 0$, и H-мода, у которой $E_r = 0$.

Моды обычно представляют собой колеблющиеся функции от r, θ и ϕ, характеризуемые радиальным, полярным и азимутальным модовыми числами n, ℓ и m соответственно. В радиальном направлении внутри сферы имеется n максимумов. Число максимумов поля в азимутальном направлении равно 2ℓ, а в полярном (между двумя полюсами) — $\ell - \mathrm{m} + 1$.

Основная мода (n = 1, m = l) имеет единственный пик в радиальном направлении внутри сферы и один пик в полярном направлении при $\theta = \pi/2$. Для больших m = l моды сильно локализованы вблизи экватора. Это происходит потому, что $P_\ell^\ell(\cos\theta) \approx \sin^\ell\theta$, быстро убывает при отклонении угла θ от $\pi/2$, а функция $J_\ell(nk_0 r)$ мала внутри сферы повсюду, кроме малой окрестности точки $r = a$. Таким образом, мода представляет собой пучок света, распространяющийся вдоль экватора, как показано на рис. 10.28, *а*, что в значительной степени напоминает моды шепчущей галереи в дисковом резонаторе, изображенном на рис 10.23. Для достаточно больших ℓ = m резонансные частоты этих мод приближенно равны $\nu_\ell \approx \ell c/2\pi a$. Этого можно было ожидать, поскольку угловое модовое число ℓ близко к числу длин волн, укладывающихся на длине экватора.

Моды шепчущей галереи можно рассматривать с точки зрения лучевой оптики с помощью квазиплоских волн, волновые векторы которых параллельны лучам (см. разд. 2.3 и рис. 9.18), зигзагообразно распространяющимся вблизи экватора, как показано на рис. 10.28, *б*. Длина волнового вектора равна

$$k = \frac{\sqrt{\ell(\ell+1)}}{a},$$

[1] Подробное математическое описание можно найти, например, в статье *Ораевского А.Н.* Волны шепчущей галереи. Квантовая электроника. Т. 32. 2002. С. 377—400.

а его азимутальная компонента $k_\phi = m/a$. Угол отклонения зигзагообразного луча принимает минимальное значение $\approx 1/\sqrt{\ell}$ для основной моды $m = \ell$, а для $m = 0$ отклонение составляет 90°

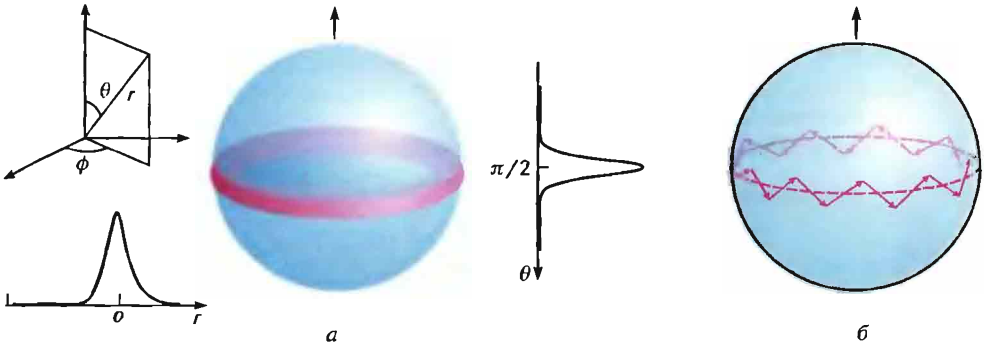


Рис. 10.28. Мода шепчущей галереи в резонаторе-микросфере (*а*); лучевая модель моды шепчущей галереи (*б*)

Микросферы, изготовленные из плавленого кварца с низкими потерями, использовались как оптические резонаторы с ультравысокими значениями Q. Как и у тороидального резонатора, изображенного на рис. 10.27, *в*, форма и качество поверхности сферы определяются поверхностным натяжением в расплавленном состоянии в процессе изготовления; в результате получается почти атомный класс чистоты поверхности. Уменьшение потерь на поверхностное рассеяние приводит к замечательно высокой добротности $Q > 10^{10}$ (см. табл. 10.2). Оптическую энергию можно ввести в микросферу, приложив ее к оптическому волокну с удаленной в месте касания оболочкой, как показано на рис. 10.29.

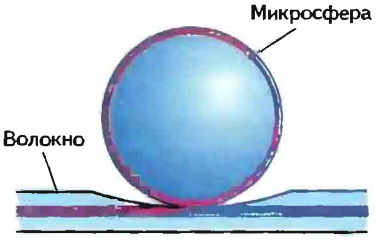


Рис. 10.29. Ввод оптической энергии из оптического волокна в сферический микрорезонатор

10.4.4. Фотонно-кристаллические микрорезонаторы

Как описано в гл. 7, фотонные кристаллы представляют собой диэлектрические периодические структуры, имеющие фотонные запрещенные зоны, т. е. спектральные полосы, внутри которых свет не может распространяться. Отражательная брэгговская решетка (ОБР) — пример одномерного фотонного кристалла, который служит отражателем на частотах, лежащих внутри запрещенной зоны. Например, в резонаторе типа микростолбика, показанном на рис. 10.27, *а*, ОБР используются в качестве зеркал. Если высота такого мик-

рорезонатора равна одному или нескольким периодам ОБР, как показано на рис. 10.30, *а*, то структура может рассматриваться как протяженный фотонный кристалл, в котором резонатор является дефектом кристаллической структуры. Такой резонатор называется **фотонно-кристаллическим резонатором.**

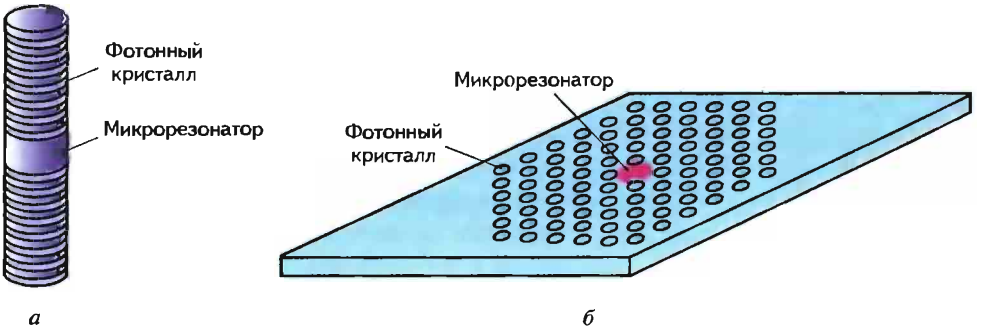


Рис. 10.30. Фотонно-кристаллические микрорезонаторы:

а — микростолбик как одномерный фотонный кристалл, в котором сам микрорезонатор является дефектом; *б* — двумерный фотонно-кристаллический резонатор можно изготовить путем сверления отверстий в диэлектрической пластине в узлах плоской гексагональной решетки; пропущенное отверстие служит микрорезонатором

Эта концепция применима и к двумерным фотонным кристаллам. Как схематически показано на рис. 10.30, *б*, **дефект** в двумерной периодической структуре кристалла — это локальное отклонение от периодичности, такое как пропущенное отверстие в периодическом наборе отверстий, просверленных в пластине и заполненных воздухом. Для длин волн, попадающих в фотонную запрещенную зону, периодическая структура, окружающая дефект, не поддерживает распространение света, так что свет захватывается и удерживается дефектом аналогично тому, как электроны или дырки захватываются дефектами решетки в полупроводниковых кристаллах. В этом случае дефект служит микрорезонатором. По-другому можно сказать, что дефект создает новые резонансные частоты, лежащие в фотонной запрещенной зоне и соответствующие модам, пространственное распределение которых сосредоточено вблизи дефекта и быстро спадает к нулю при удалении от него в глубь окружающего фотонного кристалла.

Двумерные фотонные кристаллы изготавливают с применением электронной пучковой литографии и реактивного ионного травления в полупроводниковых материалах. Микрорезонаторы с размерами, близкими к периоду фотонного кристалла, который может иметь порядок длины волны света, позволяют уменьшить модовый объем до λ^3. По сравнению с другими технологиями фотонно-кристаллические микрорезонаторы имеют наименьший модовый объем (см. табл. 10.1). Добротность Q также достаточно велика и достигает 10^4.

Рекомендуемая литература

КНИГИ

См. также книги по лазерам к гл. 15.

Hodgson N., Weber H. Laser Resonators and Beam Propagation: Fundamentals, Advanced Concepts and Applications. Springer-Verlag. 2nd ed., 2005.

Vahala K.J., ed. Optical Microcavities. World Scientific, 2004.

Staliūnas K., Sánchez-Morcillo V.J. Transverse Patterns in Nonlinear Optical Resonators. Springer-Verlag, 2003.

Chang R.K., Campillo A.J., eds. Optical Processes in Microcavities. World Scientific, 1996.

Oraevskiy A.N. Gaussian Beams and Optical Resonators. Nova, 1996.

Anan'ev Yu. Laser Resonators and the Beam Divergence Problem. Taylor & Francis, 1992.

Vaughan J.M. The Fabry-Perot Interferometer. Adam Hilger, 1989.

Hernandez G. Fabry-Perot Interferometers. Cambridge University Press, paperback ed. 1988.

Siegman A.E. Lasers. University Science, 1986.

Weinstein L.A. Open Resonators and Open Waveguides. Golem Press, 1969.

СТАТЬИ

Issue on microresonators. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 12, № 1, 2006.

Vahala K.J. Optical Microcavities. Nature. Vol. 424, 2003. P. 839—846.

Millennium issue. IEEE JournaL of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Nöckel J.U., Stone A.D. Ray and Wave Chaos in Asymmetric Resonant Optical Cavities. Nature. Vol. 385, 1997. P. 45—47.

Yamamoto Y., Slusher R.E. Optical Processes in Microcavities. Physics Today. Vol. 46, № 6, 1993. P. 66—73.

Yokoyama H. Physics and Device Applications of Optical Microcavities. Science. Vol. 256, 1992. P. 66—70.

Siegman A.E. Unstable Optical Resonators. Applied Optics. Vol. 13, 1974. P. 353—367.

Kogelnik H., Li T. Laser Beams and Resonators. Applied Optics. Vol. 5, 1966. P. 1550—1567 (published simultaneously in Proceedings of the IEEE. Vol. 54, 1966. P. 1312—1329).

Fox A.G., Li T. Resonant Modes in a Maser Interferometer. Bell System Technical Journal. Vol. 40, 1961. P. 453—488.

Boyd G.D., Gordon J.P. Confocal Multimode Resonator for Millimeter Through Optical Wavelength Masers. Bell System Technical Journal. Vol. 40, 1961. P. 489—508.

Задачи

К РАЗДЕЛУ 10.1

1. Резонансные частоты резонатора с эталоном.

а. Определите расстояние между соседними резонансными частотами в резонаторе, образованном двумя параллельными плоскими зеркалами, расположенными на расстоянии $d = 15$ см друг от друга в воздухе ($n = 1$).

б. Прозрачная пластинка толщиной $d_1 = 2{,}5$ см с показателем преломления $n = 1{,}5$ помещена внутри резонатора и слегка наклонена, чтобы свет, отраженный от пластинки, не попадал на зеркала. Определите межмодовое расстояние резонатора.

2. **Беззеркальные резонаторы.** Полупроводниковые лазеры часто изготавливаются из кристаллов, поверхности которых образуются в результате раскалывания вдоль кристаллических плоскостей. Эти поверхности действуют как отражатели и, следовательно, служат зеркалами резонатора. Коэффициент отражения по интенсивности выражается формулой (6.41). Рассмотрим кристалл, находящийся в воздухе ($n = 1$), имеющий показатель преломления $n = 3{,}6$ и коэффициент потерь $\alpha_s = 1$ см$^{-1}$. Свет отражается между двумя параллельными поверхностями, расстояние между которыми $d = 0{,}2$ мм. Определите расстояние между резонансными частотами ν_F, полный коэффициент распределенных потерь α_r, параметр резкости $\mathcal{F}$, спектральную ширину $\delta\nu$ и добротность Q. Предполагая, что длина волны генерируемого света в свободном пространстве составляет 1,55 мкм, оцените номер продольной моды q.

3. **Эталон Фабри—Перо с отражательными брэгговскими решетками.** Эталон Фабри—Перо образован слоем GaAs между двумя брэгговскими отражательными решетками из GaAs/AlAs, описанными в задаче 6 к разд. 7.1. Определите параметр резкости $\mathcal{F}$ и добротность резонатора Q. Определите коэффициент пропускания брэгговской отражательной решетки, состоящей из $N = 10$ чередующихся слоев GaAs ($n_1 = 3{,}6$) и AlAs ($n_2 = 3{,}2$) толщиной d_1 и d_2, равной четверти длины волны в каждой среде. Считайте, что свет падает из протяженной среды GaAs.

4. **Спектральный отклик резонатора.** Коэффициент пропускания симметричного резонатора Фабри—Перо был измерен с использованием света от перестраиваемого монохроматического источника. Зависимость коэффициента пропускания от частоты имеет периодические пики с периодом 150 МГц, каждый шириной 5 МГц по половине высоты. Предполагая, что среда в резонаторе — газ с $n = 1$, определите длину и параметр резкости резонатора. Предполагая далее, что единственный источник потерь связан с зеркалами, определите их коэффициенты отражения.

5. **Время затухания оптической энергии.** Сколько времени нужно для того, чтобы оптическая энергия, запасенная в резонаторе с параметром резкости $\mathcal{F} = 100$, длиной $d = 50$ см и показателем преломления среды $n = 1$, уменьшилась вдвое?

К РАЗДЕЛУ 10.2

1. **Устойчивость резонатора со сферическими зеркалами.**

а. Может ли резонатор с двумя выпуклыми зеркалами быть устойчивым?

б. Может ли резонатор с одним выпуклым и одним вогнутым зеркалами быть устойчивым?

2. **Резонатор с плоскими зеркалами, содержащий линзу.** Линза с фокусным расстоянием f помещена внутрь резонатора с плоскими зеркалами, расположенными на расстоянии d. Линза находится на расстоянии $d/2$ от каждого из зеркал.

а. Определите матрицу передачи луча, начинающегося на одном из зеркал и совершающего полный обход резонатора.

б. Определите условие устойчивости резонатора.

в. При условии устойчивости изобразите графически гауссов пучок, который может стабильно существовать в резонаторе.

3. **Самовоспроизводящиеся лучи в симметричном резонаторе.** Рассмотрим симметричный резонатор из двух вогнутых зеркал с радиусом кривизны R на расстоянии $d = 3|R|/2$ друг от друга. После скольких обходов резонатора ход луча начнет повторяться?

4. **Положение луча в неустойчивых резонаторах.** Покажите, что в неустойчивом резонаторе положение луча после m полных обходов дается формулой

$$y_m = \alpha_1 h_1^m + \alpha_2 h_2^m,$$

где α_1 и α_2 — постоянные. Здесь

$$h_1 = b + \sqrt{b^2 - 1}; \quad h_2 = b - \sqrt{b^2 - 1}; \quad b = 2\left(1 + \frac{d}{R_1}\right)\left(1 + \frac{d}{R_2}\right) - 1.$$

Указание. Используйте результаты подразд. 1.4.4.

5. **Положение луча в неустойчивом симметричном резонаторе.** Убедитесь в том, что симметричный резонатор, состоящий из двух вогнутых зеркал с радиусами $R = -30$ см на расстоянии $d = 65$ см друг от друга, неустойчив. Найдите положение луча y_1, который начинается на одном из зеркал в положении $y_0 = 0$ с углом отклонения $\theta_0 = 0{,}1°$ и совершает один полный обход. Если зеркала имеют апертуру диаметром 5 см, то после скольких полных обходов луч покинет резонатор? Напишите компьютерную программу для графического построения y_m, $m = 2, 3, \ldots$, для $d = 50$ см и $d = 65$ см. Можно использовать результат задачи 4 к разд. 10.2.

6. **Стоячие волны в виде гауссовых пучков.** Рассмотрим волну, образованную суммой двух одинаковых гауссовых пучков, распространяющихся в направлениях $+z$ и $-z$. Покажите, что в результате получается стоячая волна. Используя граничные условия на двух идеальных зеркалах, расположенных таким образом, что они совпадают с волновыми фронтами, выведите выражение (10.67) для резонансных частот.

7. **Гауссов пучок в симметричном конфокальном резонаторе.** Симметричный конфокальный резонатор с расстоянием между зеркалами $d = 16$ см, коэффициентами отражения зеркал 0,995 и $n = 1$ используется в лазере, работающем на длине волны $\lambda_0 = 1$ мкм.

а. Найдите радиус кривизны зеркал.

б. Найдите перетяжку гауссовой моды (0, 0).

в. Изобразите графически распределение интенсивности мод (1, 0) на одном из зеркал и определите расстояние между двумя пиками.

г. Определите резонансные частоты мод (0, 0) и (1, 0).

д. Предполагая, что потери связаны только с неполным отражением на зеркалах, определите коэффициент распределенных потерь резонатора α_r.

8. **Дифракционные потери в симметричном конфокальном резонаторе.** Процент дифракционных потерь за один проход для различных мод низкого по-

рядка в симметричном конфокальном резонаторе приведен на рис. 10.20 как функция числа Френеля $N_F = a^2/\lambda d$ (где d — расстояние между зеркалами; a — радиус апертуры зеркал). Используя параметры из предыдущей задачи, определите радиус зеркал, при котором потери за один проход для моды (1, 0) составляют 1 %.

К РАЗДЕЛУ 10.3

1. **Число мод в резонаторах различной размерности.** Рассмотрим свет с длиной волны $\lambda_0 = 1{,}06$ мкм и спектральной шириной $\Delta\nu = 120$ ГГц. Сколько мод имеют частоты в пределах этой спектральной полосы в следующих резонаторах ($n = 1$):

а) одномерный резонатор длиной $d = 10$ см;

б) двумерный резонатор размером 10×10 см;

в) трехмерный резонатор размером $10 \times 10 \times 10$ см?

ГЛАВА
11 СТАТИСТИЧЕСКАЯ ОПТИКА

Макс Борн (1882—1970)

Эмиль Вольф (род. в 1922)

Книга «Основы оптики», впервые опубликованная в 1959 г. Максом Борном и Эмилем Вольфом, привлекла внимание к важности когерентности в оптике. Эмилю Вольфу мы обязаны многими достижениями в развитии теории оптической когерентности.

Статистическая оптика изучает свойства случайного света. Случайность эта возникает из-за непредсказуемых флуктуаций источника света или среды, в которой свет распространяется. Естественный свет, т. е. свет, испускаемый нагретым объектом, случаен, поскольку он представляет собой суперпозицию излучения очень большого числа атомов, излучающих независимо друг от друга свет с различными частотами и фазами. Случайность света может быть также следствие рассеяния на шероховатой поверхности, матовом стекле или турбулентной жидкости, которое вносит случайные флуктуации в волновой фронт. Теория случайных флуктуаций света известна также как **теория оптической когерентности**.

В предыдущих главах предполагалось, что свет является детерминированным или «когерентным». Пример когерентного света — монохроматическая волна $u(\boldsymbol{r}, t) = \mathrm{Re}\{U(\boldsymbol{r}) \exp(j\omega t)\}$, у которой комплексная амплитуда $U(\boldsymbol{r})$ — детерминированная комплексная функция, например $U(\boldsymbol{r}) = A \exp(-jkr)/r$ в случае сферической волны (рис. 11.1, *а*). Зависимость такой волновой функции

от времени и координат идеально периодична и предсказуема. Напротив, для случайного света зависимость волновой функции от времени и координат (рис. 11.1, *б*) не полностью предсказуема и в общем случае не может быть описана без обращения к статистическим методам.

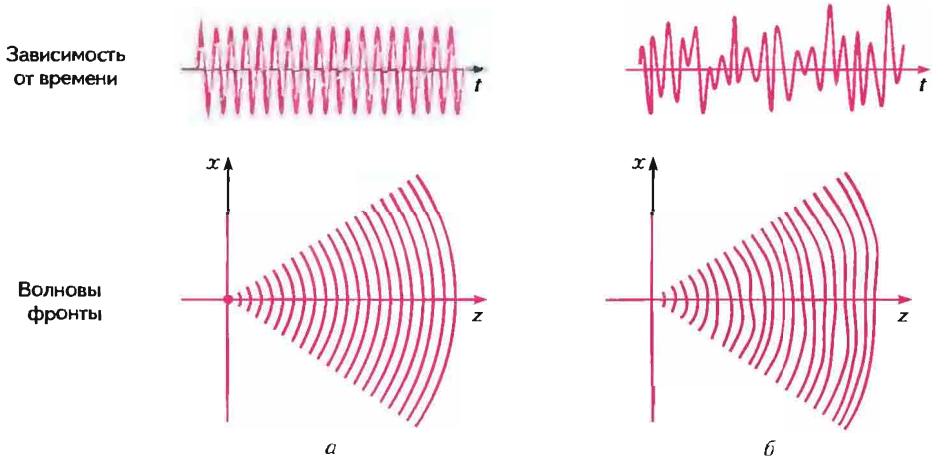


Рис. 11.1. Временна́я зависимость и волновые фронты для монохроматической сферической волны (*а*), являющейся примером когерентного света и случайного света (*б*)

Как можно извлечь из флуктуаций случайной оптической волны какие-либо полезные количественные параметры, характеризующие волну и отличающие ее от других случайных волн? Исследуем, например, три случайные оптические волны, волновые функции которых в данной точке пространства меняются со временем, как показано на рис. 11.2. Очевидно, что волна, изображенная на рис. 11.2, *б*, более «интенсивна», чем волна на рис. 11.2, *а*, и что огибающая волна (рис. 11.2, *в*) флуктуирует «быстрее», чем огибающие двух других волн.

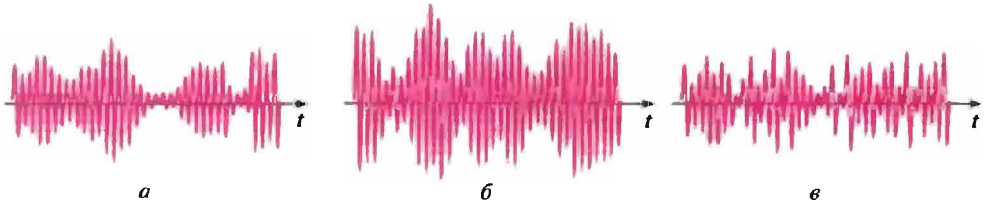


Рис. 11.2. Временна́я зависимость волновых функций для трех случайных волн

Для перевода этих качественных наблюдений в количественные меры используем понятие статистического усреднения, чтобы определить ряд неслучайных величин. Поскольку случайная функция удовлетворяет некоторым законам (волновому уравнению и граничным условиям), ее статистические средние также должны удовлетворять некоторым законам. Теория оптической когерент-

ности включает определение этих статистических средних, законы, которым они подчиняются, и меры, с помощью которых свет подразделяется на **когерентный, некогерентный** и, в общем случае, **частично когерентный.**

О данной главе

Эта глава представляет собой введение в теорию частичной когерентности. Знакомство с теорией случайных полей (случайных функций многих переменных — координат и времени) необходимо для полного понимания теории оптической когерентности. Однако круг идей, представленных в этой главе, ограничен, так что достаточно знать понятие статистического усреднения.

В разд. 11.1 дается определение двух статистических средних, используемых для описания случайного света: оптической интенсивности и функции взаимной когерентности. Вводится понятие временно́й и пространственной когерентности и устанавливается связь между временно́й когерентностью и монохроматичностью. Примеры частично когерентного света, приводимые в разд. 11.1, показывают, что пространственно когерентный свет не обязан обладать временно́й когерентностью, а монохроматический свет — быть пространственно когерентным. Одно из основных проявлений когерентности света — способность давать видимые интерференционные полосы. Разд. 11.2 посвящен законам интерференции случайного света. Прохождение частично когерентного света через свободное пространство и различные оптические системы, включая системы формирований изображений, является предметом рассмотрения в разд. 11.3. Краткое введение в теорию поляризации случайного света (частичной поляризации) дано в разд. 11.4.

11.1. СТАТИСТИЧЕСКИЕ СВОЙСТВА СЛУЧАЙНОГО СВЕТА

Произвольная оптическая волна описывается волновой функцией $u(\boldsymbol{r}, t) = \mathrm{Re}\{U(\boldsymbol{r}, t)\}$, где $U(\boldsymbol{r}, t)$ — комплексная волновая функция. Например, $U(\boldsymbol{r}, t)$ может быть выбрана в виде $U(\boldsymbol{r})\exp(j\omega t)$ для монохроматического света или быть суммой множества подобных функций с разными ω для полихроматического света (см. обсуждение комплексной волновой функции в подразд. 2.6.1). Для случайного света обе функции $u(\boldsymbol{r}, t)$ и $U(\boldsymbol{r}, t)$ являются случайными и характеризуются рядом статистических средних, которые мы определим в данном разделе.

11.1.1. Оптическая интенсивность

Интенсивность $I(\boldsymbol{r}, t)$ когерентного (детерминированного) света — это квадрат модуля комплексной волновой функции $U(\boldsymbol{r}, t)$:

$$I(\boldsymbol{r}, t) = |U(\boldsymbol{r}, t)|^2 \tag{11.1}$$

(см. подразд. 2.2.1 и 2.6.1). Для монохроматического детерминированного света интенсивность не зависит от времени, а для светового импульса — зависит.

Для случайного света $U(\boldsymbol{r}, t)$ является случайной функцией времени и координат. Поэтому интенсивность $|U(\boldsymbol{r}, t)|^2$ также является случайной. **Средняя интенсивность** определяется как

$$I(\boldsymbol{r}, t) = \left\langle |U(\boldsymbol{r}, t)|^2 \right\rangle, \tag{11.2}$$

Средняя интенсивность

где скобки $\langle\ \rangle$ обозначают усреднение по ансамблю реализаций случайной функции. Последнее означает, что волна многократно создается при одинаковых условиях, причем каждый раз получаются разные волновые функции. Вычисляется средняя интенсивность в каждой точке в каждый момент времени. Там, где это не приводит к неоднозначности, мы будем называть функцию $I(\boldsymbol{r}, t)$ просто *интенсивностью* света, подразумевая при этом *усреднение*. Величина $|U(\boldsymbol{r}, t)|^2$ называется **случайной**, или **мгновенной**, **интенсивностью**. Для детерминированного света операция усреднения не нужна, так как при каждой попытке получается одна и та же волновая функция, поэтому (11.2) эквивалентно (11.1).

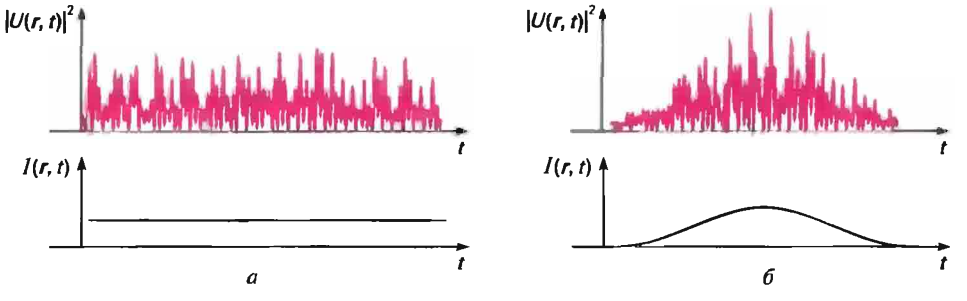


Рис. 11.3. Статистически стационарная волна имеет среднюю интенсивность, не меняющуюся со временем (*а*). Статистически нестационарная волна имеет зависящую от времени интенсивность (*б*). Графики могут представлять, например, свет лампы накаливания, питаемой постоянным электрическим током (*а*) и импульсом электрического тока (*б*)

Средняя интенсивность может не зависеть от времени или меняться со временем, как показано на рис. 11.3, *а* и *б* соответственно. Первый случай имеет место, когда волна статистически стационарна, т. е. ее статистические средние не зависят от времени. Мгновенная интенсивность $|U(\boldsymbol{r}, t)|^2$ флуктуирует случайным образом, но ее среднее значение постоянно. В этом случае мы будем обозначать его просто $I(\boldsymbol{r})$. Стационарность случайной величины не означает ее постоянства — она означает постоянство свойств в среднем. Примером случайного стационарного света может служить свет обычной лампы накаливания, питаемой постоянным электрическим током. Средняя интенсивность $I(\boldsymbol{r})$ зависит от расстояния до лампы, но не меняется со временем. Однако случай-

ная интенсивность $|U(\boldsymbol{r}, t)|^2$ флуктуирует как во времени, так и в пространстве, как показано на рис. 11.3, *а*.

Когда свет стационарен, статистическое усреднение в (11.2) обычно выполняется путем усреднения по большому промежутку времени (вместо усреднения по большому числу реализаций), тогда

$$I(\boldsymbol{r}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |U(\boldsymbol{r}, t)|^2 \mathrm{d}t. \tag{11.3}$$

11.1.2. Временна́я когерентность и спектр

Рассмотрим флуктуации *стационарного* света в фиксированной точке $\boldsymbol{r}$ как функцию времени. Стационарная случайная функция $U(\boldsymbol{r}, t)$ имеет постоянную интенсивность $I(\boldsymbol{r}) = \langle |U(\boldsymbol{r}, t)|^2 \rangle$. Для краткости мы опустим обозначение зависимости от координат, поскольку точка $\boldsymbol{r}$ фиксирована; тогда $U(\boldsymbol{r}, t) = U(t)$ и $I(\boldsymbol{r}) = I$.

Случайные флуктуации характеризуются временны́м масштабом, характеризующим «память» случайной функции. Флуктуации, разделенные временны́м промежутком, превышающим время памяти, статистически независимы, так что процесс «забывает» себя. Функция оказывается плавно меняющейся на временны́х масштабах порядка времени памяти, но резко и нерегулярно меняющейся на больших временны́х масштабах (см. рис. 11.2). Количественной мерой такого временно́го поведения является статистическое *среднее*, называемое автокорреляционной функцией. Эта функция описывает степень, до которой функция флуктуирует в унисон в два различных момента времени, разделенные определенным промежутком. Таким образом, устанавливается временно́й масштаб процесса, лежащего в основе генерации волновой функции.

Функция временно́й когерентности

Автокорреляционная функция стационарной комплексной случайной функции $U(t)$ — это среднее от произведения $U^*(t)$ и $U(t+\tau)$ как функция времени задержки τ

$$\boxed{G(\tau) = \langle U^*(t)U(t+\tau)\rangle,} \tag{11.4}$$

Функция временно́й когерентности

или

$$G(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} U^*(t)U(t+\tau)\mathrm{d}t \tag{11.5}$$

(см. разд. А.1 приложения А).

Для понимания важности определения (11.4) рассмотрим случай, когда среднее значение комплексной волновой функции равно нулю $\langle U(t)\rangle = 0$. Это возможно, если угол наклона (фаза) вектора на комплексной плоскости равновероятно принимает все значения между 0 и 2π, как показано на рис. 11.4. Фаза произведения $U^*(t)U(t+\tau)$ — это угол между векторами $U(t)$ и $U(t+\tau)$. Если между $U(t)$ и $U(t+\tau)$ нет корреляции, то угол между соответствующими векторами на комплексной плоскости случайным образом меняется от 0 до 2π. Вектор $U^*(t)U(t+\tau)$ тогда имеет совершенно случайное направление, поэтому его среднее значение — автокорреляционная функция $G(\tau)$ — обращается в нуль. Напротив, если для данного τ между $U(t)$ и $U(t+\tau)$ имеется корреляция, то их комплексные вектора сохраняют некоторую связь. Их флуктуации тогда связаны так, что вектор произведения $U^*(t)U(t+\tau)$ имеет преимущественное направление, и его среднее значение $G(\tau)$ отлично от нуля.

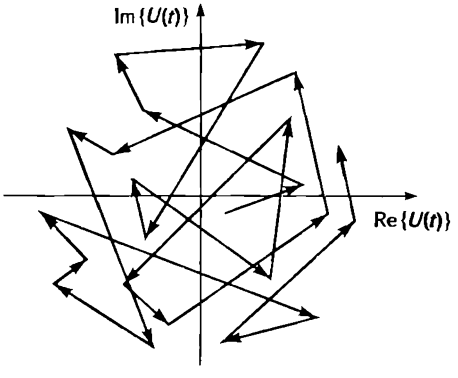


Рис. 11.4. Изменение вектора $U(t)$ на комплексной плоскости со временем в случае, когда его аргумент имеет равномерное распределение между 0 и 2π. Средние значения его мнимой и действительной частей равны нулю, поэтому $\langle U(t)\rangle = 0$

В теории оптической когерентности автокорреляционная функция $G(\tau)$ известна как **функция временно́й когерентности**. Легко показать, что $G(\tau)$ обладает свойством эрмитовой симметрии $G(-\tau) = G^*(\tau)$ и что интенсивность I, определенная выражением (11.2), равна $G(\tau)$ при $\tau = 0$,

$$I = G(0). \tag{11.6}$$

Степень временно́й когерентности

Функция временно́й когерентности $G(\tau)$ несет информацию как об интенсивности $I = G(0)$, так и о степени корреляции (когерентности) стационарного света. Мерой когерентности, нечувствительной к интенсивности, является нормированная автокорреляционная функция

$$\boxed{g(\tau) = \frac{G(\tau)}{G(0)} = \frac{\langle U^*(t)U(t+\tau)\rangle}{\langle U^*(t)U(t)\rangle},} \tag{11.7}$$

Комплексная степень временно́й когерентности

которая называется **комплексной степенью временно́й когерентности**. Ее абсолютное значение не превышает единицы

$$0 \le |g(\tau)| \le 1. \tag{11.8}$$

Значение $|g(\tau)|$ является мерой корреляции между $U(t)$ и $U(t+\tau)$. Когда свет детерминированный и монохроматический, т. е. $U(t) = A\exp(j\omega_0 t)$, где A — постоянная, формула (11.7) дает

$$g(\tau) = \exp(j\omega_0\tau), \tag{11.9}$$

так что $|g(\tau)| = 1$ для всех τ. Переменные $U(t)$ и $U(t+\tau)$ в этом случае полностью коррелированны при любых временных задержках τ. Обычно с ростом τ $|g(\tau)|$ спадает от своего максимального значения $|g(0)| = 1$, и при достаточно больших τ флуктуации становятся некоррелированными.

Время когерентности

Если $|g(\tau)|$ монотонно убывает с ростом временно́й задержки, то значение τ_c, при котором спад достигает заданной величины (например, 1/2 или $1/e$), служит мерой времени памяти флуктуаций и называется **временем когерентности** (рис. 11.5).

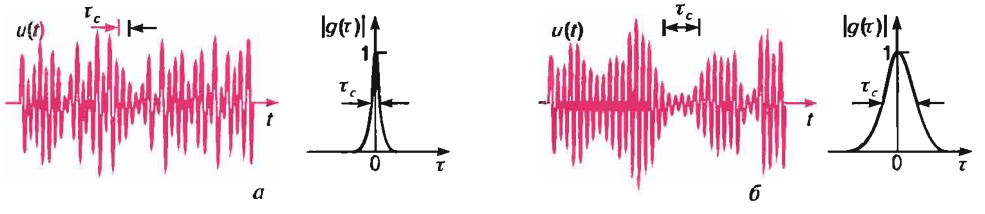


Рис. 11.5. Иллюстративные примеры волновых функций, модуля комплексной степени временно́й когерентности $|g(\tau)|$ и времени когерентности τ_c для оптических полей с малым (*а*) и большим (*б*) временем когерентности. Амплитуда и фаза волновой функции испытывают случайные временные изменения с постоянными времени, примерно равными времени когерентности. В обоих случаях время когерентности τ_c больше, чем период оптических колебаний. На протяжении времени когерентности волна достаточно предсказуема и может быть аппроксимирована синусоидой. Однако, зная амплитуду и фазу волны в некоторый момент времени, невозможно предсказать их значения через промежуток времени, превышающий время когерентности

При $\tau < \tau_c$ флуктуации сильно коррелированны, а при $\tau > \tau_c$ — слабо. В общем случае τ_c определяет ширину функции $|g(\tau)|$. Хотя определение ширины достаточно произвольно (см. разд. А.2 приложения А), как правило, в качестве определения времени когерентности выбирается ширина по интенсивности

$$\tau_c = \int_{-\infty}^{\infty} |g(\tau)|^2 \, d\tau \tag{11.10}$$

Время когерентности

[см. (А.15) и заметьте, что $g(0) = 1$]. Время когерентности монохроматического света бесконечно, поскольку $|g(\tau)| = 1$ на всей оси.

Упражнение 11.1

Время когерентности

Убедитесь, что следующие выражения для комплексной степени временной когерентности согласуются с определением τ_c (11.10):

$$g(\tau) = \begin{cases} \exp\left(-\dfrac{|\tau|}{\tau_c}\right) & \text{(экспоненциальная);} \\ \exp\left(-\dfrac{\pi\tau^2}{2\tau_c{}^2}\right) & \text{(гауссова).} \end{cases} \tag{11.11}$$

Во сколько раз уменьшается $|g(\tau)|$, если τ возрастает от 0 до τ_c в каждом случае?

Свет, у которого время когерентности τ_c много больше, чем разности временны́х задержек, встречаемых в рассматриваемой оптической системе, эффективно является полностью когерентным. Так, например, свет эффективно когерентен, если расстояние $c\tau_c$ много больше всех оптических разностей хода, встречаемых в системе. Расстояние

$$\boxed{l_c = c\tau_c} \tag{11.12}$$

Длина когерентности

называется **длиной когерентности**.

Спектральная плотность мощности

Для определения *среднего* спектра случайного света произведем разложение Фурье случайной функции $U(t)$. Амплитуда компоненты с частотой ν есть Фурье-образ (см. приложение А)

$$V(\nu) = \int_{-\infty}^{\infty} U(t)\exp(-j2\pi\nu t)\,\mathrm{d}t. \tag{11.13}$$

Средняя энергия на единицу площади у таких компонент в интервале частот от ν до $\nu + \mathrm{d}\nu$ есть $\langle|V(\nu)|^2\rangle\mathrm{d}\nu$, так что $\langle|V(\nu)|^2\rangle$ представляет собой спектральную плотность энергии света (энергию, приходящуюся на единицу площади и на единицу частоты). Заметим, что комплексная волновая функция $U(t)$ была определена так, что $V(\nu) = 0$ для отрицательных частот ν (см. подразд. 2.6.1).

Поскольку истинно стационарная функция $U(t)$ вечна и несет бесконечную энергию, рассмотрим спектральную плотность *мощности*, а не энергии. Вначале определим спектральную плотность энергии для функции $U(t)$, наблюдаемой во временно́м окне протяженностью T, посредством обрезанного преобразования Фурье:

$$V_T(\nu) = \int_{-T/2}^{T/2} U(t)\exp(-j2\pi\nu t)\,\mathrm{d}t. \tag{11.14}$$

Затем определим спектральную плотность энергии $\langle |V_T(\nu)|^2 \rangle$. Спектральная плотность мощности есть энергия в единицу времени $(1/T)\langle |V_T(\nu)|^2 \rangle$. Теперь расширим временно́е окно до бесконечности, перейдя к пределу $T \to \infty$. Результат

$$S(\nu) = \lim_{T\to\infty} \frac{1}{T} \left\langle |V_T(\nu)|^2 \right\rangle \tag{11.15}$$

называется **спектральной плотностью мощности**. Она отлична от нуля только для положительных частот. Поскольку $U(t)$ определена так, что $|U(t)|^2$ представляет собой мощность на единицу площади, или интенсивность (Вт/см^2), то $S(\nu)\mathrm{d}\nu$ представляет собой среднюю мощность на единицу площади, переносимую на частотах между ν и $\nu + \mathrm{d}\nu$, так что $S(\nu)$ на самом деле представляет собой спектральную плотность интенсивности [Вт/(см$^2 \cdot$ Гц)]. Ее часто называют просто **спектральной плотностью** или **спектром**. Полная средняя интенсивность дается интегралом

$$I = \int_0^\infty S(\nu)\mathrm{d}\nu. \tag{11.16}$$

Можно показать, что автокорреляционная функция $G(\tau)$, определенная формулой (11.4), и спектральная плотность $S(\nu)$, определенная формулой (11.15), образуют пару функций, связанных преобразованием Фурье (см. задачу 2 к разд. 11.1)

$$\boxed{S(\nu) = \int_{-\infty}^{\infty} G(\tau)\exp(-j2\pi\nu\tau)\mathrm{d}\tau.} \tag{11.17}$$

Спектральная плотность мощности

Это соотношение известно как **теорема Винера—Хинчина**.

Оптическая волна, представляющая цветное изображение, такое, например, как на рис. 11.6 (см. ув. вклейку), имеет спектр, зависящий от координаты $\boldsymbol{r}$; каждый показанный спектральный профиль соответствует воспринимаемому цвету.

Спектральная ширина

Спектр света часто ограничен узкой полосой с некоторой центральной частотой ν_0. Спектральной шириной, или шириной линии света, называется ширина $\Delta\nu$ спектральной плотности $S(\nu)$. Так как $S(\nu)$ и $G(\tau)$ связаны преобразованием Фурье, их ширины обратно пропорциональны друг другу. Источник света с широким спектром имеет малое время когерентности и наоборот, как показано на рис. 11.7. В предельном случае монохроматического света $G(\tau) = I\exp(j\omega_0\tau)$, так что соответствующая спектральная плотность интенсивности $S(\nu) = I\delta(\nu - \nu_0)$ имеет единственную частотную компоненту ν_0.

Таким образом, в этом случае $\tau_c = \infty$ и $\Delta\nu = 0$. Время когерентности источника света можно увеличить с помощью оптического фильтра, уменьшающего спектральную ширину. В этом случае когерентность приобретается за счет потери энергии света.

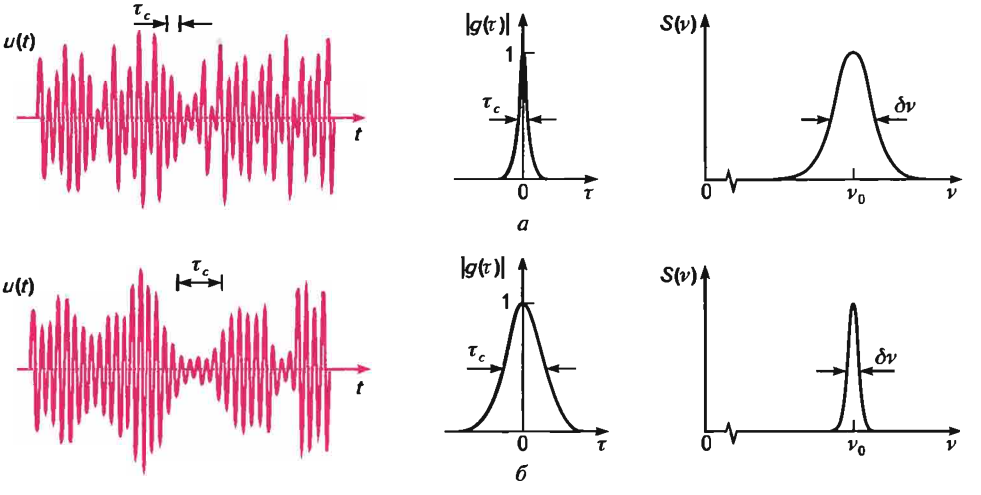


Рис. 11.7. Две случайные волны (*а*, *б*), модули их комплексных степеней временно́й когерентности и спектральные плотности

Существует несколько определений спектральной ширины. Наиболее распространенным из них является полная ширина функции $S(\nu)$ на уровне половины максимума (Full Width Half-Maximum, FWHM). Связь между временем когерентности и спектральной шириной зависит от спектрального профиля, как показано в табл. 11.1 (см. также приложение А, разд. А.2).

Таблица 11.1. **Связь между спектральной шириной и временем когерентности**

Контур спектральной плотности	Прямоугольный	Лоренцев	Гауссов
Спектральная ширина на уровне половины максимума $\Delta\nu_{\text{FWHM}}$	$\dfrac{1}{\tau_c}$	$\dfrac{1}{\pi\tau_c} \approx \dfrac{0{,}32}{\tau_c}$	$\dfrac{\sqrt{2\ln 2/\pi}}{\tau_c} \approx \dfrac{0{,}66}{\tau_c}$

Другим определением спектральной ширины является

$$\Delta\nu_c = \frac{\left(\int_0^\infty S(\nu)\,\mathrm{d}\nu\right)^2}{\int_0^\infty S^2(\nu)\,\mathrm{d}\nu}. \qquad (11.18)$$

С помощью этого определения можно показать, что

$$\boxed{\Delta\nu_c = \frac{1}{\tau_c}} \qquad (11.19)$$

Спектральная ширина

независимо от спектрального профиля (см. упражнение 11.2). Например, если $S(\nu)$ — прямоугольная функция, занимающая интервал от $(\nu_0 - B/2)$ до $(\nu_0 + B/2)$, то формула (11.18) даст $\Delta\nu_c = B$. Два определения спектральной ширины, $\Delta\nu_c \equiv \Delta\nu_{\text{FWHM}}$, различаются множителем, который меняется от $1/\pi \approx 0{,}32$ до 1 для профилей, приведенных в табл. 11.1.

Упражнение 11.2

Связь между спектральной шириной и временем когерентности

Покажите, что время когерентности τ_c, определенное формулой (11.10), связано со спектральной шириной $\Delta\nu_c$, определенной формулой (11.18), простой обратно пропорциональной зависимостью $\tau_c = 1/\Delta\nu_c$. *Указание*: используйте определения $\Delta\nu_c$ и τ_c, преобразование Фурье, связывающее $S(\nu)$ и $G(\tau)$, и теорему Парсеваля [см. (А.7) в приложении А].

Примеры спектральных ширин различных источников света и соответствующих времен когерентности и длин когерентности $l_c = c\tau_c$ приведены в табл. 11.2.

Таблица 11.2. **Спектральные ширины ряда источников света и соответствующие времена и длины когерентности в свободном пространстве**

Источник	$\Delta\nu_c$, Гц	$\tau_c = 1/\Delta\nu_c$	$l_c = c\tau_c$
Фильтрованный солнечный свет ($\lambda_0 = 0{,}4$—$0{,}8$ мкм)	$3{,}74 \cdot 10^{14}$	2,67 фс	800 нм
Светодиод ($\lambda_0 = 1$ мкм, $\Delta\lambda_0 = 50$ нм)	$1{,}5 \cdot 10^{13}$	67 фс	20 мкм
Натриевая лампа низкого давления	$5 \cdot 10^{11}$	2 пс	600 мкм
Многомодовый лазер ($\lambda_0 = 633$ нм)	$1{,}5 \cdot 10^{9}$	0,67 нс	20 см
Одномодовый лазер ($\lambda_0 = 633$ нм)	$1\ 10^{6}$	1 мкс	300 м

Пример 11.1

Волна, состоящая из случайной последовательности волновых пакетов

Свет, испускаемый некогерентным источником, можно моделировать последовательностью волновых пакетов, испускаемых в случайные моменты времени (рис. 11.8). Каждый волновой пакет имеет случайную фазу, поскольку пакеты излучаются разными атомами.

Волновые пакеты могут быть, например, синусоидальными с экспоненциально затухающей огибающей, так что пакет, испущенный в момент $t = 0$, имеет комплексную волновую функцию (в заданной точке пространства)

$$U_p(t) = \begin{cases} A_p \exp\left(-\dfrac{t}{\tau_c}\right)(j\omega_0 t), & t \geq 0; \\ 0, & t < 0. \end{cases} \tag{11.20}$$

Рис. 11.8. Свет, образованный волновыми пакетами, испущенными в разные моменты времени, имеет время когерентности, равное длительности волнового пакета

Моменты испускания совершенно случайны. То же самое можно сказать о независимых фазах различных актов испускания, включенных в A_p. Статистические свойства полного поля можно определить с помощью необходимых операций усреднения с использованием законов математической статистики. В результате получается комплексная степень когерентности

$$g(\tau) = \exp\left(-\frac{|\tau|}{\tau_c}\right)\exp(j\omega_0\tau),$$

модуль которой — симметричная экспоненциальная функция. Соответствующая спектральная плотность мощности есть функция Лоренца

$$S(\nu) = \frac{\Delta\nu/2\pi}{(\nu - \nu_0)^2 + (\Delta\nu/2)^2},$$

где $\Delta\nu = 1/\pi\tau_c$ (см. табл. А.8 в приложении А). Время когерентности τ_c в этом случае точно равно длительности пакета. Следовательно, утверждение, что такой свет коррелирован в пределах времени когерентности, означает, что он коррелирован на протяжении длительности отдельного волнового пакета.

11.1.3. Пространственная когерентность

Функция взаимной когерентности

Важной характеристикой пространственных и временны́х флуктуаций случайной функции $U(\boldsymbol{r}, t)$ является функция взаимной корреляции $U(\boldsymbol{r}_1, t)$ и $U(\boldsymbol{r}_2, t)$ между парами точек $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$

$$\boxed{G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = \langle U^*(\boldsymbol{r}_1, t)U(\boldsymbol{r}_2, t+\tau)\rangle.} \tag{11.21}$$

Функция взаимной когерентности

Эта функция времени задержки называется функцией взаимной когерентности. В нормированном виде

$$g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = \frac{G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)}{\sqrt{I(\boldsymbol{r}_1) I(\boldsymbol{r}_2)}}, \tag{11.22}$$

Комплексная степень когерентности

она называется **комплексной степенью когерентности**. Если две точки совпадают, так что $\boldsymbol{r}_1 = \boldsymbol{r}_2 = \boldsymbol{r}$, то (11.21) и (11.22) совпадают, соответственно, с функцией временно́й когерентности (11.4) и комплексной функцией временно́й когерентности (11.7) в точке $\boldsymbol{r}$.

Комплексная степень когерентности $g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$ представляет собой коэффициент взаимной корреляции случайных переменных $U^*(\boldsymbol{r}_1, t)$ и $U(\boldsymbol{r}_2, t + \tau)$. Ее модуль принимает значения в интервале от нуля до единицы

$$0 \le |g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)| \le 1. \tag{11.23}$$

Таким образом, эта величина является мерой степени корреляции между флуктуациями в точках $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$ спустя время τ.

Когда два вектора на комплексной плоскости $U(\boldsymbol{r}_1, t)$ и $U(\boldsymbol{r}_2, t)$ флуктуируют независимо и их фазы совершенно случайны (каждая с равной вероятностью принимает любое значение в интервале от 0 до 2π), $|g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)| = 0$, поскольку среднее от произведения $U^*(\boldsymbol{r}_1, t) U(\boldsymbol{r}_2, t + \tau)$ исчезает. В этом случае флуктуации в двух точках не коррелированы. Другой предельный случай $|g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)| = 1$ реализуется, когда флуктуации в точках $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$ спустя время τ полностью коррелированы. Заметим, что $|g(\boldsymbol{r}_1, \boldsymbol{r}_2, 0)|$ не обязательно равно единице, однако по определению $|g(\boldsymbol{r}, \boldsymbol{r}, 0)| = 1$.

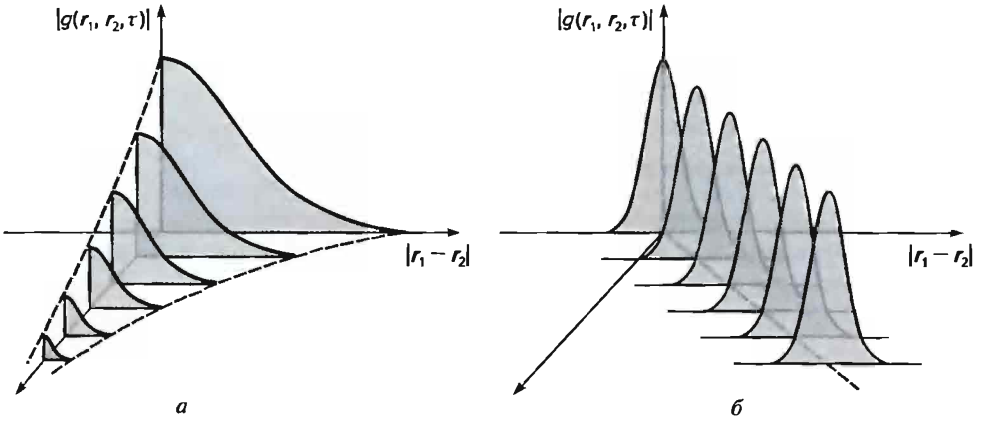


Рис. 11.9. Два примера зависимости $|g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)|$ от расстояния $|\boldsymbol{r}_1 - \boldsymbol{r}_2|$ и времени задержки τ: *а* — максимальная корреляция при заданном $|\boldsymbol{r}_1 - \boldsymbol{r}_2|$ имеет место при $\tau = 0$; *б* — максимальная корреляция имеет место при $|\boldsymbol{r}_1 - \boldsymbol{r}_2| = c\tau$

Зависимость $g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$ от времени задержки и координат характеризует временну́ю и пространственную когерентность света. Два примера зависимости $|g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)|$ от расстояния $|\boldsymbol{r}_1 - \boldsymbol{r}_2|$ и времени задержки τ показаны на рис. 11.9.

Пространственные и временны́е флуктуации света тесно связаны, поскольку свет распространяется в виде волн, и комплексная волновая функция $U(\boldsymbol{r}, t)$ должна удовлетворять волновому уравнению. Это накладывает определенные условия на функцию взаимной когерентности (см. упражнение 11.3). Для иллюстрации рассмотрим, например, плоскую волну случайного света, распространяющуюся в направлении z в однородной среде без дисперсии со скоростью c. Флуктуации в точках $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$ полностью коррелированы, если время задержки равно $\tau = \tau_0 = |z_1 - z_2|/c$, так что $|g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau_0)| = 1$. Как функция τ величина $|g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)|$ имеет максимум при $\tau = \tau_0$, как показано на рис. 11.9, *б*. Этот пример будет снова рассмотрен в подразд. 11.1.4.

Упражнение 11.3

Дифференциальные уравнения для функции взаимной когерентности

В свободном пространстве $U(\boldsymbol{r}, t)$ должна удовлетворять волновому уравнению

$$\nabla^2 U - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} = 0.$$

Используя определение (11.21), покажите, что функция взаимной когерентности $G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$ удовлетворяет двум дифференциальным уравнениям в частных производных

$$\nabla_1^2 G - \frac{1}{c^2}\frac{\partial^2 G}{\partial \tau^2} = 0; \tag{11.24а}$$

$$\nabla_2^2 G - \frac{1}{c^2}\frac{\partial^2 G}{\partial \tau^2} = 0, \tag{11.24б}$$

где ∇_1^2 и ∇_2^2 — операторы Лапласа по отношению к переменным $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$ соответственно.

Взаимная интенсивность

Пространственную когерентность можно оценить, исследуя зависимость функции взаимной когерентности от координат при фиксированном времени задержки τ. Во многих ситуациях наиболее подходящей является точка $\tau = 0$, как в примере на рис. 11.9, *а*. Это, однако, не всегда так, как показывает пример на рис. 11.9, *б*. Функция взаимной когерентности при $\tau = 0$

$$G(\boldsymbol{r}_1, \boldsymbol{r}_2, 0) = \langle U^*(\boldsymbol{r}_1, t) U(\boldsymbol{r}_2, t) \rangle \tag{11.25}$$

называется **взаимной интенсивностью** и для простоты обозначается как $G(\boldsymbol{r}_1, \boldsymbol{r}_2)$. Диагональные значения взаимной интенсивности ($\boldsymbol{r}_1 = \boldsymbol{r}_2 = \boldsymbol{r}$) дают интенсивность света $I(\boldsymbol{r}) = G(\boldsymbol{r}, \boldsymbol{r})$.

Когда разности хода, встречающиеся в оптической системе, малы по сравнению с длиной когерентности $l_c = c\tau_c$, свет можно рассматривать как эффективно обладающий полной временной когерентностью, так что функция взаимной когерентности гармонически зависит от времени

$$G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = G(\boldsymbol{r}_1, \boldsymbol{r}_2)\exp(j\omega_0\tau), \tag{11.26}$$

где ω_0 — центральная частота. В этом случае говорят, что свет **квазимонохроматический**, и взаимная интенсивность $G(\boldsymbol{r}_1, \boldsymbol{r}_2)$ полностью описывает пространственную когерентность.

Комплексная степень когерентности $g(\boldsymbol{r}_1, \boldsymbol{r}_2, 0)$ также обозначается через $g(\boldsymbol{r}_1, \boldsymbol{r}_2)$. Таким образом $g(\boldsymbol{r}_1, \boldsymbol{r}_2)$ — нормированная взаимная интенсивность,

$$g(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{G(\boldsymbol{r}_1, \boldsymbol{r}_2)}{\sqrt{I(\boldsymbol{r}_1)I(\boldsymbol{r}_2)}}. \tag{11.27}$$

Нормированная взаимная интенсивность

Модуль $|g(\boldsymbol{r}_1, \boldsymbol{r}_2)|$ заключен между нулем и единицей и используется как мера степени пространственной когерентности (когда время задержки τ равно нулю). Если комплексная волновая функция $U(\boldsymbol{r}, t)$ детерминирована, то $|g(\boldsymbol{r}_1, \boldsymbol{r}_2)| = 1$ для всех $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, так что свет полностью коррелирован во всем пространстве.

Площадь когерентности

Пространственная когерентность квазимонохроматического света в заданной плоскости в окрестности данной точки $\boldsymbol{r}_2$ описывается величиной $|g(\boldsymbol{r}_1, \boldsymbol{r}_2)|$ как функцией расстояния $|\boldsymbol{r}_1 - \boldsymbol{r}_2|$. Эта функция равна единице, когда $\boldsymbol{r}_1 = \boldsymbol{r}_2$, и спадает с ростом $|\boldsymbol{r}_1 - \boldsymbol{r}_2|$ (не обязательно монотонно). Площадь области, при нахождении в которой точки $\boldsymbol{r}_1$ величина $|g(\boldsymbol{r}_1, \boldsymbol{r}_2)|$ превышает некоторую заданную часть (например, 1/2 или $1/e$) своего максимального значения, называется **площадью когерентности**. Она характеризует пространственную протяженность величины $|g(\boldsymbol{r}_1, \boldsymbol{r}_2)|$ как функции $\boldsymbol{r}_1$ при заданном $\boldsymbol{r}_2$, как показано на рис. 11.10. В пределе абсолютно когерентного света площадь когерентности бесконечна.

Площадь когерентности — важная характеристика случайного света. Этот параметр нужно рассматривать в сравнении с другими типичными размерами оптической системы. Например, если площадь когерентности больше размеров апертуры, через которую пропускается свет, так что во всех точках, представляющих интерес $|g(\boldsymbol{r}_1, \boldsymbol{r}_2)| \approx 1$, то свет можно рассматривать как когерентный, т. е. как если бы площадь когерентности была бесконечной. Аналогично, если площадь когерентности меньше, чем разрешение оптической системы, то она может рассматриваться как бесконечно малая, т. е. $g(\boldsymbol{r}_1, \boldsymbol{r}_2) = 0$ практически для всех $\boldsymbol{r}_1 \neq \boldsymbol{r}_2$. В этом предельном случае свет называют **некогерентным**.

Свет, исходящий от протяженной излучающей раскаленной поверхности, имеет площадь когерентности порядка λ^2, где λ — центральная частота, так что

в большинстве практических случаев его можно считать некогерентным. Таким образом, когерентность и некогерентность представляют собой два предельных частных случая частичной когерентности.

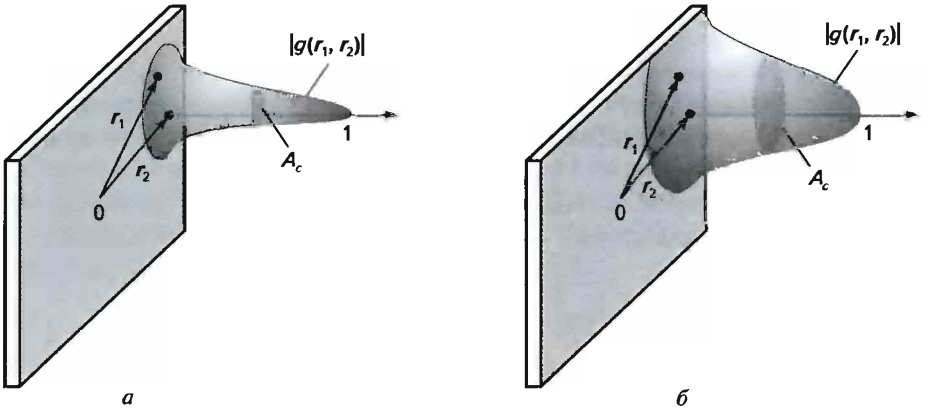


Рис. 11.10. Два иллюстрированных примера зависимости модуля нормированной взаимной интенсивности от $\boldsymbol{r}_1$ в окрестности фиксированной точки $\boldsymbol{r}_2$. Площадь когерентности в случае (*а*) меньше, чем в случае (*б*)

Взаимная спектральная плотность

Функция взаимной когерентности $G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$ описывает пространственную корреляцию при каждом времени задержки τ. Задержка $\tau = 0$ выбирается для определения взаимной интенсивности $G(\boldsymbol{r}_1, \boldsymbol{r}_2) = G(\boldsymbol{r}_1, \boldsymbol{r}_2, 0)$, которая удобна для описания пространственной когерентности квазимонохроматического света. Полезной альтернативой является описание когерентности в частотном представлении, т. е. исследование пространственной корреляции на фиксированной частоте. **Взаимная спектральная плотность** (или взаимный спектр мощности) определяется как Фурье-образ функции $G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$ относительно τ

$$S(\boldsymbol{r}_1, \boldsymbol{r}_2, \nu) = \int_{-\infty}^{\infty} G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)\exp(-j2\pi\nu\tau)\,\mathrm{d}\tau. \qquad (11.28)$$

Взаимная спектральная плотность

Когда $\boldsymbol{r}_1 = \boldsymbol{r}_2 = \boldsymbol{r}$, взаимная спектральная плотность превращается в спектральную плотность мощности $S(\nu)$ в точке $\boldsymbol{r}$, определяемую формулой (11.17).

Нормированная взаимная спектральная плотность определяется выражением

$$s(\boldsymbol{r}_1, \boldsymbol{r}_2, \nu) = \frac{S(\boldsymbol{r}_1, \boldsymbol{r}_2, \nu)}{\sqrt{S(\boldsymbol{r}_1, \boldsymbol{r}_1, \nu)S(\boldsymbol{r}_2, \boldsymbol{r}_2, \nu)}}. \qquad (11.29)$$

Можно показать, что ее абсолютная величина лежит в пределах от нуля до единицы, так что она является мерой степени пространственной когерентности на частоте ν. Она характеризует степень коррелированности флуктуаций компонент на частоте ν в точках $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$.

В некоторых случаях взаимная спектральная плотность факторизуется и ее можно представить как произведение функции координат и функции частоты, $S(\boldsymbol{r}_1, \boldsymbol{r}_2, \nu) = G(\boldsymbol{r}_1, \boldsymbol{r}_2)s(\nu)$, так что пространственные и частотные свойства можно отделить друг от друга. В этом случае свет называется **кросс-спектрально чистым**. Функция взаимной когерентности тогда тоже должна факторизоваться и представляться в виде произведения функции координат и функции времени, $G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = G(\boldsymbol{r}_1, \boldsymbol{r}_2)g(\tau)$, где $g(\tau)$ — результат обратного преобразования Фурье функции $s(\nu)$. Если при факторизации потребовать, чтобы

$$\int s(\nu)\,\mathrm{d}\nu = 1,$$

то $G(\boldsymbol{r}_1, \boldsymbol{r}_2) = G(\boldsymbol{r}_1, \boldsymbol{r}_2, 0)$, так что $G(\boldsymbol{r}_1, \boldsymbol{r}_2)$ — не что иное, как взаимная интенсивность. Кросс-спектрально чистый свет обладает двумя важными свойствами:

1. В одной точке $\boldsymbol{r}$ имеем $S(\boldsymbol{r}, \boldsymbol{r}, \nu) = G(\boldsymbol{r}, \boldsymbol{r})s(\nu) = I(\boldsymbol{r})s(\nu)$. Спектр имеет одинаковый профиль во всех точках. Если свет несет видимое изображение, то оно одноцветное, и разные точки его отличаются только яркостью.

2. Нормированная взаимная спектральная плотность

$$s(\boldsymbol{r}_1, \boldsymbol{r}_2, \nu) = \frac{G(\boldsymbol{r}_1, \boldsymbol{r}_2)}{\sqrt{G(\boldsymbol{r}_1, \boldsymbol{r}_1)G(\boldsymbol{r}_2, \boldsymbol{r}_2)}} = g(\boldsymbol{r}_1, \boldsymbol{r}_2) \tag{11.30}$$

не зависит от частоты. В этом случае нормированная взаимная интенсивность $g(\boldsymbol{r}_1, \boldsymbol{r}_2)$ описывает пространственную когерентность на всех частотах.

11.1.4. Продольная когерентность

В этом разделе мы введем понятие продольной когерентности на примерах случайных волн с фиксированными волновыми фронтами, такими как плоские и сферические волны

Частично когерентная плоская волна

Рассмотрим плоскую волну

$$U(\boldsymbol{r}, t) = a\left(t - \frac{z}{c}\right)\exp\left[j\omega_0\left(t - \frac{z}{c}\right)\right], \tag{11.31}$$

распространяющуюся в направлении z в однородной среде со скоростью c. Как показано в подразд. 2.6.1, функция $U(\boldsymbol{r}, t)$ удовлетворяет волновому уравнению при произвольной функции $a(t)$. Если $a(t)$ — случайная функция, то $U(\boldsymbol{r}, t)$

описывает частично когерентный свет. Функция взаимной когерентности, определенная формулой (11.21), принимает вид

$$G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = G_a\left(\tau - \frac{z_2 - z_1}{c}\right)\exp\left[j\omega_0\left(\tau - \frac{z_2 - z_1}{c}\right)\right], \tag{11.32}$$

где z_1 и z_2 — z-компоненты $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$, а $G_a(\tau) = \langle a^*(t)a(t+\tau)\rangle$ — автокорреляционная функция $a(t)$, которая предполагается не зависящей от t.

Интенсивность $I(\boldsymbol{r}) = G(\boldsymbol{r}, \boldsymbol{r}, 0) = G_a(0)$ постоянна во всем пространстве. Временна́я когерентность характеризуется временно́й функцией $G(\boldsymbol{r}, \boldsymbol{r}, \tau) = G_a(\tau)\exp(j\omega_0\tau)$, не зависящей от координат. Комплексная степень когерентности $g(\boldsymbol{r}, \boldsymbol{r}, \tau) = g_a(\tau)\exp(j\omega_0\tau)$, где $g_a(\tau) = G_a(\tau)/G_a(0)$. Ширина функции $|g_a(\tau)| = |g(\boldsymbol{r}, \boldsymbol{r}, \tau)|$, определяемая выражением, аналогичным (11.10), есть время когерентности τ_c. Оно одинаково во всех точках.

Спектральная плотность мощности — Фурье-образ функции $G(\boldsymbol{r}, \boldsymbol{r}, \tau)$ относительно τ. Из (11.32) видно, что $S(\nu)$ равняется Фурье-образу функции $G_a(\tau)$, сдвинутому на частоту ν_0 (в соответствии со свойством частотного сдвига Фурье-образа, описанным в разд. А.1 приложения А). Следовательно, волна имеет одинаковую спектральную плотность мощности во всех точках пространства.

Свойства пространственной когерентности описываются функцией

$$G(\boldsymbol{r}_1, \boldsymbol{r}_2, 0) = G_a\left(\frac{z_1 - z_2}{c}\right)\exp\left[j\omega_0\left(\frac{z_1 - z_2}{c}\right)\right] \tag{11.33}$$

и ее нормированной версией

$$G(\boldsymbol{r}_1, \boldsymbol{r}_2, 0) = G_a\left(\frac{z_1 - z_2}{c}\right)\exp\left[j\omega_0\left(\frac{z_1 - z_2}{c}\right)\right]. \tag{11.34}$$

Если две точки $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$ лежат в одной поперечной плоскости, т. е. $z_1 = z_2$, то $|g(\boldsymbol{r}_1, \boldsymbol{r}_2, 0)| = |g_a(0)| = 1$. Это означает, что флуктуации в точках волнового фронта (плоскости, нормальной к оси z) полностью коррелированы; площадь когерентности в любой поперечной плоскости бесконечна (рис. 11.11). С другой стороны, флуктуации в двух точках, разделенных продольным расстоянием $(z_2 - z_1)$, таким, что $|z_2 - z_1|/c > \tau_c$ или $|z_2 - z_1| > l_c$, где $l_c = c\tau_c$ — когерентная длина, почти не коррелируют друг с другом.

Подводя итог, отметим, что частично когерентная плоская волна пространственно когерентна в любой поперечной плоскости, но лишь частично когерентна в продольном направлении. Аксиальная (продольная) когерентность волны имеет взаимно-однозначное соответствие с временно́й когерентностью. Отношение длины когерентности $l_c = c\tau_c$ к максимальной разности хода l_{max} в системе определяет ту роль, которую играет когерентность. Если $l_c \gg l_{max}$, то волна эффективно полностью когерентна. Длины когерентности ряда источников света приведены в табл. 11.2.

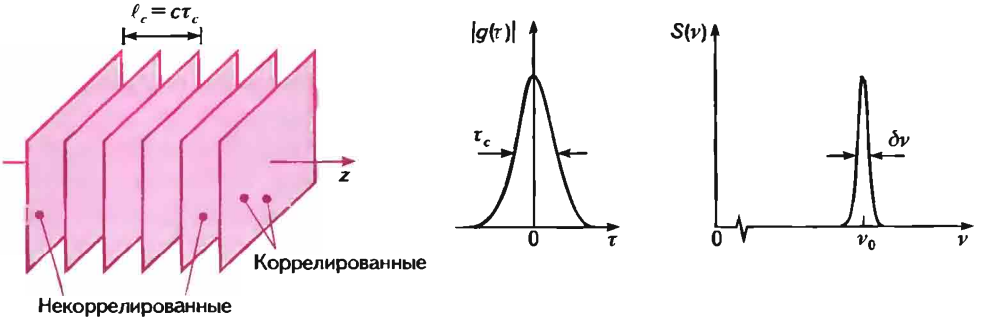


Рис. 11.11. Флуктуации частично когерентной плоской волны в точках любого волнового фронта (поперечной плоскости) полностью коррелированы, в то время как в точках разных волновых фронтов, разделенных расстоянием больше, чем длина когерентности $l_c = c\tau_c$, практически не коррелированы

Частично когерентная сферическая волна

Частично когерентная сферическая волна описывается комплексной волновой функцией (см. подразд. 2.2.2 и 2.6.1)

$$U(\boldsymbol{r},\ t) = \frac{1}{r}a\left(t - \frac{r}{c}\right)\exp\left[j\omega_0\left(t - \frac{r}{c}\right)\right], \tag{11.35}$$

где $a(t)$ — случайная функция. Соответствующая функция взаимной когерентности имеет вид

$$G(\boldsymbol{r}_1,\ \boldsymbol{r}_2,\ \tau) = \frac{1}{r_1 r_2}G_a\left(\tau - \frac{r_2 - r_1}{c}\right)\exp\left[j\omega_0\left(\tau - \frac{r_2 - r_1}{c}\right)\right], \tag{11.36}$$

где $G_a(\tau) = \langle a^*(t)a(t+\tau)\rangle$.

Интенсивность $I(\boldsymbol{r}) = G_a(0)/r^2$ меняется в соответствии с законом обратных квадратов. Время когерентности τ_c есть ширина функции $|g_a(\tau)| = |G_a(\tau)/G_a(0)|$. Оно одинаково во всех точках пространства. То же самое относится и к спектральной плотности мощности. При $\tau = 0$ флуктуации во всех точках волнового фронта (сферы) полностью коррелированы, а флуктуации в двух точках разных волновых фронтов, радиальное расстояние между которыми $|\boldsymbol{r}_2 - \boldsymbol{r}_1| \gg l_c = c\tau_c$, не коррелированы (рис. 11.12).

Рис. 11.12. Частично когерентная сферическая волна имеет полную пространственную когерентность в точках одного волнового фронта, но не в точках на различном расстоянии от центра

Произвольная частично когерентная волна, пропущенная через точечное отверстие, генерирует частично когерентную сферическую волну. Этот про-

цесс вносит когерентность в приходящую волну: точки на любой сфере с центром на отверстии становятся полностью коррелированными. Однако во времени волна остается лишь частично когерентной. *Отверстие сообщает волне пространственную когерентность, но не временну́ю.*

Предположим теперь, что в отверстии установлен очень узкополосный частотный фильтр, так что прошедшая волна становится почти монохроматической. Такая волна приобретет почти полную как пространственную, так и временну́ю когерентность. Временна́я когерентность вносится узкополосным частотным фильтром, а пространственная — точечным отверстием, которое действует как пространственный фильтр. Ценой получения этой идеальной волны является, конечно, потеря оптической энергии, происходящая в процессах пространственной и временно́й фильтрации.

11.2. ИНТЕРФЕРЕНЦИЯ ЧАСТИЧНО КОГЕРЕНТНОГО СВЕТА

Интерференция *когерентного* света обсуждалась в разд. 2.5. Этот раздел посвящен интерференции *частично когерентного* света.

11.2.1. Интерференция двух частично когерентных волн

Статистические свойства двух частично когерентных волн U_1 и U_2 описываются не только их собственными функциями взаимной когерентности, но и мерой того, до какой степени коррелируют их флуктуации между собой. В данной точке $\boldsymbol{r}$ и в данный момент времени t интенсивности обеих волн $I_1 = \langle |U_1|^2 \rangle$, $I_2 = \langle |U_2|^2 \rangle$ их кросс-корреляционная функция описывается статистическим средним $G_{12} = \langle U_1^* U_2 \rangle$, а ее нормированная версия

$$g_{12} = \frac{\langle U_1^* U_2 \rangle}{\sqrt{I_1 I_2}}. \tag{11.37}$$

При суперпозиции этих двух волн средняя интенсивность их суммы

$$\begin{aligned} I &= \langle |U_1 + U_2|^2 \rangle = \langle |U_1|^2 \rangle + \langle |U_2|^2 \rangle + \langle U_1^* U_2 \rangle \langle U_1 U_2^* \rangle = \\ &= I_1 + I_2 + G_{12} + G_{12}^* = I_1 + I_2 + 2\,\mathrm{Re}\{G_{12}\} = I_1 + I_2 + 2\sqrt{I_1 I_2}\,\mathrm{Re}\{g_{12}\}, \end{aligned} \tag{11.38}$$

откуда

$$\boxed{I = I_1 + I_2 + 2\sqrt{I_1 I_2}\,|g_{12}|\cos\varphi,} \tag{11.39}$$

Уравнение интерференции

где $\varphi = \arg\{g_{12}\}$ — фаза g_{12}. Третий член в правой части (11.39) представляет оптическую интерференцию.

Существует два важных предельных случая:

1. Для двух полностью коррелированных волн $g_{12} = \exp(j\varphi)$, $|g_{12}| = 1$, и мы возвращаемся к формуле (2.49) для двух когерентных волн с разностью фаз φ.

2. Для двух некоррелированных волн с $g_{12} = 0$ имеем $I = I_1 + I_2$, так что интерференция отсутствует.

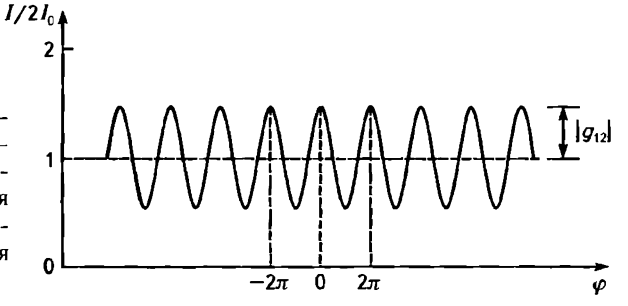


Рис. 11.13. Нормированная интенсивность $I/2I_0$ суммы двух частично когерентных волн одинаковой интенсивности ($I_1 = I_2 = I_0$) как функция фазы φ их нормированной кросс-корреляции g_{12}. Данная синусоидальная картина имеет видность $\mathcal{V} = |g_{12}|$

В общем случае нормированная интенсивность I как функция φ имеет вид синусоидальной картины, как показано на рис. 11.13. Сила интерференции измеряется **видностью** $\mathcal{V}$ (которая также называется глубиной модуляции, или контрастом интерференционной картины)

$$\mathcal{V} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{11.40}$$

где $I_{\max}$ и $I_{\min}$ — соответственно максимальное и минимальное значение, принимаемое I при изменении φ. Поскольку $\cos\varphi$ меняется между -1 и 1, подстановка (11.39) в (11.40) дает

$$\mathcal{V} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}|g_{12}|. \tag{11.41}$$

Таким образом, видность пропорциональна абсолютной величине нормированной кросс-корреляции $|g_{12}|$. В частном случае, когда $I_1 = I_2$, имеем

$$\boxed{\mathcal{V} = |g_{12}|.} \tag{11.42}$$

Видность

Уравнение интерференции (11.39) будет теперь рассмотрено в ряде частных случаев, иллюстрирующих влияние, оказываемое временно́й и пространственной когерентностью на интерференцию частично когерентного света.

11.2.2. Интерференция и временна́я когерентность

Рассмотрим частично когерентную волну $U(t)$ с интенсивностью I_0 и комплексной степенью временной когерентности

$$g(\tau) = \frac{\langle U^*(t)U(t+\tau)\rangle}{I_0}.$$

Если сложить волну $U(t)$ с ее копией $U(t+\tau)$, задержанной на время τ, какой будет интенсивность суперпозиции?

Используя интерференционную формулу (11.38) с

$$U_1 = U(t); \quad U_2 = U(t+\tau); \quad I_1 = I_2 = I_0;$$

$$g_{12} = \frac{U_1^* U_2}{I_0} = \frac{U^*(t)U(t+\tau)}{I_0} = g(\tau),$$

получаем

$$I = 2I_0\left[1 + \mathrm{Re}\{g(\tau)\}\right] = 2I_0\left[1 + |g(\tau)|\cos\varphi(\tau)\right], \tag{11.43}$$

где $\varphi(\tau) = \arg\{g(\tau)\}$. Таким образом, очевидно, что *способность волны интерферировать с собственной копией, задержанной на некоторое время, определяется комплексной степенью временно́й когерентности при данной задержке.*

Осуществить сложение волны со своей задержанной копией можно с помощью светоделителя, дающего две одинаковые волны, одну из которых заставляют пройти больший оптический путь, чем другую, а затем волны совмещают с помощью другого (или того же самого) светоделителя. Это реализуется, например, в интерферометрах Маха—Цендера или Майкельсона (см. рис. 2.24).

Рассмотрим, например, частично когерентную волну, введенную в подразд. 11.1.4 формулой (11.31), у которой комплексная степень временно́й когерентности $g(\tau) = g_a(\tau)\exp(j\omega_0\tau)$. Спектральная ширина волны $\Delta\nu_c = 1/\tau_c$, где τ_c [ширина функции $g_a(\tau)$] — время когерентности. Подставляя это в (11.43), получаем

$$I = 2I_0\left\{1 + |g_a(\tau)|\cos\left[\omega_0\tau + \varphi(\tau)\right]\right\}, \tag{11.44}$$

где $\varphi_a(\tau) = \arg\{g_a(\tau)\}$.

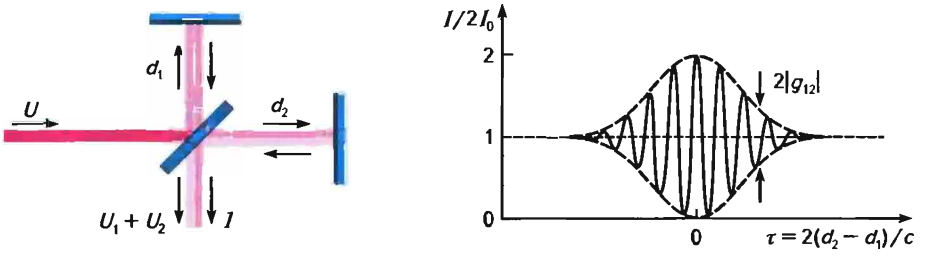


Рис. 11.14. Нормированная интенсивность $I/2I_0$ как функция времени задержки τ при введении частично когерентной плоской волны в интерферометр Майкельсона. Видность определяет модуль комплексной степени временно́й когерентности

Зависимость I от τ, называемая **интерферограммой**, показана на рис. 11.14. При $\Delta\nu_c = 1/\tau_c \ll \nu_0$ функции $|g_a(\tau)|$ и $\varphi_a(\tau)$ меняются медленно по сравнению с периодом $1/\nu_0$. Видность данной интерферограммы в окрестности некоторого времени задержки τ равна $\mathcal{V} = |g(\tau)| = |g_a(\tau)|$. Она имеет наибольшее значение

при $\tau = 0$ и спадает к нулю при $\tau \gg \tau_c$, т. е. когда оптическая разность хода много больше длины когерентности $l_c = c\tau_c$. Для интерферометра Майкельсона, изображенного на рис. 11.14, $\tau = 2(d_2 - d_1)/c$. Интерференция происходит, только когда оптическая разность хода меньше, чем длина когерентности.

Таким образом, модуль комплексной степени временной когерентности волны $|g(\tau)|$ можно измерить путем мониторинга видности интерференционной картины при изменении времени задержки. Фаза $g(\tau)$ может быть измерена по положению максимумов интерференционной картины.

Фурье-спектроскопия

Интересные следствия можно получить, выразив (11.43) через спектральную плотность мощности волны $S(\nu)$. Используя Фурье-соотношение между $G(\tau)$ и $S(\nu)$

$$G(\tau) = I_0 g(\tau) = \int_0^\infty S(\nu)\exp(j2\pi\nu\tau)\,\mathrm{d}\nu, \tag{11.45}$$

подставляя это в (11.43) и замечая, что $S(\nu)$ действительна и что

$$\int_0^\infty S(\nu)\,\mathrm{d}\nu = I_0,$$

получим

$$I = 2\int_0^\infty S(\nu)\left[1 + \cos(2\pi\nu\tau)\right]\mathrm{d}\nu. \tag{11.46}$$

Это выражение можно интерпретировать как взвешенную суперпозицию интерферограмм, производимых каждой из монохроматических компонент волны. Каждая компонента ν дает интерферограмму с периодом $1/\nu$ и видностью единица, однако составная интерферограмма имеет пониженную видность из-за различия периодов.

Уравнение (11.46) предполагает, что спектральная плотность источника света $S(\nu)$ может быть определена по измерению интерферограммы I в зависимости от τ с последующим Фурье-преобразованием результата. Этот метод известен как **Фурье-спектроскопия**.

Оптическая когерентная томография

Оптическая когерентная томография (ОКТ) — интерферометрический метод получения профилей многослойных сред, т. е. измерения коэффициентов отражения и глубины залегания каждой границы раздела слоев. В нем используется источник частично когерентного света с малой длиной когерентности и интерферометр Майкельсона. Как показано на рис. 11.15, копия исходной волны, задержанная подвижным зеркалом, накладывается на совокупность волн, отраженных от многочисленных границ раздела в образце. Информацию о профиле образца несет интерферограмма, представляющая собой интенсив-

ность, измеряемую приемником по мере перемещения подвижного зеркала. Благодаря малой длине когерентности источника, интерферограмма состоит из набора полос с центрами при таких задержках, вносимых подвижным зеркалом, которые совпадают с задержками от отражающих границ.

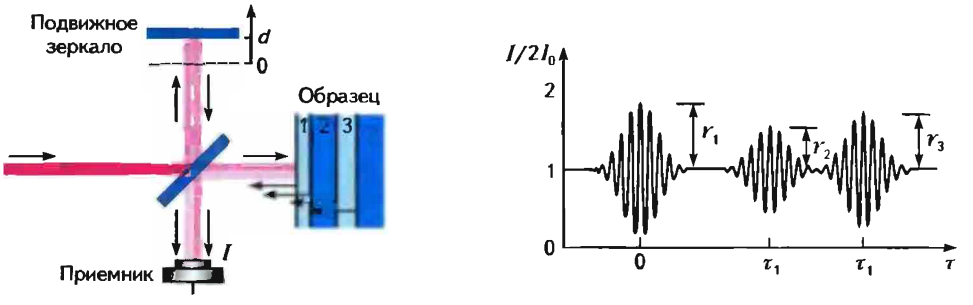


Рис. 11.15. Оптическая когерентная томография

Пусть $U(t-\tau)$ — волна, отраженная от подвижного зеркала с задержкой $\tau = d/c$, и пусть $r_i U(t-\tau_i)$, $i = 1, 2, \ldots$ — волны, отраженные от границ раздела в толще образца, где r_i — амплитудный коэффициент отражения i-й границы; соответствующие времена задержки обозначены τ_i. Для симметричного светоделителя средняя интенсивность тогда будет

$$I(\tau) = \left\langle \left| U(t-\tau) + \sum_i r_i U(t-\tau_i) \right|^2 \right\rangle,$$

что можно записать в нормированном виде как

$$\frac{I}{2I_0} = 1 + \sum_i r_i \operatorname{Re}\{g(\tau - \tau_i)\} + \sum_{ij} r_i r_j^* \operatorname{Re}\{g(\tau_j - \tau_i)\}, \qquad (11.47)$$

поскольку комплексная степень временной когерентности источника характеризуется функцией

$$g(\tau) = \frac{\langle U^*(t)U(t+\tau)\rangle}{\langle U^*(t)U^*(t)\rangle}.$$

Второй член в правой части (11.47) чрезвычайно важен, поскольку он представляет интерференцию между опорной волной от подвижного зеркала и каждой из волн, отраженных от границ раздела в образце. Третий член представляет интерференцию волн, отраженных от образца между собой; поскольку его слагаемые не зависят от перемещения подвижного зеркала и связанной с ним задержки $\tau = d/c$, их можно рассматривать как постоянный фон и не учитывать.

Для источника света с центральной частотой ν_0 имеем $g(\tau) = g_a(\tau)\exp(j\omega_0\tau)$, где ширина профиля $g_a(\tau)$ есть время когерентности τ_c. Тогда уравнение (11.47) приобретает вид

$$\frac{I}{2I_0} \approx 1 + \sum_i r_i \left|g_a(\tau - \tau_i)\right| \cos\left[\omega_0(\tau - \tau_i) + \varphi_a(\tau - \tau_i)\right], \tag{11.48}$$

где $\varphi_a(\tau) = \arg\{g_a(\tau)\}$. Если источник имеет малую длину когерентности, функция $\varphi_a(\tau)$ имеет узкий профиль. Как показано на рис. 11.15, в этом случае отражение от каждой границы в образце генерирует отдельный набор интерференционных полос небольшой протяженности τ_c с центром при соответствующей временно́й задержке. Измерение ОКТ-интерферограммы, таким образом, позволяет определить как толщину каждого из слоев образца, так и коэффициент отражения на каждой границе раздела.

Оптическая когерентная томография проявила себя как эффективный метод получения изображений в клинической медицине и технике.

11.2.3. Интерференция и пространственная когерентность

Влияние пространственной когерентности на интерференцию можно продемонстрировать, рассматривая опыт Юнга с двумя точечными отверстиями, обсуждавшийся в упражнении 2.10 для когерентного света. Частично когерентная оптическая волна $U(\boldsymbol{r}, t)$ освещает непрозрачный экран с двумя точечными отверстиями в точках $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$. Волна имеет функцию взаимной когерентности $G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = \langle U^*(\boldsymbol{r}_1, t)U(\boldsymbol{r}_2, t + \tau)\rangle$ и комплексную степень когерентности $g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$. Интенсивность на отверстиях предполагается одинаковой.

Свет дифрагирует в виде двух сферических волн с центрами на отверстиях. Эти две волны интерферируют, и интенсивность их суммы I измеряется в точке $\boldsymbol{r}$ плоскости наблюдения. Расстояние d от экрана до этой плоскости достаточно велико, чтобы можно было применять параболоидальное приближение. В декартовых координатах (рис. 11.16) $\boldsymbol{r}_1 = (-a, 0, 0)$, $\boldsymbol{r}_2 = (a, 0, 0)$ и $\boldsymbol{r} = (x, 0, d)$. Интенсивность наблюдается как функция x. Важным геометрическим параметром является угол $\theta \approx 2a/d$, под которым видны отверстия с плоскости наблюдения.

В параболоидальном (френелевском) приближении [см. (2.21)] две дифрагирующие сферические волны приближенно связаны с $U(\boldsymbol{r}, t)$ соотношениями

$$U_1(\boldsymbol{r}, t) \propto U\left(\boldsymbol{r}_1, t - \frac{|\boldsymbol{r} - \boldsymbol{r}_1|}{c}\right) \approx U\left(\boldsymbol{r}_1, t - \frac{d + (x + a)^2/2d}{c}\right); \tag{11.49а}$$

$$U_2(\boldsymbol{r}, t) \propto U\left(\boldsymbol{r}_2, t - \frac{|\boldsymbol{r} - \boldsymbol{r}_2|}{c}\right) \approx U\left(\boldsymbol{r}_2, t - \frac{d + (x - a)^2/2d}{c}\right) \tag{11.49б}$$

и имеют примерно одинаковые интенсивности $I_1 = I_2 = I$. Нормированная кросс-корреляция между двумя волнами в точке

$$g_{12} = \frac{\langle U_1^*(\boldsymbol{r},\ t)U_2(\boldsymbol{r},\ t)\rangle}{I_0} = g(\boldsymbol{r}_1,\ \boldsymbol{r}_2,\ \tau_x), \tag{11.50}$$

где τ_x — разность времен задержки двух волн,

$$\tau_x = \frac{|\boldsymbol{r} - \boldsymbol{r}_1| + |\boldsymbol{r} - \boldsymbol{r}_2|}{c} = \frac{(x+a)^2 - (x-a)^2}{2dc} = \frac{2ax}{dc} = \frac{\theta}{c}x. \tag{11.51}$$

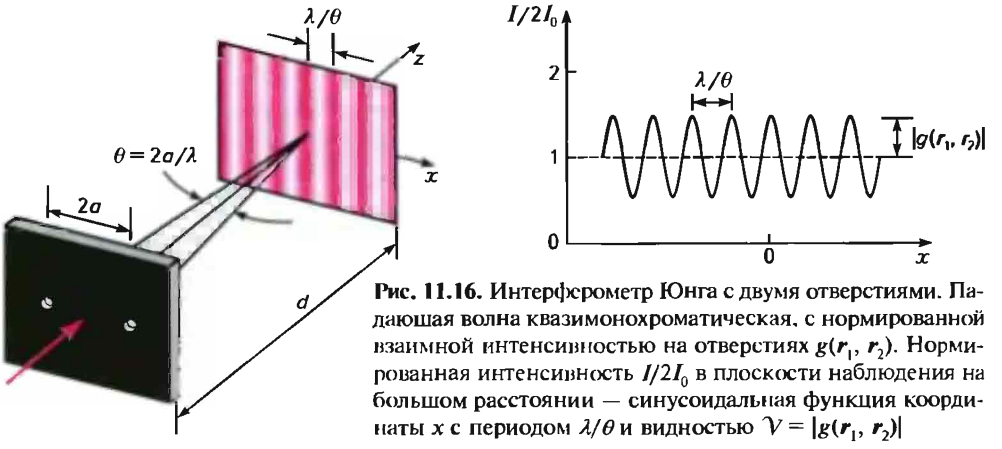


Рис. 11.16. Интерферометр Юнга с двумя отверстиями. Падающая волна квазимонохроматическая, с нормированной взаимной интенсивностью на отверстиях $g(\boldsymbol{r}_1, \boldsymbol{r}_2)$. Нормированная интенсивность $I/2I_0$ в плоскости наблюдения на большом расстоянии — синусоидальная функция координаты x с периодом λ/θ и видностью $\mathcal{V} = |g(\boldsymbol{r}_1, \boldsymbol{r}_2)|$

Подставляя (11.50) в формулу интерференции (11.39), получаем наблюдаемую интенсивность:

$$\boxed{I(x) = 2I_0\left[1 + |g(\boldsymbol{r}_1,\ \boldsymbol{r}_2,\ \tau_x)|\cos\varphi_x\right]}, \tag{11.52}$$

где $\varphi_x = \arg\{g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau_x)\}$. Это уравнение выражает картину наблюдаемой интерференции в зависимости от координаты в плоскости наблюдения через модуль и фазу комплексной степени когерентности на отверстиях при временной задержке $\tau_x = \theta_x/c$.

Квазимонохроматический свет

Если свет квазимонохроматический с центральной частотой $\nu_0 = \omega_0/2\pi$, т. е. если $g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) \approx g(\boldsymbol{r}_1, \boldsymbol{r}_2)\exp(j\omega_0\tau)$, то (11.52) дает

$$\boxed{I(x) = 2I_0\left[1 + \mathcal{V}\cos\left(\frac{2\pi\theta}{\lambda}x + \varphi\right)\right]}, \tag{11.53}$$

где $\lambda = c/\nu_0$; $\mathcal{V} = |g(\boldsymbol{r}_1, \boldsymbol{r}_2)|$; $\tau_x = \theta x/c$; $\varphi_x = \arg\{g(\boldsymbol{r}_1, \boldsymbol{r}_2)\}$. Картина интерференционных полос синусоидальная с пространственным периодом λ/θ и видностью $\mathcal{V}$.

По аналогии с временным случаем, *видность интерференционной картины равна модулю комплексной степени пространственной когерентности на двух отверстиях* (см. рис. 11.16). Положение максимумов зависит от фазы.

Интерференция света от протяженных источников

Если падающая волна в интерферометре Юнга — плоская когерентная волна, бегущая вдоль оси z, $U(\boldsymbol{r}, t) = \exp(-jkz)\exp(j\omega_0 t)$, то $g(\boldsymbol{r}_1, \boldsymbol{r}_2) = 1$, так что $|g(\boldsymbol{r}_1, \boldsymbol{r}_2)| = 1$ и $\arg\{g(\boldsymbol{r}_1, \boldsymbol{r}_2)\} = 0$. Тогда интерференционная картина имеет видность, равную единице, и максимум при $x = 0$. Однако, если освещающая плоская волна падает наклонно под малым углом θ_x к оси z в плоскости x—z, т. е. $U(\boldsymbol{r}, t) \approx \exp[-j(kz + k\theta_x x)]\exp(j\omega_0 t)$, то $g(\boldsymbol{r}_1, \boldsymbol{r}_2) = \exp(-jk\theta_x 2a)$. Видность остается равной единице, $\mathcal{V} = 1$, но наклон приводит к фазовому сдвигу $\varphi = -k\theta_x 2a$, так что интерференционная картина смещается в поперечном направлении на долю периода $2a\theta_x/\lambda$. Когда $\varphi = 2\pi$, картина смещается на один период.

Предположим теперь, что падающий свет представляет собой набор независимых плоских волн от источника, видного под углом θ_s с плоскости отверстий (рис. 11.17). Тогда фазовый сдвиг φ принимает значения, лежащие в интервале

$$\pm 2\pi \frac{\theta_s}{2}\frac{2a}{\lambda} = \pm 2\pi\theta_s \frac{a}{\lambda},$$

а картина полос представляет собой суперпозицию смещенных синусоид. Если, $\theta_s = \lambda/2a$, то φ принимает значения в интервале $\pm\pi$, что достаточно для смазывания интерференционной картины и сведения ее видности к нулю.

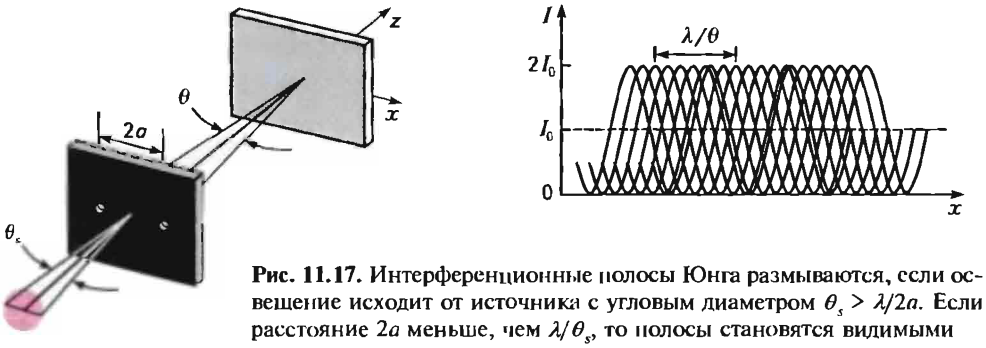


Рис. 11.17. Интерференционные полосы Юнга размываются, если освещение исходит от источника с угловым диаметром $\theta_s > \lambda/2a$. Если расстояние $2a$ меньше, чем λ/θ_s, то полосы становятся видимыми

Мы делаем вывод, что степень пространственной когерентности на двух отверстиях очень мала, когда угловой размер источника $\theta_s = \lambda/2a$ или больше. Следовательно, расстояние

$$\boxed{\rho_c \approx \frac{\lambda}{\theta_s}} \qquad (11.54)$$

Расстояние когерентности

является мерой пространственной когерентности в плоскости экрана; величина A_c — мерой площади когерентности света, испущенного источником с угловым размером θ_s:

$$\boxed{A_c \approx \left(\frac{\lambda}{\theta_s}\right)^2}. \tag{11.55}$$

Например, угловой размер Солнца составляет 0,5°, так что расстояние когерентности для фильтрованного солнечного света с длиной волны λ равно $\rho_c \approx \lambda/\theta_s \approx 115\lambda$. При $\lambda = 0{,}5$ мкм $\rho_c \approx 57{,}5$ мкм.

Более строгий анализ (см. подразд. 11.3.3) показывает, что поперечное расстояние когерентности ρ_c для кругового источника некогерентного света с равномерным распределением интенсивности равно

$$\rho_c = 1{,}22\frac{\lambda}{\theta_s}. \tag{11.56}$$

Влияние спектральной ширины на интерференцию

Наконец, исследуем влияние спектральной ширины на интерференцию в схеме Юнга с двумя отверстиями. Предположим, что спектральная плотность мощности падающей волны имеет вид узкого пика шириной $\Delta\nu_c$ с центром при ν_0 и что $\Delta\nu_c \ll \nu_0$. Тогда комплексная степень когерентности принимает вид

$$g(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = g_a(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)\exp(j\omega_0\tau), \tag{11.57}$$

где $g_a(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$ — медленно меняющаяся функция (по сравнению с периодом $1/\nu_0$). Подставляя (11.57) в (11.52), получаем

$$\boxed{I(x) = 2I_0\left[1 + \mathcal{V}_x \cos\left(\frac{2\pi\theta}{\bar{\lambda}}x + \varphi_x\right)\right]}, \tag{11.58}$$

где $\mathcal{V}_x = |g_a(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau_x)|$; $\varphi_x = \arg\{g_a(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau_x)\}$; $\tau_x = \theta x/c$ и $\bar{\lambda} = c/\nu_0$.

Таким образом, интерференционная картина синусоидальная с периодом $\bar{\lambda}/\theta$, но с переменной видностью $\mathcal{V}_x$ и фазой φ_x, которые равны модулю и фазе комплексной степени когерентности на двух отверстиях, соответственно, при времени задержки $\tau_x = \theta x/c$. Если $|g_a(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)| = 1$ при $\tau = 0$, она убывает с ростом τ и исчезает при $\tau \gg \tau_c$. Соответственно этому видность $\mathcal{V}_x$, равная единице при $x = 0$, убывает с ростом x и исчезает при $x \gg x_c = c\tau_c/\theta$. Таким образом, интерференционная картина видна на протяжении

$$x_c = \frac{l_c}{\theta}, \tag{11.59}$$

где l_c — длина когерентности; θ — угол, под которым видны отверстия с плоскости наблюдения (рис. 11.18).

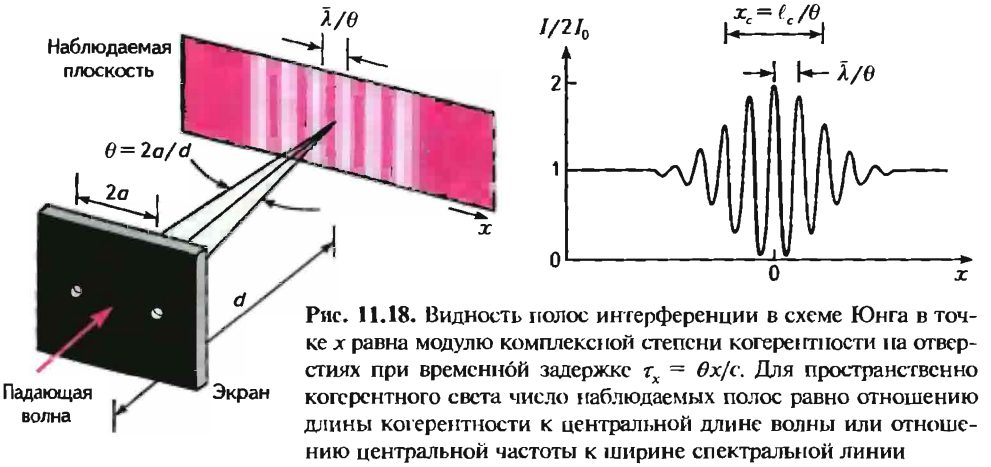


Рис. 11.18. Видность полос интерференции в схеме Юнга в точке x равна модулю комплексной степени когерентности на отверстиях при временной задержке $\tau_x = \theta x/c$. Для пространственно когерентного света число наблюдаемых полос равно отношению длины когерентности к центральной длине волны или отношению центральной частоты к ширине спектральной линии

Таким образом, число наблюдаемых полос

$$\frac{x_c}{\bar{\lambda}/\theta} = \frac{l_c}{\bar{\lambda}} = \frac{c\tau_c}{\bar{\lambda}} = \frac{\nu_0}{\Delta\nu_c}.$$

Оно равно отношению $l_c/\bar{\lambda}$ длины когерентности к центральной длине волны или отношению $\nu_0/\Delta\nu_c$ центральной частоты к ширине спектральной линии. Ясно, что если $|g_a(\boldsymbol{r}_1, \boldsymbol{r}_2, 0)| < 1$, т. е. источник не является пространственно когерентным, то видность еще больше понизится и число видимых полос будет еще меньше.

11.3. ПРОХОЖДЕНИЕ ЧАСТИЧНО КОГЕРЕНТНОГО СВЕТА ЧЕРЕЗ ОПТИЧЕСКИЕ СИСТЕМЫ

Прохождение когерентного света через тонкие элементы оптических систем, диафрагмы и свободное пространство обсуждалось в гл. 2 и 4. В данном разделе мы преследуем ту же цель для квазимонохроматического частично когерентного света. Мы предполагаем, что спектральная ширина достаточно мала, так что длина когерентности намного больше оптических разностей хода в системе. Функцию взаимной когерентности тогда можно аппроксимировать выражением $G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) \approx G(\boldsymbol{r}_1, \boldsymbol{r}_2) \exp(j\omega_0\tau)$, где $G(\boldsymbol{r}_1, \boldsymbol{r}_2)$ — взаимная интенсивность; ν_0 — центральная частота.

Заметим, что законы прохождения света, применявшиеся для детерминированной функции $U(\boldsymbol{r})$, представляющей когерентный свет, применимы также и к случайной функции $U(\boldsymbol{r})$, представляющей частично когерентный свет.

Однако для частично когерентного света предметом нашего интереса будут законы, которым подчиняются статистические средние: интенсивность $I(\boldsymbol{r})$ и взаимная интенсивность $G(\boldsymbol{r}_1, \boldsymbol{r}_2)$.

11.3.1. Распространение частично когерентного света

Прохождение через тонкие оптические элементы

Когда частично когерентная волна проходит через тонкий элемент оптической системы, характеризуемый амплитудным коэффициентом пропускания $t(x, y)$, падающая и прошедшая волны связаны соотношением $U_2(\boldsymbol{r}) = t(\boldsymbol{r})U_1(\boldsymbol{r})$, где $\boldsymbol{r} = (x, y)$ — координаты точки в плоскости оптического элемента (рис. 11.19). Используя определение взаимной интенсивности $G(\boldsymbol{r}_1, \boldsymbol{r}_2) = \langle U^*(\boldsymbol{r}_1)U(\boldsymbol{r}_2)\rangle$, получаем

$$G_2(\boldsymbol{r}_1, \boldsymbol{r}_2) = t^*(\boldsymbol{r}_1)t(\boldsymbol{r}_2)G_1(\boldsymbol{r}_1, \boldsymbol{r}_2), \tag{11.60}$$

где $G_1(\boldsymbol{r}_1, \boldsymbol{r}_2)$ и $G_2(\boldsymbol{r}_1, \boldsymbol{r}_2)$ — взаимные интенсивности падающего и прошедшего света соответственно.

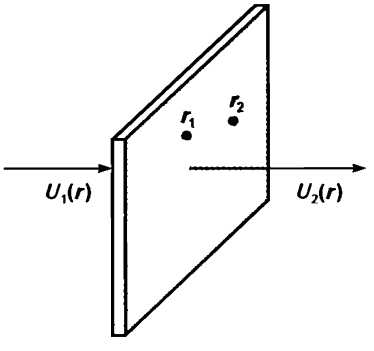


Рис. 11.19. Модуль степени пространственной когерентности не меняется при прохождении тонкого оптического элемента

Поскольку интенсивность в точке $\boldsymbol{r}$ равна взаимной интенсивности при $\boldsymbol{r}_1 = \boldsymbol{r}_2 = \boldsymbol{r}$,

$$I_2(\boldsymbol{r}) = |t(\boldsymbol{r})|^2 I_1(\boldsymbol{r}). \tag{11.61}$$

Следовательно, нормированные взаимные интенсивности, определенные формулой (11.27), удовлетворяют соотношению

$$|g_2(\boldsymbol{r}_1, \boldsymbol{r}_2)| = |g_1(\boldsymbol{r}_1, \boldsymbol{r}_2)|. \tag{11.62}$$

Хотя при прохождении через тонкий оптический элемент может измениться интенсивность частично когерентного света, он не может изменить модуля степени пространственной когерентности. Естественно, если комплексный амплитудный коэффициент пропускания элемента сам является случайной функцией, то когерентность света при прохождении изменится.

Прохождение через произвольную оптическую систему

Теперь рассмотрим произвольную оптическую систему, включающую прохождение участков свободного пространства или толстых оптических элементов. В гл. 4 было показано, что комплексная амплитуда $U_2(\boldsymbol{r})$ в точке $\boldsymbol{r} = (x, y)$ выходной плоскости такой системы в общем случае представляет собой интеграл, выражающий взвешенную суперпозицию вкладов от комплексной амплитуды $U_1(\boldsymbol{r}')$ в точках $\boldsymbol{r}' = (x', y')$ входной плоскости (рис. 11.20):

$$U_2(\boldsymbol{r}) = \int h(\boldsymbol{r};\ \boldsymbol{r}')U_1(\boldsymbol{r}')\mathrm{d}\boldsymbol{r}', \tag{11.63}$$

где $h(\boldsymbol{r};\ \boldsymbol{r}')$ — функция отклика системы на импульсное воздействие. Интеграл в (11.63) — двойной интеграл по координатам точек $\boldsymbol{r}' = (x', y')$ всей входной плоскости.

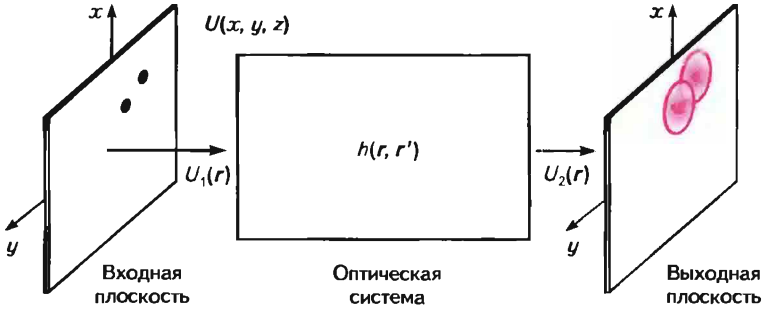


Рис. 11.20. Оптическая система характеризуется своей функцией отклика на импульсное воздействие $h(\boldsymbol{r};\ \boldsymbol{r}')$

Чтобы перевести это соотношение между случайными функциями $U_2(\boldsymbol{r})$ и $U_1(\boldsymbol{r})$ в соотношение между их взаимными интенсивностями, подставим (11.63) в определение $G_2(\boldsymbol{r}_1,\ \boldsymbol{r}_2) = \langle U_2^*(\boldsymbol{r}_1)U_2(\boldsymbol{r}_2)\rangle$ и, используя определение $G_1(\boldsymbol{r}_1,\ \boldsymbol{r}_2) = \langle U_1^*(\boldsymbol{r}_1)U_1(\boldsymbol{r}_2)\rangle$, получим

$$\boxed{G_2(\boldsymbol{r}_1,\ \boldsymbol{r}_2) = \iint h^*(\boldsymbol{r}_1;\ \boldsymbol{r}_1')h(\boldsymbol{r}_2;\ \boldsymbol{r}_2')G_1(\boldsymbol{r}_1',\ \boldsymbol{r}_2')\mathrm{d}\boldsymbol{r}_1'\mathrm{d}\boldsymbol{r}_2'.} \tag{11.64}$$

Взаимная интенсивность

Если взаимная интенсивность $G_1(\boldsymbol{r}_1,\ \boldsymbol{r}_2)$ входящего света и функция отклика системы на импульсное воздействие $h(\boldsymbol{r};\ \boldsymbol{r}')$ известны, то взаимную интенсивность света на выходе $G_2(\boldsymbol{r}_1,\ \boldsymbol{r}_2)$ можно определить посредством вычисления интегралов в (11.64).

Интенсивность света на выходе получается с помощью определения $I_2(\boldsymbol{r}) = G_2(\boldsymbol{r},\ \boldsymbol{r})$, тогда (11.64) сводится к

$$\boxed{I_2(\boldsymbol{r}) = \iint h^*(\boldsymbol{r};\ \boldsymbol{r}_1')h(\boldsymbol{r};\ \boldsymbol{r}_2')G_1(\boldsymbol{r}_1',\ \boldsymbol{r}_2')\mathrm{d}\boldsymbol{r}_1'\mathrm{d}\boldsymbol{r}_2'.} \tag{11.65}$$

Интенсивность изображения

Для определения интенсивности света на выходе нужно знать взаимную интенсивность входящего света. *Знание интенсивности входящего света* $I_1(\boldsymbol{r})$ *недостаточно для определения интенсивности света на выходе* $I_2(\boldsymbol{r})$.

11.3.2. Формирование изображений в некогерентном свете

Рассмотрим частный случай, когда свет на входе некогерентный. Взаимная интенсивность $G_1(\boldsymbol{r}_1, \boldsymbol{r}_2)$ исчезает, когда $\boldsymbol{r}_2$ хотя бы немного отличается от $\boldsymbol{r}_1$, так что расстояние когерентности намного меньше, чем другие характерные размеры в системе (например, разрешение в системе для получения изображений). Тогда взаимную интенсивность можно записать в виде

$$G_1(\boldsymbol{r}_1, \boldsymbol{r}_2) \approx \sigma\sqrt{I_1(\boldsymbol{r}_1) I_1(\boldsymbol{r}_2)}\, g(\boldsymbol{r}_1 - \boldsymbol{r}_2),$$

где $g(\boldsymbol{r}_1 - \boldsymbol{r}_2)$ — очень узкая функция. Когда $G_1(\boldsymbol{r}_1, \boldsymbol{r}_2)$ появляется под знаком интеграла в (11.64) или (11.65), удобно заменить $g(\boldsymbol{r}_1 - \boldsymbol{r}_2)$ дельта-функцией $g(\boldsymbol{r}_1 - \boldsymbol{r}_2) = \sigma\delta(\boldsymbol{r}_1 - \boldsymbol{r}_2)$, где

$$\sigma = \int g(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r},$$

так что

$$G_1(\boldsymbol{r}_1, \boldsymbol{r}_2) \approx \sigma\sqrt{I_1(\boldsymbol{r}_1) I_1(\boldsymbol{r}_2)}\, \delta(\boldsymbol{r}_1 - \boldsymbol{r}_2). \tag{11.66}$$

Поскольку взаимная интенсивность должна оставаться конечной, а $\delta(0) \to \infty$, это выражение, очевидно, не вполне корректное. Оно пригодно только для вычисления интегралов, таких как в формуле (11.65). При подстановке (11.66) в (11.65) дельта-функция сводит двойной интеграл к однократному, и мы получаем

$$\boxed{I_2(\boldsymbol{r}) = \int I_1(\boldsymbol{r}')\, h_i(\boldsymbol{r};\ \boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}',} \tag{11.67}$$

Уравнение изображения (некогерентное освещение)

где

$$\boxed{h_i(\boldsymbol{r};\ \boldsymbol{r}') = \sigma\left|h(\boldsymbol{r};\ \boldsymbol{r}')\right|^2,} \tag{11.68}$$

Функция отклика на импульсное воздействие (некогерентное освещение)

При этих условиях соотношение между интенсивностями во входной и выходной плоскостях описывает линейную систему с функцией отклика на импульсное воздействие $h_i(\boldsymbol{r};\ \boldsymbol{r}')$, которая также называется **функцией рассеяния точки**. Таким образом, когда входной свет полностью некогерентен, интенсивность света в каждой точке $\boldsymbol{r}$ выходной плоскости является взвешенной суперпозицией вкладов от интенсивностей во множестве точек $\boldsymbol{r}'$ входной плоскости; интерференция не происходит, и интенсивности просто складываются (рис. 11.21). Необходимо подчеркнуть принципиальное отличие от полностью когерентной

системы, где интегралом суперпозиции типа (11.63) связаны комплексные амплитуды, а не интенсивности.

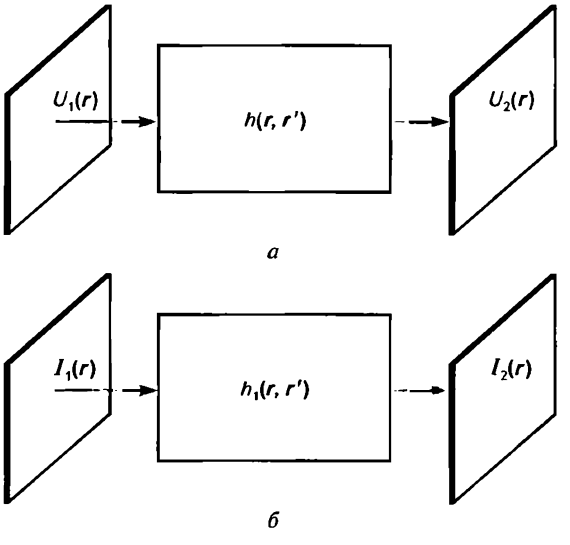


Рис. 11.21. Комплексные *амплитуды* света на входной и выходной плоскостях оптической системы, освещаемой *когерентным* светом, связаны линейной системой с функцией отклика на импульсное воздействие $h(\boldsymbol{r}; \boldsymbol{r}')$ (*а*). *Интенсивности* света на входной и выходной плоскостях оптической системы, освещаемой *некогерентным* светом, связаны линейной системой с функцией отклика на импульсное воздействие $h_i(\boldsymbol{r}; \boldsymbol{r}') = \sigma|h(\boldsymbol{r}; \boldsymbol{r}')|^2$ (*б*)

В некоторых оптических системах функция отклика на импульсное воздействие зависит от $(\boldsymbol{r} - \boldsymbol{r}')$. Обозначим ее $h(\boldsymbol{r} - \boldsymbol{r}')$. Такая система называется **изопланарной** (см. приложение Б). В этом случае $h_i(\boldsymbol{r}; \boldsymbol{r}') = h_i(\boldsymbol{r} - \boldsymbol{r}')$. Интегралы в (11.63) и (11.67) тогда являются двумерными свертками, и системы можно описывать передаточными функциями $H(\nu_x, \nu_y)$ и $H_i(\nu_x, \nu_y)$, представляющими собой Фурье-образы функций $h(\boldsymbol{r}) = h(x, y)$ и $h_i(\boldsymbol{r}) = h_i(x, y)$ соответственно.

В качестве примера применим полученные соотношения к системе, дающей изображение. В подразд. 4.4.3 было показано, что при когерентном освещении функция отклика на импульсное воздействие однолинзовой сфокусированной изображающей системы, показанной на рис. 11.22, в приближении Френеля дается выражением

$$h(\boldsymbol{r}) \propto P\left(\frac{x}{\lambda d_2}, \frac{y}{\lambda d_2}\right), \tag{11.69}$$

где $P(\nu_x, \nu_y)$ — Фурье-образ функции зрачка $p(x, y)$; d_2 — расстояние от линзы до плоскости изображения. Функция зрачка равна единице в пределах апертуры и нулю в остальных точках.

Когда освещение квазимонохроматическое и пространственно некогерентное, интенсивности света в плоскостях предмета и изображения линейно связаны системой с функцией отклика на импульсное воздействие

$$h_i(\boldsymbol{r}) = \sigma|h(\boldsymbol{r})|^2 \propto \left|P\left(\frac{x}{\lambda d_2}, \frac{y}{\lambda d_2}\right)\right|^2 \tag{11.70}$$

где λ — длина волны, соответствующая центральной частоте ν_0.

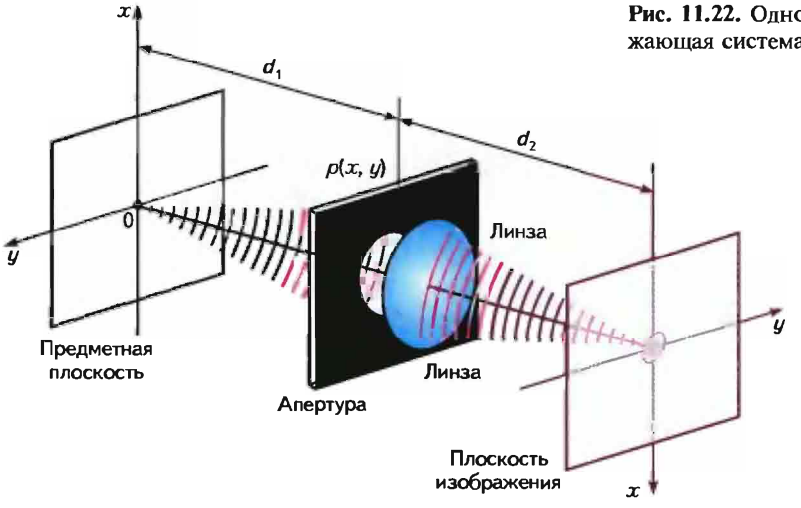


Рис. 11.22. Однолинзовая изображающая система

Пример 11.2

Изображающая система с круговой апертурой

Если апертура имеет форму круга радиусом a, то функция зрачка $p(x, y) = 1$ внутри круга и 0 вне его. Фурье-образ этой функции

$$P(\nu_x, \nu_y) = \frac{aJ_1(2\pi\nu_\rho a)}{\nu_\rho}; \quad \nu_\rho = \sqrt{\nu_x^2 + \nu_y^2}, \tag{11.71}$$

где J_1 — функция Бесселя (см. приложение А, разд. А.3). Функция отклика на импульсное воздействие когерентной системы получается подстановкой в (11.36)

$$h(x, y) \propto \frac{J_1(2\pi\nu_s\rho)}{\pi\nu_s\rho}; \quad \rho = \sqrt{x^2 + y^2}, \tag{11.72}$$

где

$$\nu_s = \frac{\theta}{2\lambda}; \quad \theta = \frac{2a}{d_2}. \tag{11.73}$$

Тогда для некогерентного освещения функция отклика на импульсное воздействие

$$h_i(x, y) \propto \left[\frac{J_1(2\pi\nu_s\rho)}{\pi\nu_s\rho}\right]^2 \tag{11.74}$$

Графики функций отклика $h(x, y)$ и $h_i(x, y)$ показаны на рис. 11.23. Обе функции достигают своего первого нуля, когда $2\pi\nu_s\rho = 3{,}832$, или

$$\rho = \rho_s \approx \frac{3{,}832}{2\pi\nu_s} = \frac{3{,}832\lambda}{\pi\theta},$$

откуда

$$\boxed{\rho_s \approx 1{,}22\frac{\lambda}{\theta}.} \qquad (11.75)$$

Разрешение двух точек

Таким образом, изображение точки (импульса) во входной плоскости представляет собой пятно с интенсивностью $h_i(x, y)$ радиусом ρ_s. Когда входное распределение состоит из двух точек (импульсов), разделенных расстоянием, изображение одной точки исчезает в центре изображения другой точки. Таким образом, расстояние ρ_s является мерой разрешения изображающей системы.

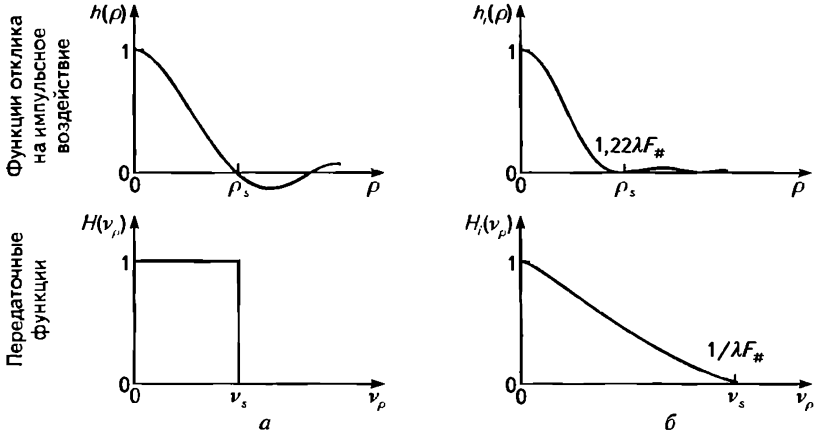


Рис. 11.23. Функции отклика на импульсное воздействие и передаточные функции для однолинзовой сфокусированной дифракционно-ограниченной изображающей системы с круговой апертурой и F — числом $F_\#$ при когерентном (*а*) и некогерентном (*б*) освещении

Передаточные функции линейных систем (см. приложение Б) с функциями отклика на импульсное воздействие $h(x, y)$ и $h_i(x, y)$ получаются преобразованием Фурье (см. приложение А)

$$H(\nu_x, \nu_y) = \begin{cases} 1, & \nu_\rho < \nu_s; \\ 0, & \nu_\rho \geq \nu_s; \end{cases} \qquad (11.76)$$

$$H_i(\nu_x, \nu_y) = \begin{cases} \dfrac{2}{\pi}\left[\arccos\dfrac{\nu_\rho}{2\nu_s} - \dfrac{\nu_\rho}{2\nu_s}\sqrt{1-\left(\dfrac{\nu_\rho}{2\nu_s}\right)^2}\right], & \nu_\rho < 2\nu_s; \\ 0, & \nu_\rho \geq 2\nu_s, \end{cases} \qquad (11.77)$$

где

$$\nu_\rho = \sqrt{\nu_x^2 + \nu_y^2}.$$

Обе функции нормированы так, что их значения при $\nu_r = 0$ равны 1. Графики эти функции показаны на рис. 11.23. Для когерентного освещения передаточная функция постоянна и имеет частоту отсечки $\nu_s = \theta/2\lambda$ штрихов/мм. Для некогерентного освещения передаточная функция спадает приблизительно линейно с ростом пространственной частоты и имеет частоту отсечки $2\nu_s = \theta/\lambda$ штрихов/мм.

Если предмет помещен на бесконечность, т. е. $d_1 = \infty$, то $d_2 = f$, т. е. изображение находится в фокальной плоскости линзы. Тогда угол $\theta = 2a/f$ равен обратному F-числу линзы, $F_\# = f/2a$. Частоты отсечки ν_s и $2\nu_s$ связаны с F — числом линзы формулой

$$\text{Частота отсечки (штрих/мм)} = \begin{cases} \dfrac{1}{2\lambda F_\#} & \text{(когерентное освещение);} \\ \dfrac{1}{\lambda F_\#} & \text{(некогерентное освещение).} \end{cases} \tag{11.78}$$

Не следует делать ложного заключения, что некогерентное освещение превосходит когерентное, так как оно имеет удвоенную полосу пространственных частот. Передаточные функции обеих систем нельзя сравнивать непосредственно, поскольку одна описывает отображение комплексной амплитуды, а другая — интенсивности.

11.3.3. Приобретение пространственной когерентности при распространении

Уравнение (11.64) описывает изменение взаимной интенсивности при распространении света через оптическую систему с функцией отклика на импульсное воздействие $h(\boldsymbol{r}; \boldsymbol{r}')$. Если падающий свет является некогерентным, взаимную интенсивность $G_1(\boldsymbol{r}_1, \boldsymbol{r}_2)$ можно заменить на $\sigma\sqrt{I_1(\boldsymbol{r}_1)I_1(\boldsymbol{r}_2)}\,\delta(\boldsymbol{r}_1 - \boldsymbol{r}_2)$ и подставить в двойной интеграл (11.64), что дает однократный интеграл

$$\boxed{G_2(\boldsymbol{r}_1, \boldsymbol{r}_2) = \sigma\int h^*(\boldsymbol{r}_1; \boldsymbol{r})h(\boldsymbol{r}_2; \boldsymbol{r})I_1(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}.} \tag{11.79}$$

Взаимная интенсивность

Очевидно, что принимаемый свет не остается некогерентным. В общем случае свет приобретает пространственную когерентность просто за счет распространения. Это не удивительно. Хотя флуктуации света в различных точках входной плоскости не коррелированы, излучение от каждой точки расходится и перекрывается с излучением от соседних точек. Свет, достигающий двух точек в плоскости выхода, приходит от многих точек плоскости входа, причем некоторые из них являются общими (рис. 11.24). Эти общие вклады создают частичную корреляцию между флуктуациями в точках выходной плоскости.

Эти явления напоминают прохождение некоррелированного по времени сигнала (белого шума) через фильтр нижних частот. Фильтр сглаживает функ-

цию и сужает ее спектральную полосу, так что время когерентности увеличивается, и сигнал более не является некоррелированным. Распространение света через оптическую систему является разновидностью пространственной фильтрации, которая уменьшает ширину полосы пространственных частот и, следовательно, увеличивает площадь когерентности.

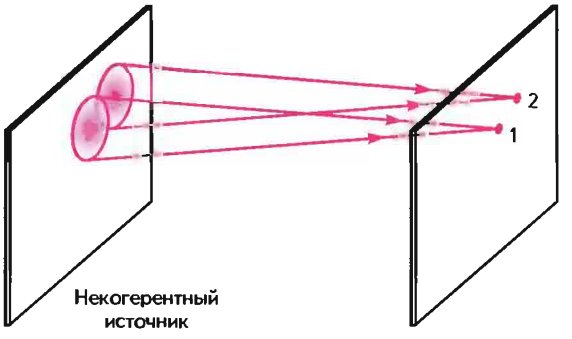


Рис. 11.24. Приобретение когерентности при распространении является результатом расплывания света. Хотя свет источника полностью некоррелирован, флуктуации света в точках 1 и 2 имеют общее происхождение (заштрихованная область) и, следовательно, частично коррелированы

Теорема Ван Циттерта—Цернике

Существует математическая аналогия между приобретением когерентности исходно некогерентным светом при распространении через оптическую систему и изменением амплитуды когерентного света, проходящего через ту же систему. В соответствии с (11.79), если точка наблюдения $\boldsymbol{r}_1$ фиксирована, например в начале координат $\mathbf{0}$, а взаимная интенсивность $G_2(\mathbf{0}, \boldsymbol{r}_2)$ исследуется как функция $\boldsymbol{r}_2$, то

$$G_2(\mathbf{0},\ \boldsymbol{r}_2) = \sigma \int h^*(\mathbf{0};\ \boldsymbol{r}) h(\boldsymbol{r}_2;\ \boldsymbol{r}) I_1(\boldsymbol{r}) \mathrm{d}\boldsymbol{r}. \tag{11.80}$$

Обозначая $U_2(\boldsymbol{r}_2) = G_2(\mathbf{0}, \boldsymbol{r}_2)$ и $U_1(\boldsymbol{r}) = \sigma h^*(\mathbf{0}; \boldsymbol{r}) I_1(\boldsymbol{r})$, можно переписать (11.80) в привычном виде

$$U_2(\boldsymbol{r}_2) = \int h(\boldsymbol{r}_2;\ \boldsymbol{r}) U_1(\boldsymbol{r}) \mathrm{d}\boldsymbol{r}, \tag{11.81}$$

что точно совпадает с интегралом (11.63), описывающим распространение когерентного света. Таким образом, наблюдаемая взаимная интенсивность $G(\mathbf{0}, \boldsymbol{r}_2)$ на выходе оптической системы, облучаемой некогерентным светом на входе, математически идентична наблюдаемой комплексной амплитуде при облучении той же системы когерентной волной с комплексной амплитудой $U_1(\boldsymbol{r}) = \sigma h^*(\mathbf{0}; \boldsymbol{r}) I_1(\boldsymbol{r})$.

В качестве примера предположим, что некогерентная волна на входе имеет равномерное распределение интенсивности в пределах апертуры $p(\boldsymbol{r})$ [$p(\boldsymbol{r}) = 1$ в пределах апертуры и 0 в остальных точках], т. е. $I_1(\boldsymbol{r}) = p(\boldsymbol{r})$. Предположим, что оптическая система — свободное пространство, т. е.

$$h(\boldsymbol{r}';\ \boldsymbol{r}) = \frac{\exp\left(-jk|\boldsymbol{r}' - \boldsymbol{r}|\right)}{|\boldsymbol{r}' - \boldsymbol{r}|}.$$

Взаимная интенсивность $G_2(\mathbf{0}, \boldsymbol{r}_2)$ тогда идентична амплитуде $U_2(\boldsymbol{r}_2)$, получаемой при прохождении через ту же систему когерентной волны с амплитудой

$$U_1(\boldsymbol{r}) = \sigma h^*(\mathbf{0};\ \boldsymbol{r})\, p(\boldsymbol{r}) = \sigma(\boldsymbol{r})\frac{\exp(jkr)}{r}.$$

Это сферическая волна, сходящаяся в точку **0** выходной плоскости и прошедшая через апертуру.

Аналогия между дифракцией когерентного света и приобретением пространственной когерентности некогерентным светом, проходящим через ту же систему, известна как **теорема Ван Циттерта—Цернике**.

Приобретение когерентности в свободном пространстве

Рассмотрим в качестве оптической системы свободное пространство между двумя параллельными плоскостями, расположенными на расстоянии d друг от друга (рис. 11.25). Свет на входной плоскости квазимонохроматический, пространственно некогерентный и имеет интенсивность $I(x, y)$ в пределах конечной области. Расстояние d достаточно велико, так что для интересующих нас точек выходной плоскости справедливо приближение Фраунгофера. При этих условиях функция отклика на импульсное воздействие оптической системы описывается дифракционной формулой Фраунгофера [см. (4.4)]

$$h(\boldsymbol{r};\ \boldsymbol{r}') = h_0 \exp\left(-j\pi\frac{x^2+y^2}{\lambda d}\right)\exp\left(j2\pi\frac{xx'+yy'}{\lambda d}\right), \tag{11.82}$$

где $\boldsymbol{r} = (x, y, d)$ и $\boldsymbol{r}' = (x', y', 0)$ — координаты точек в выходной и входной плоскостях соответственно; h_0 — постоянная,

$$h_0 = \frac{j}{\lambda d}\exp\left(\frac{-j2\pi d}{\lambda}\right).$$

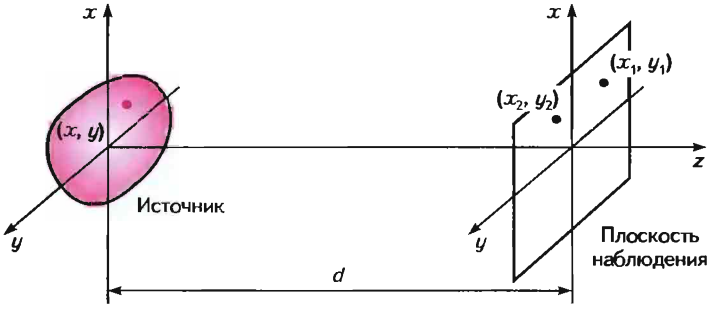


Рис. 11.25. Излучение некогерентного источника в свободном пространстве

Для определения функции взаимной когерентности $G(x_1, y_1; x_2, y_2)$ в двух точках (x_1, y_1) и (x_2, y_2) выходной плоскости подставим (11.82) в (11.79) и получим

$$\begin{aligned} &\left|G(x_1, y_1; x_2, y_2)\right| = \\ &= \sigma_1 \left| \int_{-\infty}^{\infty}\int \exp\left\{ j\frac{2\pi}{\lambda d}\left[(x_2 - x_1)x + (y_2 - y_1)y\right]\right\} I(x, y)\,dx\,dy \right|, \end{aligned} \tag{11.83}$$

где σ_1 — еще одна постоянная,

$$\sigma_1 = \sigma\left|h_0\right|^2 = \frac{\sigma}{\lambda^2 d^2}.$$

При заданной $I(x, y)$ можно легко выразить $|G(x_1, y_1; x_2, y_2)|$ через двумерный Фурье-образ $I(x, y)$

$$J(\nu_x, \nu_y) = \int_{-\infty}^{\infty}\int \exp\left[j2\pi(\nu_x x + \nu_y y)\right] I(x, y)\,dx\,dy, \tag{11.84}$$

взятый при

$$\nu_x = \frac{x_2 - x_1}{\lambda d}; \quad \nu_y = \frac{y_2 - y_1}{\lambda d}.$$

Модуль соответствующей нормированной взаимной интенсивности равен

$$\boxed{\left|g(x_1, y_1; x_2, y_2)\right| = \frac{\left|J\left[(x_2 - x_1)/\lambda d,\ (y_2 - y_1)/\lambda d\right]\right|}{J(0, 0)}}, \tag{11.85}$$

Это Фурье-соответствие между профилем интенсивности некогерентного источника и степенью пространственной когерентности поля в его дальней зоне аналогично Фурье-соотношению между амплитудами когерентного света в плоскостях входа и выхода (см. подразд. 4.2.1). Эта аналогия вполне ожидаема на основании теоремы Ван Циттерта—Цернике.

Из (11.85) вытекают выдающиеся следствия. Если площадь источника, т. е. пространственная протяженность $I(x, y)$, мала, то ее Фурье-образ $J(\nu_x, \nu_y)$ широк, так что взаимная интенсивность в плоскости выхода отлична от нуля в обширной области, и площадь когерентности велика. В предельном случае света, исходящего из одной точки плоскости входа, площадь когерентности становится бесконечной, а поле излучения — полностью когерентным пространственно. Это подтверждает результаты наших предыдущих обсуждений в подразд. 11.1.4, касающихся когерентности сферических волн. С другой стороны, если входной когерентный свет исходит от источника большой протяженности, то после распространения свет имеет малую площадь когерентности.

Пример 11.3

Излучение от некогерентного кругового источника

Для входного света с равномерным распределением интенсивности внутри круговой апертуры формула (11.85) дает

$$|g(x_1, y_1; x_2, y_2)| = \left|\frac{2J_1(\pi\rho\theta_s/\lambda)}{\pi\rho\theta_s/\lambda}\right|, \tag{11.86}$$

где ρ — расстояние между двумя точками,

$$\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2};$$

$\theta_s = 2a/d$ — угловой размер источника; J_1 — функция Бесселя. График этого соотношения показан на рис. 11.26. Функция Бесселя достигает своего первого нуля, когда ее аргумент равен 3,832. Следовательно, мы можем определить площадь когерентности как площадь круга радиусом $\rho_c = 3{,}832(\lambda/\pi\theta_s)$, так что

$$\boxed{\rho_c = 1{,}22\frac{\lambda}{\theta_s}.} \tag{11.87}$$

Расстояние когерентности

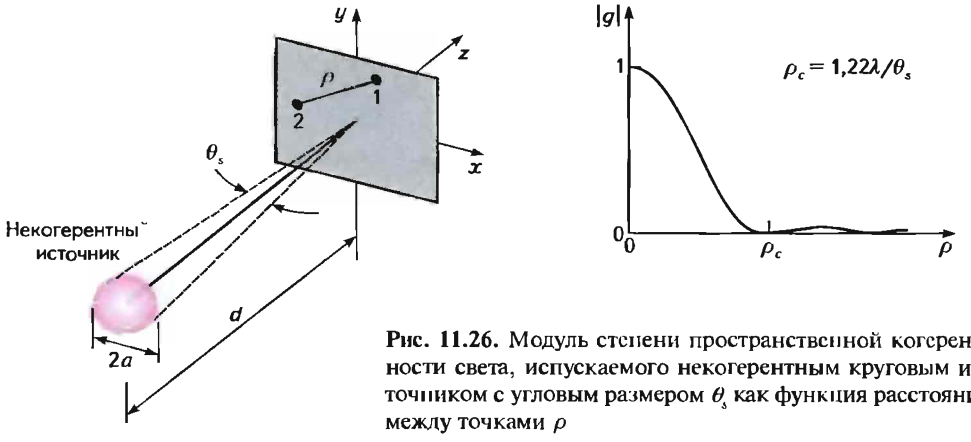


Рис. 11.26. Модуль степени пространственной когерентности света, испускаемого некогерентным круговым источником с угловым размером θ_s как функция расстояния между точками ρ

Такой же результат (11.54) был получен в рамках менее строгого анализа. Площадь когерентности обратно пропорциональна θ_s^2. Например, некогерентный источник света с длиной волны $\lambda = 0{,}6$ мкм радиусом 1 см, наблюдаемый с расстояния $d = 100$ м, имеет расстояние когерентности $\rho_c \approx 3{,}7$ мм.

Измерение углового диаметра звезд: звездный интерферометр Майкельсона

Уравнение (11.87) лежит в основе метода измерения угловых диаметров звезд. Если звезду рассматривать как некогерентный диск диаметром $2a$ с равномерной яркостью, то в плоскости наблюдения на расстоянии d от звез-

ды функция когерентности падает до нуля, если расстояние между двумя точками наблюдения достигает $\rho_c = 1{,}22\lambda/\theta_s$. Измерение ρ_c при данном λ позволяет определить угловой диаметр $\theta_s = 2a/d$.

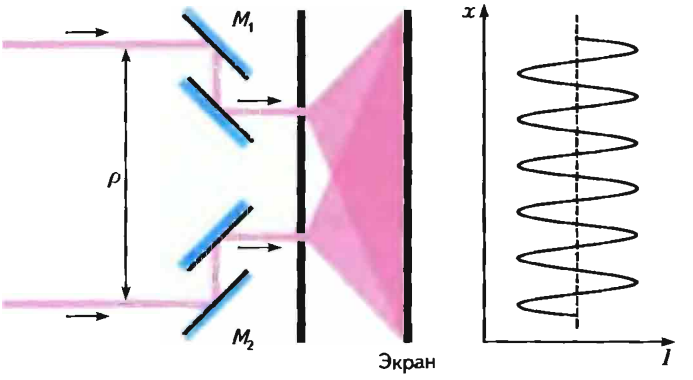


Рис. 11.27. Звездный интерферометр Майкельсона. Угловой диаметр звезды вычисляется по данным измерения взаимной интенсивности в двух точках с переменным расстоянием r между ними с использованием двухщелевого интерферометра Юнга. При изменении расстояния ρ между зеркалами M_1 и M_2 измеряется видность интерференционных полос. Когда $\rho = \rho_c = 1{,}22\lambda/\theta_s$, видность обращается в нуль

В качестве примера примем угловой диаметр Солнца равным 0,5°, тогда $\theta_s = 8{,}7 \quad 10^{-3}$ радиан. Предположим, что интенсивность распределена равномерно, тогда получим $\rho_c = 140\lambda$. При $\lambda = 0{,}5$ мкм $\rho_c = 70$ мкм. Для наблюдения интерференционных полос в схеме Юнга с двумя щелями нужно, чтобы расстояние между щелями было меньше, чем 70 мкм. Звезды с меньшим угловым диаметром имеют соответственно большие площади когерентности. Например, первая звезда, угловой диаметр которой был измерен данным методом (Альфа Ориона), имеет угловой диаметр $\theta_s = 22{,}6 \cdot 10^{-8}$, так что при $\lambda = 0{,}57$ мкм $\rho_c = 3{,}1$ м. Интерферометр Юнга можно модифицировать, чтобы приспособить к таким большим расстояниям между щелями, с помощью подвижных зеркал, как показано на рис. 11.27.

11.4. ЧАСТИЧНАЯ ПОЛЯРИЗАЦИЯ

Как мы видели в гл. 6, скалярная теория света часто неадекватна и требуется векторная теория, учитывающая поляризацию. В данном разделе кратко обсуждается статистическая теория случайного света, включающая эффекты поляризации. Теория частичной поляризации основана на характеристике компонент вектора светового поля с помощью корреляций и кросс-корреляций, аналогичных уже рассмотренным выше в данной главе.

Для упрощения описания не будем рассматривать пространственные эффекты. Ограничим наше рассмотрение плоской поперечной электромагнитной (ТЕМ) волной, распространяющейся вдоль оси z. Вектор электрического поля имеет две компоненты $U_x(t)$ и $U_y(t)$ в направлениях x и y, которые в общем случае представляют собой случайные функции. Каждая из них характеризуется своей автокорреляционной функцией (временной функцией когерентности)

$$G_{xx}(\tau) = \langle U_x^*(t)U_x(t+\tau)\rangle; \tag{11.88}$$

$$G_{yy}(\tau) = \langle U_y^*(t)U_y(t+\tau)\rangle. \tag{11.89}$$

Дополнительное описание волны дает функция взаимной корреляции $U_x(t)$ и $U_y(t)$

$$G_{xy}(\tau) = \langle U_x^*(t)U_y(t+\tau)\rangle. \tag{11.90}$$

Нормированная функция

$$g_{xy}(\tau) = \frac{G_{xy}(\tau)}{\sqrt{G_{xx}(0)G_{yy}(0)}} \tag{11.91}$$

представляет собой коэффициент взаимной корреляции $U_x(t)$ и $U_y(t+\tau)$. Он удовлетворяет неравенству $0 \le |g_{xy}(\tau)| \le 1$. Если обе компоненты в любой момент времени не коррелированы, то $|g_{xy}(\tau)| = 0$, а при наличии полной корреляции $|g_{xy}(\tau)| = 1$.

Спектральные свойства света в общем случае связаны с его поляризационными свойствами, при этом авто- и кросс-корреляционные функции могут иметь различную зависимость от τ. Однако для квазимонохроматического света все зависимости от τ в формулах (11.88)—(11.91) приближенно имеют один и тот же вид $\exp(j\omega_0\tau)$, так что поляризационные свойства описываются их значениями при $\tau = 0$. Тогда для описания волны будем использовать три числа $G_{xx}(0)$, $G_{yy}(0)$ и $G_{xy}(0)$, которые далее будем обозначать через G_{xx}, G_{yy} и G_{xy}. Заметим, что $G_{xx} = I_x$ и $G_{yy} = I_y$ — действительные числа, описывающие интенсивности x- и y-компонент, однако G_{xy} — комплексное число, причем $G_{xy} = G_{yx}^*$, что видно из его определения.

Матрица когерентности

Четыре переменные G_{xx}, G_{yy}, G_{xy} и G_{yx} удобно записывать в виде эрмитовой матрицы размерности 2×2

$$\mathbf{G} = \begin{bmatrix} G_{xx} & G_{xy} \\ G_{yx} & G_{yy} \end{bmatrix}, \tag{11.92}$$

которая называется **матрицей когерентности**. Диагональные ее элементы — интенсивности I_x и I_y, а недиагональные — кросс-корреляции. След этой матрицы $\mathrm{Sp}\,\mathbf{G} = I_x + I_y \equiv \bar{I}$ — полная интенсивность.

Матрицу когерентности можно записать также через вектор Джонса

$$\boldsymbol{J} = \begin{bmatrix} U_x \\ U_y \end{bmatrix},$$

определенный через комплексные волновые функции и комплексные амплитуды (а не комплексные огибающие, как в подразд. 6.1.2):

$$\langle \boldsymbol{J}^* \boldsymbol{J}^\dagger \rangle = \left\langle \begin{bmatrix} U_x^* \\ U_y^* \end{bmatrix} \begin{bmatrix} U_x & U_y \end{bmatrix} \right\rangle = \begin{bmatrix} \langle U_x^* U_x \rangle & \langle U_x^* U_y \rangle \\ \langle U_y^* U_x \rangle & \langle U_y^* U_y \rangle \end{bmatrix} = \mathbf{G}, \tag{11.93}$$

где † обозначает транспонирование матрицы, а U_x и U_y обозначают $U_x(t)$ и $U_y(t)$ соответственно.

Вектор Джонса преобразуется поляризационными устройствами, такими как поляризаторы и фазовые пластинки, в соответствии с формулой $\boldsymbol{J}' = \mathbf{T}\boldsymbol{J}$ [см. (6.17)], где $\mathbf{T}$ — матрица Джонса, описывающая устройство [см. (6.18)—(6.25)]. Тогда матрица когерентности преобразуется в соответствии с $\mathbf{G}' = \langle \mathbf{T}^* \boldsymbol{J}^* (\mathbf{T}\boldsymbol{J})^\dagger \rangle = \langle \mathbf{T}^* \boldsymbol{J}^* \boldsymbol{J}^\dagger \mathbf{T}^\dagger \rangle = \langle \mathbf{T}^* (\boldsymbol{J}^* \boldsymbol{J}^\dagger) \mathbf{T}^\dagger \rangle$, так что

$$\boxed{\mathbf{G}' = \mathbf{T}^* \mathbf{G} \mathbf{T}^\dagger.} \tag{11.94}$$

Таким образом, мы имеем формализм для определения влияния поляризационных устройств на матрицу когерентности частично поляризованного света

Параметры Стокса и представление на сфере Пуанкаре

Параметры Стокса были определены в подразд. 6.1.1 для когерентного света как набор четырех действительных параметров, связанных с произведениями компонент x и y комплексной огибающей [см. (6.9)]. Это определение легко обобщается на случай частично когерентного света путем взятия средних от указанных произведений

$$\boxed{S_0 = \langle |U_x|^2 \rangle + \langle |U_y|^2 \rangle = G_{xx} + G_{yy};} \tag{11.95а}$$

$$\boxed{S_1 = \langle |U_x|^2 \rangle - \langle |U_y|^2 \rangle = G_{xx} - G_{yy};} \tag{11.95б}$$

$$\boxed{S_2 = 2\,\mathrm{Re}\{\langle U_x^* U_y \rangle\} = 2\,\mathrm{Re}\{G_{xy}\};} \tag{11.95в}$$

$$\boxed{S_3 = 2\,\mathrm{Im}\{\langle U_x^* U_y \rangle\} = 2\,\mathrm{Im}\{G_{xy}\}.} \tag{11.95г}$$

Параметры Стокса

Таким образом, параметры Стокса непосредственно связаны с элементами матрицы когерентности $\mathbf{G}$. Первый параметр S_0 — просто сумма диагональных элементов, которая представляет собой полную интенсивность $\bar{I}$. Второй па-

раметр S_1 — разность диагональных элементов, т. е. разность интенсивностей двух поляризационных компонент. Третий и четвертый параметры, S_2 и S_3, пропорциональны действительной и мнимой части недиагонального элемента, т. е. функции взаимной корреляции. Пользуясь этими соотношениями, можно легко показать, что неравенство $|G_{xy}|^2 \le G_{xx}G_{yy}$ приводит к условию $S_1^2 + S_2^2 + S_3^2 \le S_0^2$. Для когерентного света эти неравенства превращаются в равенства.

Состояние поляризации частично поляризованного света можно представить геометрически как точку с декартовыми координатами $(S_1/S_0,\ S_2/S_0,\ S_3/S_0)$. Поскольку $S_1^2 + S_2^2 + S_3^2 \le S_0^2$, такая точка лежит внутри или на поверхности сферы Пуанкаре.

Для понимания важности матрицы когерентности и параметров Стокса исследуем два предельных случая.

Неполяризованный свет

Свет с интенсивностью $\bar{I}$ называется неполяризованным, если две его компоненты имеют одинаковые интенсивности и не коррелированы,

$$I_x = I_y = \frac{1}{2}\bar{I}; \quad G_{xy} = 0.$$

Тогда матрица когерентности

$$\boxed{\mathbf{G} = \frac{1}{2}\bar{I}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.} \qquad (11.96)$$

Неполяризованный свет

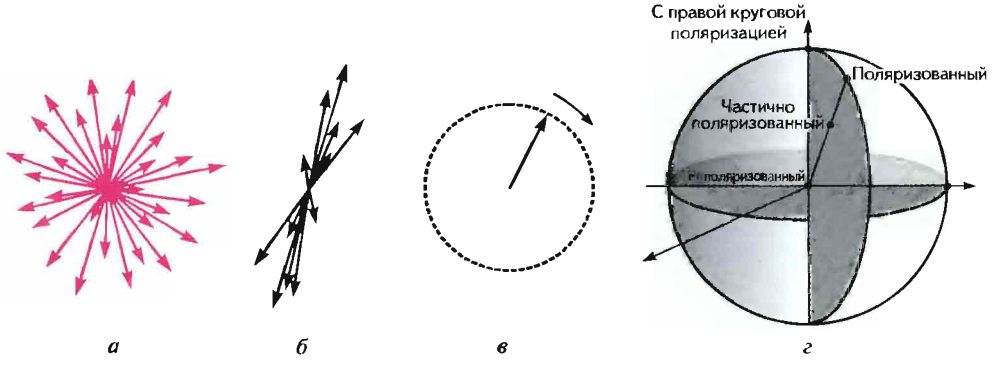


Рис. 11.28. Флуктуации вектора электрического поля для неполяризованного (*а*), частично поляризованного (*б*), поляризованного света с правой круговой (ПК) поляризацией (*в*) и представление на сфере Пуанкаре (*г*)

С помощью (11.94) и (6.22) можно показать, что выражение (11.96) инвариантно относительно вращений системы координат, так что две компоненты всегда имеют одинаковые интенсивности и не коррелированы. Следовательно

у неполяризованного света вектор электрического поля статистически изотропен; он с равной вероятностью может иметь любое направление в плоскости x—y, как показано на рис. 11.28, *а*.

При прохождении через поляризатор неполяризованный свет становится линейно-поляризованным, однако остается случайным со средней интенсивностью $\bar{I}/2$. Фазовая пластинка не оказывает действия на неполяризованный свет, поскольку она только вносит фазовый сдвиг между компонентами, которые изначально имеют совершенно случайные фазы. Аналогично, неполяризованный свет остается таковым при пропускании через вращатель плоскости поляризации. Эти эффекты можно формально продемонстрировать с использованием формул (11.94) и (11.96) вместе с формулами (6.18), (6.40) и (6.10).

Параметры Стокса неполяризованного света $(S_0, S_1, S_2, S_3) = (\bar{I}, 0, 0, 0)$, как легко показать с использованием (11.95) и (11.96). Соответствующее представление на сфере Пуанкаре — точка с декартовыми координатами

$$\left(\frac{S_1}{S_0}, \frac{S_2}{S_0}, \frac{S_3}{S_0}\right) = (0, 0, 0),$$

т. е. центр сферы.

Поляризованный свет

Если коэффициент взаимной корреляции

$$g_{xy} = \frac{G_{xy}}{\sqrt{\bar{I}_x \bar{I}_y}}$$

имеет модуль, равный единице, $|g_{xy}| = 1$, то обе компоненты вектора электрического поля абсолютно коррелированны, и свет называется полностью поляризованным, или просто **поляризованным**. Тогда матрица когерентности принимает вид

$$\mathbf{G} = \begin{bmatrix} I_x & \sqrt{I_x I_y}\, e^{i\varphi} \\ \sqrt{I_x I_y}\, e^{i\varphi} & I_y \end{bmatrix}, \tag{11.97}$$

где φ — аргумент g_{xy}. Определяя $U_x = \sqrt{I_x}$; $U_y = \sqrt{I_y}\, e^{i\varphi}$, получаем

$$\mathbf{G} = \begin{bmatrix} U_x^* U_x & U_x^* U_y \\ U_y^* U_x & U_y^* U_y \end{bmatrix} = \boldsymbol{J}^* \boldsymbol{J}^\dagger, \tag{11.98}$$

где $\boldsymbol{J}$ — вектор Джонса с компонентами U_x и U_y. Таким образом, $\mathbf{G}$ имеет тот же вид, что и матрица когерентности когерентной волны. Используя векторы

Джонса, приведенные в табл. 6.1, можно определить матрицы когерентности для различных состояний поляризации. Вот два примера:

Линейная поляризация в направлении x	$G = \bar{I}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	Правая круговая поляризация	$G = \frac{1}{2}\bar{I}\begin{bmatrix} 1 & j \\ -j & 1 \end{bmatrix}$.

Параметры Стокса, соответствующие (11.98), удовлетворяют соотношению $S_1^2 + S_2^2 + S_3^2 = S_0^2$, поэтому поляризованный свет представляется точками на поверхности сферы Пуанкаре, а не внутри нее.

Поучительно исследовать различие между неполяризованным светом и светом с круговой поляризацией. В обоих случаях интенсивности x- и y-компонент равны ($I_x = I_y$). Для круговой поляризации обе компоненты полностью коррелированы, а для неполяризованного света — абсолютно не коррелированы. Свет с круговой поляризацией можно преобразовать в линейно-поляризованный свет с помощью фазовой пластинки, однако неполяризованный свет при этом останется неполяризованным. Состояниям круговой поляризации на сфере Пуанкаре соответствуют северный и южный полюсы, а неполяризованному свету — центр сферы.

Степень поляризации

Частичная поляризация — общее состояние случайной поляризации, лежащее между двумя предельными состояниями полностью поляризованного и полностью неполяризованного света. Одна из мер **степени поляризации** определяется через детерминант и след матрицы когерентности:

$$\mathbb{P} = \sqrt{1 - \frac{4 \det \mathbf{G}}{(\mathrm{Sp}\,\mathbf{G})^2}} = \tag{11.99}$$

$$= \sqrt{1 - 4\left[\frac{I_x I_y}{(I_x + I_y)^2}\right]\left(1 - |g_{xy}|^2\right)}. \tag{11.100}$$

Значимость этой меры подтверждается следующими соображениями.

- Она удовлетворяет условию $0 \le \mathbb{P} \le 1$.
- Для поляризованного света $\mathbb{P}$ имеет наибольшее значение, равное единице, что легко показать, подставляя $|g_{xy}| = 1$ в (11.100). Для неполяризованного света она имеет наименьшее значение $\mathbb{P} = 0$, поскольку $I_x = I_y$ и $g_{xy} = 0$.
- Она инвариантна по отношению к вращениям системы координат (поскольку определитель и след матрицы инвариантны относительно унитарных преобразований).
- Степень поляризации можно выразить через параметры Стокса как

$$\mathbb{P} = \frac{\sqrt{S_1^2 + S_1^2 + S_1^2}}{S_0}, \tag{11.101}$$

так что в представлении на сфере Пуанкаре она равна расстоянию до центра сферы.

• Можно показать (упражнение 11.4), что частично поляризованная волна всегда может быть представлена как смесь двух некоррелированных волн: полностью поляризованной волны и неполяризованной волны с отношением интенсивности поляризованной компоненты к полной интенсивности, равным степени поляризации $\mathbb{P}$.

Упражнение 11.4

Частично поляризованный свет

Покажите, что суперпозиция неполяризованного света с интенсивностью $(I_x + I_y)(1 - \mathbb{P})$ и поляризованного света с интенсивностью $(I_x + I_y)\mathbb{P}$, где $\mathbb{P}$ дается формулой (11.100), образует свет с интенсивностями компонент I_x и I_y и нормированным коэффициентом взаимной корреляции $|g_{xy}|$.

Рекомендуемая литература

ОБЩАЯ

Kokhanovsky A.A. Polarization Optics of Random Media. Springer-Verlag, 2003.

O'Neill E.L. Introduction to Statistical Optics. Addison-Wesley, 1963; Dover, reissued 2003.

Lauterborn W., Kurz T Coherent Optics: Fundamentals and Applications. Springer, 2nd ed. 2003.

Born M., Wolf E. Principles of Optics. Cambridge University Press, 7th expanded and corrected ed. 2002. Chapter 10.

Frieden B.R. Probability, Statistical Optics, and Data Testing. A Problem Solving Approach. Springer-Verlag, 1983; 3rd ed. 2001.

Goodman J.W. Statistical Optics. Wiley, 1985; paperback ed. 2000.

Rowe H.E. Electromagnetic Propagation in Multi-Mode Random Media. Wiley, 1999.

Brosseau C. Fundamentals of Polarized Light. A Statistical Optics Approach. Wiley, 1998.

Mandel L., Wolf E. Optical Coherence and Quantum Optics. Cambridge University Press, 1995.

Lefèvre H. The Fiber-Optic Gyroscope. Artech, 1993.

Reynolds G., DeVelis J.B., Parrent G.B., Thompson B.J. The New Physical Optics Notebook: Tutorials in Fourier Optics. SPIE Optical Engineering Press, 1989.

Peřina J. Coherence of Light. Reidel, 1971, 2nd ed. 1985.

Dainty J.C., ed. Laser Speckle and Related Phenomena. Springer-Verlag, 1975; 2nd ed. 1984.

Marathay A.S. Elements of Optical Coherence Theory. Wiley, 1982.

Saleh B.E.A. Photoelectron Statistics with Applications to Spectroscopy and Optical Communication. Springer-Verlag, 1978.

Crosignani B., Di Porto P., Bertolotti M. Statistical Properties of Scattered Light. Academic Press, 1975.

Beran M.J., Parrent, Jr. G.B. Theory of Partial Coherence. Prentice Hall, 1964; SPIE Optical Engineering Press, reissued 1974.

Hanbury-Brown R. The Intensity Interferometer. Its Application to Astronomy. Taylor & Francis, 1974.

Troup G.J. Optical Coherence Theory. Methuen, 1967.

КНИГИ ПО СЛУЧАЙНЫМ ФУНКЦИЯМ

Papoulis A., Pillai S.U. Probability, Random Variables, and Stochastic Processes. McGraw-Hill, 1965; 4th ed. 2002.

Parzen E. Stochastic Processes. Holden-Day, 1962; Society for Industrial and Applied Mathematics (SIAM), reissued 1999.

Parzen E. Modern Probability Theory and Its Applications. Wiley, 1960; paperback ed. 1992.

Helstrom C.W. Probability and Stochastic Processes for Engineers and Scientists. Macmillan, 2nd. ed. 1991.

Davenport, Jr. W.B., Root W.L. An Introduction to the Theory of Random Signals and Noise. McGraw-Hill, 1958; IEEE Press, reissued 1987.

Vanmarcke E. Random Fields. MIT Press, 1983.

Thomas J.B. An Introduction to Applied Probability and Random Processes. Wiley, 1971.

КНИГИ ПО ОПТИЧЕСКОЙ КОГЕРЕНТНОЙ ТОМОГРАФИИ

Brezinski M.E. Optical Coherence Tomography: Principles and Applications. Academic Press, 2006.

Drexler W., ed. Optical Coherence Tomography and Coherence Techniques. Vol. 2: Progress in Biomedical Optics and Imaging. SPIE Optical Engineering Press, 2005.

Drexler W., ed. Optical Coherence Tomography and Coherence Techniques. Vol. 1: Progress in Biomedical Optics and Imaging. SPIE Optical Engineering Press, 2003.

Bouma B.E., Teamey G.J., eds. Handbook of Optical Coherence Tomography. Marcel Dekker, 2002.

СТАТЬИ

Tomlins P.H., Wang R.K. Theory, Developments and Applications of Optical Coherence Tomography. Journal of Physics D: Applied Physics. Vol. 38, 2005. P. 2519—2535.

Fercher A.F., Drexler W., Hitzenberger C.K., Lasser T. Optical Coherence Tomography Principles and Applications. Reports on Progress in Physics. Vol. 66, 2003. P. 239—303.

Mandel L., Wolf E., eds. Selected Papers on Coherence and Fluctuations of Light (1850—1966). SPIE Optical Engineering Press (Milestone Series. Vol. 19), 1990.

Smith R.B., ed. Selected Papers on Fiber Optic Gyroscopes. SPIE Optical Engineering Press (Milestone Series. Vol. 8), 1989.

Feature issues on applications of coherence and statistical optics. Journal of the Optical Society of America. № 7, 1986 and № 8, 1986.

Yu F.T.S. Principles of Optical Processing with Partially Coherent Light. In Progress in Optics. Vol. 23. E. Wolf, ed. North-Holland, 1986.

Tango W.J., Twiss R.Q. Michelson Stellar Interferometry. In Progress in Optics. Vol. 17. E. Wolf, ed. North-Holland, 1980.

Reynolds G.O., DeVelis J.B. Review of Optical Coherence Effects in Instrument Design. SPIE Proceedings. Vol. 194, 1979. P. 2—33.

Baltes H.P., Geist J., Walther A. Radiometry and Coherence. In Inverse Source Problems in Optics. H.P. Baltes, ed. Springer-Verlag, 1978.

Wolf E. Coherence and Radiometry, Journal of the Optical Society of America. Vol. 68, 1978. P. 6—17.

Mandel L., Wolf E., eds. Selected Papers on Coherence and Fluctuations of Light. Vol. 1 and 2. Dover, 1970.

Thompson B.J. Image Formation with Partially Coherent Light. In Progress in Optics. Vol. 7. E. Wolf, ed. North-Holland, 1969.

Mandel L., Wolf E. Coherence Properties of Optical Fields. Reviews of Modern Physics. Vol. 37, 1965. P. 231—287.

Задачи

К РАЗДЕЛУ 11.1

1. **Лоренцев спектр.** Светоизлучающий диод (СИД) излучает свет с лоренцевым спектром с шириной линии по половине максимума $\Delta\nu = 10^{13}$ Гц с центром на частоте, соответствующей длине волны $\lambda_0 = 0{,}7$ мкм. Определите ширину линии $\Delta\lambda_0$ (в нанометрах), время когерентности τ_c и длину когерентности l_c. Каков максимальный промежуток времени, в пределах которого модуль комплексной степени временной когерентности $|g(\tau)|$ больше 0,5?

2. **Доказательство теоремы Винера—Хинчина.** Используя формулы (11.4), (11.14) и (11.15), докажите, что спектральная плотность $S(\nu)$ является Фурье-образом автокорреляционной функции $G(\tau)$. Докажите, что интенсивность I равна интегралу от спектральной плотности мощности $S(\nu)$.

3. **Взаимная интенсивность.** Взаимная интенсивность оптической волны в точках оси x дается выражением

$$G(x_1, x_2) = I_0 \exp\left[-\frac{(x_1^2 + x_2^2)}{W_0^2}\right]\exp\left[-\frac{(x_1 - x_2)^2}{\rho_c^2}\right],$$

где I_0, W_0 и ρ_c — постоянные. Изобразите распределение интенсивности как функцию x. Выведите выражение для нормированной взаимной интенсивности $g(x_1, x_2)$ и постройте ее график как функции $(x_1 - x_2)$. Каков физический смысл параметров I_0, W_0 и ρ_c?

4. **Функция взаимной когерентности.** Оптическая волна имеет функцию взаимной когерентности в точках оси x

$$G(x_1, x_2, \tau) = \exp\left(-\frac{\pi\tau^2}{2\tau_c^2}\right)\exp\left[j2\pi u(x_1, x_2)\tau\right]\exp\left[-\frac{(x_1 - x_2)^2}{\rho_c^2}\right],$$

где $u(x_1, x_2) = 5 \cdot 10^{14}$ с$^{-1}$ при $x_1 + x_2 > 0$ и $6 \cdot 10^{14}$ с$^{-1}$ при $x_1 + x_2 < 0$, $\rho_c = 1$ мм и $\tau_c = 1$ мкс. Определите интенсивность, спектральную плотность мощности, длину когерентности и расстояние когерентности в поперечной плоскости. Какая из этих величин зависит от положения? Если эту волну записать на цветную пленку, как будет выглядеть изображение?

5. **Длина когерентности.** Покажите, что свет с узким частотным спектром имеет длину когерентности $l_c \approx \lambda^2/\Delta\lambda$, где $\Delta\lambda$ — ширина линии в единицах длины волны. Покажите, что свет с широким однородным спектром, простирающимся от λ_{min} до $\lambda_{max} = 2\lambda_{min}$, имеет длину когерентности $l_c = \lambda_{max}$.

6. **Влияние спектральной ширины на пространственную когерентность.** Точечный источник, расположенный в начале (0,0,0) декартовой системы координат, испускает свет с лоренцевым спектром и временем когерентности $\tau_c = 10$ пс. Найдите выражение для нормированной взаимной интенсивности света в точках $(0, 0, d)$ и $(x, 0, d)$, где $d = 10$ см. Постройте график модуля нормированной взаимной интенсивности как функции x.

7. **Гауссова взаимная интенсивность.** Оптическая волна в свободном пространстве имеет функцию взаимной когерентности $G(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = J(\boldsymbol{r}_1 - \boldsymbol{r}_2)\exp(j\omega_0\tau)$.

а. Покажите, что функция $J(\boldsymbol{r})$ должна удовлетворять уравнению Гельмгольца $\nabla^2 J + k_0^2 J = 0$, где $k_0 = \omega_0/c$.

б. Приближенное решение уравнения Гельмгольца имеет вид гауссова пучка

$$J(\boldsymbol{r}) = \frac{1}{q(z)}\exp\left[-\frac{jk_0(x^2+y^2)}{2q(z)}\right]\exp(-jk_0 z),$$

где $q(z) = z + jz_0$ и z_0 — постоянные. Это решение подробно исследовалось в гл. 3 в связи с гауссовыми пучками. Найдите выражение для площади когерентности вблизи оси z и покажите, что она растет с ростом $|z|$, так что волна приобретает когерентность с удалением от начала координат.

К РАЗДЕЛУ 11.2

1. **Влияние спектральной ширины на видность полос.** Свет от натриевой лампы с шириной лоренцева спектра $\Delta\nu = 5 \cdot 10^{11}$ Гц используется в интерферометре Майкельсона. Определите максимальную разность хода, при которой видность интерферограммы $\mathcal{V} > 1/2$.

2. **Число наблюдаемых полос в интерферометре Юнга.** Определите число наблюдаемых полос в интерферометре Юнга при использовании каждого из источников из табл. 11.2. Во всех случаях считайте пространственную когерентность полной.

3. **Спектр суперпозиции двух волн.** Оптическая волна представляет собой суперпозицию двух волн $U_1(t)$ и $U_2(t)$ с одинаковыми гауссовыми спектрами $S_1(\nu) = S_2(\nu)$ с шириной $\Delta\nu$ и центральной частотой ν_0. Волны не обязательно некоррелированные. Найдите выражение для спектральной плотности мощности $S(\nu)$ суперпозиции $U(t) = U_1(t) + U_2(t)$. Исследуйте возможность того, что $S(\nu)$ также гауссова и имеет сдвинутую центральную частоту $\nu_1 \neq \nu_0$. Если бы это оказалось возможным, наша вера в доплеровский сдвиг как способ определения скорости звезд была бы поколеблена, поскольку у сдвига частоты могли бы быть другие причины, чем эффект Доплера.

К РАЗДЕЛУ 11.3

1. **Частично когерентный гауссов пучок.** Квазимонохроматическая световая волна с длиной волны λ распространяется в свободном пространстве в направлении z. Ее интенсивность в плоскости $z = 0$ описывается гауссовой функцией

$$I(x) = I_0\exp\left(-\frac{2x^2}{W_0^2}\right),$$

а ее нормированная взаимная интенсивность — также гауссова функция

$$g(x_1, x_2) = \exp\left[-\frac{(x_1 + x_2)^2}{\rho_c^2}\right].$$

Покажите, что интенсивность на расстоянии z, удовлетворяющем условиям применимости приближения Фраунгофера, также является гауссовой функцией

$$I_z(x) \propto \exp\left(-\frac{2x^2}{W^2(z)}\right)$$

и получите выражение для ширины пучка $W(z)$ как функции z и параметров W_0, ρ_c и λ. Обсудите влияние пространственной когерентности на расходимость пучка.

2. **Фурье-линза.** Квазимонохроматический пространственно-некогерентный свет с равномерным распределением интенсивности освещает транспарант с коэффициентом пропускания по интенсивности $f(x, y)$ и получающийся свет проходит от передней до задней фокальной плоскости линзы. Найдите выражение для интенсивности наблюдаемого света. Сравните результат со случаем когерентного света, когда линза осуществляет преобразование Фурье (см. разд. 4.2).

3. **Свет от двухточечного некогерентного источника.** Пространственно-некогерентный квазимонохроматический источник света излучает только в двух точках, разделенных расстоянием $2a$. Найдите выражение для нормированной взаимной интенсивности на расстоянии d от источника (используйте приближение Фраунгофера).

4. **Когерентность света, прошедшего через оптическую систему, осуществляющую преобразование Фурье.** Свет от квазимонохроматического пространственно некогерентного источника с равномерным распределением интенсивности пропускается через тонкую щель шириной $2a$ и распространяется от передней до задней фокальной плоскости линзы. Найдите выражение для нормированной взаимной интенсивности на задней фокальной плоскости.

К РАЗДЕЛУ 11.4

1. **Частично поляризованный свет.** Интенсивности двух компонент частично поляризованной волны равны $I_x = I_y = 1/2$, а аргумент коэффициента взаимной корреляции равен $\pi/2$.

   а. Постройте график зависимости степени поляризации $\mathbb{P}$ от модуля коэффициента взаимной корреляции $|g_{xy}|$.

   б. Определите матрицу когерентности, если $\mathbb{P} = 0$; 0,5 и 1, и опишите природу света в каждом случае.

   в. Если свет пропускается через поляризатор, ось которого направлена вдоль оси x, то какова интенсивность прошедшего света?

ГЛАВА
12 ОПТИКА ФОТОНОВ

Макс Планк (1858—1947) предположил, что испускание и поглощение света веществом происходит в форме квантов энергии.

Альберт Эйнштейн (1879—1955) выдвинул гипотезу о том, что свет сам по себе состоит из квантов энергии.

Электромагнитная оптика (см. гл. 5) дает наиболее полное описание света в рамках **классической оптики**. Она включает в себя волновую оптику, которая, в свою очередь, включает оптику лучей (рис. 12.1). Хотя классическая электромагнитная теория и способна дать объяснение множеству эффектов в оптике, как видно из предшествующих глав данной книги, существуют явления, которые невозможно объяснить на ее основе. Это стало очевидно в начале прошлого столетия и привело к созданию квантовой электромагнитной теории, известной как **квантовая электродинамика**. Применительно к оптическим явлениям эта теория называется также **квантовой оптикой**. Квантовая электродинамика (КЭД) в настоящее время принята в качестве теории, способной описать почти все известные оптические явления.

В рамках КЭД электрическое и магнитное поля $\boldsymbol{E}$ и $\boldsymbol{H}$ рассматриваются математически как операторы в гильбертовом пространстве, а не просто векторы. Предполагается, что они удовлетворяют некоторым операторным уравнениям и коммутационным соотношениям, которые управляют их временно́й динамикой и взаимной зависимостью. Уравнения КЭД описывают взаимодей-

ствие электромагнитных полей с веществом таким же образом, как это делают уравнения Максвелла в классической электродинамике. КЭД приводит к результатам, которые являются принципиально квантовыми и не могут быть объяснены классически. Однако, несмотря на свои выдающиеся успехи, КЭД не является судьей последней инстанции для *всех* оптических явлений. На сегодняшний день ограниченность КЭД проявляет себя в эффектах, описываемых **электрослабой теорией**, сочетающей квантовую электродинамику и теорию слабого взаимодействия. Электрослабая теория успешно объясняет неожиданные (не сохраняющие четность) малые повороты плоскости поляризации света при прохождении через некоторые материалы[1]. Есть надежда, что сочетание электрослабой теории с теориями сильного и гравитационного взаимодействия в конечном счете приведет к общей единой теории, включающей все силы, известные в природе.

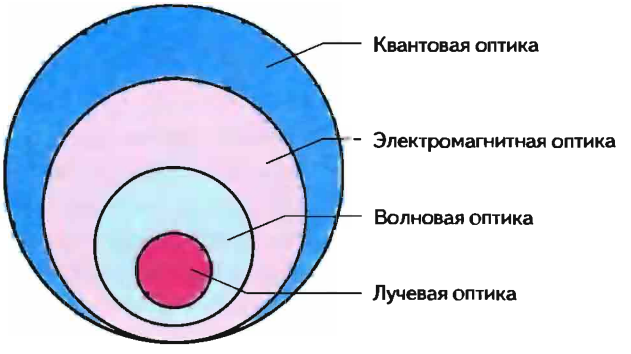


Рис. 12.1. Квантовая оптика дает теоретическое объяснение всех возможных оптических явлений. Она является более общей, чем электромагнитная оптика, которая, как было показано ранее, охватывает волновую и лучевую оптику

О данной главе

Формальное построение КЭД выходит за рамки настоящей книги. Тем не менее можно описать многие квантовые свойства света и его взаимодействия с веществом путем введения в электромагнитную оптику нескольких простых соотношений, заимствованных из КЭД, которые выражают корпускулярность, локализацию и флуктуации квантовых полей и энергии. Этот набор правил, который мы называем **оптикой фотонов**, позволяет рассматривать оптические явления, лежащие вне пределов классической теории, при этом классическая оптика сохраняется в качестве предельного случая. Тем не менее оптика фотонов не является теорией, способной дать объяснение всем эффектам, которые объясняет квантовая оптика.

В разд. 12.1 вводится понятие фотона и исследуются его свойства. Мы используем электромагнитную оптику в качестве отправной точки, накладывая ряд правил, определяющих поведение энергии, импульса, поляризации, по-

[1] См., например, *Vetter P.A., Meekhof D.M., Majumder P.K., Lamoreaux S.K., Fortson E.N.* Precise Test of Electroweak Theory from a New Measurement of Parity Nonconservation in Atomic Thallium. Physical Review Letters. Vol. 74, 1995. P. 2658—2661.

ложения, временно́й локализации и интерференции фотонов. Эти правила имеют вид обманчиво простых соотношений, приводящих к далеко идущим последствиям. Далее, в разд. 12.2, следует обсуждение свойств коллектива фотонов и фотонных потоков. Число фотонов, испускаемых источником света за данный промежуток времени, почти всегда случайно и демонстрирует статистические свойства, зависящие от природы источника. Устанавливается статистика числа фотонов для нескольких важных оптических источников, включая лазеры и тепловые излучатели. Исследуется также действие простых оптических элементов (таких как светоделители и фильтры) на случайность фотонного потока. В разд. 12.3 мы используем квантовую оптику для изучения квантовых флуктуаций модуля и фазы электромагнитного поля. Дано краткое введение в теорию когерентных, сжатых и бифотонных состояний света. Взаимодействие фотонов с атомами и полупроводниками описано в гл. 13 и 16 соответственно.

12.1. ФОТОН

С квантовой точки зрения свет состоит из частиц, называемых фотонами. Фотон несет электромагнитную энергию и импульс, а также обладает собственным угловым моментом (спином), который связан с его поляризационными свойствами. Он может нести также орбитальный угловой момент. Фотон имеет нулевую массу покоя и распространяется со скоростью света в вакууме c_0; в веществе его скорость уменьшается до $c < c_0$. Фотон обладает также волновыми свойствами, которые определяют свойства его локализации в пространстве и времени и закономерности его интерференции и дифракции.

Понятие фотона первоначально возникло из попытки Макса Планка в 1900 г. разрешить давно стоящую проблему, касающуюся спектра излучения черного тела (эта тема обсуждается в гл. 13). Ему удалось добиться цели посредством квантования допустимых значений энергии каждой из электромагнитных мод в полости, из которой исходит излучение. В 1905 г. Альберт Эйнштейн обобщил понятие квантования, представив свет как таковой в виде коллектива фотонов. Это позволило ему успешно объяснить фотоэлектрический эффект (эта тема обсуждается в гл. 18).

Концепция фотона и правила оптики фотонов вводятся путем рассмотрения света внутри оптического резонатора (полости). Этот выбор удобен тем, что рассматриваемое пространство ограничивается простой геометрией. Однако присутствие резонатора оказывается не важным, и можно показать, что результаты от него не зависят.

Электромагнитно-оптическая теория света в резонаторе

В соответствии с законами электромагнитной оптики свет внутри резонатора без потерь объема V полностью характеризуется электромагнитным полем, которое принимает вид суперпозиции дискретных ортогональных мод с

различным пространственным распределением, различными частотами и поляризациями. Вектор электрического поля равен $\mathcal{E}(\boldsymbol{r}, t) = \mathrm{Re}\{\boldsymbol{E}(\boldsymbol{r}, t)\}$, где

$$\boldsymbol{E}(\boldsymbol{r}, t) = \sum_{q} A_q U_q(\boldsymbol{r}) \exp\left(j2\pi\nu_q t\right)\hat{\boldsymbol{e}}_q. \tag{12.1}$$

Мода q имеет комплексную огибающую A_q, частоту ν_q, поляризацию вдоль направления единичного вектора $\hat{\boldsymbol{e}}_q$ и пространственное распределение, характеризуемое комплексной функцией $U_q(\boldsymbol{r})$, нормированной условием

$$\int_V |U_q(\boldsymbol{r})|^2 \, \mathrm{d}\boldsymbol{r} = 1.$$

Выбор функций $U_q(\boldsymbol{r})$, $\exp(j2\pi\nu_q t)$ и векторов $\hat{\boldsymbol{e}}_q$ для разложения — не единственный, их можно выбрать и по-другому, включая полихроматические моды.

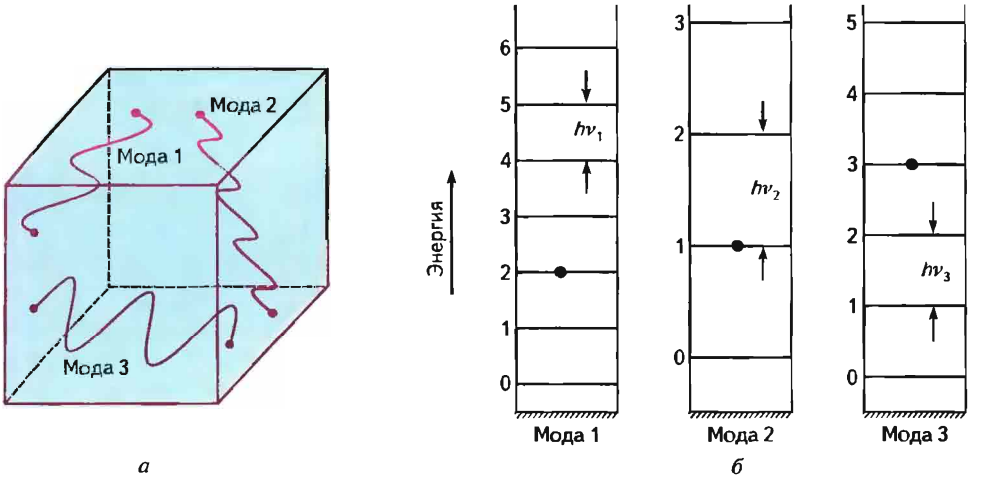


Рис. 12.2. Три моды с различными частотами и направлениями в кубическом резонаторе (*а*). Допустимые значения энергии трех мод с частотами ν_1, ν_2 и ν_3 (*б*). Сплошные кружки представляют число фотонов в каждой моде: моды 1, 2 и 3 содержат 2, 1 и 3 фотона соответственно

В кубическом резонаторе со стороной d удобным выбором пространственных функций для разложения является набор стоячих волн

$$U_q(\boldsymbol{r}) = \left(\frac{2}{d}\right)^{3/2} \sin\left(q_x \frac{\pi}{d} x\right)\sin\left(q_y \frac{\pi}{d} y\right)\sin\left(q_z \frac{\pi}{d} z\right), \tag{12.2}$$

где q_x, q_y и q_z — целые числа, все вместе обозначаемые индексом $q = (q_x, q_y, q_z)$ (см. разд. 10.3 и рис. 12.2, *а*). В соответствии с (5.71) плотность энергии моды q равна

$$q = \frac{1}{2}\varepsilon |A_q|^2 |U_q(r)|^2$$

так что энергия, содержащаяся в моде, составляет

$$E_q = \frac{1}{2}\varepsilon\int_V |A_q|^2 |U_q(\boldsymbol{r})|^2 \, \mathrm{d}\boldsymbol{r} = \frac{1}{2}\varepsilon |A_q|^2 \tag{12.3}$$

где V — модовый объем. В классической электромагнитной теории энергия E_q может принимать произвольные неотрицательные значения, которые могут быть как угодно малыми. Полная энергия равна сумме энергий всех мод.

Фотонно-оптическая теория света в резонаторе

Электромагнитно-оптическая теория, описанная выше, поддерживается в фотонной оптике, однако на энергию, которую может иметь каждая из мод, накладывается ограничение. Вместо произвольных значений из непрерывного диапазона энергия должна принимать дискретный набор значений, отстоящих друг от друга на определенную величину. Энергия моды, таким образом, может принимать только целые кратные значения этой величины. Каждая порция энергии переносится одним фотоном; мода может нести произвольное число фотонов.

Свет в резонаторе состоит из набора мод, каждая из которых содержит целое число одинаковых фотонов. Характеристики моды, такие как частота, пространственное распределение, направление распространения и поляризация, приписываются фотону.

12.1.1. Энергия фотона

В фотонной оптике энергия моды электромагнитного поля квантуется на дискретные уровни, отличающиеся на энергию фотона (см. рис. 12.2). Энергия фотона в моде с частотой ν равна

$$\boxed{E = h\nu = \hbar\omega,} \tag{12.4}$$

Энергия фотона

где $h = 6{,}63 \cdot 10^{-34}$ Дж с — постоянная Планка; $\hbar \equiv h/2\pi$. Энергия может отдаваться или получаться модой только порциями $h\nu$.

Мода, не содержащая фотонов, тем не менее обладает энергией $E = h\nu/2$, которая называется нулевой энергией. При наличии фотонов полная энергия моды

$$E_n = \left(n + \frac{1}{2}\right)h\nu, \quad n = 0,\ 1,\ 2, \tag{12.5}$$

В большинстве экспериментов нулевая энергия непосредственно не наблюдаема, поскольку измеряются только разности энергий [например, $E_2 - E_1$ в (12.5)]. Однако присутствие нулевой энергии проявляется в слабых эффектах

при воздействии статических полей на вещество. Наличие нулевой энергии играет также определяющую роль в процессе спонтанного излучения света атомом, что обсуждается в гл. 13.

Порядок величины энергии фотона легко оценить. Инфракрасный фотон с длиной волны $\lambda_0 = 1$ мкм в свободном пространстве имеет частоту

$$\nu = \frac{c_0}{\lambda_0} = 3 \cdot 10^{14} \text{ Гц}.$$

Следовательно, его энергия составляет

$$h\nu = 1{,}99 \ 10^{-19} \text{ Дж} = \frac{1{,}99 \cdot 10^{19}}{1{,}6 \cdot 10^{-19}} \text{ эВ} = 1{,}24 \text{ эВ (электронвольт)},$$

что равно кинетической энергии электрона, ускоренного разностью потенциалов 1,24 В. Формула перевода длины волны (мкм) в энергию фотона (эВ) имеет простой вид

$$E \text{ (эВ)} = \frac{1{,}24}{\lambda_0 \text{ (мкм)}}. \tag{12.6}$$

В качестве другого примера возьмем микроволновый фотон с длиной волны $\lambda_0 = 1$ см. Его энергия в 10 раз меньше, а именно $1{,}24 \cdot 10^{-4}$ эВ.

Обратная длина волны часто используется в качестве единицы энергии, особенно в химии. Она выражается в см^{-1} и определяется путем выражения длины волны в см и простого вычисления обратной величины. Так, 1 см^{-1} соответствует 1,24/10 000 эВ, а 1 эВ — 8068,1 см^{-1}. Связь между частотой, длиной волны, энергией и обратной длиной волны фотона иллюстрирует рис. 12.3 (см. цв. вклейку).

Поскольку фотоны больших частот несут большую энергию, корпускулярная природа света становится все более важной по мере того, как частота его растет. Волновые эффекты, такие как интерференция и дифракция, трудно наблюдать, когда длина волны становится короче. Рентгеновское и гамма-излучения почти всегда ведут себя как набор частиц, в противоположность радиоволнам, которые почти всегда ведут себя как волны. Частота света в оптическом диапазоне такова, что появляются и корпускулярные, и волновые свойства, что делает оптику фотонов особенно необходимой.

12.1.2. Поляризация фотона

Как отмечалось выше, свет характеризуется набором мод с различными частотами, направлениями и поляризациями, каждая из которых может заниматься определенным числом фотонов. Для каждой монохроматической волны, распространяющейся в некотором направлении, имеется две поляризационные моды. Поляризация фотона — это поляризация моды, к которой он

относится. Например, фотон может быть линейно поляризован в направлении x или поляризован по правому кругу.

Поскольку поляризационные моды свободного пространства вырождены, они не единственны. Можно использовать моды с линейной поляризацией вдоль осей x и y, с линейной поляризацией вдоль двух других взаимно-ортогональных направлений x' и y', или моды с правой и левой круговой поляризацией. Выбор того или иного базиса — дело удобства. Проблема возникает, когда фотон, относящийся к данной поляризационной моде (скажем, поляризованный в направлении x), нужно наблюдать в другом базисе (скажем, базисе линейных поляризаций вдоль осей x' и y'). Поскольку энергия фотона не может делиться между двумя модами, необходима вероятностная интерпретация.

В классической электромагнитной оптике состояние поляризации описывается вектором Джонса, компоненты которого (A_x, A_y) являются компонентами комплексной огибающей в направлениях x и y (см. подразд. 6.1.1). Та же самая волна может быть представлена в другой системе координат (x', y'), например, повернутой на угол 45° относительно первоначальной, с компонентами вектора Джонса

$$A_{x'} = \frac{1}{\sqrt{2}}\left(A_x - A_y\right); \quad A_{y'} = \frac{1}{\sqrt{2}}\left(A_x + A_y\right), \tag{12.7}$$

как описано в подразд. 6.1.2. Волна, линейно поляризованная в направлении x, описывается вектором Джонса с компонентами $(A_0, 0)$ в системе координат x—y, где A_0 — комплексная огибающая. В системе координат (x', y') вектор Джонса имеет компоненты $\left(\frac{1}{\sqrt{2}}A_0, \frac{1}{\sqrt{2}}A_0\right)$.

> В фотонной оптике состояние поляризации одного фотона описывается вектором Джонса с комплексными компонентами (A_x, A_y), нормированными условием $|A_x|^2 + |A_y|^2 = 1$. Коэффициенты A_x и A_y интерпретируются как комплексные амплитуды вероятности, а квадраты их модулей $|A_x|^2$ и $|A_y|^2$ — как вероятности того, что фотон будет наблюдаться в поляризационной моде с линейной поляризацией вдоль осей x и y, соответственно.

Компоненты (A_x, A_y) преобразуются от одной системы координат к другой как компоненты обычного вектора Джонса, и новые компоненты представляют собой комплексные амплитуды вероятности для новых мод. Таким образом, один фотон может существовать с определенной вероятностью более чем в одной моде. Эта концепция иллюстрируется следующими примерами.

Линейно поляризованные фотоны

Фотон поляризован линейной в направлении x. В базисе состояний линейной поляризации x—y такой фотон описывается вектором Джонса с компонентами (1, 0). В базисе линейно поляризованных мод, направления по-

ляризации которых x', y' повернуты на 45°, компоненты этого вектора равны $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, так что вероятность наблюдения фотона, поляризованного линейно вдоль любой из осей, равна 1/2. Это схематически изображено на рис. 12.4.

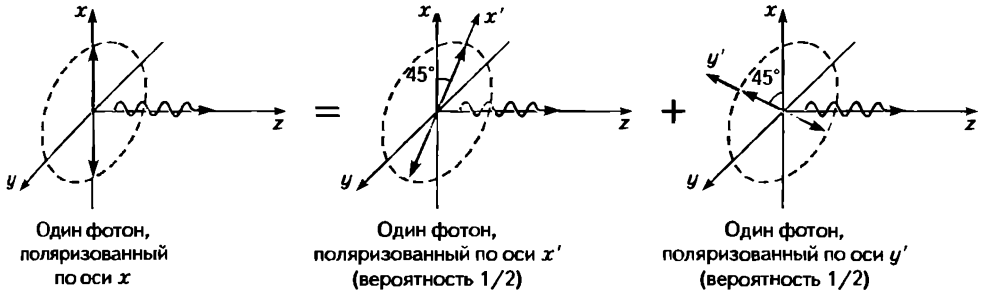


Рис. 12.4. Фотон, принадлежащий моде, поляризованной по оси x — то же самое, что фотон в состоянии, представляющем собой суперпозицию мод с поляризациями по x' и y' с вероятностью каждой 1/2

Пример 12.1

Прохождение линейно поляризованного фотона через поляризатор

Рассмотрим прохождение фотона, линейно поляризованного в направлении x, через поляризатор, ось которого x' образует угол θ с осью x, как показано на рис. 12.5. Поляризатор пропускает свет, поляризованный вдоль оси x', и не пропускает свет, поляризованный вдоль ортогонального направления y' Для определения вероятности прохождения фотона через поляризатор запишем вектор Джонса поляризационного состояния фотона в системе координат (x', y') как $(\cos\theta, -\sin\theta)$ [см. (6.21) и (6.22)]. Вероятность наблюдения фотона в моде с линейной поляризацией равна $\cos^2\theta$, что представляет собой вероятность прохождения фотона через поляризатор $p(\theta) = \cos^2\theta$. Из классической оптики известно, что коэффициент пропускания поляризатора по интенсивности в этом случае равен $\cos^2\theta$ (см. подразд. 6.1.2). Это говорит о том, что вероятность пропускания одного фотона равна классическому коэффициенту пропускания, $p(\theta) = \mathcal{T}(\theta)$.

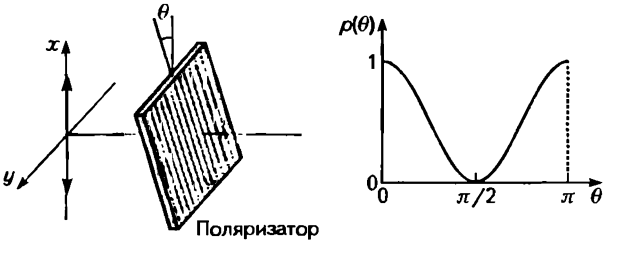


Рис. 12.5. Вероятность прохождения линейно поляризованного фотона через поляризатор. Ось поляризатора образует угол θ с направлением поляризации фотона

Фотоны с круговой поляризацией

Фотон, поляризованный по кругу, описывается вектором Джонса с компонентами $\frac{1}{\sqrt{2}}(1, \pm j)$, где знаки «+» и «—» соответствуют правой и левой поляризациям. Такое описание основано на использовании системы координат (x, y), т. е. линейно поляризованных мод. Следовательно, вероятность прохождения одного фотона через линейный поляризатор, ось которого направлена по оси x или y, равна 1/2. Можно также показать, что вероятность будет такой же при любом направлении оси поляризатора. Таким образом, можно рассматривать фотон с круговой поляризацией как вероятностную суперпозицию фотонов с линейной поляризацией по осям x и y с вероятностью каждой 1/2.

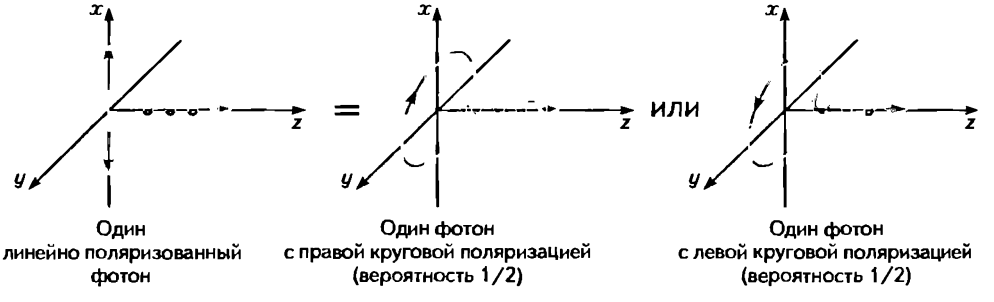


Рис. 12.6. Линейно поляризованный фотон эквивалентен суперпозиции фотонов с правой и левой круговой поляризацией с вероятностью 1/2

Состояния с правой и левой круговой поляризацией также можно использовать как моды (систему координат). В таком описании линейно поляризованный фотон можно рассматривать как вероятностную суперпозицию фотонов с правой и левой круговой поляризацией с вероятностью каждой 1/2, как показано на рис. 12.6.

12.1.3. Положение фотона

С каждым фотоном частоты ν связана волна, описываемая комплексной волновой функцией моды $U(\boldsymbol{r})\exp(j2\pi\nu t)$. Однако, когда фотон взаимодействует с детектором малой площади dA, расположенным перпендикулярно направлению распространения в точке $\boldsymbol{r}$, он либо целиком детектируется, либо нет. Положение, в котором регистрируется фотон, не определено однозначно. Оно определяется интенсивностью света $I(\boldsymbol{r}) \propto |U(\boldsymbol{r})|^2$ в соответствии со следующим вероятностным законом.

Вероятность $p(\boldsymbol{r})\mathrm{d}A$ наблюдения фотона в точке $\boldsymbol{r}$ в пределах бесконечно малой площади $\mathrm{d}A$ в любой момент времени пропорциональна локальной интенсивности света $I(\boldsymbol{r}) \propto |U(\boldsymbol{r})|^2$, так что

$$p(\boldsymbol{r})\mathrm{d}A \propto I(\boldsymbol{r})\mathrm{d}A. \tag{12.8}$$

Положение фотона

Таким образом, фотон можно с большей вероятностью обнаружить там, где выше интенсивность света. Например, фотон, относящийся к моде, описываемой стоячей волной с распределением интенсивности

$$I(x,\ y,\ z) \propto \sin^2\left(\frac{\pi z}{d}\right),$$

где $0 \le z \le d$, будет наиболее вероятно обнаружен в точке $z = d/2$, но никогда — в точках $z = 0$ или $z = d$. В противоположность волнам, распределенным непрерывно в пространстве, и частицам, сосредоточенным в точках, фотоны ведут себя *одновременно* как распределенные и сосредоточенные объекты. Такое поведение называется **корпускулярно-волновым дуализмом**. Локализованная природа фотона становится очевидной при его детектировании.

Упражнение 12.1

Фотоны в гауссовом пучке

а. Рассмотрим один фотон, описываемый гауссовым пучком (модой $\mathrm{TEM}_{0,0}$ резонатора со сферическими зеркалами, см. подразд. 3.1.2, 5.4.1 и 10.2.2). Какова вероятность детектирования фотона внутри круга, радиус которого равен радиусу перетяжки пучка W_0? Напомним, что в перетяжке ($z = 0$)

$$I(\rho,\ z = 0) \propto \exp\left(-\frac{2\rho^2}{W_0^2}\right),$$

где ρ — радиальная координата.

б. При условии, что пучок содержит большое число n независимых фотонов, оцените среднее число фотонов внутри указанного круга.

Прохождение отдельного фотона через светоделитель

Идеальный светоделитель — оптическое устройство, которое расщепляет пучок света на два пучка, выходящих под прямым углом друг к другу. Он характеризуется коэффициентом пропускания $\mathcal{T}$ и коэффициентом отражения $\mathcal{R} = 1 - \mathcal{T}$ по интенсивности. Интенсивность прошедшей волны I_t и интенсивность отраженной волны I_r можно вычислить, зная интенсивность падающей волны I, по формулам электромагнитной теории $I_t = \mathcal{T}I$ и $I_r = (1 - \mathcal{T})I$.

Поскольку фотон неделим, он должен выбирать между двумя возможными направлениями, предоставляемыми светоделителем. Единичный фотон, падающий на делитель, следует одному из двух возможных путей в соответствии с вероятностным правилом (12.8), определяющим его положение. Вероятность прохождения фотона пропорциональна I_t и, следовательно, равна $\mathcal{T} = I_t/I$, а вероятность отражения $1 - \mathcal{T} = I_r/I$. С точки зрения теории вероятностей задача идентична задаче о бросании несимметричной монеты (рис. 12.7).

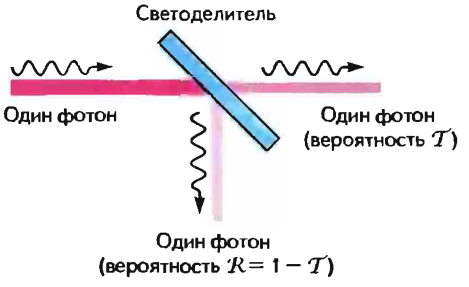


Рис. 12.7. Вероятностное отражение и пропускание фотона на светоделителе

12.1.4. Импульс фотона

Как обсуждалось в подразд. 5.4.1, в классической электромагнитной оптике плоская электромагнитная волна обладает импульсом, объемная плотность которого равна $(W/c)\hat{\boldsymbol{k}}$, где W — объемная плотность энергии; $\hat{\boldsymbol{k}}$ — единичный вектор в направлении волнового вектора $\boldsymbol{k}$.

В оптике фотонов импульс фотона равен

$$\boldsymbol{p} = \frac{E}{c}\hat{\boldsymbol{k}},$$

где $E = \hbar\omega = \hbar ck$ — энергия фотона. Следовательно:

Импульс, связанный с фотоном, принадлежащим к моде в виде плоской волны с волновым вектором $\boldsymbol{k}$, составляет

$$\boldsymbol{p} = \hbar\boldsymbol{k}. \tag{12.9}$$

Его модуль

$$p = \hbar k = \frac{\hbar\omega}{c} = \frac{\hbar 2\pi}{\lambda},$$

так что

$$p = \frac{E}{c} = \frac{h}{\lambda}. \tag{12.10}$$

*Импульс локализованной волны

Волна более общего вида, чем плоская волна, с комплексной волновой функцией вида $U(\boldsymbol{r})\exp(j2\pi\nu t)$, может быть разложена в сумму плоских волн с различными волновыми векторами с помощью методов Фурье-оптики

(см. гл. 4). Компонента с волновым вектором $\boldsymbol{k}$ может быть записана в виде $A(\boldsymbol{k}) \exp(-j\boldsymbol{k} \cdot \boldsymbol{r}) \exp(j2\pi\nu t)$, где $A(\boldsymbol{k})$ — амплитуда.

Импульс фотона, описываемого произвольной комплексной волновой функцией $U(\boldsymbol{r}) \exp(j2\pi\nu t)$, не определен. Он принимает значения (12.9) с вероятностью, пропорциональной $|A(\boldsymbol{k})|^2$, где $A(\boldsymbol{k})$ — амплитуда Фурье-компоненты $U(\boldsymbol{r})$ с волновым вектором $\boldsymbol{k}$.

Если $f(x, y) = U(x, y, 0)$ — комплексная амплитуда в плоскости $z = 0$, то Фурье-компонента с волновым вектором $\boldsymbol{k} = (k_x, k_y, k_z)$ имеет амплитуду

$$A(\boldsymbol{k}) = F\left(\frac{k_x}{2\pi}, \frac{k_y}{2\pi}\right),$$

где $F(\nu_x, \nu_y)$ — двумерный Фурье-образ функции $f(x, y)$. (см. гл. 4). Поскольку функции $f(x, y)$ и $F(\nu_x, \nu_y)$ образуют пару, связанную Фурье-преобразованием, их ширины обратно пропорциональны друг другу и связаны так же, как длительность и ширина полосы импульса [см. приложение А, формулу (А.13)]. В результате устанавливается соотношение неопределенностей между положением фотона и направлением его импульса, поскольку положение фотона в плоскости $z = 0$ вероятностным образом определяется величиной $|U(\boldsymbol{r})|^2 = |f(x, y)|^2$, а направление импульса вероятностным образом определяется функцией

$$|A(\boldsymbol{k})|^2 = \left|F\left(\frac{k_x}{2\pi}, \frac{k_y}{2\pi}\right)\right|^2$$

Таким образом, если в плоскости $z = 0$ σ_x является неопределенностью положения в направлении x, а σ_θ — неопределенность угла отклонения от оси z, который предполагается малым,

$$\sigma_\theta = \arcsin\left(\frac{\sigma_{kx}}{k}\right) \approx \frac{\lambda}{2\pi}\sigma_{kx},$$

то соотношение неопределенностей $\sigma_x\sigma_{kx} \geq 1/2$ эквивалентно $\sigma_x\sigma_\theta \geq \lambda/4\pi$.

Фотон плоской волны имеет известный импульс (фиксированной величины и направления), так что $\sigma_\theta = 0$, но его положение является полностью неопределенным ($\sigma_x = \infty$). Он с равной вероятностью может быть обнаружен в любой точке плоскости $z = 0$. При прохождении фотона плоской волны через апертуру его поперечное положение становится определенным за счет появления поперечного импульса, т. е. расходимости пучка. Таким образом, соотношение неопределенностей между координатой и импульсом фотона повторяет выводы теории дифракции, описанной в гл. 4. Противоположным плоской волне предельным случаем является фотон сферической волны. Он точно локализован в центре волны, но его импульс является полностью неопределенным.

Радиационное давление

Поскольку фотон переносит импульс и импульс сохраняется, то атом, испускающий фотон, испытывает отдачу величины $h\nu/c$. Более того, импульс, связанный с фотоном, может передаваться объектам конечной массы, создавая силу, вызывающую механическое движение. Например, лучи света можно использовать для отклонения атомных пучков, направленных перпендикулярно потоку фотонов. Для описания этого явления часто употребляется термин **радиационное давление** или **давление света** (давление = сила/площадь).

Упражнение 12.2

Эффект отдачи

Рассчитайте скорость отдачи атома ^{198}Hg, испустившего фотон с энергией 4,88 эВ. Сравните результат со среднеквадратичной тепловой скоростью v атомов при температуре $T = 300$ °K (которая получается, если приравнять среднюю кинетическую энергию к средней тепловой энергии, т. е.

$$\frac{mV^2}{2} = \frac{3k_B T}{2},$$

где $k_B = 1{,}38 \cdot 10^{-23}$ Дж/К — постоянная Больцмана).

Спиновый момент импульса фотона

Фотоны обладают собственным внутренним моментом импульса, который связан с состояниями круговой поляризации. Величина спина фотона принимает два значения

$$\boxed{\mathbb{S} = \pm\hbar,} \qquad (12.11)$$

Спин фотона

где знак «плюс»/«минус» связан с правой/левой круговой поляризацией. Вектор спина параллелен (антипараллелен) вектору импульса или, что то же самое, волновому вектору. Линейно поляризованный фотон имеет равную вероятность параллельной и антипараллельной ориентации спина.

Так же, как фотоны переносят импульс, фотоны с круговой поляризацией сообщают объекту крутящий момент. Например, фотон с круговой поляризацией сообщает крутящий момент полуволновой пластинке.

Орбитальный момент импульса фотона

Кроме спинового момента импульса, связанного с круговой поляризацией, электромагнитная волна может нести угловой момент, связанный с ее пространственным распределением. Например, пучок Гаусса—Лагерра, описываемый волновой функцией $U_{l,m}(\rho, \phi, z)$ [формула (3.69)] с азимутальной зависимостью $\exp(jl\phi)$ и винтовым волновым фронтом, имеет момент импульса

(при $l \neq 0$), не зависящий от состояния поляризации. Чтобы отличить его от спинового момента, его называют **орбитальным моментом импульса**. Фотон в указанной пространственной моде имеет орбитальный момент импульса $L = l\hbar$.

Другой пример — фотон моды типа шепчущей галереи в цилиндрическом резонаторе (см. подразд. 10.3.2). С точки зрения лучевой оптики такая мода описывается лучом, распространяющимся вдоль круговой границы резонатора. В рамках волновой оптики длина такой волны удовлетворяет условию $2\pi a = q\lambda$, где a — радиус окружности; $q = 1, 2,$ Импульс фотона тогда равен

$$p = \hbar k = \frac{\hbar 2\pi}{\lambda} = \frac{q\hbar}{a},$$

а его орбитальный момент импульса $ap = q\hbar$. Фотон аналогичной моды в микросферическом резонаторе (см. подразд. 10.4.3) радиусом a имеет орбитальный момент $L = \ell\hbar$, где целое число ℓ связано с резонансной длиной волны для оптического пути вдоль границы большого круга сферы. Это число можно рассматривать как квантовое число орбитального углового момента, которое используется при описании атома водорода (см. подразд. 13.1.1).

12.1.5. Интерференция фотона

Опыт Юнга с интерференцией от двух точечных отверстий обычно служит классическим примером для демонстрации интерференции света (см. упражнение 2.10). На самом деле опыт Юнга можно провести даже в том случае, когда в данный момент времени в установке имеется лишь один фотон. В рамках фотонной оптики результат этого эксперимента можно понять с помощью правила определения положения фотона. Интенсивность в плоскости наблюдения рассчитывается с помощью электромагнитной (волновой) оптики, а результат преобразуется в функцию плотности вероятности, определяющую случайное положение детектируемого фотона. Интерференция возникает из разности фаз, связанной с двумя возможными оптическими путями.

Рассмотрим плоскую волну, освещающую экран с двумя точечными отверстиями, как показано на рис. 12.8. С другой стороны экрана получаются две сферические волны, интерферирующие в плоскости наблюдения. В приближении параболоидальных волн получается синусоидальное распределение интенсивности, описываемое формулой (см. упражнение 2.10):

$$I(x) \approx 2I_0\left(1 + \cos\frac{2\pi x\theta}{\lambda}\right), \tag{12.12}$$

где I_0 — интенсивность каждой из волн в плоскости наблюдения; λ — длина волны; θ — угол, под которым видны два отверстия с плоскости наблюдения (см. рис. 12.8). Линия, соединяющая отверстия, принята за ось x. Формула (12.12) описывает картину интенсивности, экспериментально наблюдаемую при сильном падающем свете.

Теперь пусть в установке присутствует лишь один фотон. Тогда, в соответствии с (12.8), вероятность его детектирования в положении x пропорциональна $I(x)$. Детектирование наиболее вероятно произойдет при тех значениях x, для которых $I(x)$ максимальна. Фотон никогда не будет обнаружен в точках, где $I(x) = 0$. Если построить гистограмму положений детектируемого фотона при многократном повторении эксперимента, как сделал Тейлор в 1909 г., то получится классическая интерференционная картина, которая наблюдается в однократном эксперименте с интенсивным светом. Таким образом, интерференционная картина представляет собой распределение вероятности положений, в которых обнаруживается фотон.

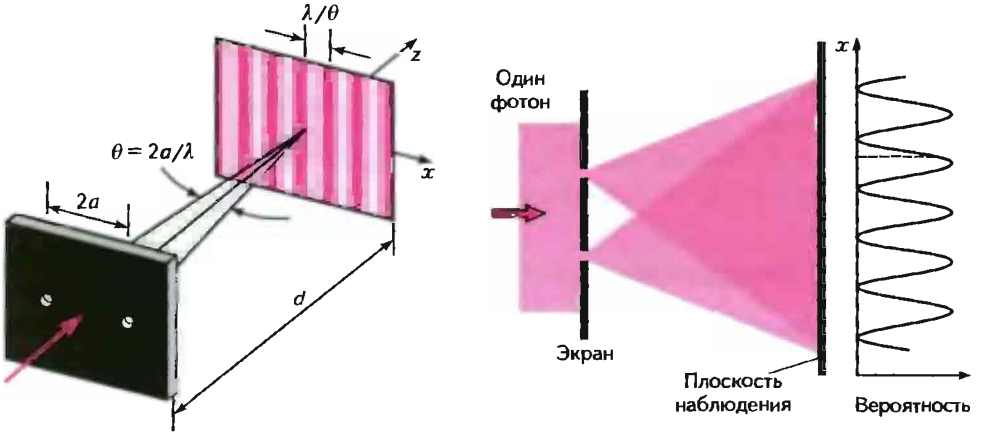


Рис. 12.8. Эксперимент Юнга по интерференции от двух точечных отверстий с использованием единичного фотона. Распределение интенсивности $I(x)$ определяет плотность вероятности детектирования фотона в точке x

Интерференция возможна благодаря пространственно протяженной природе фотона, которая позволяет ему проходить через *оба* отверстия в экране. Это сообщает ему «знание» полной геометрии эксперимента к моменту достижения плоскости наблюдения, где он обнаруживает себя как единая сущность. Если одно из отверстий перекрыть, интерференция пропадет, поскольку фотон будет вынужден проходить только через другое отверстие, что лишает его «знания» всей экспериментальной установки.

Упражнение 12.3

Единичный фотон в интерферометре Маха—Цендера

Рассмотрим плоскую световую волну с длиной волны λ, которая расщепляется на две части светоделителем (см. подразд. 12.1.3) с последующим их совмещением в интерферометре Маха—Цендера, как показано на рис. 12.9 (см. также рис. 2.24, *а*). Для случая, когда волна содержит всего один фотон, постройте график вероятности обнаружения его детектором как функции d/λ

(для $0 \le d/\lambda \le 1$), где d — разность двух оптических длин пути света. Считайте, что зеркала и светоделители идеально плоские и не вносят потерь, причем у светоделителей $\mathcal{T} = \mathcal{R} = 1/2$. Где мог бы находиться фотон, когда вероятность его обнаружения на детекторе не равна единице?

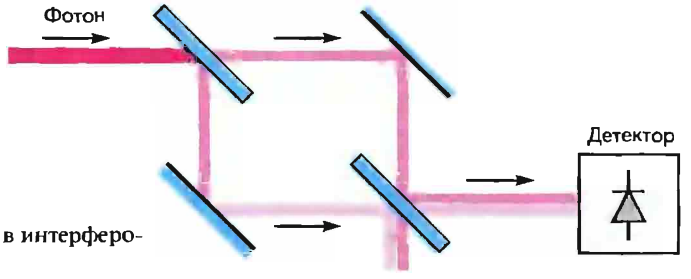


Рис. 12.9. Единичный фотон в интерферометре Маха—Цендера

12.1.6. Временна́я локализация фотона

Выражение для моды (12.1) представляет монохроматическую (одночастотную) волну, которая является «вечной» гармонической функцией времени. Фотон в монохроматической моде с одинаковой вероятностью может быть обнаружен в любой момент времени. Однако, как отмечалось выше, разложение поля излучения внутри (или вне) резонатора по модам не единственно. Более общее разложение может быть выполнено по полихроматическим модам (например, локализованным во времени волновым пакетам). Вероятность детектирования фотона, описываемого комплексной волновой функцией $U(\boldsymbol{r}, t)$ (см. подразд. 2.6.1) в любом месте пространства в течение бесконечно малого интервала времени между t и $t + \mathrm{d}t$, пропорциональна $I(\boldsymbol{r}, t)\mathrm{d}t \propto |U(\boldsymbol{r}, t)|^2 \mathrm{d}t$.

Таким образом, правило (12.8), определяющее пространственное положение фотона, обобщается так, чтобы включить и его локализацию во времени:

Вероятность наблюдения фотона в точке $\boldsymbol{r}$ в пределах бесконечно малой области $\mathrm{d}A$ и в течение бесконечно малого интервала времени между t и $t + \mathrm{d}t$, пропорциональна интенсивности моды при $\boldsymbol{r}$ и t, т. е.

$$\boxed{p(\boldsymbol{r}, t)\,\mathrm{d}A\,\mathrm{d}t \propto I(\boldsymbol{r}, t)\,\mathrm{d}A\,\mathrm{d}t \propto |U(\boldsymbol{r}, t)|^2\,\mathrm{d}A\,\mathrm{d}t.} \qquad (12.13)$$

Положение фотона в пространстве и времени

Соотношение неопределенностей между временем и энергией

Время обнаружения фотона в монохроматической моде частоты ν является полностью неопределенным, в то время как его частота ν (и, следовательно, энергия $h\nu$), напротив, полностью определена. С другой стороны, фо-

тон из моды в виде волнового пакета с функцией интенсивности $I(t)$ шириной σ_t должен быть локализован в этом промежутке времени. Ограничение временно́го промежутка, в котором существует фотон, приводит к неопределенности его частоты и, следовательно, энергии, что вытекает из свойств преобразования Фурье. Результатом является полихроматический фотон. Для простоты исключая зависимость от r, легко найдем неопределенность частоты посредством разложения функции $U(t)$ по гармоникам

$$U(t) = \int_{-\infty}^{\infty} V(\nu)\exp(j2\pi\nu t)\,\mathrm{d}\nu, \tag{12.14}$$

где $V(\nu)$ — Фурье-образ функции $U(t)$ (см. разд. А.1 приложения А). Ширина σ_ν функции $|V(\nu)|^2$ представляет собой ширину спектра. Если σ_t — среднеквадратичная ширина функции $|U(t)|^2$, то σ_t и σ_ν должны удовлетворять условию обратной пропорциональности длительности импульса и его спектральной ширины $\sigma_t\sigma_\nu \geq \pi/4$ или, что равносильно, $\sigma_t\sigma_\omega \geq 1/2$ (строгие определения σ_t и σ_ν, приводящие к этому соотношению неопределенностей, приведены в разд. А.2 приложения А).

Таким образом, энергия фотона $\hbar\omega$ не может быть определена с точностью, превышающей $\sigma_E = \hbar\sigma_\omega$. Следовательно, между неопределенностью энергии фотона и временем, в течение которого он может быть зарегистрирован, существует соотношение

$$\boxed{\sigma_E\sigma_t \geq \frac{\hbar}{2},} \tag{12.15}$$

Соотношение неопределенностей «время—энергия»

которое известно как **соотношение неопределенностей «время—энергия»**. Это соотношение аналогично соотношению между неопределенностями координаты и импульса, которое устанавливает предел точности, с которой обе эти величины могут быть измерены одновременно. Средняя энергия полихроматического фотона равна $\bar{E} = h\bar{\nu} = \hbar\bar{\omega}$.

Итак, монохроматический фотон ($\sigma_\nu \to 0$) имеет бесконечное время, в течение которого он может быть зарегистрирован ($\sigma_t \to \infty$). Фотон же, связанный с оптическим волновым пакетом, локализован во времени и, следовательно, полихроматичен с соответствующей неопределенностью в энергии. Таким образом, фотон волнового пакета можно рассматривать как локализованный распространяющийся сгусток энергии.

Упражнение 12.4

Единичный фотон гауссова волнового пакета

Рассмотрим пакет из плоских волн (см. подразд. 2.6.1), содержащий один фотон, распространяющийся в направлении z и имеющий комплексную волновую функцию

$$U(\boldsymbol{r}, t) = a\left(t - \frac{z}{c}\right), \tag{12.16}$$

где

$$a(t) = \exp\left(-\frac{t^2}{4\tau^2}\right)\exp(j2\pi\nu_0 t). \tag{12.17}$$

а. Покажите, что неопределенности времени и координаты z равны $\sigma_t = \tau$ и $\sigma_z = c\sigma_t$ соответственно.

б. Покажите, что неопределенности энергии и импульса удовлетворяют минимальным соотношениям неопределенности

$$\sigma_E \sigma_t = \frac{\hbar}{2}; \tag{12.18}$$

$$\sigma_z \sigma_p = \frac{\hbar}{2}. \tag{12.19}$$

Равенство (12.19) есть минимальный предел соотношения неопределенностей Гейзенберга между координатой и импульсом [см. (А.14) в приложении А].

Резюме

Электромагнитное излучение можно описать как сумму мод, например, однородных монохроматических плоских волн вида

$$\boldsymbol{E}(\boldsymbol{r}, t) = \sum_{q} A_q \exp(-j\boldsymbol{k}_q \cdot \boldsymbol{r})\exp(j2\pi\nu_q t)\hat{\boldsymbol{e}}_q. \tag{12.20}$$

Каждая плоская волна имеет два ортогональных состояния поляризации (например, состояния вертикальной и горизонтальной линейной поляризации, правой и левой круговой поляризации), описываемые векторами $\hat{\boldsymbol{e}}_q$. При измерении энергии моды получается целое (в общем случае случайное) число квантов энергии — фотонов. Каждый фотон, связанный с модой $\boldsymbol{q}$, обладает следующими свойствами:

- энергией $E = h\nu_q$;
- импульсом $\boldsymbol{p} = \hbar\boldsymbol{k}_q$;
- спином $\mathbb{S} = \pm\hbar$ при наличии круговой поляризации;
- фотон с равной вероятностью может быть обнаружен во всех точках пространства и в любой момент времени, поскольку волновая функция моды есть плоская монохроматическая волна.

Выбор мод не однозначен. Модовое разложение по немонохроматическим (квазимонохроматическим) и неплоским волнам также возможно

$$\boldsymbol{E}(\boldsymbol{r}, t) = \sum_{q} A_q U_q(\boldsymbol{r}, t)\hat{\boldsymbol{e}}_q. \tag{12.21}$$

Тогда фотон, связанный с модой $\boldsymbol{q}$, обладает следующими свойствами.

- Положение фотона в пространстве и времени управляется комплексной волновой функцией $U_q(\boldsymbol{r}, t)$. Вероятность обнаружения фотона в те-

чение бесконечно малого промежутка времени от t до $t + dt$ в пределах бесконечно малой области пространства dA вблизи точки $\boldsymbol{r}$ пропорциональна $|U_q(\boldsymbol{r}, t)|^2 dA\, dt$.

- Если $U_q(\boldsymbol{r}, t)$ имеет конечную временну́ю протяженность σ_t, т. е. фотон локализован во времени, то его энергия имеет неопределенность $h\sigma_\nu \geq h/4\pi\sigma_t$.
- Если $U_q(\boldsymbol{r}, t)$ имеет конечную пространственную протяженность в поперечной ($z = 0$) плоскости, т. е. если фотон локализован, например, в направлении x, то направление импульса фотона является неопределенным. Разброс момента фотона можно определить, разлагая $U_q(\boldsymbol{r}, t)$ на сумму плоских волн, волновые векторы которых $\boldsymbol{k}$ соответствуют импульсам $\hbar\boldsymbol{k}$. Пространственная локализация фотона в поперечной плоскости приводит к неопределенности направления импульса фотона.

12.2. ПОТОКИ ФОТОНОВ

В разд. 12.1 мы сосредоточили свое внимание на свойствах и поведении отдельных фотонов. Рассмотрим теперь свойства света, состоящего из множества фотонов. В результате процессов, создающих фотоны (например, испускания фотонов атомами), число фотонов в каждой моде, как правило, случайно. Распределение вероятностей, которому подчиняется число фотонов, определяется квантовым состоянием света (см. разд. 12.3). Потоки фотонов часто содержат множество распространяющихся мод, каждая из которых несет случайное число фотонов.

Если поставить эксперимент, в котором на светочувствительную поверхность падает слабый поток фотонов, то фотоны будут регистрироваться в случайные моменты времени и в случайных точках поверхности в соответствии с (12.13). Этот пространственно-временно́й процесс при слабом освещении объекта можно разглядеть невооруженным адаптированным к темноте глазом.

Временна́я картина такой регистрации фотонов может быть подчеркнута посредством раздельного изучения пространственного и временно́го поведения. Пусть детектор с хорошим временны́м разрешением собирает свет с конечной площади A, как показано на рис. 12.10. Уравнение (12.13) говорит нам, что вероятность детектирования фотона в течение бесконечно малого промежутка времени от t до $t + dt$ пропорциональна оптической мощности в момент t:

$$P(t) = \int_A I(\boldsymbol{r}, t)\, dA.$$

Фотоны регистрируются в случайные моменты времени.

С другой стороны, *пространственная картина* регистрации фотонов легко проявляется при использовании детектора с хорошим пространственным раз-

решением, интегрирующим свет за некоторое время экспозиции T (например, фотографической пленки). В соответствии с (12.13) вероятность наблюдения фотона в бесконечно малой пространственной области dA вблизи точки $\boldsymbol{r}$ пропорциональна проинтегрированной локальной интенсивности $\int_0^T I(\boldsymbol{r}, t)\,dt$. Иллюстрацией служит зернистая фотография Макса Планка на рис. 12.11. Это изображение представляет фотографическую репродукцию портрета Макса Планка на первой странице данной главы, полученную при очень слабом освещении. Каждая светлая точка на фотографии представляет случайный акт регистрации фотона; плотность числа таких актов следует распределению локальной интенсивности.

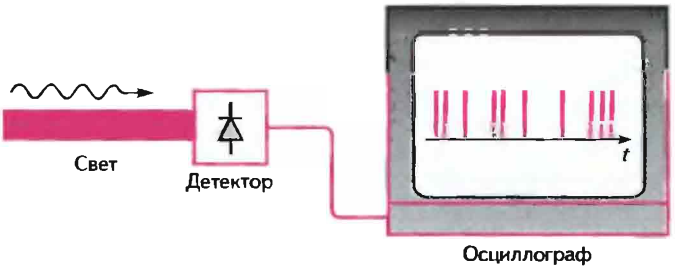


Рис. 12.10. Регистрация фотонов в случайные моменты времени детектором, собирающим свет, падающий на поверхность A

Рис. 12.11. Случайные акты регистрации фотонов имеют пространственную плотность, повторяющую распределение локальной оптической интенсивности. Это изображение Макса Планка, полученное с помощью разреженного потока фотонов, следует сравнить с фотографией в интенсивном свете на первой странице данной главы

12.2.1. Средний поток фотонов

Начнем с ряда определений, связывающих средний поток фотонов с классической интенсивностью электромагнитного излучения, его мощностью и энергией. Эти определения вытекают из формулы (12.13), определяющей место и время наблюдения отдельного фотона. Затем мы обсудим случайность фотонного потока и статистику числа фотонов для различных источников света. Наконец, мы рассмотрим случайное деление фотонного потока светоделителем или детектором.

Средняя плотность потока фотонов

Монохроматический свет частоты ν с классической интенсивностью $I(\boldsymbol{r})$ (Вт/см^2) характеризуется средней плотностью потока фотонов

$$\boxed{\phi(\boldsymbol{r}) = \frac{I(\boldsymbol{r})}{h\nu}.} \qquad (12.22)$$

Средняя плотность потока фотонов

Так как каждый фотон несет энергию $h\nu$, это равенство обеспечивает прямой перевод классической меры (в единицах энергии на квадратный сантиметр в секунду) в квантовую меру (в единицах числа фотонов на квадратный сантиметр в секунду). Для квазимонохроматического света с центральной частотой $\bar{\nu}$ все фотоны имеют приблизительно одинаковую энергию $h\bar{\nu}$, так что средняя плотность потока фотонов примерно равна

$$\phi(\boldsymbol{r}) \approx \frac{I(\boldsymbol{r})}{h\bar{\nu}}. \qquad (12.23)$$

Типичные значения $\phi(\boldsymbol{r})$ для некоторых распространенных источников света приведены в табл. 12.1. Из приведенных данных видно, что каждую секунду на квадратный сантиметр поверхности падают триллионы фотонов.

Таблица 12.1. **Средняя плотность потока фотонов для различных источников света**

Источник	Средняя плотность потока фотонов, число фотонов/(см$^2 \cdot$ с)
Свет звезд	10^6
Свет Луны	10^8
Сумерки	10^{10}
Комнатный свет	10^{12}
Солнечный свет	10^{14}
Лазерный свет*	10^{22}

* He—Ne-лазер мощностью 10 мВт на $\lambda_0 = 633$ нм с размером пятна 20 мкм.

Средний поток фотонов

Средний поток фотонов (в единицах числа фотонов в секунду) получается интегрированием средней плотности потока фотонов по выбранной поверхности

$$\boxed{\Phi = \int_A \phi(\boldsymbol{r})\,\mathrm{d}A = \frac{P}{h\bar{\nu}},} \qquad (12.24)$$

Средний поток фотонов

где $h\bar{\nu}$ — средняя энергия фотона; P — оптическая мощность (в ваттах),

$$P = \int_A I(\boldsymbol{r})\,\mathrm{d}A. \tag{12.25}$$

Например, 1 нВт оптической мощности на длине волны $\lambda_0 = 0{,}2$ мкм соответствует потоку $\Phi \approx 10^9$ фотонов в секунду. Грубо говоря, каждую наносекунду на объект падает один фотон, т. е.

$$\boxed{1\ \text{нВт}/(\lambda_0 = 0{,}2\ \text{мкм}) \to 1\ \text{фотон/нс}.} \tag{12.26}$$

Фотон с длиной волны $\lambda_0 = 1$ мкм несет в пять раз меньшую энергию, так что в этом случае 1 нВт соответствует 5 фотонам/нс.

Среднее число фотонов

Среднее число фотонов $\bar{n}$, детектируемых в области A за время T, получается умножением среднего потока фотонов Ф (12.25) на время, в результате чего

$$\bar{n} = \Phi T = \frac{E}{h\bar{\nu}}, \tag{12.27}$$

Среднее число фотонов

где $E = PT$ — оптическая энергия в джоулях.

Итак, связь между классическими и квантовыми мерами такова:

Классическая величина		Квантовая величина	
Оптическая интенсивность	$I(\boldsymbol{r})$	Плотность потока фотонов	$\phi(\boldsymbol{r}) = \dfrac{I(\boldsymbol{r})}{h\bar{\nu}}$
Оптическая мощность	P	Поток фотонов	$\Phi = \dfrac{P}{h\bar{\nu}}$
Оптическая энергия	E	Число фотонов	$\bar{n} = \dfrac{E}{h\bar{\nu}}$

Спектральные плотности потока фотонов

Для полихроматического света с шириной спектра, которой нельзя пренебречь, полезно определить спектральные плотности классической интенсивности, мощности и энергии, а также их квантовые аналоги.

Например, $P_\nu \mathrm{d}\nu$ представляет собой оптическую мощность в частотном диапазоне от ν до $\nu + \mathrm{d}\nu$, тогда как $\Phi_\nu \mathrm{d}\nu$ выражает поток фотонов, частоты которых лежат между ν и $\nu + \mathrm{d}\nu$.

Спектральную плотность потока фотонов, спектральный поток фотонов и спектральную плотность числа фотонов можно определить с помощью таблицы:

Классическая величина		Квантовая величина	
I_ν	[Вт/(см^2 Гц)]	$\phi_\nu = \dfrac{I_\nu}{h\nu}$	[фотонов/(с · см^2 · Гц)]
P_ν	[Вт/Гц]	$\Phi_\nu = \dfrac{P_\nu}{h\nu}$	[фотонов/(с · Гц)]
E_ν	[Дж/Гц]	$\bar{n}_\nu = \dfrac{E_\nu}{h\nu}$	[фотонов/Гц]

Свет, меняющийся во времени

Если интенсивность света меняется во времени, то плотность потока фотонов в (12.22) является функцией времени

$$\phi(\boldsymbol{r},\ t) = \frac{I(\boldsymbol{r},\ t)}{h\bar{\nu}}. \tag{12.28}$$

Средняя плотность потока фотонов

Тогда поток фотонов и оптическая мощность также являются функциями времени

$$\Phi(t) = \int_A \phi(\boldsymbol{r},\ t)\,\mathrm{d}A = \frac{P(t)}{h\bar{\nu}}, \tag{12.29}$$

Средний поток фотонов

где

$$P(t) = \int_A I(\boldsymbol{r},\ t)\,\mathrm{d}A. \tag{12.30}$$

Следовательно, среднее число фотонов, регистрируемых в течение времени от $t = 0$ до $t = T$, которое получается интегрированием потока фотонов, также меняется со временем:

$$\bar{n} = \int_0^T \Phi(t)\,\mathrm{d}t = \frac{E}{h\bar{\nu}}, \tag{12.31}$$

Среднее число фотонов

где E — оптическая энергия (интенсивность, проинтегрированная по площади и времени),

$$E = \int_0^T P(t)\,\mathrm{d}t = \int_0^T \int_A I(\boldsymbol{r},\ t)\,\mathrm{d}A\,\mathrm{d}t. \tag{12.32}$$

12.2.2. Случайность потока фотонов

Если классическая интенсивность $I(\boldsymbol{r}, t)$ постоянна, время и место регистрации отдельного фотона определяются формулой (12.13), согласно которой плотность вероятности регистрации фотона в пространственно-временно́й точке $(\boldsymbol{r}, t)$ пропорциональна $I(\boldsymbol{r}, t)$. Классическая электромагнитная ин-

тенсивность управляет поведением как фотонных потоков, так и отдельных фотонов, однако при этом интерпретация различна:

Для фотонных потоков классическая интенсивность $I(r, t)$ определяет среднюю плотность потока фотонов $\phi(r, t)$. Свойства источника света определяют характер флуктуации $\phi(r, t)$.

Рассмотрим детектор, интегрирующий по пространству, как показано на рис. 12.10. Если интенсивность I постоянна во времени, постоянна и мощность P. Средняя плотность потока фотонов в этом случае равна

$$\phi = \frac{I}{h\bar{\nu}},$$

а средний поток фотонов

$$\Phi = \frac{P}{h\bar{\nu}}.$$

Однако моменты времени регистрации отдельных фотонов случайны, и их статистическое поведение определяется источником, как показано на рис. 12.12, *а*. Например, при $\lambda_0 = 1$ мкм оптическая мощность $P = 1$ нВт соответствует *в среднем* $\Phi = 5$ фотонов/нс, или 0,005 фотонов за каждую пикосекунду. Конечно же, зарегистрировать можно только целое число фотонов. Среднее число 0,005 фотонов/пс означает, что если рассмотреть 10^5 временны́х интервалов длительностью $T = 1$ пс каждый, то большинство окажутся пустыми (не будет зарегистрировано ни одного фотона), около 500 будут содержать один фотон и очень немного — два и более фотонов.

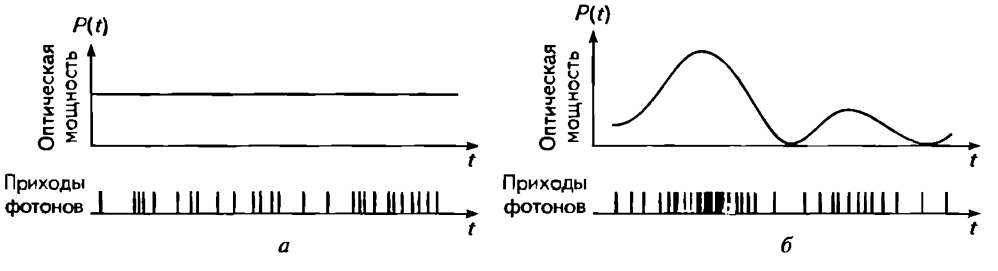


Рис. 12.12. Постоянная оптическая мощность и соответствующие случайные времена прихода фотонов (*а*). Переменная мощность и соответствующие случайные времена прихода фотонов (*б*)

Если оптическая мощность $P(t)$ меняется во времени, средняя плотность событий регистрации фотонов следует изменениям $P(t)$, как схематически показано на рис. 12.12, *б*. Средний поток равен

$$\Phi(t) = \frac{P(t)}{h\bar{\nu}},$$

что отражает факт прихода большего числа фотонов, когда мощность больше, и наоборот. Эти изменения добавляются к флуктуациям времен регистрации фотонов, связанным с источником.

Портрет Макса Планка на рис. 12.11 иллюстрирует аналогичное поведение в пространстве. Координаты точек регистрации фотонов в общем случае следуют классическому распределению интенсивностей, так что плотность фотонов выше там, где интенсивность больше, и наоборот. Однако на изображении видна значительная зернистость изображения (шум), соответствующая флуктуациям координат регистрации фотонов, связанным с источником. Эти флуктуации особенно заметны, когда средняя плотность числа фотонов мала, как в случае рис. 12.11. Когда средняя плотность потока фотонов становится большой по всему изображению, как на портрете Макса Планка, показанном на первой странице данной главы, зернистость исчезает и восстанавливается классическое распределение интенсивности.

12.2.3. Статистика числа фотонов

Понимание статистики числа фотонов важно для приложений, таких как снижение шумов в низкоинтенсивных изображениях и оптимизация передачи оптической информации. В оптоволоконных системах связи, например, информация передается в виде импульсов света (см. гл. 24). У источника контролируется только среднее число фотонов в импульсе. Действительное число испускаемых фотоном непредсказуемо и меняется от импульса к импульсу, что приводит к ошибкам при передаче информации.

Статистическое распределение числа фотонов зависит от природы источника света и, вообще говоря, должно рассматриваться с использованием квантовой теории, как кратко описано в разд. 12.3. Однако при определенных условиях акты регистрации фотонов можно рассматривать как независимые реализации последовательности случайных событий с частотой, равной потоку фотонов, пропорциональному оптической мощности. Оптическая мощность может быть детерминированной (как у когерентного света) или случайной функцией времени (как у частично когерентного света). Для частично когерентного света (см. гл. 11) флуктуации мощности коррелированны, так что акты регистрации фотонов уже не являются последовательностью независимых событий. В этом случае статистика фотонов значительно меняется.

Когерентный свет

Когерентный свет имеет постоянную оптическую мощность P. Соответствующий средний поток фотонов

$$\Phi = \frac{P}{h\bar{\nu}} \quad \text{(фотонов/с)}$$

также постоянен, однако времена актов регистрации отдельных фотонов случайны, как показано на рис. 12.12, *а* и 12.13. Пусть за время число

зарегистрированных фотонов равно n. Мы уже знаем, что среднее значение n равно

$$\bar{n} = \Phi T = \frac{PT}{h\bar{\nu}}.$$

Найдем выражение для распределения вероятностей $p(n)$, т. е. $p(0)$ — вероятность того, что не зарегистрировано ни одного фотона, $p(1)$ — вероятность регистрации одного фотона и т. д.

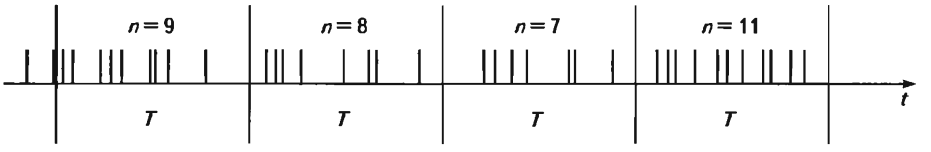


Рис. 12.13. Случайные приходы фотонов в пучке света с постоянной мощностью P в течение равных интервалов времени T. Хотя оптическая мощность постоянна, число фотонов n в каждом интервале случайно

Выражение для распределения вероятностей можно вывести в предположении, что акты регистрации фотонов статистически независимы. В результате получается распределение Пуассона

$$p(n) = \frac{\bar{n}^n \exp(-\bar{n})}{n!}, \quad n = 0,\ 1,\ 2, \ldots . \tag{12.33}$$

Распределение Пуассона

Этот результат, известный как распределение Пуассона, изображен в виде полулогарифмического графика на рис. 12.14 для нескольких значений среднего $\bar{n}$. Ширина кривых растет с ростом $\bar{n}$.

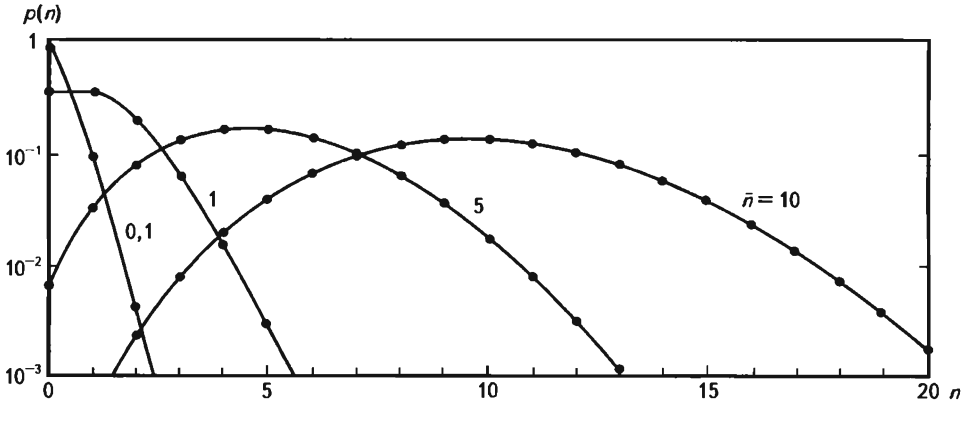
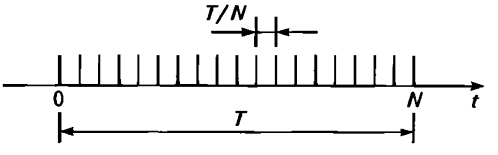


Рис. 12.14. Распределение Пуассона $p(n)$ числа фотонов n

Вывод распределения Пуассона

Разделим интервал времени T на большое число N отрезков достаточно малой длины T/N каждый, чтобы в каждый интервал попадал один фотон с вероятностью $p = \bar{n}/N$ и ни одного фотона с вероятностью $(1 - p)$. Вероятность обнаружить n независимых фотонов в N интервалах следует биномиальному распределению, как в задаче о бросании несимметричной монеты



$$p(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} = \frac{N!}{n!(N-n)!} \left(\frac{\bar{n}}{N}\right)^n \left(1 - \frac{\bar{n}}{N}\right)^{N-n}$$

В пределе при $N \to \infty$

$$\frac{N!}{(N-n)N^n} \to 1; \quad \left[1 - \frac{\bar{n}}{n}\right]^{N-n} \to \exp(-\bar{n}),$$

что и приводит к (12.33).

Среднее значение и дисперсия

Любое случайное число характеризуют два важных параметра. Это среднее значение

$$\bar{n} = \sum_{n=0}^{\infty} n p(n) \tag{12.34}$$

и дисперсия

$$\sigma_n^2 = \sum_{n=0}^{\infty} (n - \bar{n})^2 p(n), \tag{12.35}$$

представляющая собой средний квадрат отклонения от среднего. Стандартное отклонение σ_n (квадратный корень из дисперсии) является мерой ширины распределения. Величины $p(n)$, $\bar{n}$ и σ_n вместе называются статистикой числа фотонов. Хотя функция $p(n)$ содержит больше информации, чем среднее значение и дисперсия, последние являются полезными мерами.

Нетрудно показать [с использованием (12.34), (12.35) и (12.36)], что среднее распределения Пуассона действительно равно $\bar{n}$, а дисперсия равна среднему

$$\boxed{\sigma_n^2 = \bar{n}.} \tag{12.36}$$

Дисперсия, распределение Пуассона

Например, когда $\bar{n} = 100$, $\sigma_n = 10$, т. е. присутствие 100 фотонов сопровождается неточностью около ± 10 фотонов.

Пуассоновское распределение числа фотонов применимо для идеального лазера, излучающего пучок монохроматического когерентного света в одной моде (см. гл. 15). Это распределение соответствует квантовому состоянию света, известному как когерентное состояние (см. подразд. 12.3.1). Это распределение также обеспечивает отличное приближение для статистики фотонов от других источников света, включая многомодовое тепловое излучение.

Отношение сигнал—шум

Случайность числа фотонов представляет собой фундаментальный источник шума, с которым надо считаться при использовании света для передачи сигнала. Если представить среднее значение сигнала величиной $\bar{n}$, а шума — σ_n, то полезной характеристикой света как среды, переносящей информацию, будет отношение сигнал—шум (ОСШ). ОСШ случайной величины определяется как

$$\text{ОСШ} = \frac{(\text{среднее})^2}{\text{дисперсия}} = \frac{\bar{n}^2}{\sigma_n^2}. \tag{12.37}$$

Для распределения Пуассона

$$\boxed{\text{ОСШ} = \bar{n},} \tag{12.38}$$

Отношение сигнал—шум, распределение Пуассона

так что ОСШ линейно растет с увеличением среднего числа регистрируемых фотонов.

Хотя ОСШ — полезная мера случайности сигнала, в некоторых приложениях необходимо знать само распределение вероятностей. Например, если посылать сигнал со средним числом фотонов $\bar{n} = 20$, то в соответствии с (12.33), вероятность того, что ни один фотон не будет получен, равна $p(0) \approx 2 \cdot 10^{-9}$. Эта величина представляет собой вероятность ошибки в передаче информации. Данная тема рассматривается в гл. 24.

Тепловое излучение

Когда времена прихода фотонов не независимы, как в случае теплового излучения, статистика числа фотонов может подчиняться распределениям, отличным от распределения Пуассона. Рассмотрим оптический резонатор, температура стенок которого поддерживается равной T (°K), так что фотоны испускаются в моды резонатора. В соответствии с законами статистической механики в условиях теплового равновесия вероятность заселения энергетического уровня E_n в моде удовлетворяет **распределению Больцмана**

$$\boxed{P(E_n) \propto \exp\left(-\frac{E_n}{k_B T}\right),} \tag{12.39}$$

Распределение Больцмана

где k_B — **постоянная Больцмана** ($k_B = 1{,}38 \cdot 10^{-23}$ Дж/К). Происхождение этого распределения более детально обсуждается в разд. 13.2.

В состоянии теплового равновесия энергия, связанная с каждой модой, — случайная величина. Более высокие значения энергии менее вероятны по сравнению с более низкими, как предписывает этот простой экспоненциальный закон с параметром $k_B T$. Распределение Больцмана схематически представлено на рис. 12.15 с температурой в качестве параметра. Чем меньше $k_B T$, тем менее вероятно наблюдение больших энергий. При комнатной температуре ($T = 300$ °K) имеем $k_B T = 0{,}026$ эВ, что эквивалентно 208 см$^{-1}$. Если рассматривать фотоны в моде резонатора с частотой ν как газ в состоянии теплового равновесия при температуре T, то из распределения Больцмана (12.40) и условия квантования энергии (12.5), $E_n = (n + 1/2)h\nu$, следует, что вероятность обнаружения n фотонов равна

$$p(n) \propto \exp\left(-\frac{nh\nu}{k_B T}\right) = \left[\exp\left(-\frac{h\nu}{k_B T}\right)\right]^n \qquad n = 0,\ 1,\ 2, \qquad (12.40)$$

Используя условие, что сумма вероятностей распределения равна единице, т. е.

$$\sum_{n=0}^{\infty} p(n) = 1,$$

находим постоянную нормировки, равную $[1 - \exp(-h\nu/k_B T)]$. Нулевая энергия $E_n = (1/2)h\nu$ исчезает при нормировке и не влияет на результаты, как обсуждалось в подразд. 12.1.1.

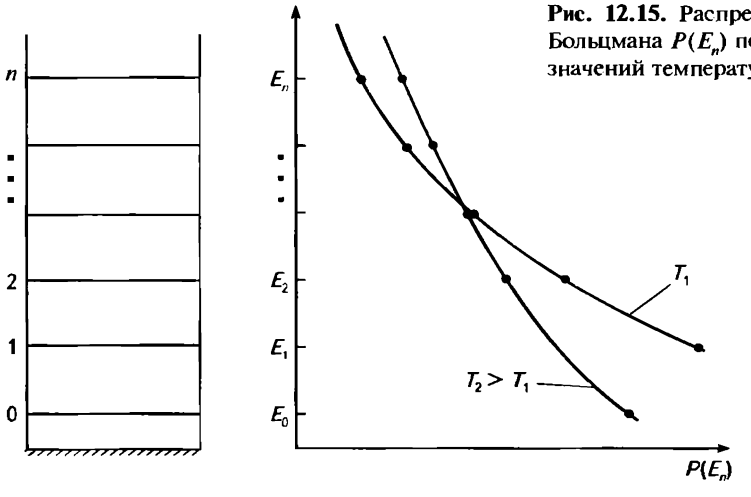


Рис. 12.15. Распределение вероятностей Больцмана $P(E_n)$ по энергиям E_n для двух значений температуры T

Распределение вероятностей наиболее просто записывается через среднее

$$p(n) = \frac{1}{\bar{n} + 1}\left(\frac{\bar{n}}{\bar{n} + 1}\right)^n, \qquad (12.41)$$

Распределение Бозе—Эйнштейна

где

$$\bar{n} = \frac{1}{\exp(h\nu/k_B T) - 1} \tag{12.42}$$

в соответствии с (12.34). В теории вероятностей такое распределение называется **геометрическим распределением**, поскольку $p(n)$ — геометрически убывающая функция n. В физике оно известно как **распределение Бозе—Эйнштейна**. Выражение (12.42) согласуется со средним, рассчитанным для коллектива фотонов, взаимодействующих с атомами в состоянии теплового равновесия (13.71).

Распределение Бозе—Эйнштейна показано на рис. 12.16 для нескольких значений $\bar{n}$ или, что равносильно, для нескольких значений температуры T [см. (12.42)]. Его экспоненциальный характер очевиден из линейности этого полулогарифмического графика. Сравнение рис. 12.16 и 12.14 показывает, что распределение числа фотонов у теплового излучения монотонно убывает и намного шире, чем у когерентного света.

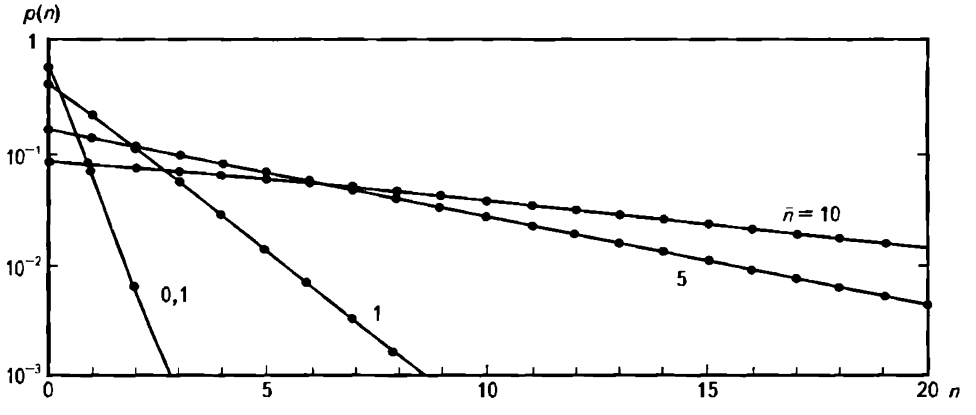


Рис. 12.16. Распределение Бозе—Эйнштейна $p(n)$ по числу фотонов n

С учетом (12.36) дисперсия числа фотонов получается равной

$$\boxed{\sigma_n^2 = \bar{n} + \bar{n}^2.} \tag{12.43}$$

Дисперсия.
Распределение Бозе—Эйнштейна

Сравнивая это выражение с дисперсией распределения Пуассона, которая равна просто $\bar{n}$ согласно (12.38), мы видим, что тепловое излучение имеет большую дисперсию. Это соответствует большей неопределенности и большему диапазону флуктуаций числа фотонов. Отношение сигнал—шум для распределения Бозе—Эйнштейна равно

$$\text{ОСШ} = \frac{\bar{n}}{\bar{n} + 1}. \tag{12.44}$$

Оно всегда меньше единицы, независимо от оптической мощности. Амплитуда и фаза теплового излучения ведут себя как случайные величины, как описано в гл. 11. Случайность эта приводит к уширению распределения числа фотонов. Действительно, этот вид света имеет слишком высокий уровень шума, чтобы использоваться в высокоскоростных системах передачи информации.

Упражнение 12.5

Средняя энергия в моде резонатора

Покажите, что средняя энергия в моде резонатора с частотой ν в условиях теплового равновесия при температуре T дается выражением

$$\bar{E} = k_B T \frac{h\nu/k_B T}{\exp(h\nu/k_B T) - 1}. \tag{12.45}$$

Изобразите зависимость $\bar{E}$ от ν для нескольких значений $k_B T/h$. Используйте разложение знаменателя в ряд Тейлора для получения упрощенного выражения для $\bar{E}$ в пределе $h\nu/k_B T \ll 1$. Объясните физический смысл результата.

---

*Другие источники света

Как отмечалось выше, для некоторых источников света регистрацию фотонов можно рассматривать как последовательность независимых событий, происходящих с частотой, пропорциональной оптической мощности. Для когерентного света оптическая мощность P является детерминированной, а число фотонов подчиняется распределению Пуассона

$$p(n) = \frac{w^n e^{-w}}{n!},$$

где

$$w = \frac{1}{h\nu}\int_0^T P(t)\,dt = \frac{1}{h\nu}\int_0^T\int_A I(r,\ t)\,dA\,dt. \tag{12.46}$$

Величина w — интегральный поток фотонов (нормированная интегральная мощность) — постоянная, представляющая среднее число фотонов $\bar{n}$.

Для источников света, у которых интенсивность $I(r, t)$ сама флуктуирует случайным образом во времени и/или в пространстве, оптическая мощность $P(t)$ также претерпевает случайные флуктуации (см. рис. 12.12, *б*), и ее интеграл w также случаен. В результате случайным является не только число фотонов, но и его среднее значение w. Благодаря этому дополнительному источнику случайности статистика числа фотонов частично когерентного света отличается от распределения Пуассона. Если флуктуации среднего числа фотонов w описываются функцией плотности вероятности $p(w)$, то распределение безус-

ловной вероятности для частично когерентного света получается усреднением условного распределения Пуассона

$$p(n \mid w) = \frac{w^n e^{-w}}{n!}$$

по всем возможным значениям w с функцией плотности вероятности $p(w)$. Получающееся распределение числа фотонов имеет вид

$$\boxed{p(n) = \int_0^\infty \frac{w^n e^{-w}}{n!} p(w)\,dw,} \qquad (12.47)$$

Формула Манделя

известный как формула Манделя. Выражение (12.47) также называют дважды случайным пуассоновским распределением отсчетов, поскольку в него вносят вклад два источника случайности: фотоны сами по себе (что дает локально пуассоновское распределение) и флуктуации интенсивности, возникающие из-за некогерентной природы света (требующей уточнения).

Заметим, что данная теория статистики фотонов применима только к некоторым состояниям света — так называемому классическому свету. Более общая теория, основанная на квантовом описании состояний света, кратко описана в разд. 12.3.

Среднее значение и дисперсия числа фотонов для частично когерентного света получаются с помощью (12.34) и (12.25) в сочетании с (12.47):

$$\bar{n} = \bar{w}; \qquad (12.48)$$

$$\sigma_n^2 = \bar{n} + \sigma_w^2 \qquad (12.49)$$

соответственно. Здесь σ_w^2 обозначает дисперсию w. Отметим, что дисперсия числа фотонов состоит из суммы двух членов, первый из которых представляет основной вклад от распределения Пуассона, а второй — дополнительный вклад от классических флуктуаций оптической мощности.

Рассмотрим другой важный пример, в котором флуктуации нормированной интегральной оптической мощности подчиняются экспоненциальной функции плотности вероятности

$$p(w) = \begin{cases} \dfrac{1}{\bar{w}} \exp\left(-\dfrac{w}{\bar{w}}\right), & w \geq 0; \\ 0, & w < 0. \end{cases} \qquad (12.50)$$

Это распределение соответствует квазимонохроматическому пространственно когерентному свету, у которого действительная и мнимая части комплексной амплитуды поля независимы и имеют гауссовы распределения вероятностей. Оно применимо, когда спектральная ширина достаточно мала, так что время когерентности τ_c намного больше времени счета T, а площадь когерентности A_c намного больше площади детектора A (см. гл. 11).

Распределение числа фотонов $p(n)$, соответствующее (12.50), можно получить подстановкой этого выражения в (12.47) с последующим вычислением интеграла. В результате получается распределение Бозе—Эйнштейна (12.41). Следовательно, оптическое поле с гауссовым распределением имеет статистику фотонов, идентичную статистике одномодового теплового излучения. Если площадь A и время T не малы, статистика модифицируется и описывает многомодовое тепловое излучение (см. задачи 5—7 к разд. 12.2).

12.2.4. Случайное разбиение фотонных потоков

О разбиении фотонного потока говорят, если из него удаляется часть фотонов. Эти фотоны могут либо отклоняться, либо разрушаться. Процесс называется случайным разбиением, когда фотоны отклоняются, и случайным выбором, когда они разрушаются. Это может происходить различными путями. Быть может, простейший пример случайного разбиения дает идеальный светоделитель без потерь. Фотоны выбираются случайным образом, чтобы присоединиться к одному из двух возникающих потоков (рис. 12.17). Примером случайного выбора является действие оптического поглощающего фильтра на пучок света. Фотоны случайным образом выбираются, чтобы либо пройти через фильтр, либо разрушиться (превратиться в тепло).

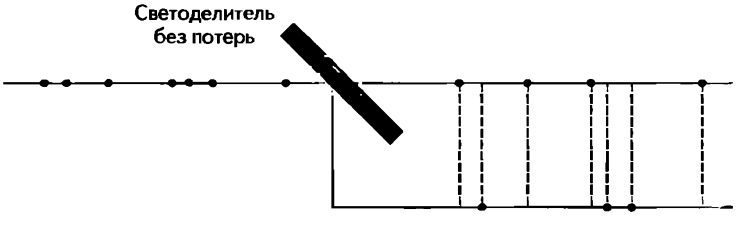


Рис. 12.17. Случайное разбиение фотонов светоделителем

Ограничим наше рассмотрение ситуациями, в которых удаление каждого фотона является независимым случайным испытанием. Применительно к светоделителю это требование удовлетворяется, если поток фотонов попадает только на один из входных портов (см. рис. 12.17). Это устраняет возможность интерференции, которая в общем случае делает неприменимым предположение о независимых испытаниях. Хотя результаты, полученные ниже, сформулированы для случайного разбиения, они равным образом применимы и для случайного выбора.

Рассмотрим светоделитель без потерь с коэффициентами пропускания $\mathcal{T}$ и отражения $\mathcal{R} = 1 - \mathcal{T}$. В электромагнитной оптике интенсивность прошедшей волны I_t связана с интенсивностью падающей волны I соотношением $I_t = \mathcal{T}I$. Результат попадания одиночного фотона на светоделитель исследован в подразд. 12.1.3. Было показано, что вероятность прохождения фотона равна коэффициенту пропускания $\mathcal{T}$. Теперь мы рассмотрим случай, когда на светодели-

тель падает средний поток фотонов Ф, так что среднее число фотонов, попадающих на светоделитель за время T, $\bar{n} = \Phi T$. В соответствии с (12.28) среднее число фотонов в пучке пропорционально оптической энергии. Следовательно, среднее число пропущенных и отраженных за это время фотонов должно быть $\mathcal{T}\bar{n}$ и $(1 - \mathcal{T})\bar{n}$ соответственно.

Рассмотрим теперь более общий вопрос: что происходит со статистикой числа фотонов $p(n)$ фотонного потока при разбиении на светоделителе?

Отдельный фотон, падая на светоделитель, пропускается с вероятностью $\mathcal{T}$ и отражается с вероятностью $1 - \mathcal{T}$ (см. рис. 12.7). Если падающий пучок содержит ровно n фотонов, то вероятность $p(m)$ того, что m фотонов будет пропущено, такая же, как при бросании монеты, у которой вероятность выпадения «орла» равна $\mathcal{T}$. Из элементарной теории вероятностей мы знаем, что в результате получается биномиальное распределение

$$p(m) = \binom{n}{m} \mathcal{T}^n (1 - \mathcal{T})^{n-m}, \quad m = 0, 1, \ldots, n, \tag{12.51}$$

где

$$\binom{n}{m} = \frac{n!}{m!}(n-m)!.$$

Легко показать, что среднее число прошедших фотонов

$$\bar{m} = \mathcal{T}n. \tag{12.52}$$

Дисперсия для биномиального распределение дается формулой

$$\sigma_m^2 = \mathcal{T}(1 - \mathcal{T})n = (1 - \mathcal{T})\bar{m}. \tag{12.53}$$

Из-за симметрии задачи результаты для отраженного пучка получаются немедленно. По мере того, как среднее число прошедших фотонов $\bar{m}$ растет, отношение сигнал—шум, даваемое формулой

$$\frac{\bar{m}^2}{\sigma_m^2} = \frac{\bar{m}}{1 - \mathcal{T}},$$

тоже растет. Следовательно, для больших интенсивностей фотоны будут делиться между двумя потоками в хорошем согласии с $\mathcal{T}$ и $1 - \mathcal{T}$, отражая факт восстановления законов классической оптики.

Приведенные выше выражения позволяют вычислять влияние светоделителя на фотоны, число которых подчиняется произвольной статистике. Решение получается при учете того факта, что число фотонов n на входе светоделителя является случайным, а не фиксированным. Пусть вероятность присутствия n фотонов равна $p_0(n)$. Если рассматривать фотоны как независимые события, то распределение их числа в прошедшем потоке будет взвешенной суммой биномиальных распределений с n, принимающим случайные значения. Вес соответствует вероятности присутствия n фотонов. Вероятность обна-

ружения m фотонов, прошедших через светоделитель, при условии, что на входе распределение числа фотонов есть $p_0(n)$, тогда выражается формулой

$$p(m) = \sum_n p(m \mid n)\, p_0(n),$$

где $p(m|n)$ — биномиальное распределение,

$$p(m \mid n) = \binom{n}{m} \mathcal{T}^m \left(1 - \mathcal{T}\right)^{(n-m)}$$

В явном виде имеем:

$$\boxed{p(m) = \sum_{n=m}^{\infty} \binom{n}{m} \mathcal{T}^m \left(1 - \mathcal{T}\right)^{(n-m)} p_0(n).} \qquad (12.54)$$

Статистика числа фотонов при случайном разбиении

Если $p_0(n)$ — распределение Пуассона (когерентный свет) или Бозе—Эйнштейна (одномодовое тепловое излучение), то результат оказывается весьма простым: $p(m)$ имеет точно такой же вид, как и $p_0(n)$. Оба эти распределения сохраняют свой вид при случайном разбиении. Таким образом, свет одномодового лазера, прошедшего через светоделитель, остается пуассоновским, а тепловое излучение продолжает подчиняться распределению Бозе—Эйнштейна; при этом, конечно, среднее число фотонов уменьшается на множитель. Свет с детерминированным числом фотонов (см. подразд. 12.3.2), напротив, не сохраняет свой вид при случайном разбиении, что делает его недостаточно устойчивым к внешним воздействиям.

Отношение сигнал—шум для m легко вычисляется в случае фотонных потоков, претерпевающих случайное разбиение или случайный выбор. Для когерентного света и одномодового теплового излучения результаты имеют вид

$$\text{ОСШ} = \begin{cases} \mathcal{T}\bar{n} & \text{когерентный свет;} \qquad (12.55) \\ \dfrac{\mathcal{T}\bar{n}}{\mathcal{T}\bar{n} + 1} & \text{тепловое излучение.} \qquad (12.56) \end{cases}$$

Поскольку $\mathcal{T} \leq 1$, ясно, что случайное разбиение уменьшает отношение сигнал—шум. Другими словами, случайное разбиение вносит шум. Этот эффект особенно силен для света с определенным числом фотонов.

Полученные результаты применимы к *детектированию* фотонов. Если каждый фотон имеет независимую возможность быть зарегистрированным, то из n фотонов будут зарегистрированы m, причем $p(m)$ связано с $p_0(n)$ соотношением (12.54). Мы убедимся в полезности этого результата при рассмотрении теории детектирования фотонов (см. гл. 18).

*12.3. КВАНТОВЫЕ СОСТОЯНИЯ СВЕТА

Число фотонов в моде электромагнитного поля в общем случае является случайной величиной. В данном разделе будет показано, что в квантовой оптике само электрическое поле, вообще говоря, также случайно. Рассмотрим моду электромагнитного поля в виде плоской монохроматической волны в объеме V, описываемую электрическим полем $\mathcal{E}(\boldsymbol{r}, t) = \mathrm{Re}\{\boldsymbol{E}(\boldsymbol{r}, t)\}$, где

$$\boldsymbol{E}(\boldsymbol{r}, t) = A\exp(-j\boldsymbol{k}\cdot\boldsymbol{r})\exp(j2\pi\nu t)\hat{\boldsymbol{e}}. \tag{12.57}$$

В соответствии с классической электромагнитной оптикой [формула (12.3)] энергия моды имеет фиксированное значение $\varepsilon|A|^2V/2$. Определим комплексную переменную a так, что

$$\frac{\varepsilon|A|^2V}{2} = h\nu|a|^2,$$

что позволяет интерпретировать $|a|^2$ как энергию моды в единицах числа фотонов. Электрическое поле тогда можно записать как

$$\boldsymbol{E}(\boldsymbol{r}, t) = \sqrt{\frac{2h\nu}{\varepsilon V}}a\exp(-j\boldsymbol{k}\cdot\boldsymbol{r})\exp(j2\pi\nu t)\hat{\boldsymbol{e}}, \tag{12.58}$$

где комплексная переменная a определяет комплексную амплитуду поля.

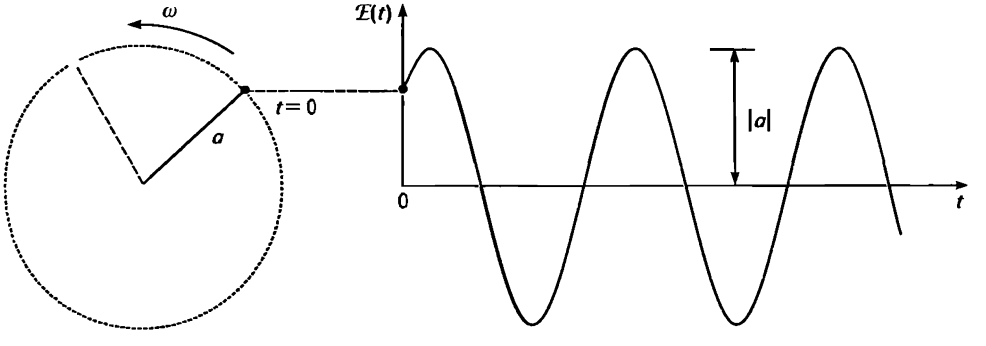


Рис. 12.18. Действительная и мнимая части переменной $a\exp(j2\pi\nu t)$, описывающей комплексную амплитуду классического электромагнитного поля частоты ν. Временная динамика такая же, как у классического гармонического осциллятора с угловой частотой $\omega = 2\pi\nu$

В классической электромагнитной оптике $a\exp(j2\pi\nu t)$ — вращающийся вектор на комплексной плоскости, проекция которого на действительную ось определяет синусоидальное электрическое поле (рис. 12.18). Действительная и мнимая части $a = x + jp$, равные $x = \mathrm{Re}\{a\}$ и $p = \mathrm{Im}\{a\}$ соответственно, называются квадратурными компонентами комплексного вектора a, поскольку

по фазе они сдвинуты относительно друг друга на четверть полного периода (90°). Они определяют амплитуду и фазу синусоидальной волны, которая описывает изменение электрического поля во времени. Вращающийся вектор $a\exp(j2\pi\nu t)$ на комплексной плоскости описывает также движение гармонического осциллятора, при этом действительная часть x пропорциональна координате, а мнимая часть p — импульсу. С математической точки зрения классическая монохроматическая мода электромагнитного поля и классический гармонический осциллятор ведут себя одинаково.

Аналогично можно показать, что квантовая монохроматическая электромагнитная мода ведет себя так же, как одномерный квантово-механический гармонический осциллятор. Для облегчения такого сравнения рассмотрим вначале квантовую теорию простого гармонического осциллятора.

Квантовая теория гармонического осциллятора

Частица массы m с координатой x, импульсом p и потенциальной энергией

$$V(x) = \frac{1}{2}\kappa x^2,$$

где κ — постоянная упругости, представляет собой гармонический осциллятор с полной энергией

$$\frac{1}{2}\frac{p^2}{m} + \frac{1}{2}\kappa x^2$$

и частотой колебаний

$$\omega = \sqrt{\frac{k}{m}}.$$

Без потери общности можно положить $m = 1$, так что полная энергия составляет

$$\frac{1}{2}\left(p^2 + \omega^2 x^2\right).$$

Согласно квантовой механике, его поведение в стационарном состоянии описывается комплексной волновой функцией $\psi(x)$, удовлетворяющей стационарному уравнению Шредингера

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + V(x)\psi(x) = E\psi(x), \tag{12.59}$$

где E — энергия частицы. Решение уравнения Шредингера для гармонического осциллятора ($V(x) = \omega^2 x^2/2$) порождает набор дискретных значений энергии

$$E_n = \left(n + \frac{1}{2}\right)h\nu, \quad n = 0,\ 1,\ 2, \tag{12.60}$$

Соседние уровни энергии отстоят друг от друга на квант энергии $h\nu = \hbar\omega$. Соответствующие волновые функции $\psi_n(x)$ представляют собой нормированные функции Эрмита—Гаусса

$$\psi_n(x) = \frac{1}{\sqrt{2^n n!}}\left(\frac{\omega}{\pi\hbar}\right)^{1/4} \mathbb{H}_n\left[\sqrt{\frac{\omega}{\hbar}}x\right]\exp\left(-\frac{\omega x^2}{2\hbar}\right), \tag{12.61}$$

где $\mathbb{H}_n(x)$ — полином Эрмита порядка n [см. (3.62)—(3.64) и (3.67)].

Произвольная волновая функция $\psi(x)$ может быть разложена по ортогональным собственным функциям $\{\psi_n(x)\}$:

$$\psi(x) = \sum_n c_n\psi_n(x).$$

Если дана волновая функция $\psi(x)$, определяющая состояние системы, то поведение частицы определяется следующим образом.

- Вероятность $p(n)$ того, что гармонический осциллятор несет n квантов энергии, определяется коэффициентом $|c_n|^2$.
- Плотность вероятности обнаружить частицу в положении x дается величиной $|\psi(x)|^2$.
- Плотность вероятности обнаружения частицы с импульсом p дается величиной $|\phi(p)|^2$, где $\phi(p)$ пропорционально результату обратного преобразования Фурье функции $\psi(x)$ на (пространственной) частоте p/h,

$$\phi(p) = \frac{1}{\sqrt{h}}\int_{-\infty}^{\infty}\psi(x)\exp\left(j2\pi\frac{p}{h}x\right)dx. \tag{12.62}$$

Как показано в разд. А.2 приложения А, связь между $\psi(x)$ и $\phi(p)$ через преобразование Фурье приводит к соотношению неопределенностей между среднеквадратичными ширинами величин x и p/h вида

$$\frac{\sigma_x\sigma_p}{h} \geq \frac{1}{4\pi}, \quad \text{или} \quad \sigma_x\sigma_p \geq \frac{\hbar}{2}. \tag{12.63}$$

Это соотношение представляет собой хорошо известное соотношение неопределенностей Гейзенберга между координатой и импульсом.

Аналогия между оптической модой и гармоническим осциллятором

Энергия моды электромагнитного поля равна $h\nu|a|^2 = h\nu(x^2 + p^2)$. Аналогия с гармоническим осциллятором, энергия которого $(1/2)(p^2 + \omega^2x^2)$, устанавливается посредством подстановки

$$x = \frac{1}{\sqrt{2\hbar\omega}}\omega x; \quad p = \frac{1}{\sqrt{2\hbar\omega}}p. \tag{12.64}$$

Энергия моды тогда становится равной

$$h\nu\left(x^2 + p^2\right) = \frac{1}{2}\left(p^2 + \omega^2 x^2\right),$$

что совпадает с энергией гармонического осциллятора. Ввиду полной аналогии, мы приходим к заключению, что энергия квантовой электромагнитной моды, подобно энергии квантово-механического гармонического осциллятора, принимает дискретные значения $(n + 1/2)h\nu$, как указывалось ранее. С точностью до соответствующих масштабных множителей поведение координаты x и импульса p гармонического осциллятора также описывает квадратурные компоненты электромагнитного поля x и p.

Свойства

Мода электромагнитного поля частоты ν описывается комплексной волновой функцией $\psi(x)$, которая управляет неопределенностями квадратурных компонент x и p электромагнитного поля, а также статистикой числа фотонов в моде.

- Вероятность $p(n)$ того, что мода содержит n фотонов, дается величиной $|c_n|^2$, где c_n — коэффициенты разложения $\psi(x)$ по собственным функциям $\psi_n(x)$, т. е.

$$\psi(x) = \sum_n c_n \psi_n(x).$$

- Функции плотности вероятности квадратурных компонент x и p даются величинами $|\psi(x)|^2$ и $|\phi(p)|^2$ соответственно, где $\psi(\ )$ и $\phi(\ )$ связаны соотношением

$$\phi(p) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \psi(x) \exp\left(j2\pi px\right) \mathrm{d}x. \tag{12.65}$$

Это соотношение выводится из (12.62) подстановкой (12.64) с учетом того, что интегралы от $|\psi(x)|^2$ и $|\phi(p)|^2$ должны равняться единице.

- Между среднеквадратичными ширинами квадратурных компонент существует соотношение неопределенностей вида

$$\boxed{\sigma_x \sigma_p \geq \frac{1}{4}}, \tag{12.66}$$

Неопределенность квадратур

так что эти компоненты не могут быть определены одновременно с произвольной точностью

12.3.1. Когерентные состояния света

Произведение неопределенностей $\sigma_x \sigma_p$ достигает своего минимума 1/4, когда функция $\psi(x)$ гауссова (см. разд. А.2 приложения А). В этом случае

$$\psi(x) \propto \exp\left[-(x-\alpha_x)^2\right], \tag{12.67}$$

и ее Фурье-образ тоже гауссов, так что

$$\phi(p) \propto \exp\left[-(p-\alpha_p)^2\right]. \tag{12.68}$$

Здесь α_x и α_p — произвольные величины, представляющие средние значения x и p соответственно. Неопределенности квадратурных компонент, вычисленные с помощью $|\psi(x)|^2$ и $|\phi(p)|^2$, тогда равны

$$\sigma_x = \sigma_p = \frac{1}{2}. \tag{12.69}$$

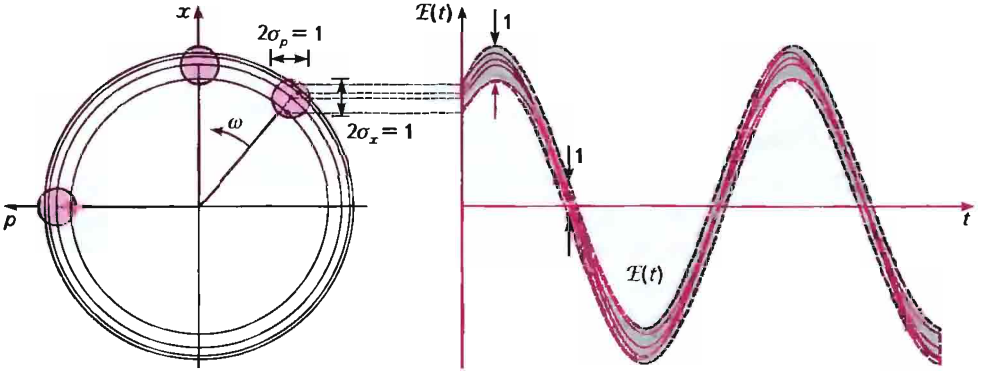


Рис. 12.19. Неопределенности для когерентного состояния. Примеры поведения $\mathcal{E}(t) \propto \propto a \exp(j2\pi\nu t)$ изображены путем выбора нескольких произвольных точек внутри круга неопределенности. Коэффициент пропорциональности выбран равным единице

В этом случае говорят, что электромагнитное поле находится в **когерентном состоянии**. Неопределенности в пределах одного стандартного отклонения для квадратурных компонент x и p, а также для комплексной амплитуды a и электрического поля $\mathcal{E}(t)$ света в когерентном состоянии показаны на рис. 12.19. Квадрат модуля $|c_n|^2$ коэффициента разложения $\psi(x)$ в базисе функций Эрмита—Гаусса равен

$$\frac{\bar{n}^n \exp(-\bar{n})}{n!},$$

где $\bar{n} = \alpha_x^2 + \alpha_p^2$. Таким образом, вероятность счета фотонов $p(x)$ есть распределение Пуассона. В отличие от классической электромагнитной оптики в квантовой оптике когерентный свет *не является* детерминированным.

Неопределенность в когерентном состоянии наиболее выражена, когда α_x и α_p малы. Поведение электрического поля во времени показано на рис. 12.20 в пределе, когда $\alpha_x = \alpha_p = 0$. Это соответствует случаю, когда мода содержит нулевое число фотонов и имеет только остаточную нулевую энергию $h\nu/2$. Такое состояние называется **вакуумным состоянием**.

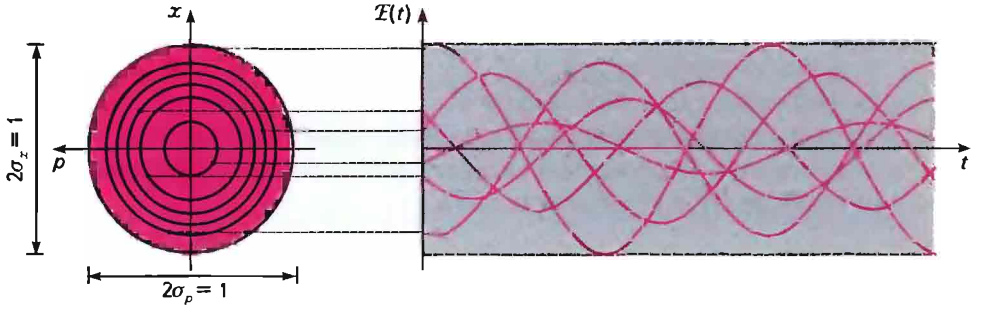


Рис. 12.20. Репрезентативные неопределенности вакуумного состояния

12.3.2. Сжатые состояния света

Квадратурно сжатый свет

Хотя произведение неопределенностей $\sigma_x\sigma_p$ не может быть меньше своего минимального значения 1/4, неопределенность одной из квадратурных компонент может быть уменьшена (сжата) до значений меньше 1/2; это влечет за собой увеличение неопределенности другой компоненты. Этот вид света, который, очевидно, является неклассическим, называется **квадратурно сжатым**. Например, состояние, в котором $\psi(x)$ — гауссова функция с (растянутой) шириной $\sigma_x = s/2$ ($s > 1$), соответствует гауссовой функции $\phi(p)$ с (сжатой) шириной $\sigma_p = 1/2s$. Произведение $\sigma_x\sigma_p$ остается равным минимальному значению 1/4, но круг неопределенности для комплексного вектора сжимается в эллипс, как показано на рис. 12.21.

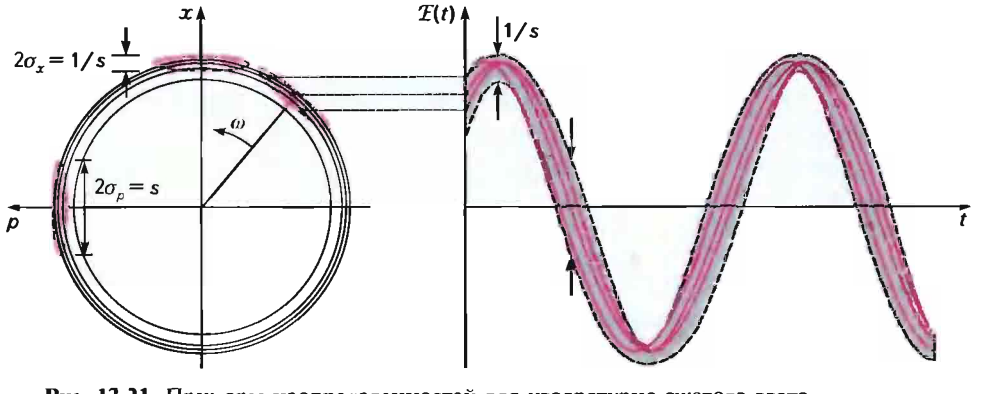


Рис. 12.21. Примеры неопределенностей для квадратурно сжатого света

Асимметрия неопределенностей двух квадратурных компонент проявляется во временно́м поведении электрического поля в виде периодического увеличения и уменьшения неопределенности каждые четверть периода. Если поле измерять только в те моменты времени, когда его неопределенность минимальна, то его шум будет снижен по сравнению с когерентным состоянием. Выбор этих моментов можно достичь с помощью гетеродинирования сжатого поля когерентным оптическим полем с соответствующей фазой (см. разд. 24.5). Из-за пониженных шумов сжатый свет нашел свое применение в точных измерениях. Однако он неустойчив в присутствии потерь.

Свет, сжатый по числу фотонов

Квадратурно сжатый свет имеет неопределенность одной из своих квадратурных компонент, сжатую по сравнению с неопределенностью в когерентном состоянии. Другим видом неклассического света является свет, **сжатый по отношению к числу фотонов**, или **субпуассоновский** свет. Он имеет дисперсию числа фотонов, которая «сжата» по сравнению с ее значением для когерентного (пуассоновского) состояния, так что $\sigma_n^2 < \bar{n}$. Флуктуации числа фотонов, подчиняющиеся этому соотношению, являются неклассическими, поскольку (12.49) не выполняется. Как и квадратурно сжатый свет, субпуассоновский свет находит применение в прецизионных измерениях и неустойчив по отношению к наличию потерь. Такой свет можно генерировать путем помещения квантовой точки в специально сконструированный микрорезонатор (см. пример 17.13) или с применением «пучков-близнецов» (twin-beam light), как описано ниже.

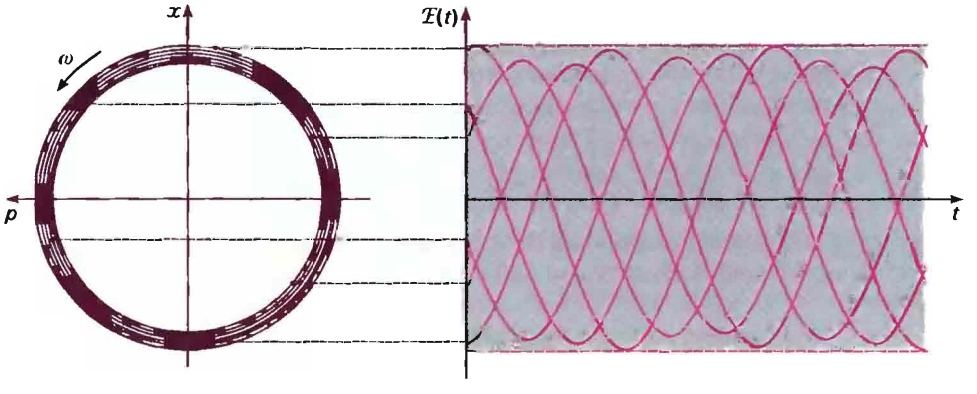


Рис. 12.22. Примеры неопределенностей для состояния с определенным числом фотонов. Это состояние — сжатое по отношению к числу фотонов, но оно не является квадратурно сжатым

Мода электромагнитного поля, описываемая собственным состоянием гармонического осциллятора $\psi(x) = \psi_{n_0}(x)$, дает наиболее яркий пример света, сжатого по отношению к числу фотонов. Это состояние называется состоянием с определенным числом фотонов, поскольку $p(n) = |c_n|^2 = 1$ для $n = n_0$, а все

другие коэффициенты исчезают ($c_n = 0$ для $n \neq n_0$). Число фотонов в моде детерминировано; оно в точности равно n_0. Среднее число фотонов, очевидно, $\bar{n} = n_0$, а дисперсия равна нулю (поскольку флуктуации числа фотонов отсутствуют). Случай $n_0 = 1$ соответствует наличию ровно одного фотона. Многие другие состояния также проявляют сжатие по отношению к числу фотонов.

Неопределенности, присущие состояниям света с заданным числом фотонов, иллюстрирует рис. 12.22. Квадратурные компоненты, а также модуль и фаза вектора на комплексной плоскости являются неопределенными, в то время как число фотонов абсолютно определено.

«Пучки-близнецы»

Возникает естественный вопрос, нельзя ли генерировать свет в состоянии, сжатом по отношению к числу фотонов, посредством какой-либо манипуляции со светом, находящимся в когерентном состоянии. Первая мысль состоит в отслеживании числа фотонов от когерентного источника в последовательные интервалы времени, а затем использовании только тех интервалов, в течение которых это число соответствует заданному. К сожалению, такой подход обречен на неудачу, поскольку сам акт наблюдения фотонов уничтожает их и, следовательно, делает невозможным их дальнейшее использование.

Однако с помощью нелинейной оптики и **«пучков-близнецов»** можно селективно манипулировать когерентным светом и получать сжатие по отношению к числу фотонов. Фотоны можно генерировать коррелированными парами посредством спонтанного параметрического рассеяния — нелинейного процесса, в котором некоторая часть фотонов, падающих на кристалл, расщепляется на пары фотонов с сохранением энергии и импульса (см. подразд. 21.2.3 и 21.4.3). Поскольку в обоих «пучках-близнецах» генерируется одинаковое число фотонов, полное распределение числа фотонов имеет ширину, сжатую по сравнению с ее классическим значением, а генерируемый таким образом свет можно рассматривать как **двухмодовый сжатый по отношению к числу фотонов**. Имея такой свет в виде «пучков-близнецов», можно получать информацию из одного пучка, производя над ним измерения. Хотя фотоны этого пучка будут уничтожены в процессе измерения, полученная информация может быть использована для управления числом фотонов (и другими свойствами) оставшегося «пучка-близнеца».

Рекомендуемая литература

КНИГИ ПО КВАНТОВОЙ ОПТИКЕ И КВАНТОВОЙ МЕХАНИКЕ

Fox M. Quantum Optics. An Introduction. Oxford University Press, 2006.

Vogel W., Welsch D.-G. Quantum Optics. Akademie-Verlag, 1994; Wiley-VCH, 3rd ed. 2006.

Klauder J.R., Sudarshan E.C.G. Fundamentals of Quantum Optics. Benjamin, 1968; Dover, reissued 2006.

Walls D.F., Milburn G.J. Quantum Optics. Springer-Verlag, 1995; paperback 2nd ed. 2006.

Vedral V. Introduction to Quantum Information Science. Oxford University Press, 2006.

Feynman R.P. QED: The Strange Theory of Light and Matter. Princeton University Press, 1985; reissued 2006.

Feynman R.P., Leighton R.B., Sands M. The Feynman Lectures on Physics. Vol. 3. Quantum Mechanics, 1965. Vol. 1. Mainly Mechanics, Radiation and Heat, 1963; Addison-Wesley, 2nd ed. 2006.

Sergienko A.V., ed. Quantum Communications and Cryptography. Taylor & Francis, 2006.

Planck M. Planck's Columbia Lectures: Abridged and Unabridged Versions, with commentary by W. Vlasak. Adaptive Enterprises, 2005.

Ficek Z., Swain S. Quantum Inteiference and Coherence: Theory and Experiments. Springer-Verlag, 2005.

Gerry C.C., Knight P.L. Introductory Quantum Optics. Cambridge University Press, 2005.

Paul H. Introduction to Quantum Optics: From Light Quanta to Quantum Teleportation. Cambridge University Press, 2004.

Bachor H.-A., Ralph T.C. A Guide to Experiments in Quantum Optics. Wiley-VCH, paperback 2nd ed. 2004.

Carmichael H.J. Statistical Methods in Quantum Optics 1: Master Equations and Fokker-Planck Equations. Springer-Verlag, 2003.

Nielsen M.A., Chuang I.L. Quantum Computation and Quantum Information. Cambridge University Press, 2002.

Baierlein R. Newton to Einstein: The Trail of Light. Cambridge University Press, paperback ed. 2001.

Kim J., Somani S., Yamamoto Y. Nonclassical Light from Semiconductor Lasers and LEDs. Springer-Verlag, 2001.

Schleich W. Quantum Optics in Phase Space. Wiley-VCH, 2001.

Loudon R. The Quantum Theory of Light. Oxford University Press, 3rd ed. 2000.

Yamamoto Y., Tassone F., Cao H. Semiconductor Cavity Quantum Electrodynamics. Springer-Verlag, 2000.

Mead C.A. Collective Electrodynamics: Quantum Foundations of Electromagnetism. MIT Press, 2000.

Meystre P., Sargent III M. Elements of Quantum Optics. Springer-Verlag, 3rd ed. 1999.

Yamamoto Y., Imamoglu A. Mesoscopic Quantum Optics. Wiley, 1999.

Feynman R.P. Quantum Electrodynamics. Benjamin, 1962; Addison-Wesley, 1998.

Peřinová V., Luks A., Peřina J. Phase in Optics. World Scientific, 1998.

Barnett S.M., Radmore P.M. Methods in Theoretical Quantum Optics. Clarendon, 1997.

Scully M.O., Zubairy M.S. Quantum Optics. Cambridge University Press, 1997.

Mandel L., Wolf E. Optical Coherence and Quantum Optics. Cambridge University Press, 1995.

Peřina J. Quantum Statistics of Linear and Nonlinear Optical Phenomena. Reidel, 2nd ed. 1991.

Louisell W.H. Quantum Statistical Properties of Radiation. Wiley, 1973, reprinted 1990.

Pike E.R., Walther H., eds. Photons and Quantum Fluctuations. Adam Hilger, 1988.

Peřina J. Coherence of Light. Reidel, 1971, 2nd ed. 1985.

Heitler W. The Quantum Theory of Radiation. Clarendon, 3rd ed. 1954; Dover, reissued 1984.

Goldin E. Waves and Photons: An Introduction to Quantum Optics. Wiley, 1982.

Haken H. Light: Waves, Photons, Atoms. Vol. 1. North-Holland, 1981.

Saleh B. Photoelectron Statistics. Springer-Verlag, 1978.

Louisell W.H. Radiation and Noise in Quantum Electronics. McGraw-Hill, 1964; Krieger, reissued 1977.

ter Haar D. The Old Quantum Theory.Pergamon, 1967; contains English translations of key early papers by Planck, Einstein, Rutherford, and Bohr.

DeWitt C., Blandin A., Cohen-Tannoudji C., eds. Quantum Optics and Electronics. Gordon and Breach, 1965.

КНИГИ ПО СТАТИСТИЧЕСКОЙ ФИЗИКЕ И ТЕРМОДИНАМИКЕ

Blundell S., Blundell K. Concepts in Thermal Physics. Oxford University Press, paperback ed. 2006.

Plischke M., Bergersen B. Equilibrium Statistical Physics. World Scientific, 3rd ed. 2006.

Kittel C. Elementary Statistical Physics. Wiley, 1958; Dover, reissued 2004.

Reif F. Customized Complete Statistical Physics: Berkeley Physics Course. Vol. 5. McGraw-Hill, 1998.

СТАТЬИ

Glauber R.J. One Hundred Years of Light Quanta. In K. Grandin, ed. The Nobel Prizes 2005; Nobel Foundation, 2006. P. 75—98.

Jacques V., Wu E., Toury T., Treussart F., Aspect A., Grangier P., Roch J.-F. Single-Photon Wavefront-Splitting Interference. An Illustration of the Light Quantum in Action. The European Physical Journal D. Vol. 35, 2005. P. 561—565.

Special issue on trends in quantum optics. Journal of Optics B. Vol. 6, № 3, 2004.

Roychoudhuri C., Roy R., eds. The Nature of Light: What is a Photon? OPN Trends (Optics & Photonics News). Vol. 3, № 1, 2003.

Regelman D.V., Mizrahi U., Gershoni D., Ehrenfreund E., Schoenfeld W.V., Petroff P.M. Semiconductor Quantum Dot. A Quantum Light Source of Multicolor Photons with Tunable Statistics. Physical Review Letters. Vol. 87, 2001. 257401.

Agarwal G.S., ed. Selected Papers on Fundamentals of Quantum Optics. SPIE Optical Engineering Press (Milestone Series Volume 103), 1995.

Mandel L., Wolf E., eds. Selected Papers on Coherence and Fluctuations of Light (1850—1966). SPIE Optical Engineering Press (Milestone Series. Vol. 19), 1990.

Teich M.C., Saleh B.E.A. Squeezed and Antibunched Light. Physics Today. Vol. 43, № 6, 1990. P 26—34.

Teich M.C., Saleh B.E.A. Squeezed States of Light. Quantum Optics. Vol. 1, 1989. P. 153—191.

Teich M.C., Saleh B.E.A. Photon Bunching and Antibunching. E. Wolf, ed. In Progress in Optics. North-Holland. Vol. 26, 1988. P. 1—104.

Special issue on squeezed states of the electromagnetic field, Journal of the Optical Society of America B. Vol. 4, № 10, 1987.

Special issue on squeezed light, Journal of Modern Optics. Vol. 34, № 6/7, 1987.

Special issue on quantum-limited imaging and image processing, Journal of the Optical Society of America A. Vol. 3, № 12, 1986.

Teich M.C., Saleh B.E.A. Observation of Sub-Poisson Franck-Hertz Light at 253.7 nm, Journal of the Optical Society of America B. Vol. 2, 1985. P. 275—282.

Wolf E. Einstein's Researches on the Nature of Light. Optics News. Vol. 5, № 1, 1979. P. 24—39.

Weinberg S. Light as a Fundamental Particle. Physics Today. Vol. 28, № 6, 1975. P. 32—37.

Mandel L., Wolf E., eds. Selected Papers on Coherence and Fluctuations of Light. Vol. 1 and 2. Dover, 1970.

Mandel L., Wolf E. Coherence Properties of Optical Fields. Reviews of Modern Physics. Vol. 37, 1965. P. 231—287.

Mandel L. Fluctuations of Light Beams. E. Wolf, ed. In Progress in Optics North-Holland. Vol. 2, 1963.

Задачи

К РАЗДЕЛУ 12.1

1. **Энергия фотона.**

а. Какое напряжение нужно приложить, чтобы разогнать электрон с нулевой скоростью до такой же энергии, как у фотона с длиной волны $\lambda_0 = 0{,}87$ мкм?

б. Фотон с длиной волны 1,06 мкм комбинирует с фотоном, длина волны которого 10,6 мкм, и образует фотон, энергия которого равна сумме энергий обоих фотонов. Какова длина волны результирующего фотона? Взаимодействия фотонов такого типа обсуждаются в гл. 21.

2. **Положение одиночного фотона на экране.** Рассмотрим монохроматический пучок света с длиной волны λ_0, падающий на бесконечный экран в плоскости $z = 0$ и имеющий интенсивность $I(\rho) = I_0 \exp(-\rho/\rho_0)$, где $\rho = \sqrt{x^2 + y^2}$ Предположим, что интенсивность источника снижается до такого значения, что на экран попадает единственный фотон.

а. Найдите вероятность того, что фотон попадет в круг радиусом ρ_0 с центром в начале координат.

б. Если пучок содержит ровно 10^6 фотонов, сколько фотонов попадет в этот круг в среднем?

3. **Импульс свободного фотона.** Сравните полный импульс фотонов в лазерном импульсе, имеющем энергию 10 Дж, с импульсом тела массой 1 г, движущегося со скоростью 1 см/с, и с импульсом электрона, движущегося со скоростью $c_0/10$.

4. **Импульс фотона в гауссовом пучке.**

а. Какова вероятность того, что вектор импульса фотона, принадлежащего гауссову пучку с радиусом перетяжки W_0, лежит в пределах угла расходимости θ_0? Необходимые определения см. в разд. 3.1.

б. Справедливо ли в этом случае соотношение $p = E/c_0$?

5. **Левитация за счет светового давления.** Рассмотрим изолированный атом водорода массой $1{,}66 \cdot 10^{-27}$ кг.

а. Найдите силу тяжести, действующую на атом вблизи поверхности Земли, предполагая, что на уровне моря ускорение свободного падения $g = 9{,}8$ м/с^2.

б. Пусть лазерный пучок с энергией фотонов 1 эВ направлен вертикально вверх и сфокусирован таким образом, чтобы полный импульс каждого фотона передавался атому. Найдите среднюю силу, действующую на атом, при условии, что каждую секунду атомом поглощается один фотон.

в. Найдите число фотонов в секунду и соответствующую оптическую мощность, необходимые для того, чтобы атом не падал в поле тяжести в условиях идеального вакуума.

г. Сколько фотонов в секунду потребовалось бы для этого, если бы атом был идеально отражающим?

6. **Одиночный фотон в резонаторе Фабри—Перо.** Рассмотрим резонатор Фабри—Перо длиной $d = 1$ см с идеально отражающими зеркалами, заполненный

непоглощающей средой с показателем преломления $n = 1{,}5$. Предположим, что в моде, описываемой стоячей волной $\sin(10^5 \pi x/d)$, имеется единственный фотон.

а. Определите длину волны фотона и его энергию (в эВ).

б. Оцените неопределенность координаты и импульса фотона (по величине и направлению). Сравните со значениями, полученными из соотношения $\sigma_x \sigma_p = \hbar/2$.

7. **Однофотонные биения (временна́я интерференция).** Рассмотрим приемник, освещаемый полихроматической плоской волной, состоящей из двух монохроматических волн, наложенных друг на друга и распространяющихся в одном направлении. Составляющие волны имеют комплексные волновые функции

$$U_1(t) = \sqrt{I_1}\exp(j2\pi\nu_1 t); \quad U_2(t) = \sqrt{I_2}\exp(j2\pi\nu_2 t)$$

с частотами ν_1 и ν_2 и интенсивностями I_1 и I_2 соответственно. Согласно волновой оптике (см. подразд. 2.6.2), интенсивность такой волны дается формулой

$$I(t) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\left[2\pi(\nu_2 - \nu_1)t\right].$$

Предположим, что обе составляющие волны имеют одинаковые интенсивности ($I_1 = I_2$) и что волны достаточно слабые, так что лишь один полихроматический фотон достигает приемника в течение отрезка времени $T = 1/|\nu_2 - \nu_1|$.

а. Постройте график плотности вероятности $p(t)$ времени детектирования фотона для $0 \le t \le 1/|\nu_2 - \nu_1|$. В какой момент времени в этом промежутке вероятность регистрации фотона равна нулю?

б. Попытка определить, из какой из составляющих волн пришел фотон, требует измерения энергии с точностью лучше чем $\sigma_E < \hbar|\nu_2 - \nu_1|$. Используйте соотношение неопределенностей между временем и энергией, чтобы показать, что время, необходимое для такого измерения, имеет порядок периода биений. Таким образом, сам процесс измерения смазывает картину интерференции и исключает возможность ее наблюдения

8. **Изменение импульса фотона в светоделителе.** Рассмотрим единичный фотон в моде, описываемой плоской волной, падающей на идеальный светоделитель без потерь. Каков вектор импульса фотона до того, как он попадет на зеркало? Каковы возможные значения импульса фотона и вероятности наблюдения этих значений после прохождения светоделителя?

К РАЗДЕЛУ 12.2

1. **Поток фотонов.** Покажите, что мощность оптического монохроматического пучка, несущего в среднем один фотон на оптический период, обратно пропорциональна квадрату длины волны.

2. **Распределение Пуассона.** Проверьте, что распределение Пуассона (12.34) нормировано на единицу и имеет среднее $\bar{n}$ и дисперсию $\sigma_n^2 = \bar{n}$.

3. **Статистика фотонов когерентного гауссова пучка.** Предположим, что одномодовый He–Ne-лазер мощностью 100 пВт излучает свет на 633 нм в виде гауссова пучка $\mathrm{TEM}_{0,0}$ (см. гл. 3).

а. Каково среднее число фотонов, пересекающих круг, радиус которого равен радиусу перетяжки пучка W_0 за время $T = 100$ нс?

б. Каково среднеквадратичное значение числа фотонных отсчетов в случае (а)?

в. Какова вероятность того, что в случае (а) не будет ни одного фотонного отсчета?

4. **Распределение Бозе—Эйнштейна.**

а. Проверьте, что распределение Бозе—Эйнштейна (12.41) нормировано и имеет среднее значение $\bar{n}$ и дисперсию $\sigma_n^2 = \bar{n} + \bar{n}^2$.

б. При условии, что подчиняющийся статистике Бозе—Эйнштейна пучок несет в среднем $\Phi = 1$ фотон в наносекунду, какова вероятность того, что ни одного фотона не будет зарегистрировано в течение 20 нс?

5. **Отрицательно-биномиальное распределение.** В литературе по теории вероятностей хорошо известно, что сумма $\mathcal{M}$ идентично распределенных случайных переменных, каждое из которых имеет геометрическое (Бозе—Эйнштейна) распределение, подчиняется отрицательному биномиальному распределению

$$p(n) = \binom{n + \mathcal{M} - 1}{n} \frac{(\bar{n}/\mathcal{M})^n}{(1 + \bar{n}/\mathcal{M})^{n+\mathcal{M}}}.$$

Проверьте, что это распределение сводится к распределению Бозе—Эйнштейна при $\mathcal{M} = 1$ и к распределению Пуассона при $\mathcal{M} \to \infty$.

6. **Статистика фотонов для многомодового теплового излучения в резонаторе.** Рассмотрим $\mathcal{M}$ мод теплового излучения, достаточно близких друг к другу по частоте, чтобы можно было считать каждую из них заселенной, в соответствии с распределением Бозе—Эйнштейна, одинаковым средним числом фотонов $1/[\exp(h\nu/k_B T) - 1]$. Покажите, что дисперсия полного числа фотонов связана с его средним значением формулой

$$\sigma_n^2 = \bar{n} + \frac{\bar{n}^2}{\mathcal{M}},$$

отражающей тот факт, что многомодовое тепловое излучение имеет меньшую дисперсию, чем одномодовое. Присутствие многих мод обеспечивает усреднение, снижающее шумовой характер света.

7. **Статистика фотонов в пучке многомодового теплового излучения.** Источник многомодового теплового излучения генерирует $\mathcal{M}$ одинаковых мод, каждая с экспоненциально распределенной (случайной) проинтегрированной скоростью счета, имеет полную плотность вероятности $p(w)$, описываемую гамма-распределением

$$p(w) = \frac{1}{(\mathcal{M}-1)!}\left(\frac{\mathcal{M}}{\langle w\rangle}\right)^{\mathcal{M}} w^{\mathcal{M}-1} \exp\left(-\frac{\mathcal{M}w}{\langle w\rangle}\right), \quad w \geq 0.$$

С помощью формулы Манделя (12.47) покажите, что результирующее распределение числа фотонов приобретает вид отрицательно-биномиального распределения, введенного в задаче 5 к разд. 12.2.

8. **Среднее значение и дисперсия дважды случайного распределения Пуассона.** Докажите (12.49) и (12.50).

9. **Случайное разбиение когерентного света.**

а. С помощью (12.54) покажите, что распределение числа фотонов при случайном разбиении когерентного света сохраняет пуассоновский вид.

б. Покажите явно, что среднее число фотонов для света, отраженного от светоделителя без потерь, равно $(1 - \mathcal{T})\bar{n}$.

в. Докажите (12.55) для когерентного света.

10. **Случайное разбиение одномодового теплового излучения.**

а. С помощью (12.54) покажите, что распределение числа фотонов при случайном разбиении одномодового теплового излучения сохраняет вид распределения Бозе—Эйнштейна.

б. Покажите явно, что среднее число фотонов для света, отраженного от светоделителя без потерь, равно $(1 - \mathcal{T})\bar{n}$.

в. Докажите (12.56) для когерентного света.

11. **Экспоненциальное затухание среднего числа фотонов в поглотителе.**

а. Рассмотрим поглощающий материал толщины d с коэффициентом поглощения α (см$^{-1}$). Среднее число фотонов, входящих в среду, равно $\bar{n}_0$. Напишите дифференциальное уравнение для нахождения среднего числа фотонов $\bar{n}(x)$ на глубине x $(0 \le x \le d)$.

б. Решите это дифференциальное уравнение. Объясните, почему в результате получается закон экспоненциального затухания, ранее полученный из электромагнитной оптики (см. подразд. 5.5.1).

в. Запишите выражение для распределения числа фотонов $p(n)$ в произвольной точке на расстоянии x от границы поглощающей среды, если на нее падает когерентный свет.

г. Какова вероятность того, что единичный фотон, падающий на границу поглощающей среды, пройдет через нее?

К РАЗДЕЛУ 12.3

1. **Статистика биномиального распределения числа фотонов.** Биномиальное распределение вероятностей записывается как

$$p(n) = \frac{M!}{(M-n)!\,n!}\,p^n(1-p)^{M-n}$$

Оно описывает статистику отсчетов для некоторых источников света, сжатого по отношению к числу фотонов.

а. Укажите возможный механизм преобразования света с определенным числом фотонов в свет, описываемый биномиальной статистикой фотонов.

б. Докажите, что биномиальное распределение вероятностей нормировано на единицу.

в. Найдите среднее число отсчетов $\bar{n}$ и дисперсию σ_n^2 для биноминального распределения вероятностей, выразив их через два параметра, p и M.

г. Найдите выражение для отношения сигнал—шум через $\bar{n}$ и p. Вычислите его для двух предельных случаев, $p \to 0$ и $p \to \infty$. Каким типам света соответствуют эти предельные случаи?

2. **Шумы гипотетического источника фотонов**. Рассмотрим гипотетический источник света, генерирующий фотонный поток с кусочно-однородным распределением вида

$$p(n) = \begin{cases} \dfrac{1}{2\bar{n}+1}, & 0 \le n \le 2\bar{n}; \\ 0 & \text{в остальных случаях.} \end{cases}$$

а. Убедитесь, что распределение нормировано на единицу и имеет среднее значение $\bar{n}$. Вычислите дисперсию числа фотонов σ_n^2 и отношение сигнал—шум (ОСШ) и сравните их с полученными для распределений Бозе—Эйнштейна и Пуассона с такими же средними.

б. Используя ОСШ, сравните шумовые свойства данного источника и идеального одномодового лазера при $\bar{n} < 2$, $\bar{n} = 2$ и $\bar{n} > 2$.

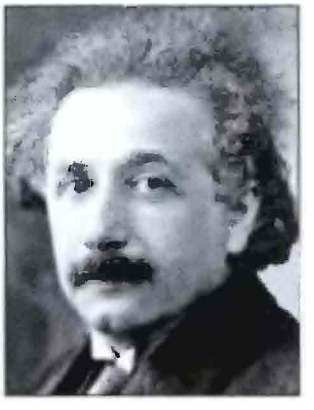
в. Во сколько раз ОСШ данного источника больше, чем для одномодового теплового излучения?

Полезные формулы:

$$1 + 2 + 3 + \ldots + j = \frac{j(j+1)}{2}; \quad 1^2 + 2^2 + 3^2 + \ldots + j^2 = \frac{j(j+1)(2j+1)}{6}.$$

ГЛАВА 13 ФОТОНЫ И АТОМЫ

Нильс Бор (1885—1962)



Альберт Эйнштейн (1879—1955)

Бор и Эйнштейн заложили теоретические основы описания взаимодействия света с веществом.

Свет взаимодействует с веществом, поскольку вещество содержит электрические заряды. Переменное электрическое поле света действует на электрические заряды и диполи атомов, молекул и твердых тел с некоторой силой, заставляя их колебаться, так что они испытывают ускорение. Колеблющиеся электрические заряды поглощают и излучают свет.

Атомы, молекулы и твердые тела имеют специфические разрешенные уровни энергии и энергетические зоны, определяемые законами квантовой механики. Фотон может взаимодействовать с атомом, если его энергия совпадает с разностью двух энергетических уровней. Если атом первоначально находился на нижнем из уровней, то фотон может отдать атому свою энергию, в результате чего атом перейдет на более высокий уровень энергии; в этом случае говорят, что фотон поглощается (уничтожается). Если же атом находится на более высоком уровне энергии, то фотон может стимулировать переход атома на более низкий уровень с испусканием (рождением) второго фотона, энергия которого равна разности энергий между начальным и конечным уровнями. При определенных условиях вынужденное излучение может приводить к генерации лазерного света.

Тепловые возбуждения вынуждают атомы вещества постоянно совершать переходы вверх и вниз между разрешенными уровнями энергии, испуская и

поглощая фотоны. Для черных тел при тепловом равновесии в стационарных условиях получающийся коллектив атомов и фотонов порождает тепловое излучение. Все черные тела, температура которых выше абсолютного нуля, испускают тепловое излучение, частотное распределение которого известно как спектр излучения абсолютно черного тела. С ростом температуры объекта заселяются более высокие его уровни, сдвигая пик спектра излучения в сторону более высоких частот (более коротких волн).

Испускание фотонов может возникать от иных источников энергии, чем тепловые возбуждения. Ультрафиолетовое облучение, звуковые волны, электрический ток и химические реакции могут заставлять атомы испускать свет, называемый люминесценцией. Другие процессы также могут приводить к испусканию света; например, к ним относится движение заряженных частиц со скоростью, превышающей скорость света в среде (черенковское излучение), а также торможение заряженных частиц в веществе (тормозное излучение).

Фотон, падающий на вещество, может также изменить направление и энергию в результате рассеяния — процесса, по которому можно судить об уровнях энергии вещества, например, связанных с молекулярными колебаниями.

Применение законов, изложенных в данной главе, к действию лазерных усилителей и генераторов рассмотрено в гл. 14 и 15 соответственно.

О данной главе

Цель данной главы состоит в формулировке законов, которым подчиняется взаимодействие света с веществом. Эти законы отвечают за генерацию лазерного, теплового и люминесцентного излучения. Глава начинается с краткого обзора характерных энергетических уровней, связанных с различными видами вещества (см. разд. 13.1), и заселения этих уровней (см. разд. 13.2). В разд. 13.3 мы обсудим поглощение и испускание фотонов атомами. Взаимодействие большого числа фотонов с большим числом атомов в стационарных условиях при тепловом равновесии рассмотрено в разд. 13.4. Наконец, элементарное описание люминесценции и рассеяния света дано в разд. 13.5.

13.1. УРОВНИ ЭНЕРГИИ

Атомы вещества могут существовать в относительной изоляции, как, например, в разреженном атомном газе, или взаимодействовать с соседними атомами, образуя молекулы, жидкости и твердые тела. Составные части вещества подчиняются законам квантовой механики.

Поведение одной нерелятивистской частицы массой m (например, электрона) в поле с потенциальной энергией $V(\boldsymbol{r}, t)$ описывается комплексной волновой функцией $\Psi(\boldsymbol{r}, t)$, которая удовлетворяет уравнению Шредингера:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(\boldsymbol{r}, t)+V(\boldsymbol{r}, t)\Psi(\boldsymbol{r}, t)=-j\nabla\frac{\partial\Psi(\boldsymbol{r}, t)}{\partial t}. \qquad (13.1)$$

Потенциальная энергия характеризует окружение частицы, включая вклады от приложенных извне оптических полей. Дифференциальное уравнение в частных производных (13.1) имеет множество решений, зависящих от конкретного вида потенциала $V(\boldsymbol{r}, t)$. Системы многих частиц, такие как атомы, молекулы, жидкости и твердые тела, описываются более сложным уравнением Шредингера, у которого потенциал содержит члены, описывающие взаимодействие между частицами. Уравнение (13.1) математически эквивалентно параксиальному уравнению Гельмгольца (2.7) в волновой оптике и уравнению (22.24) для медленной огибающей в оптике ультракоротких импульсов.

Постулат Борна утверждает, что вероятность обнаружить частицу в бесконечно малом объеме $\mathrm{d}V$ вблизи точки $\boldsymbol{r}$ в промежуток времени между t и $t + \mathrm{d}t$ равна

$$p(\boldsymbol{r}, t)\,\mathrm{d}V\,\mathrm{d}t = \left|\Psi(\boldsymbol{r}, t)\right|^2 \mathrm{d}V\,\mathrm{d}t. \tag{13.2}$$

Уравнение (13.2) напоминает (12.4) для вероятности обнаружения фотона в бесконечно малом объеме за бесконечно малый промежуток времени.

В отсутствие временно́й зависимости потенциала возможные значения энергии частицы определяются с помощью метода разделения переменных. Он приводит к решению уравнения (13.1) вида

$$\Psi(\boldsymbol{r}, t) = \psi(\boldsymbol{r}) \exp\left(j\frac{E}{\hbar}t\right),$$

где $\psi(\boldsymbol{r})$ — удовлетворяет стационарному уравнению Шредингера

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\boldsymbol{r}) + V(\boldsymbol{r}, t)\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}). \tag{13.3}$$

Уравнение (13.3), которое совпадает с уравнением Гельмгольца, можно рассматривать как задачу на собственные значения, в которой допустимые значения энергии E являются собственными значениями, а решения $\psi(\boldsymbol{r})$ — собственными функциями.

Системы многих частиц описываются обобщением уравнения (13.3). Решения дают допустимые значения энергии системы. Эти значения могут быть дискретными (как в случае атома), или непрерывными (как у свободной частицы), или образовывать плотные группы близких дискретных уровней, называемые зонами (как у полупроводников). Тепловые возбуждения или внешние поля, например свет, освещающий вещество, могут индуцировать переходы с одного энергетического уровня на другой. Именно таким путем система обменивается энергией с внешним миром.

В следующих разделах мы рассмотрим некоторые типичные схемы энергетических уровней атомов, молекул и твердых тел.

13.1.1. Атомы

Уровни энергии атомов определяются потенциальной энергией электронов в поле ядра и других электронов, а также силами, связанными с орбитальным и спиновым угловыми моментами, которые обычно намного сла-

бее, чем взаимодействие зарядов. Многие атомы и ионы используются в качестве активных сред лазеров (см. разд. 15.3).

Водород

Уровни энергии водородоподобного атома, состоящего из ядра с зарядом $+Ze$ и единственного электрона зарядом $-e$ и массой m, где Z — **атомный номер**, определяются путем подстановки кулоновского потенциала $V(r) = -Ze^2/r$ в стационарное уравнение Шредингера (13.3). Поскольку $V(r)$ зависит только от радиальной координаты, оператор Лапласа можно записать в сферических координатах, после чего уравнение в частных производных сводится к трем обыкновенным дифференциальным уравнениям с помощью метода разделения переменных. Это дает возможность решить задачу на собственные значения точно. Собственные значения представляют собой бесконечный набор энергетических уровней

$$E_n = -\frac{M_r Z^2 e^4}{(4\pi\varepsilon_0)^2 2\hbar^2}\frac{1}{n^2}, \quad n = 1,\ 2,\ 3, \tag{13.4}$$

где введение приведенной массы атома M_r вместо массы электрона m позволяет учесть конечность массы ядра. Уровни энергии (13.4), характеризуемые единственным квантовым числом n, которое называется главным квантовым числом, показаны на рис. 13.1 для H и C^{5+} Эти уровни можно получить также приравнивая кулоновскую силу притяжения к центростремительной силе, необходимой для удержания электрона на круговой орбите, при условии, что орбитальный момент импульса электрона принимает значения, кратные $\hbar$. Радиусы соответствующих боровских орбит оказываются равными

$$r_n = 4\pi\varepsilon_0 \frac{n^2\hbar^2}{mZe^2}, \quad n = 1,\ 2,\ 3,$$

радиус орбиты с n = 1 обозначен через r_1. Такое рассмотрение лежит в основе **теории Бора**, являющейся частью «старой квантовой теории».

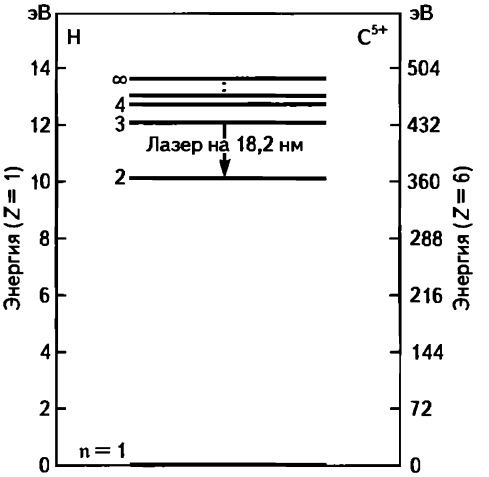


Рис. 13.1. Энергетические уровни водорода ($Z = 1$, левая ось ординат) и C^{5+} (водородоподобный ион с $Z = 6$; правая ось ординат). Переход между уровнями n = 3 и n = 2, отмеченный стрелкой, соответствует рабочему переходу лазера на C^{5+} с длиной волны 18,2 нм в вакуумном ультрафиолетовом диапазоне (см. подразд. 15.3.3). Произвольное начало отсчета энергии совпадает с уровнем n = 1

Собственные функции уравнения Шредингера получаются в виде произведения трех функций

$$\psi_{n\ell m}(r,\ \theta,\ \phi) = \mathbb{R}_{n\ell}(r)\Theta_{\ell m}(\theta)\Phi_m(\phi),$$

где $n = 1, 2, 3, \ldots$ — **главное квантовое число**, $\ell = 0, 1, 2, \ldots, n-1$ — **азимутальное квантовое число**; $m = 0, \pm1, \pm2, \ldots, \pm l$ — **магнитное квантовое число**. Здесь $\mathbb{R}_{n\ell}(r)$ представляют собой присоединенные функции Лагерра (они тесно связаны с обобщенными полиномами Лагерра, которые обсуждались в сноске к гл. 3), $\Theta_{\ell m}(\theta)$ — присоединенные функции Лежандра, а $\Phi_m(\phi)$ — фазовые функции. Эти решения аналогичны решениям для поля в микрорезонаторе, которое обсуждалось в подразд. 10.4.3.

Для учета собственного (спинового) момента электрона требуется еще одно квантовое число — **спиновое квантовое число** $s = \pm1/2$. Взаимодействие спинового момента с орбитальным угловым моментом, которое называется спин-орбитальным взаимодействием, расщепляет энергетические уровни на близко расположенные, но отдельные компоненты, которые называются **тонкой структурой**. Спин электрона взаимодействует также с магнитным моментом ядра, что приводит к еще более тонкому расщеплению уровней, называемому **сверхтонкой структурой**. Эти эффекты приводят к отличию энергетических уровней тома водорода от значений, даваемых формулой (13.4), но очень небольшому. Релятивистские поправки к уровням энергии также можно учесть в рамках формализма, известного как теория Дирака. Эта теория автоматически приводит к понятию спина в силу ее релятивистской инвариантности.

Многоэлектронные атомы

Многоэлектронные атомы состоят из ядра с зарядом $+Ze$, окруженного Z-электронами, каждый с зарядом $-e$. Уровни энергии многоэлектронных атомов можно определить с использованием теории Шредингера, поскольку можно пренебречь релятивистскими эффектами (эти эффекты значительны только для наиболее легких атомов). Для того чтобы учесть взаимодействие множества электронов друг с другом, уравнение Шредингера решается приближенным методом, известным как метод Хартри. Каждый электрон рассматривается как независимо движущийся в сферически симметричном среднем потенциале $V(r)$, который складывается из сферически симметричного потенциала кулоновского притяжения электрона к ядру и сферически симметричного потенциала отталкивания, представляющего усредненное действие кулоновских сил отталкивания со стороны всех других электронов. При этих предположениях уравнение Шредингера для Z электронов расщепляется на Z одноэлектронных уравнений Шредингера, причем полная волновая функция представляет собой произведение волновых функций отдельных электронов, а полная энергия — сумму энергий отдельных электронов. Затем для учета отклонений потенциала отталкивания от сферической симметрии, а также взаимодействий с участием электронного спина используется теория возму-

шений. Получающиеся одноэлектронные волновые функции тесно связаны с собственными функциями электрона в атоме водорода и записываются в таком же виде.

С ростом атомного номера Z-заселение последовательных одноэлектронных состояний происходит в соответствии с двумя принципами: минимальность полной энергии и **принцип запрета Паули**, согласно которому два и более электрона не могут иметь одинаковые наборы всех четырех квантовых чисел. Состояния заполняются в виде оболочек (нумеруемых главным квантовым числом n). Внутри оболочки выделяются **подоболочки**, нумеруемые парой квантовых чисел $n\ell$, где значение ℓ обычно указывается в спектроскопических обозначениях (буквам s, p, d, f, g, h, i соответствуют значения $\ell = 0, 1, 2, 3, 4, 5, 6$). **Электронная конфигурация** $n\ell^u$ представляет распределение электронов по подоболочкам, индекс u указывает число электронов в каждой из них. Например, электронная конфигурация основного состояния He ($Z = 2$) $1s^2$, его два электрона полностью заполняют оболочку с n = 1 (для которой $\ell = 0$). Конфигурации низколежащих возбужденных состояний He включают $1s2s$ и $1s2p$, в которых один из электронов возбуждается в оболочку с n = 2. Для Ne ($Z = 10$) конфигурация основного состояния $1s^22s^22p^6$; его 10 электронов полностью заполняют оболочки с n = 1 и n = 2.

Каждая из таких электронных конфигураций соответствует группе близко расположенных энергетических уровней тонкой структуры, которые называются **мультиплетом.** Примеры мультиплетов для атомов He и Ne схематически показаны на рис. 13.2. Расщепление в мультиплете обусловлено главным образом спин-орбитальным взаимодействием, т. е. взаимодействием полного спина с полным орбитальным моментом атома. Такая схема связи, известная как связь Рассела—Саундерса (РС), работает для всех атомов, кроме наиболее тяжелых. Для данной электронной конфигурации различные угловые моменты атома отображаются символом терма $^{2\mathcal{S}+1}\mathcal{L}_j$, где $\mathcal{S}$ — квантовое число полного спина, $2\mathcal{S} + 1$ — спиновая мультиплетность (например, синглет, триплет), $\mathcal{L}$ — квантовое число полного орбитального момента в спектроскопическом обозначении (заглавные буквы S, P, D, F,... для $\mathcal{L} = 0, 1, 2, 3$ соответственно), а j — квантовое число полного углового момента атома. Символ терма атома или иона часто приводится сразу после электронной конфигурации. Например, наиболее низко лежащие возбужденные синглетные и триплетные состояния He обозначаются $1s2s\,^1S_0$ и $1s2s\,^3S_0$, соответственно, как показано на рис. 13.2. Когда все заселенные подоболочки заполнены (как в случае основных состояний всех инертных газов, а также ряда других атомов, таких как Ca, Cd, Yb и Hg), символ терма имеет вид $^1S_0$. Типичная величина сдвига энергетических уровней из-за спин-орбитального взаимодействия составляет 1 часть из 10^4 для H и растет с увеличением атомного номера Z. Релятивистские эффекты для водорода дают сдвиги уровней того же порядка, но они не зависят от атомного номера и поэтому могут быть опущены при рассмотрении всех атомов, кроме наиболее легких. Сверхтонкие сдвиги еще примерно в 10^3 раз меньше. Существуют и другие взаимодействия, например, спин-спиновые, но они пренебрежимо малы.

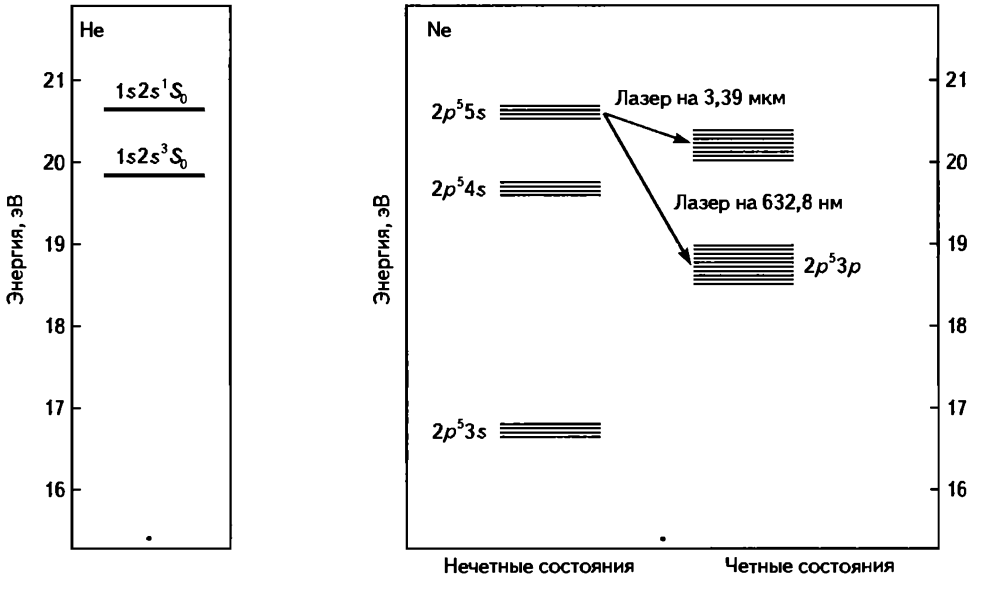


Рис. 13.2. Избранные уровни энергии атомов He и Ne. Показаны символы электронных конфигураций и термов (у атома Ne конфигурации заполненных оболочек $1s^22s^2$ опущены для краткости). Энергетические расстояния между компонентами тонкого расщепления показаны схематически и сильно преувеличены. Переходы в Ne, отмеченные стрелками, соответствуют длинам волн 3,39 мкм и 632,8 нм. Эти переходы, лежащие в средней ИК- и видимой области спектра, соответственно, обычно используются в He—Ne-лазерах (см. подразд. 14.3.5 и 15.3.4). Близкое совпадение возбужденных уровней He и Ne позволяет осуществить возбуждение атомов Ne при столкновениях с He в газовом разряде, отсюда название «He—Ne-лазер»

Чем больше значение n, тем менее прочно связан электрон в атоме, поскольку кулоновское экранирование внутренними электронами уменьшает потенциал, создаваемый ядром. В результате оболочки в типичном случае заполняются в порядке $n = 1, 2, 3, \ldots$. Аналогично, чем больше значение ℓ, тем менее прочно связан электрон, поскольку при этом плотность вероятности обнаружения электрона постепенно сдвигается в периферийную область. Поэтому подоболочки обычно заполняются в последовательности s, p, d, f, Вследствие такой последовательности заполнения многие свойства химических элементов являются периодическими функциями Z, что демонстрирует периодическая таблица элементов, приведенная на рис. 13.3. Следующие друг за другом строки таблицы соответствуют последовательным значениям главного квантового числа n. Каждый столбец таблицы содержит элементы, физические и химические свойства которых имеют определенное сходство, поскольку у них одно и то же число электронов на внешних оболочках (валентных электронов). Например, столбец содержит благородные газы, включая He и Ne, которые одноатомны и химически инертны, так как имеют заполненные внешние

оболочки и большую разность энергий между заполненными p-подоболочками и следующими более высокими s-подоболочками. Столбцы I и VII, напротив, содержат элементы, которые химически высокоактивны и легко образуют молекулы. Например, каждый атом щелочного металла в столбце I имеет один внешний электрон. Этим электроном он легко делится с любым оказавшимся вблизи атомом галогена из столбца VII, который нуждается в таком электроне для заполнения своей внешней оболочки.

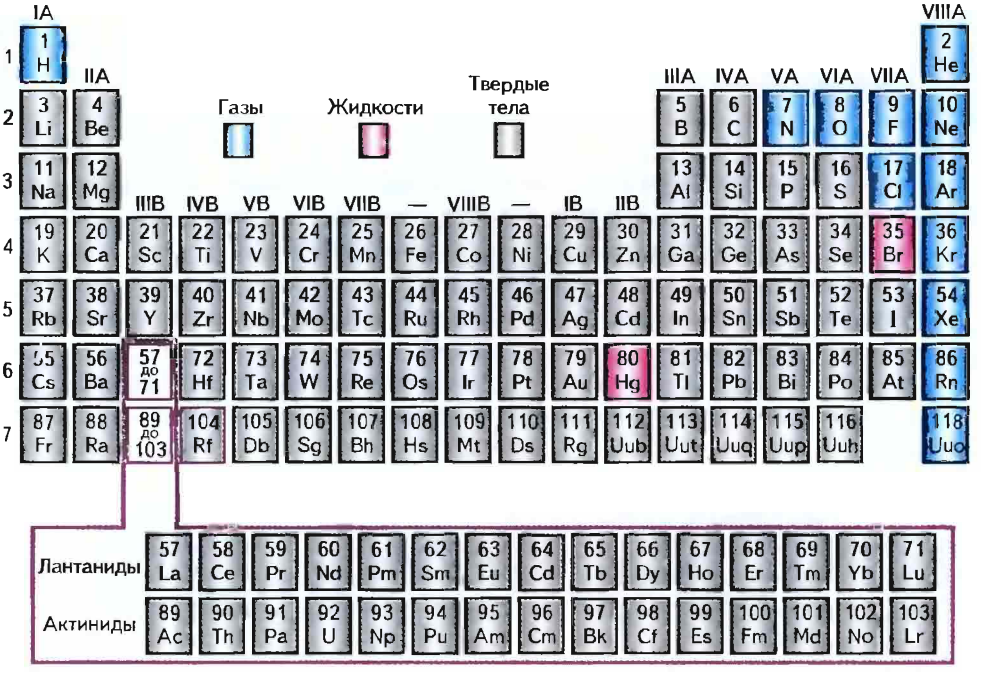


Рис. 13.3. Периодическая таблица элементов с указанием их атомных номеров Z и сокращенных обозначений. Последовательные строки таблицы содержат элементы, у которых валентные электроны имеют одно и то же главное квантовое число n, показанное арабскими цифрами слева. Каждый столбец таблицы, традиционно нумеруемый римскими цифрами, содержит элементы со сходными физическим и химическими свойствами. Элементы, которые при комнатной температуре существуют в виде газов, жидкостей и твердых тел, отмечены синим, розовым и серебристым цветом соответственно

Вообще говоря, многоэлектронные атомы и ионы обладают огромным многообразием уровней энергии, и формулы типа (13.4) справедливы только для оптически активных электронов, например валентных электронов в атомах щелочных металлов. Хотя оптические переходы обычно затрагивают только валентные электроны, это многообразие энергетических уровней порождает изобилие переходов, многие из которых пригодны для лазерной генерации (см. подразд. 14.3.5 и 15.3.4). Например, разности энергий между возбужденными

уровнями атома Ne, показанными на рис. 13.2, соответствуют длинам волн, лежащим в инфракрасной и видимой областях спектра, и в типичном случае достигают значений в несколько электронвольт (связь между различными единицами энергии см. на рис. 12.3).

13.1.2. Молекулы

Молекулы могут образовываться при соединении двух и более атомов. Молекулярные уровни энергии частично определяются потенциальной энергией, связанной с силами притяжения, связывающими атомы. Устойчивая молекула образуется, когда разделение валентных электронов между составляющими атомами приводит к понижению полной энергии системы. Два основных типа связи в молекулах — это **ионная связь** и **ковалентная связь**. Для молекул с ионной связью (таких как HF) области положительного и отрицательного зарядов пространственно разделены, так что молекула обладает постоянным дипольным моментом. Для молекул с ковалентной связью (таких как H_2) составляющие атомы полностью делят между собой валентные электроны, так что в результате молекула не имеет постоянного дипольного момента. Для неодинаковых ядер связь может быть частично ионной, частично ковалентной. Тип связи играет важную роль в определении структуры энергетических уровней молекулы.

Энергетические уровни молекулы возникают в результате трех основных взаимодействий и лежат в трех разных областях длин волн: вращательные переходы лежат в микроволновой и дальней инфракрасной области, колебательные — в инфракрасной и электронные — в видимой и ультрафиолетовой. Временные масштабы соответствующих процессов также сильно различаются, что позволяет приближенно рассматривать их по отдельности.

Различные молекулы, от простых газов до органических красителей в растворах, служат активными средами лазеров (см. подразд. 15.3.2 и 15.3.3).

Вращение двухатомной молекулы

Вращение двухатомной молекулы с моментом инерции J вокруг ее центра масс можно рассматривать как вращение жесткого ротатора вокруг оси, перпендикулярной по отношению к межъядерной оси. Классическая вращательная энергия такой системы равна $E_r = L2/2J$, где L — момент импульса системы относительно оси вращения. Согласно квантовой механике, квадрат модуля углового момента такой системы квантуется в соответствии с формулой $L^2 = r(r + 1)\hbar^2$, где r — вращательное квантовое число. Тогда для уровней энергии вращающейся двухатомной молекулы получаем

$$E_r = \frac{1}{2J} r(r+1)\hbar^2, \quad r = 0,\ 1,\ 2, \tag{13.5}$$

Расстояние между уровнями энергии $\hbar\omega$ в типичном случае лежит в диапазоне 10^{-4}—10^{-2} эВ, что соответствует фотонам микроволновой и дальней инф-

ракрасной областей спектра. Расстояние между последовательными энергетическими уровнями растет с увеличением вращательного квантового числа r, в противоположность расстоянию между последовательными электронными уровнями энергии в атоме водорода, которое убывает с ростом главного квантового числа в соответствии с формулой (13.4). Двухатомные молекулы с одинаковыми ядрами (такие как N_2) не имеют постоянного дипольного момента. По этой причине они не обладают чисто вращательными спектрами.

Колебания двухатомной молекулы

Колебания двухатомной молекулы (такой как N_2, CO, или HCl) определяются внутримолекулярным притяжением и происходят под действием возвращающей силы, приблизительно пропорциональной изменению межъядерного расстояния x. Такую систему можно моделировать двумя материальными точками с массами M_1 и M_2, соединенными пружиной. Приведенная масса системы равна

$$M_r = \frac{M_1 M_2}{M_1 + M_2}.$$

Молекулярная постоянная упругости κ может быть определена так, что потенциальная энергия равна

$$V(x) = \frac{1}{2}\kappa x^2.$$

Таким образом, молекулярным колебаниям соответствует модель квантовомеханического гармонического осциллятора. Как указывалось в разд. 12.3, энергетические уровни такой системы описываются формулой

$$E_v = \left(v + \frac{1}{2}\right)\hbar\omega, \quad v = 0,\ 1,\ 2, \tag{13.6}$$

где ω — угловая частота колебаний,

$$\omega = \sqrt{\frac{\kappa}{M_r}};$$

$\hbar\omega/2$ — нулевая энергия. Формула (13.6) аналогична выражению для энергии моды электрического поля (12.5). Типичные значения $\hbar\omega$ для молекулярных колебаний лежат в интервале 0,05—0,5 эВ, соответствующем инфракрасному диапазону спектра (см. уровни энергии N_2 на рис. 13.4).

В отличие от уровней энергии атома водорода и вращательных уровней двухатомной молекулы, колебательные уровни энергии расположены на равном расстоянии друг от друга. На практике, однако, кривые потенциальной энергии большинства молекул становятся ангармоническими с ростом энергии (см. разд. 21.7), в результате чего расстояние между колебательными уровнями

уменьшается с увеличением υ. В процессе колебательного перехода молекула может изменить и свое вращательное состояние, т. е. меняются и υ и r, что порождает колебательно-вращательные спектры.

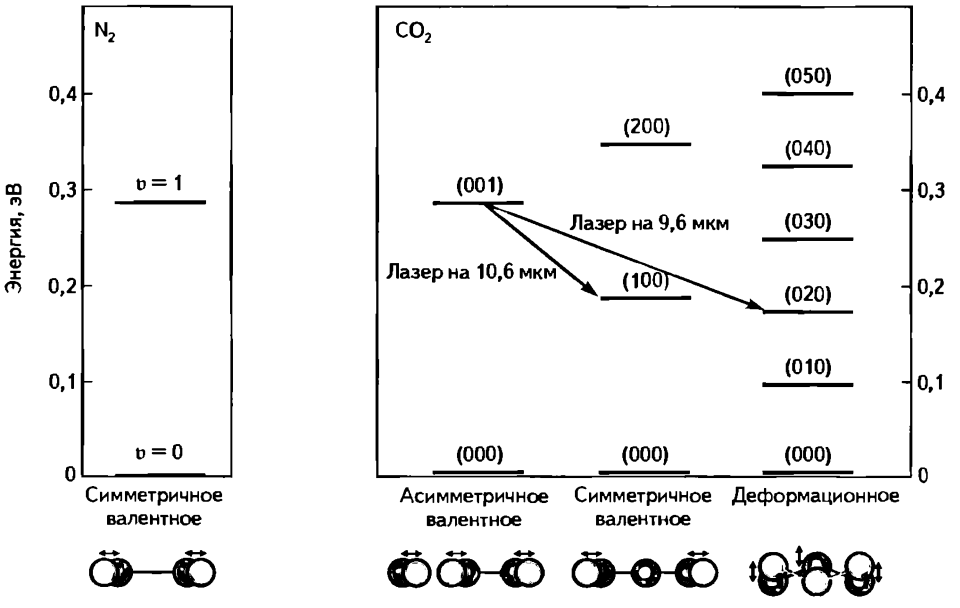


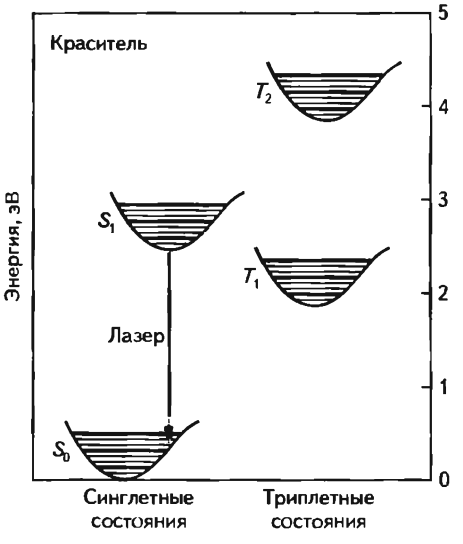
Рис. 13.4. Низшие колебательные уровни энергии молекул N_2 и CO_2 (начало отсчета выбрано при $\upsilon = 0$). Переходы, отмеченные стрелками, сопровождаются обменом энергией между различными нормальными модами и соответствуют $\lambda_0 = 10{,}6$ мкм и $\lambda_0 = 9{,}6$ мкм. Эти переходы используются в лазерах на CO_2 (см. подразд. 14.3.5 и 15.3.4). Каждый колебательный уровень CO_2 имеет множество близко расположенных вращательных подуровней энергии (не показаны)

Колебания трехатомной молекулы

Трехатомной молекулой, имеющей важное значение для фотоники, является двуокись углерода, широко используемая в качестве активной среды лазеров. Эта линейная трехатомная молекула может претерпевать независимые колебания трех типов, показанные на рис. 13.4: асимметричное валентное (АСВ), симметричное валентное (СВ) и деформационное (Д). Каждая из этих нормальных мод ведет себя как независимый квантово-механический гармонический осциллятор со своей постоянной упругости и, соответственно, значением $\hbar\omega$. Допустимые значения энергии являются суммой энергий трех членов вида (13.6) со своими квантовыми числами (υ_1, υ_2, υ_3), характеризующими указанные колебания (см. рис. 13.4). Как и для двухатомных молекул, каждый колебательный уровень расщепляется на близко расположенные вращательные подуровни (не показаны на рис. 13.4), энергии которых приближенно даются формулой (13.5).

Молекулы красителей

Органические красители — большие и сложные молекулы. В них могут происходить электронные, колебательные и вращательные переходы между огромным числом энергетических уровней. Эти уровни существуют как в синглетных (S), так и в триплетных (T) состояниях (см. рис. 12.6). В синглетных состояниях спин возбужденного электрона антипараллелен спину оставшейся части молекулы, а в триплетных — параллелен.



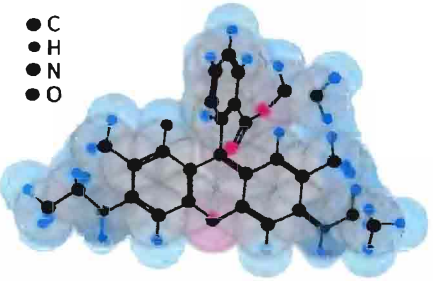


Рис. 13.5. Структура иона родамина 6Ж, химическая формула которого $C_{28}H_{31}N_2O_3^+$:

слева — схематическая иллюстрация лазерного перехода между двумя синглетными электронными состояниями, для которых некоторое различие в строении молекулы отмечено сдвигом по горизонтали. Колебательные и вращательные уровни энергии изображены толстыми и тонкими линиями, соответственно

Разности энергий уровней соответствуют длинам волн излучения, покрывающим широкие области в видимом и ультрафиолетовом диапазонах. На рис. 13.5 схематически представлена часть структуры энергетических уровней родамина 6Ж, который становится ионом при растворении в таких жидкостях, как этиловый спирт или вода. Именно этот краситель используется в качестве лазерной среды в желтой части спектра. Лазеры на органических красителях кратко обсуждаются в подразд. 15.3.3.

13.1.3. Твердые тела

Молекулы или атомы твердых тел расположены в непосредственной близости друг от друга и, как правило, объединяются в периодическую структуру, образуя **кристаллическую решетку**. Величина сил, удерживающих атомы в решетке, по грубой оценке такая же, как и в молекулах. Таким образом, энергетические уровни электрона в твердом теле определяются не только потенциалами отдельных атомов, но и потенциалами соседних атомов решетки. Некристаллические твердые тела, такие как стекла, также имеют упорядоченную структуру, как у кристаллов, но на малых пространственных масштабах.

В обычных твердых телах существует три основных типа связи: ионная, ковалентная и металлическая. Твердые тела с **ионной связью** (такие как CaF_2) имеют решетку, состоящую из положительных и отрицательных ионов со сферически симметричными заполненными оболочками. Поэтому в них отсутствуют свободные электроны, которые могли бы переносить электрический ток, и они являются изоляторами. Они, как правило, прозрачны в видимом диапазоне спектра, так как их запрещенные зоны лежат в ультрафиолетовой области (см. рис. 5.14). Твердые тела с **ковалентной связью**, как и молекулы с ковалентной связью, состоят из атомов, связанных объединенными валентными электронами. Они часто являются изоляторами и могут быть прозрачны (как алмаз) или непрозрачны (как графит) в видимом диапазоне. Твердые тела с ковалентной связью могут быть полупроводниками (как GaAs), которые непрозрачны в видимом и прозрачны в инфракрасном диапазоне (см. рис. 5.14). **Металлические твердые тела** имеют валентные электроны, принадлежащие всем положительным ионам решетки и движущиеся в их комбинированном потенциале. Способность электронов свободно перемещаться по решетке отвечает за высокую электропроводность металлических тел. Металлы сильно отражают свет и непрозрачны в видимом диапазоне.

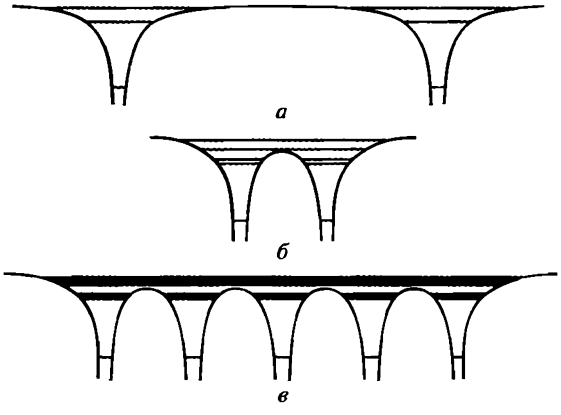


Рис. 13.6. Схема энергетических уровней для:

а — двух изолированных атомов; *б* — тех же атомов, образующих двухатомную молекулу; *в* — пяти одинаковых атомов, образующих рудиментарный одномерный кристалл

Полезно исследовать, как меняются энергетические уровни изолированного атома при сближении с соседними атомами в процессе образования кристаллической решетки. Изолированные атомы и молекулы (например, в газе) имеют дискретные энергетические уровни (см. рис. 13.1—13.5). Каждый отдельный атом в собрании таких идентичных изолированных атомов имеет идентичный набор дискретных энергетических уровней. При сближении атомов в процессе формирования твердого тела возрастает роль обменного взаимодействия (проистекающего из квантово-механического требования неразличимости тождественных частиц), равно как и различных полей, действующих со стороны соседних атомов. Первоначально резкие энергетические уровни энергии валентных электронов изолированного атома постепенно расширяются и пре-

вращаются в наборы из большого числа тесно расположенных энергетических уровней, образующих энергетические зоны. Этот процесс иллюстрируется рис. 13.6, где схематически показаны электронные уровни для двух изолированных атомов (рис. 13.6, *а*), для молекулы, содержащей два таких атома (рис. 13.6, *б*), и для рудиментарной одномерной решетки, состоящей из пяти таких атомов (рис. 13.6, *в*). Самые низкие уровни энергии остаются резкими, поскольку электроны внутренних подоболочек экранируются от воздействия соседних атомов, в то время как резкие энергетические уровни внешних электронов атомов превращаются в зоны по мере того, как атомы приходят в состояние непосредственной близости друг к другу.

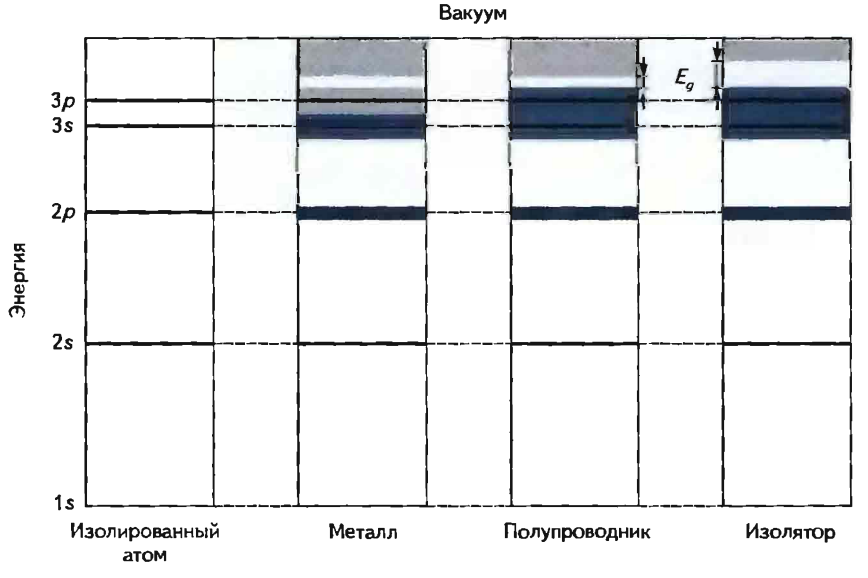


Рис. 13.7. Расщепление дискретных уровней изолированного атома с образованием энергетических зон при сближении атомов и формировании твердого тела. Полностью занятые зоны изображены синим цветом, светло-серым показаны незаполненные зоны или части зон. Запрещенные зоны — белого цвета. Типичные значения проводимости для металлов, полупроводников и изоляторов при комнатной температуре составляют 10^8 $(Ом \cdot м)^{-1}$, 10^{-4}—10^5 $(Ом \cdot м)^{-1}$ и 10^{-10} $(Ом \cdot м)^{-1}$ соответственно

Эта картина более детально уточняется на рис. 13.7, где схематически сравниваются энергетические уровни изолированного атома и трех различных типов твердых тел с решеткой из таких атомов: металла, полупроводника и диэлектрика. Низшие уровни энергии электронов с электронными конфигурациями 1*s*, 2*s* и 2*p* напоминают уровни изолированного атома, поскольку внутренние электроны экранируются от действия межатомных сил. Напротив, более высокие уровни энергии валентных электронов атома с конфигурациями 3*s* и 3*p* расщепляются с образованием энергетических зон. Наиболее низкая из неза-

нятых или частично занятых зон называется **зоной проводимости**, а наиболее высокая из полностью занятых зон — **валентной зоной**. Между ними лежит **запрещенная зона** с шириной E_g. Как и в изолированном атоме, электроны в первую очередь заселяют состояния с наименьшей возможной энергией, подчиняясь при этом принципу Паули.

Металлы имеют частично заполненные зоны проводимости при любых температурах (см. рис. 13.7). Существование множества незаполненных состояний в этой зоне отвечает за их высокую электропроводность (см. подразд. 5.5.4). Металлы составляют большинство элементов периодической таблицы (см. рис. 13.3). Полуметаллы характеризуются перекрыванием валентной зоны и зоны проводимости.

Собственные полупроводники имеют заполненную валентную зону и свободную зону проводимости при $T = 0$ °K. Ввиду отсутствия доступных свободных состояний в валентной зоне и отсутствия электронов в зоне проводимости, проводимость идеального собственного полупроводника при $T = 0$ °K равна нулю. По мере увеличения температуры все большее число электронов из валентной зоны приобретает достаточную тепловую энергию, чтобы оказаться в зоне проводимости, и дает вклад в электропроводность материала.

Изоляторы также имеют полностью занятую валентную зону и незаполненную зону проводимости. Они отличаются от полупроводников большей шириной запрещенной зоны (в типичном случае $E_g > 3$ эВ). Например, ширина запрещенной зоны у кремния (полупроводник) $E_g \approx 1{,}1$ эВ, а у алмаза (изолятор) $E_g \approx 5{,}5$ эВ. Поэтому у изоляторов значительно меньшее число электронов имеет тепловую энергию, достаточную для попадания в зону проводимости и участия в электропроводности. Нужно, однако, отметить, что такие обстоятельства, как степень перекрывания зон, также играют роль в определении принадлежности материала к металлам, полупроводникам или диэлектрикам.

Твердотельные материалы играют целый ряд важных ролей в фотонике. Металлы с их высокой отражательной способностью в видимом и инфракрасном диапазонах часто используются для изготовления зеркал. Многие типы лазерных усилителей и генераторов основаны на легированных диэлектрических материалах. Легированные полупроводники в виде p—n-переходов, гетероструктур и квантовых ям также широко используются как активные среды лазеров и приемники излучения.

Перейдем к изучению энергетических уровней некоторых легированных диэлектрических сред и полупроводниковых материалов, используемых в качестве активных сред лазеров.

Легированные диэлектрические материалы

Ионные или ковалентные твердые тела, являющиеся изоляторами и прозрачные в определенной области спектра, называются прозрачными диэлектрическими средами. При наличии подходящих оптических, тепловых и механических свойств такие материалы часто служат основой для внедрения лазерно-активных ионов. Оптические свойства этих основных материалов рас-

сматривались в подразд. 5.5.3 в связи с осцилляторной моделью Лоренца. Хотя эта модель классическая, она адекватно описывает основной материал, поскольку он прозрачен в представляющем интерес диапазоне длин волн (его резонансы лежат вне диапазона). Напротив, резонансы легирующих ионов необходимо определять с помощью квантовой механики или, что на практике бывает чаще, эмпирически. Наиболее распространенными добавками являются ионы переходных металлов и лантанидов.

Степень воздействия среды-хозяина на энергетические уровни лазерно-активных ионов в принципе определяется тем, насколько хорошо их оптически активные электроны экранированы от воздействия соседних атомов решетки. Как будет видно ниже, энергетические уровни ионов переходных металлов существенно модифицируются за счет эффектов кристаллического поля, а уровни энергии редкоземельных ионов — нет. В качестве примеров мы рассмотрим энергетические уровни четырех распространенных лазерных веществ: $Cr^{3+}:Al_2O_3$ (рубин), $Cr^{3+}:BeAl_2O_4$ (александрит), $Nd^{3+}:Y_3Al_5O_{12}$ ($Nd^{3+}:YAG$) и Nd^{3+}: стекло.

Ионы переходных металлов

Наиболее часто в качестве легирующих добавок в лазерах используются ионы Cr^{3+} и Ti^{3+}. Электронные конфигурации и символы термов этих элементов и их трехвалентных ионов приведены в табл. 13.1. Ni^{2+} и Co^{2+} также часто используются в качестве легирующих добавок.

Таблица 13.1. **Переходные металлы и лантаниды (редкоземельные элементы), имеющие важное значение в качестве легирующих добавок в твердотельных лазерах: электронные конфигурации* и символы термов**

Атомный номер Z	Атом			Ион		
	Элемент	Конфигурация	Терм	Ион	Конфигурация	Терм
Переходные металлы						
22	Ti	$3d^24s^2$	$^3F_2$	Ti^{3+}	$3d^1$	$^2D_{3/2}$
24	Cr	$3d^54s^1$	$^7S_3$	Cr^{3+}	$3d^3$	$^4F_{3/2}$
Лантаниды						
60	Nd	$4f^46s^2$	$^5I_4$	Nd^{3+}	$4f^3$	$^4I_{9/2}$
68	Er	$4f^{12}6s^2$	$^3H_6$	Er^{3+}	$4f^{11}$	$^4I_{15/2}$
70	Yb	$4f^{14}6s^2$	$^1S_0$	Yb^{3+}	$4f^{13}$	$^2F_{7/2}$

* По соглашению при записи конфигурации опущены заполненные подоболочки, включая $5s^25p^6$ в оболочке с $n = 5$ у лантанидов.

Мы рассмотрим энергетические уровни двух диэлектрических сред, легированных ионами Cr^{3+}, а именно, рубин и александрит (рис. 13.8). Рубин знаменит тем, что на нем был создан первый лазер, а александрит привлек к себе внимание возможностью перестройки длины волны лазерного излучения в широком диапазоне. Энергетические уровни титан-сапфира, важного лазерного материала на переходных ионах, будут рассмотрены в подразд. 15.3.1.

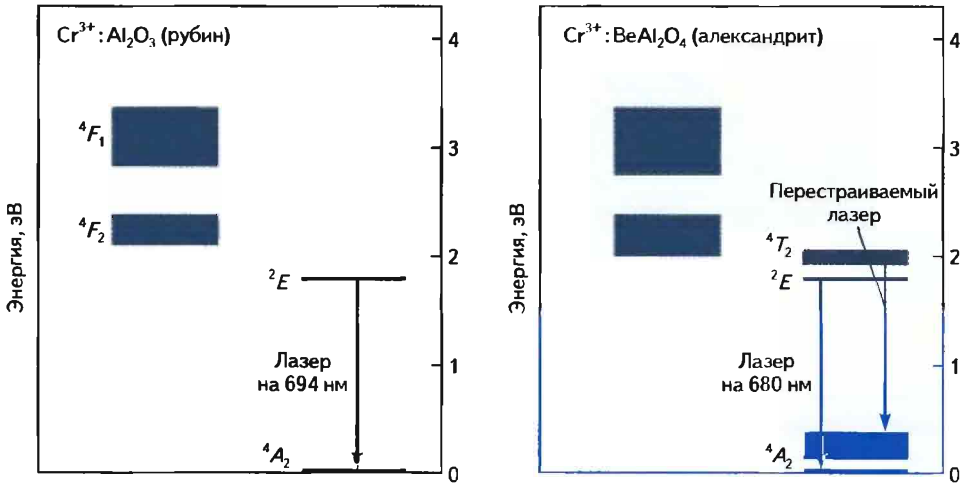


Рис. 13.8. Некоторые уровни энергии и энергетические зоны для $Cr^{3+}:Al_2O_3$ (рубин) и Cr^{3+} $BeAl_2O_4$ (александрит). Красными стрелками показаны лазерные переходы. Обычно лазер излучает на фиксированной длине волны. Однако александрит обеспечивает лазерную генерацию в довольно широком диапазоне длин волн. Переход от темной к светлой окраске в нижней зоне лазерного перехода в александрите отражает убывание относительной заселенности уровней

Рубин ($Cr^{3+}:Al_2O_3$) — диэлектрическая среда с показателем преломления $n \approx 1{,}76$, состоящая, в основном, из сапфира (Al_2O_3, известного также как окись алюминия, глинозем или корунд), в котором небольшое число ионов Al^{3+} (~0,05 %) замещено ионами Cr^{3+} **Александрит** ($Cr^{3+}:BeAl_2O_4$) получается добавлением небольшого количества окиси хрома (~0,1 %) в основной материал — хризоберилл ($BeAl_2O_4$). Этот материал имеет показатель преломления, близкий к показателю преломления рубина ($n \approx 1{,}74$). Хризоберилл — двуосный кристалл, а сапфир — одноосный.

Поскольку $3d$-электроны в ионах Cr^{3+} в обоих материалах подвержены действию соседних ионов, их энергетические уровни в большой степени определяются окружающими кристаллическими полями и поэтому очень чувствительны к основному материалу. В частности, каждый ион хрома окружен атомами кислорода, которые создают значительный пространственно-неоднородный потенциал. Этот потенциал, наилучшим образом описываемый **теорией кристаллического поля** (или **теорией поля лигандов**), вместе с потенциалом ядра Cr^{3+} опре-

деляет энергетические уровни рубина и александрита через уравнение Шредингера. Вследствие этого энергетические состояния ионов переходных металлов в диэлектрическом материале, как правило, обозначаются символами теории групп (неприводимых представлений группы симметрии основного кристалла — *примеч. пер.*), а не символами термов.

В результате спектр энергии представляет собой смесь дискретных уровней и энергетических зон, некоторые из них показаны на рис. 13.8 вместе с их теоретико-групповыми символами. Энергетические уровни этих материалов совершенно различны, хотя легирующая добавка одна и та же. В частности, энергетическая зона $^4A_2$ у александрита представляет собой набор **вибронных состояний**, которые возникают в результате взаимодействия электронных состояний иона с колебаниями решетки кристалла. Эти состояния похожи на состояния молекул органических красителей (см. рис. 13.5). Вследствие этого александрит лазерно-активен в довольно широкой полосе длин волн, чего нет у рубина (см. подразд. 14.3.5 и 15.3.4). Тем не менее у обоих материалов есть и дискретные лазерные переходы, причем с довольно близкими длинами волн (694 и 680 нм для рубина и александрита соответственно).

Лантаниды в качестве легирующих добавок

Лантаниды, включающие элементы от ^{58}Ce до ^{71}Lu, расположены в шестой строке периодической таблицы (см. рис. 13.3). Эти элементы называют также **редкоземельными**, поскольку их долгое время считали редкими, что впоследствии оказалось не так. Последовательности лантанидов соответствует постепенное заполнение подоболочки $4f$, которая лежит между заполненными подоболочками $5s^25p^6$ и $6s^2$. Лантаниды обычно существуют в виде трехвалентных катионов, конфигурация их валентных электронов принимает вид $4f^u$, где u принимает значения от 1 (Ce^{3+}) до 14 (Lu^{3+}).

Nd^{3+}, Er^{3+} и Yb^{3+} — особенно важные легирующие добавки для лазерных усилителей и генераторов. Nd^{3+} в стекле и Er^{3+} в кварцевом волокне широко используются в лазерных усилителях, как подчеркивается далее в подразд. 14.3.2 и 14.3.3. Волокна Nd^{3+} : ИАГ, Nd^{3+} : YVO_4 и Nd^{3+} : кварц часто служат лазерными генераторами, как обсуждается в подразд. 15.3.1. Среди других лантанидов широкое применение в качестве лазерно-активных ионов имеют Tm^{3+} и Ho^{3+}.

Поведение трехвалентных ионов лантанидов в основном диэлектрике и в изолированном состоянии довольно похожее. Это происходит потому, что электроны $4f$ достаточно хорошо экранируются от внешнего воздействия решетки заполненными подоболочками $5s$ и $5p$ (см. табл. 13.1). Это резко контрастирует с поведением ионов переходных металлов. Таким образом, в отличие от случаев рубина и александрита, энергетические уровни редкоземельных ионов в достаточной степени независимы от окружающего материала. Этот факт иллюстрируется рис. 13.9 для ионов Nd^{3+} в двух существенно различных основных веществах: иттрий-алюминиевом гранате (ИАГ) и стекле. Основные лазерные переходы в ближнем ИК-диапазоне в обоих материалах, соответствующие разности энергий между уровнями $^4F_{3/2}$ и $^4I_{11/2}$, имеют замечательно близкие

длины волн: 1,064 мкм для Nd^{3+} в ИАГ и 1,053 мкм для Nd^{3+} в стекле. Мультиплеты тонкой структуры, возникающие из-за расщепления уровней кристаллическим полем, различны в обоих материалах, однако это обстоятельство не играет важной роли, а расщепления слишком малы, чтобы можно было показать их на рисунке.

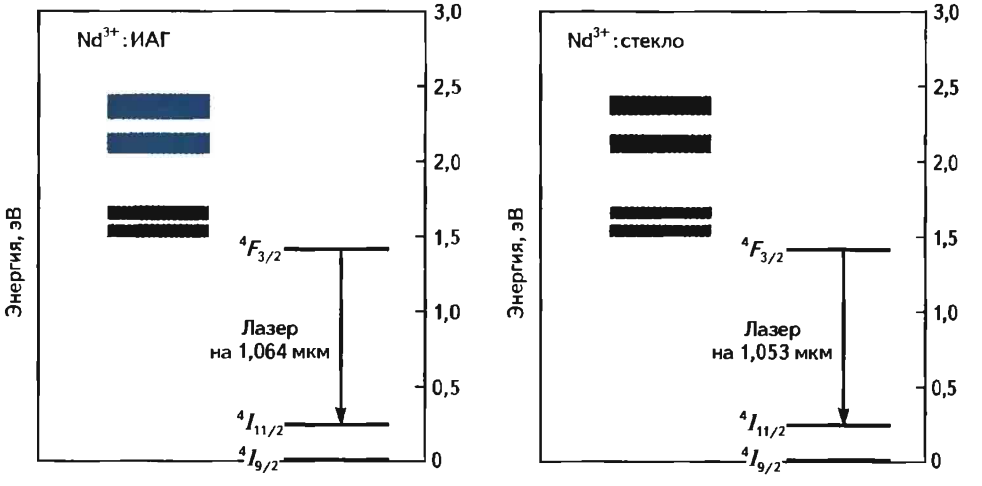


Рис. 13.9. Некоторые энергетические уровни Nd^{3+} в ИАГ и стекле. Стрелками показан главный лазерный переход в ближнем ИК-диапазоне на длине волны 1,064 мкм в ИАГ и на 1,053 мкм в фосфатном стекле. Тонкая структура энергетических уровней двух материалов различна (см., например, рис. 15.24), но на данном рисунке она не видна

Актиниды в качестве легирующих добавок

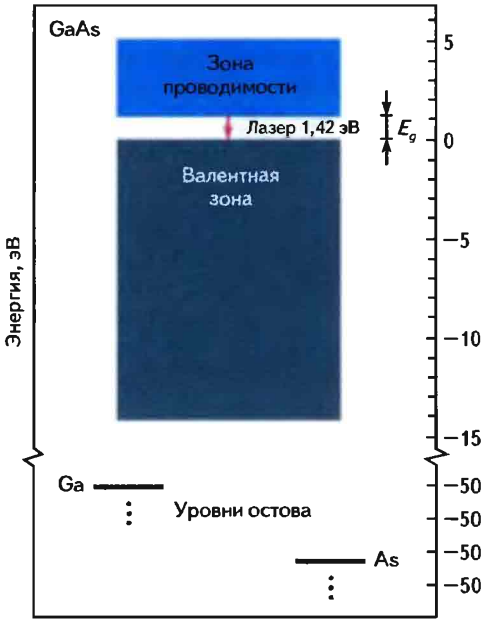
Актиниды, которые иногда объединяют с редкоземельными элементами, характеризуются заполнением подоболочки $5f$, которая лежит глубоко внутри заполненной подоболочки $7s^2$. Химические свойства актинидов аналогичны свойствам их редкоземельных гомологов (см. рис. 13.3). На основе ионов актинидов в диэлектрических средах (например, $U^{3+}:CaF_2$) были созданы лазеры, однако основным препятствием для развития в данном направлении является радиоактивность указанных ионов.

Полупроводники

Полупроводники широко распространены в фотонике. Они используются в источниках, таких как светодиоды и диодные лазеры, в качестве приемников, а также играют много других важных ролей. Мы дадим краткое введение в строение энергетических уровней объемных полупроводников, квантовых ям, квантовых проволочек и квантовых точек. Более подробная информация, касающаяся энергетических уровней полупроводников, приведена в гл. 16.

Объемные полупроводники

Бинарный полупроводник GaAs раньше всего нашел применение в фотонике. Этот материал имеет кристаллическую структуру типа цинковой обманки, состоящую из двух гранецентрированных кубических решеток, одна из атомов Ga, другая из атомов As, смещенных друг относительно друга на 1/4 длины диагонали (рис. 13.10). В элементарную ячейку кристалла, имеющую форму куба, входят четыре молекулы GaAs. Каждый атом окружен четырьмя атомами другого типа, расположенными на равных расстояниях в вершинах правильного тетраэдра.



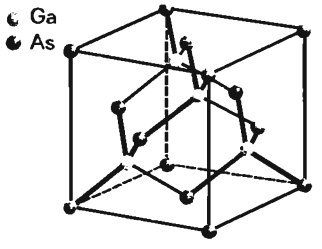


Рис. 13.10. Полупроводник GaAs имеет кристаллическую структуру цинковой обманки, включающую две гранецентрированные кубические решетки, одну для Ga, другую для As. Более высокие уровни энергии расположены близко друг от друга и образуют зоны. Нуль энергии выбран (произвольно) на верхнем крае валентной зоны. Лазерный диод на GaAs работает на электронном переходе между зоной проводимости и валентной зоной в ближнем инфракрасном диапазоне спектра (см. гл. 17)

Полупроводники имеют множество близко расположенных электронных уровней энергии, образующих зоны, как показано на рис. 13.10 для GaAs. Ширина запрещенной зоны E_g между валентной зоной и зоной проводимости GaAs составляет 1,42 эВ при комнатной температуре. Уровни остова Ga и As ($3d$) являются дискретными, как видно на рис. 13.10. Валентная зона GaAs формируется из уровней $4s$ и $4p$, как схематически показано на рис. 13.7. Свойства полупроводников более детально рассмотрены в гл. 16.

Квантовые ямы

Техника выращивания кристаллов, например молекулярно-пучковая или газофазная эпитаксия, позволяет выращивать материалы с заданной структурой энергетических зон. В полупроводниковых структурах с квантовыми ямами ширина запрещенной зоны создается зависящей от координат особым об-

разом, что придает материалу уникальные электронные и оптические свойства. Примером является **квантово-размерная решетка**, показанная на рис. 13.11. Она состоит из сверхтонких (1—15 нм) слоев GaAs, чередующихся с тонкими (20 нм) слоями AlGaAs. Ширина запрещенной зоны GaAs меньше, чем у AlGaAs. Для движения перпендикулярно слоям возможные уровни энергии электронов в зоне проводимости и дырок в валентной зоне дискретны и достаточно удалены друг от друга, как в квантово-механической задаче о частице в прямоугольной потенциальной яме (см. упражнение 16.5). Наиболее низкие из этих уровней схематически показаны в каждой из квантовых ям. Барьеры из AlGaAs можно также сделать сверхтонкими (<1 нм). В этом случае электроны соседних ям легко взаимодействуют друг с другом посредством квантово-механического туннелирования, и их дискретные уровни расщепляются в миниатюрные энергетические зоны, называемые **мини-зонами**. В этом случае материал называется **сверхрешеткой**, поскольку мини-зоны создаются решеткой, постоянная которой превышает естественный период атомной решетки среды.

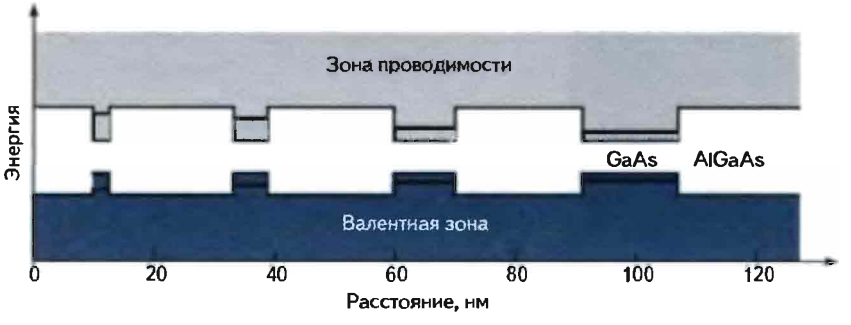


Рис. 13.11. Квантованные энергии в монокристаллической квантово-размерной решетке AlGaAs/GaAs. Ширина ям может быть периодической или произвольной

Квантовые проволоки

Полупроводниковый материал в форме тонкой проволоки, окруженной материалом с большей шириной запрещенной зоны, называется **квантовой проволокой**. Она ведет себя как потенциальная яма, сильно ограничивающая движение электронов и дырок в поперечных направлениях, но не вдоль оси проволочки. Квантовые проволоки легко изготовить из III—V и II—VI материалов, таких как InP и CdSe соответственно. Они могут иметь прямоугольное или круглое поперечное сечение. Нанотрубки и нанопроволочки, изготовленные из широкого набора материалов, могут служить в качестве квантовых проволок. В частности, углеродные нанотрубки — цилиндрические молекулы углерода диаметром один или несколько нанометров — демонстрируют замечательные свойства. Молекулы углерода организуются в тонкие полые трубочки за счет притяжения Ван-дер-Ваальса. Нанотрубки с одной или несколькими стенками обладают уникальными оптическими, механическими и электричес-

кими свойствами. Они могут вести себя как полупроводник или металл с высокой проводимостью, в зависимости от их конкретной структуры. Существуют многообразные применения нанотрубок в фотонике, начиная с нитей накаливания в источниках света и кончая фотоэлектрическими приемниками.

Квантовые точки

Известные также как **нанокристаллы** и **квантовые ящики**, **квантовые точки** представляют собой полупроводниковые частицы, размеры которых обычно составляют от 1 до 10 нм, но могут достигать и нескольких мкм. Они могут быть изготовлены из множества различных полупроводников и иметь разнообразную геометрическую форму (например, куба, сферы или пирамиды). Такие частицы обычно внедряют в полупроводниковый материал с большей шириной запрещенной зоны, стекло или полимер. Иногда их изготавливают в виде структур в форме диска с использованием молекулярно-пучковой эпитаксии или осаждения из газовой фазы. В этом случае движение электронов ограничено плоскостью и обнаруживает свойства электронных оболочек двумерного атома, напоминающих структуру поля в тороидальных резонаторах (см. подразд. 10.4.2). Квантовые точки можно создать также с помощью электронно-лучевой литографии, где заданная структура протравливается на полупроводниковой пластине, а затем на нее наносится слой металла с высокой электропроводностью. Квантовые точки легко вырастить в жидкой фазе с помощью методов коллоидной химии; коллоидные нанокристаллы затем располагаются в жидкой суспензии или в пластиковом композите.

Размеры квантовых точек и, следовательно, число атомов, содержащихся в них, может меняться в широких пределах. Число электронов может быть совсем мало (несколько штук), а может достигать миллионов. Кубик размером 10 нм из GaAs содержит более 40 000 атомов. Все электроны принадлежат квантовой точке в целом; энергетические уровни представляют собой уровни ее экситонов, т. е. электронно-дырочных пар, созданных внутри нее и ограниченных ее объемом. Вследствие ограничения движения электрона малым объемом пространства, как и в атоме, возникает набор дискретных уровней энергии. Поэтому квантовые точки иногда называют **искусственными атомами**.

Однако, в отличие от атома, квантовая точка, изготовленная из определенного материала, имеет необычное свойство: ее энергетические уровни сильно зависят от ее размера. Например, цвет свечения квантовых точек CdSe при возбуждении светом можно плавно перестраивать от красного при диаметре точек 5 нм до фиолетового при 1,5 нм (рис. 13.12, см. цв. вклейку). Энергия испускаемого фотона растет по мере уменьшения размера точек, поскольку для «запирания» возбуждений полупроводника в меньшем объеме требуется большая энергия. Длина волны возбуждающего света произвольна, но должна быть сдвинута в синюю сторону по отношению к длине волны испускания. Квантовые точки из InP люминесцируют в ближней ИК-области, а из InAs — в телекоммуникационном окне 1300—1600 нм сетей на основе кварцевого оптоволокна. Возбуждаемые светом квантовые точки из Si также излучают в

широком спектральном диапазоне, простирающемся от ИК до видимого (см. подразд. 17.4.1). Квантовые точки можно изготовить также из органических соединений.

Квантовые точки, покрытые полупроводником с большей шириной запрещенной зоны, известны как квантовые точки с архитектурой «ядро—оболочка»; покрытие может осуществляться несколькими слоями полупроводника с чередующимися большими и меньшими ширинами запрещенной зоны и, следовательно, обладать свойствами квантовых ям. Покрытия существенно улучшают возможности перестройки и эффективность фотолюминесценции наноструктуры. Упорядоченные структуры — **кристаллы из квантовых точек** — можно вырастить разными способами, включая самосборку сферических нанокристаллов в плотно упакованную конфигурацию. Так же как в сверхрешетках, состоящих из множества квантовых ям, туннелирование происходит и в кристаллах из квантовых точек, которые называются **нанокристаллическими сверхрешетками**.

Квантово-точечные структуры иногда бывают достаточно большими, чтобы к ним можно было присоединить электроды, тогда они могут служить миниатюрными фотонными устройствами. Специально созданные наборы квантовых точек с заданным расположением могут поддерживать токи и работать в широком или специально выбранном диапазоне длин волн. Квантовые точки полезны в качестве спектральных меток в биологических, промышленных и военных приложениях, а также при распознавании подделок. Они используются в широком круге приложений, таких как высокоэффективные лазеры, широкополосные светоизлучающие диоды, источники единичных фотонов, элементы памяти, фотоприемники, солнечные батареи, плоские дисплеи и поглотители в материалах, где путем фильтрации желательно убрать ультрафиолетовый свет.

13.2. ЗАСЕЛЕННОСТЬ УРОВНЕЙ ЭНЕРГИИ

Как отмечалось выше, каждый атом или молекула ансамбля совершает случайные переходы между своими энергетическими уровнями. Эти переходы подчиняются законам статистической физики. Определяющим фактором как для поведения заселенностей уровней в среднем, так и для их флуктуаций является температура.

13.2.1. Распределение Больцмана

Рассмотрим ансамбль различимых объектов, таких как атомы или молекулы, образующих разреженный газ. Каждый атом находится на одном из своих энергетических уровней E_1, E_2, Если система находится в тепловом равновесии при температуре T (т. е. атомы находятся в контакте с тепловым резервуаром, поддерживаемым при температуре T, а их движение достигло ус-

тановившегося состояния, в котором флуктуации в среднем не зависят от времени), то вероятность $P(E_m)$ того, что произвольный атом находится на уровне энергии E_m, дается распределением Больцмана

$$P(E_m) \propto \exp\left(-\frac{E_m}{k_B T}\right), \quad m = 1,\ 2,\ 3, \tag{13.7}$$

где k_B — постоянная Больцмана. Коэффициент пропорциональности выбирается так, чтобы

$$\sum_m P(E_m) = 1.$$

Вероятность заселения $P(E_m)$ — экспоненциально убывающая функция энергии E_m, как показано на рис. 13.13.

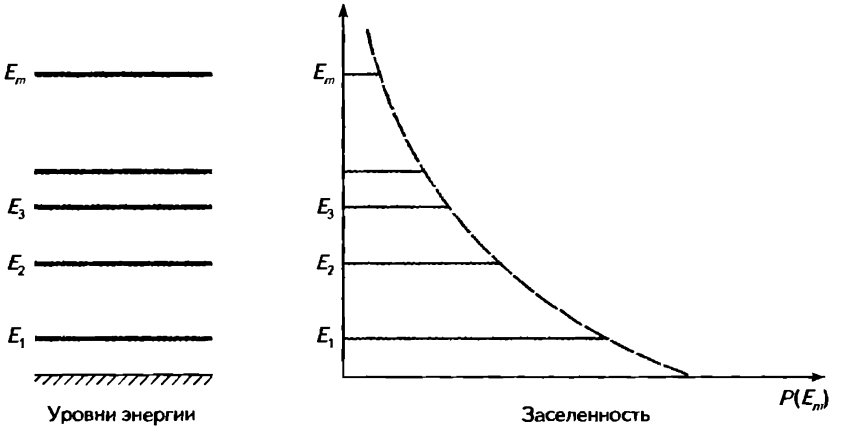


Рис. 13.13. Распределение Больцмана утверждает, что вероятность заселения энергетического уровня E_m произвольного атома — экспоненциально убывающая функция E_m

Происхождение распределения Больцмана можно понять, рассматривая систему большого числа одинаковых элементов, полная энергия которой фиксирована и равна E. Элементы изолированы от своего окружения, но находятся в тепловом равновесии, обмениваясь между собой энергией через тепловой резервуар при температуре T. Способы деления энергии считаются различимыми, если они включают различные энергетические состояния, и все возможные способы деления полной энергии предполагаются равновероятными. Если один из элементов ансамбля заберет больше энергии, то другим достанется меньше энергии и, следовательно, меньше способов ее деления. Значит, большая энергия у одного члена ансамбля — менее вероятное событие, чем малая. Для количественного описания рассмотрим два элемента ансамбля. Поскольку они независимы, вероятность того, что энергия первого из них равна E_1, а второго —

E_2, равна произведению вероятностей $P(E_1)P(E_2)$. Если сумма энергии фиксирована на уровне $E_1 + E_2$, то $P(E_1)P(E_2)$ должно быть функцией $E_1 + E_2$, что однозначным образом характеризует экспоненциальную функцию. Энергия $k_B T$ при равномерном распределении для двух степеней свободы, связанных с гармонической модой, непосредственно приводит в распределению Больцмана.

Рассмотрим распределение Больцмана для большого числа атомов N. Если N_m — число атомов, занимающих энергетический уровень E_m, то отношение $N_m/N \approx P(E_m)$. Если N_1 атомов занимают энергетический уровень E_1, а N_2 атомов — лежащий выше уровень E_2, то отношение заселенностей в среднем равно

$$\frac{N_2}{N_1} = \exp\left(-\frac{E_2 - E_1}{k_B T}\right). \tag{13.8}$$

Эта величина зависит от температуры T. При $T = 0\,^\circ\mathrm{K}$ все атомы находятся в состоянии с наименьшей энергией (основном состоянии). При повышении температуры вероятность заселения верхних уровней увеличивается. В состоянии теплового равновесия заселенность нижнего уровня всегда больше, чем верхнего. Это условие не обязано выполняться в неравновесных условиях, при которых заселенность верхнего уровня может быть больше, чем у нижнего. Такая ситуация, называемая **инверсией заселенностей**, составляет основу действия лазеров (см. гл. 14 и 15).

Выше предполагалось, что атом единственным образом заселяет данный уровень энергии. Однако часто бывает, что один уровень энергии может заселяться несколькими способами. Это бывает в тех случаях, когда несколько разных состояний (например, состояния с различными значениями углового момента) имеют одно и то же значение энергии. С учетом такого вырождения (13.8) записывается в более общем виде

$$\frac{N_2}{N_1} = \frac{g_2}{g_1}\exp\left(-\frac{E_2 - E_1}{k_B T}\right). \tag{13.9}$$

Кратности вырождения g_2 и g_1 представляют собой число состояний, соответствующих энергиям E_2 и E_1.

13.2.2. Распределение Ферми—Дирака

Квантовые системы с перекрывающимися волновыми функциями, такие как многоэлектронные атомы и полупроводники, подчиняются принципу запрета Паули. В этом случае в каждом состоянии может находиться самое большее один электрон, т. е. число электронов N_m в состоянии m может быть или 0, или 1. Вероятность заселения состояния с энергией E тогда описывается **распределением Ферми—Дирака** (или **функцией Ферми**)

$$f(E) = \frac{1}{\exp\left[(E - E_F)/k_B T\right] + 1}, \tag{13.10}$$

где E_F — энергия Ферми. Функция Ферми имеет максимальное значение, равное 1, отвечающее тому, что состояние с энергией E с достоверностью занято. С ростом E она монотонно убывает, принимая значение 1/2 при энергии Ферми $E = E_F$.

Важно заметить, что $f(E)$ не является ни функцией плотности вероятности, ни функцией распределения вероятностей, а представляет собой распределение (последовательность) вероятностей для различных значений E, каждая из которых принимает значения в интервале от 0 до 1. Тем не менее, когда $E \gg E_F$ (и $E \gg k_B T$), функция Ферми ведет себя как функция распределения Больцмана

$$P(E) \propto \exp\left(-\frac{E}{k_B T}\right), \tag{13.11}$$

как видно из (13.10). Распределения Ферми—Дирака и Больцмана сравниваются на рис. 13.14. Поскольку в типичном случае для атомных электронов на внешних подоболочках $E \gg E_F$, уровни энергии, соответствующие оптическому переходу, часто характеризуются распределением Больцмана. Функция Ферми более детально рассматривается в гл. 16.

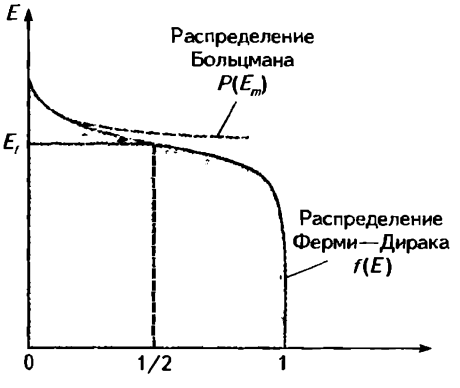


Рис. 13.14. Распределение Ферми—Дирака $f(E)$ хорошо аппроксимируется распределением Больцмана $P(E_m)$ при $E \gg E_f$

13.3. ВЗАИМОДЕЙСТВИЕ ФОТОНОВ С АТОМАМИ

13.3.1. Взаимодействие одномодового света с атомом

Как известно из теории атома, он может испускать (рождать) или поглощать (уничтожать) фотон, совершая переход вниз или вверх между энергетическими уровнями с сохранением полной энергии системы атом + поле. Законы, которым подчиняются эти процессы, описаны в настоящем разделе. Взаимодействие фотонов с электронами и дырками в полупроводниках рассмотрены в разд. 16.2.

Взаимодействие между атомом и электромагнитной модой

Рассмотрим энергетические уровни E_1 и E_2 атома, помещенного в оптический резонатор объема V, в котором может существовать много электромагнитных мод. Нас особенно интересует взаимодействие между атомом и фо-

тонами *заданной* моды излучения с частотой $\nu \approx \nu_0$, где $h\nu_0 = E_2 - E_1$, поскольку такая энергия фотона совпадает с разностью энергий атомных уровней. Такие взаимодействия формально исследуются с использованием квантовой электродинамики. Ключевые результаты приведены ниже без доказательства. Возможны три вида взаимодействия: спонтанное излучение, поглощение и вынужденное излучение.

Спонтанное излучение

Если атом первоначально находился на более высоком уровне энергии, он может спонтанно перейти на более низкий уровень и испустить энергию в виде фотона (рис. 13.15). Энергия фотона $h\nu$ добавляется к энергии электромагнитной моды. Процесс называется спонтанным излучением, поскольку переход происходит независимо от числа фотонов, уже имеющихся в моде.

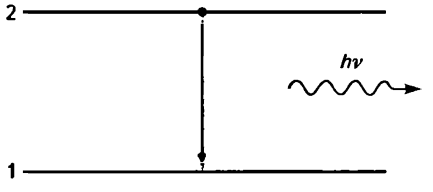


Рис. 13.15. Спонтанное излучение фотона в моду частоты ν при переходе атома с энергетического уровня 2 на уровень 1. Энергия фотона равна $h\nu \approx E_2 - E_1$

В резонаторе объемом V плотность вероятности (в секунду) или скорость этого спонтанного испускания зависит от ν способом, характеризующим данный атомный переход:

$$p_{\text{сп}} = \frac{c}{V}\sigma(\nu). \qquad (13.12)$$

Спонтанное излучение в предписанную моду

Функция $\sigma(\nu)$, зависящая от ν, отлична от нуля вблизи частоты атомного резонанса ν_0 и называется **сечением перехода**. Важность этого определения станет ясна впоследствии, однако уже сейчас видно, что размерность сечения — см^2, поскольку размерность $p_{\text{сп}}$ — с$^{-1}$. В принципе $\sigma(\nu)$ можно найти из уравнения Шредингера, однако вычисления, как правило, достаточно сложны, и обычно $\sigma(\nu)$ определяют экспериментально. Формула (13.12) отдельно применяется к каждой моде, при этом сечение задается формулой

$$\sigma = \sigma_{\max} \cos^2 \theta, \qquad (13.13)$$

где θ — угол между дипольным моментом атома и направлением поля моды; максимальное сечение $\sigma_{\max}$ достигается, когда дипольный момент и поле направлены одинаково.

Термин «плотность вероятности» означает, что вероятность испускания в течение малого отрезка времени от t до $t + \Delta t$ равна $p_{\text{сп}}\Delta t$. Как плотность вероятности, величина $p_{\text{сп}}$ может принимать числовые значения больше 1 с$^{-1}$, при

этом, конечно, сама вероятность $p_{сп}\Delta t$ всегда меньше единицы. Таким образом, если имеется большое число N таких атомов, то примерно $\Delta N = (p_{сп}\Delta t)N$ атомов совершат переход за время Δt. Можно записать $dN/dt = -p_{сп}N$, так что число возбужденных атомов $N(t) = N(0)\exp(-p_{сп}t)$ экспоненциально спадает с постоянной времени $1/p_{сп}$, как показано на рис. 13.16.

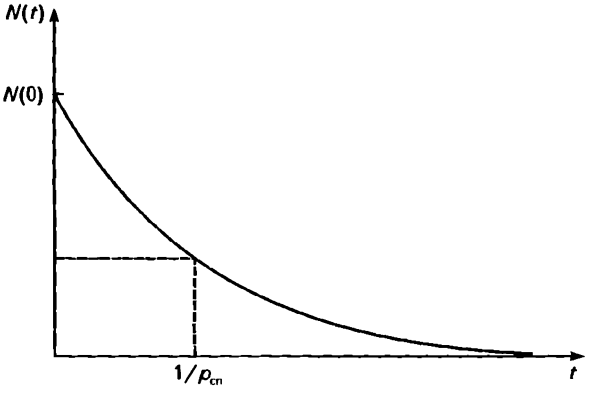


Рис. 13.16. Спонтанное излучение в одну моду приводит к экспоненциальному спаду числа возбужденных атомов с постоянной времени $1/p_{сп}$

Поглощение

Если атом первоначально находился на нижнем уровне энергии, а мода излучения содержит фотон, то фотон может исчезнуть, а атом перейти на более высокий уровень энергии (рис. 13.17). Этот процесс **индуцирован** фотоном и называется **поглощением**. Он может происходить лишь при условии, что в моде имеется фотон.

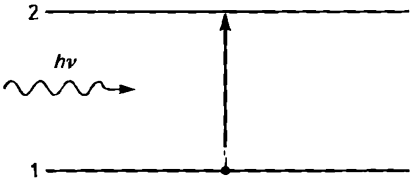


Рис. 13.17. Поглощение фотона с энергией $h\nu$ вызывает переход атома с нижнего энергетического уровня 1 на верхний уровень 2

Плотность вероятности поглощения фотона из заданной моды с частотой ν в резонаторе объемом V подчиняется *тому же закону*, что и спонтанное излучение в эту моду:

$$p_{погл} = \frac{c}{V}\sigma(\nu). \tag{13.14}$$

Однако, если в моде n фотонов, то плотность вероятности того, что атом поглотит *один* фотон в n раз больше, поскольку события взаимно исключающие, т. е.

$$\boxed{P_{погл} = n\frac{c}{V}\sigma(\nu).} \tag{13.15}$$

Поглощение одного фотона из моды с n фотонами

Вынужденное излучение

Наконец, если атом находится на верхнем уровне энергии, а мода содержит фотон, то атом может быть вынужден испустить еще один фотон в ту же моду. Этот процесс называется **вынужденным излучением**. Он является обратным по отношению к поглощению. Присутствие фотона в моде с определенной частотой, направлением распространения и поляризацией стимулирует испускание точной копии («клонирование») фотона с теми же характеристиками (рис. 13.18). Этот процесс лежит в основе действия лазерных усилителей и генераторов, как будет выяснено в последующих главах.

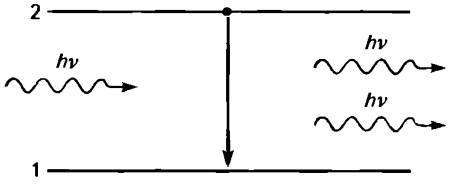


Рис. 13.18. Вынужденное излучение — процесс, при котором фотон с энергией $h\nu$ вынуждает атом испустить точно такой же фотон с переходом на нижний уровень энергии

Плотность вероятности $p_{\text{в.исп}}$, с которой этот процесс происходит в резонаторе объемом V, подчиняется *тому же закону*, что и спонтанное излучение и поглощение:

$$p_{\text{в.исп}} = \frac{c}{V}\sigma(\nu). \tag{13.16}$$

Если мода исходно содержала n фотонов, то плотность вероятности вынужденного испускания еще одного фотона, как и в случае поглощения, равна

$$\boxed{P_{\text{в.исп}} = n\frac{c}{V}\sigma(\nu).} \tag{13.17}$$

Вынужденное испускание одного фотона в моду с n фотонами

Поскольку $P_{\text{в.исп}} = P_{\text{погл}}$, используем общее обозначение W_i для плотности вероятности вынужденного излучения и поглощения.

Поскольку в дополнение к вынужденному излучению присутствует спонтанное излучение, полная плотность вероятности испускания фотона атомом в моду дается выражением

$$P_{\text{в.исп}} + p_{\text{в.исп}} = (n+1)\frac{c}{V}\sigma(\nu).$$

С точки зрения квантовой электродинамики спонтанное излучение можно трактовать как испускание, индуцированное вакуумными флуктуациями, связанными с данной модой (см. подразд. 12.1.1). Поскольку энергия этих флуктуаций не годится для поглощения, $P_{\text{погл}}$ пропорционально n, а не $(n + 1)$.

Три возможных взаимодействия между атомом и модой резонатора (спонтанное излучение, поглощение и вынужденное излучение) подчиняются фундаментальным соотношениям, сформулированным выше. Их следует рассматривать как законы, определяющие взаимодействие фотонов с атомами и дополняющие правила оптики фотонов, приведенные в гл. 12.

Перейдем к более детальному обсуждению характера и следствий этих простых соотношений.

Функция формы линии

Сечение перехода $\sigma(\nu)$ характеризует взаимодействие атома с излучением. Площадь под графиком этой функции

$$S = \int_0^\infty \sigma(\nu)\,\mathrm{d}\nu, \tag{13.18}$$

измеряется в [см^2 · Гц]. Она называется **силой перехода** или **силой осциллятора** и характеризует величину взаимодействия. Форма зависимости $\sigma(\nu)$ характеризует относительную величину взаимодействия с фотонами различных частот. Эту форму удобно описывать с помощью **функции формы линии** $g(\nu) = \sigma(\nu)/S$, измеряемой в Гц$^{-1}$ и имеющей единичную площадь

$$\int_0^\infty g(\nu)\,\mathrm{d}\nu = 1.$$

Сечение перехода выражается через функцию формы линии как

$$\sigma(\nu) = Sg(\nu). \tag{13.19}$$

Функция формы линии $g(\nu)$ центрирована на частоте резонанса ν_0, где $\sigma(\nu)$ имеет максимум, и быстро спадает к нулю при удалении ν от ν_0. Таким образом, переходы наиболее вероятны при условии, что $\nu \approx \nu_0$. Ширина функции $g(\nu)$ называется **шириной линии** перехода. Она обычно определяется как полная ширина $\Delta\nu$ пика функции $g(\nu)$ на уровне половины ее максимального значения (full-width at half-maximum, FWHM) (см. разд. А.2 приложения А). Поскольку площадь под графиком $g(\nu)$ равна единице, ширина линии обратно пропорциональна высоте максимума

$$\Delta\nu \propto \frac{1}{g(\nu_0)}. \tag{13.20}$$

Полезно также ввести пиковое значение сечения на резонансной частоте $\sigma_0 = \sigma(\nu_0)$. Тогда функция $\sigma(\nu)$ характеризуется высотой σ_0, шириной $\Delta\nu$, площадью S и профилем $g(\nu)$, как показано на рис. 13.19.

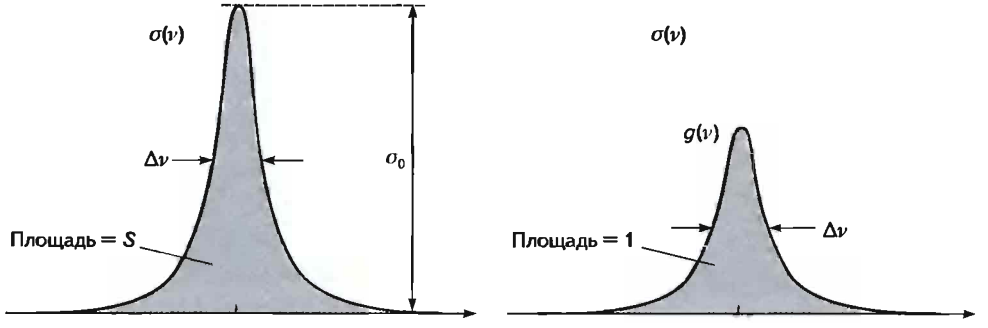


Рис. 13.19. Сечение перехода $\sigma(\nu)$ и функция формы линии $g(\nu)$

13.3.2. Спонтанное излучение

Полное спонтанное излучение во все моды

Выражение (13.12) дает плотность вероятности для спонтанного излучения в *заданную* моду частоты ν (независимо от того, есть ли в этой моде фотоны). Согласно (10.82), плотность мод в трехмерном резонаторе квадратично растет как

$$M(\nu) = \frac{8\pi\nu^2}{c^3}.$$

Эта величина аппроксимирует число мод с частотой ν на единицу объема резонатора и на единицу частоты при условии, что число мод достаточно велико, так что для их подсчета можно пользоваться непрерывным приближением. Атом может испустить *один* фотон частоты ν в *любую* из этих мод, как схематически показано на рис. 13.20.

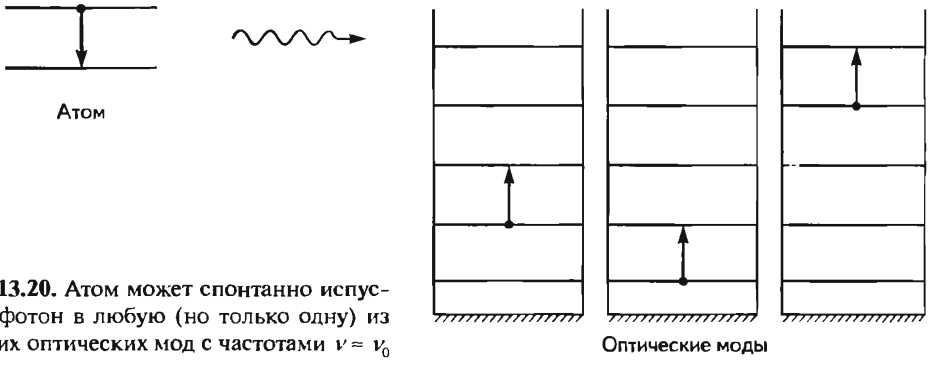


Рис. 13.20. Атом может спонтанно испустить фотон в любую (но только одну) из многих оптических мод с частотами $\nu \approx \nu_0$

Следовательно, плотность вероятности спонтанного излучения во все моды дается взвешенной суммой плотностей вероятности спонтанного излучения в

каждую из них с плотностью мод в качестве весовой функции. Поскольку моды на каждой частоте имеют изотропное распределение по направлениям и каждая при этом имеет две поляризации, определим среднее сечение перехода $\bar{\sigma}(\nu)$. Если θ — угол между дипольным моментом атома и направлением поля, то из (13.13) следует

$$\bar{\sigma}(\nu) = \frac{1}{3}\sigma_{\max}, \tag{13.21}$$

поскольку $\langle \cos^2\theta \rangle = 1/3$, где $\langle \cdot \rangle$ обозначает усреднение по ориентациям в трехмерном пространстве. Полная вероятность спонтанного излучения приобретает вид

$$P_{\text{ср}} = \int_0^\infty \left[\frac{c}{V}\bar{\sigma}(\nu)\right]\left[V M(\nu)\right] \mathrm{d}\nu = c\int_0^\infty \bar{\sigma}(\nu) M(\nu) \mathrm{d}\nu. \tag{13.22}$$

Функция $\bar{\sigma}(\nu)$ имеет форму острого пика при $\nu \approx \nu_0$, очень узкого по сравнению с плавно меняющейся квадратичной функцией $M(\nu)$. Поэтому под интегралом можно заменить $M(\nu)$ на $M(\nu_0)$ и вынести эту величину из-под знака интеграла. Тогда плотность вероятности спонтанного излучения одного фотона в любую из мод получается равной

$$P_{\text{сп}} = M(\nu_0) c\bar{S} = \frac{8\pi\bar{S}}{\lambda^2}, \tag{13.23}$$

где $\lambda = c/\nu_0$ — длина волны света в среде;

$$\bar{S} = \int_0^\infty \bar{\sigma}(\nu) \mathrm{d}\nu.$$

Определим временну́ю постоянную $t_{\text{сп}}$, называемую временем жизни спонтанного перехода $2 \to 1$, как

$$\frac{1}{t_{\text{сп}}} \equiv P_{\text{сп}} = M(\nu_0) c\bar{S}.$$

Тогда

$$\boxed{P_{\text{сп}} = \frac{1}{t_{\text{сп}}}} \tag{13.24}$$

Спонтанное излучение одного фотона в любую моду

независимо от объема резонатора V. Это позволяет выразить $\bar{S}$ как

$$\bar{S} = \frac{\lambda^2}{8\pi t_{\text{сп}}}, \tag{13.25}$$

что позволяет определить силу перехода из экспериментально измеренного спонтанного времени жизни $t_{\text{сп}}$.

Выражение (13.25) полезно, поскольку аналитическое вычисление $\bar{S}$ потребовало бы глубокого знания квантово-механического поведения системы, которое не всегда доступно. В типичном случае $t_{сп} \approx 10^{-8}$ с для атомных переходов, например, из первого возбужденного состояния водорода; однако для разных переходов $t_{сп}$ может меняться в очень широких пределах — от субпикосекунд до минут.

Упражнение 13.1

Частота спонтанно испущенных фотонов

Покажите, что плотность вероятности того, что возбужденный атом спонтанно испустит фотон с частотой между ν и $\nu + d\nu$, равна

$$P_{сп}(\nu)d\nu = \frac{1}{t_{сп}} g(\nu)d\nu.$$

Объясните, почему спектр спонтанного излучения атома после излучения большого числа фотонов пропорционален функции формы линии $g(\nu)$.

Соотношение между сечением перехода и спонтанным временем жизни

Использование (13.25) и формулы $\bar{\sigma}(\nu) = \bar{S}g(\nu)$ показывает, что среднее сечение перехода связано со спонтанным временем жизни и функцией формы линии как

$$\boxed{\bar{\sigma}(\nu) = \frac{\lambda^2}{8\pi t_{сп}} g(\nu).} \qquad (13.26)$$

Среднее сечение перехода

Среднее сечение перехода на центральной частоте ν_0 равно

$$\bar{\sigma}_0 \equiv \bar{\sigma}(\nu_0) = \frac{\lambda^2}{8\pi t_{сп}} g(\nu_0). \qquad (13.27)$$

Поскольку $g(\nu_0)$ обратно пропорционально $\Delta\nu$, в соответствии с (13.20), максимальное значение сечения $\bar{\sigma}_0$ обратно пропорционально ширине линии $\Delta\nu$ при данном значении $t_{сп}$. Сечение перехода σ для вынужденного излучения в определенную моду и связанное с ним максимальное сечение σ_0 подчиняются соотношениям, совпадающим с (13.26) и (13.27), за исключением того, что эффективное время спонтанного излучения уменьшается в соответствии с (13.13) и (13.21). Для простоты мы далее не всегда будем делать различие между $t_{сп}$ для спонтанного излучения и его эффективным значением для вынужденного излучения.

13.3.3. Вынужденное излучение и поглощение

Переходы, индуцированные монохроматическим светом

Рассмотрим теперь взаимодействие одномодового света с атомом, при котором на атом падает поток фотонов, в отличие от взаимодействия с модой резонатора объема V, рассмотренного выше. Пусть монохроматический свет с частотой ν, интенсивностью I и средней плотностью потока фотонов [фотон/(см^2 · с)]

$$\phi = \frac{I}{h\nu} \tag{13.28}$$

взаимодействует с атомом, резонансная частота которого равна ν_0. Мы хотим определить плотности вероятности вынужденного излучения и поглощения $W_i \equiv P_{\text{погл}} = P_{\text{в.исп}}$ в этих условиях.

Для определения числа фотонов n, участвующих в процессе взаимодействия, построим цилиндр с площадью основания A, высотой c и объемом cA. Ось цилиндра параллельна направлению распространения света $\boldsymbol{k}$. Поток фотонов через основание цилиндра равен $\Phi = \phi A$ (фотонов/с). Поскольку фотоны распространяются со скоростью света c, все фотоны, заключенные внутри цилиндра, пересекут его основание в течение одной секунды. Отсюда следует, что в каждый момент времени цилиндр содержит $n = \phi A = \phi V/c$ фотонов, так что

$$\phi = n\frac{c}{V}. \tag{13.29}$$

Для определения W_i подставим (13.29) в (13.14) и получим

$$\boxed{W_i = \phi\sigma(\nu).} \tag{13.30}$$

Очевидно, что $\sigma(\nu)$ — коэффициент пропорциональности между плотностью вероятности вынужденного излучения и плотностью потока фотонов. Название «сечение перехода» теперь приобретает следующий смысл: ϕ — поток фотонов через 1 см^2; $\sigma(\nu)$ — эффективная площадь поперечного сечения атома (см^2), а $\phi\sigma(\nu)$ — часть фотонного потока «захваченного» атомом с целью поглощения или вынужденного излучения.

Если скорость спонтанного излучения увеличивается из-за наличия множества мод, в которые атом может испустить фотон, то вынужденное излучение происходит только в моды, в которых уже имеются фотоны. Его скорость увеличивается за счет возможного присутствия большего числа фотонов в нескольких модах.

Переходы, индуцируемые широкополосным светом

Теперь рассмотрим атом в резонаторе объема V, содержащем многомодовый полихроматический свет со спектральной плотностью $\rho(\nu)$ (энергия на единицу частоты и на единицу объема), причем его спектральная ширина

на велика по сравнению с атомной шириной линии. Среднее число фотонов в полосе частот от ν до $\nu + d\nu$ есть $\rho(\nu)V d\nu/h\nu$, каждый из них имеет плотность вероятности $(c/V)\sigma(\nu)$ инициировать атомный переход, так что полная вероятность поглощения или вынужденного излучения равна

$$W_i = \int_0^\infty \frac{\rho(\nu)V}{h\nu}\left[\frac{c}{V}\sigma(\nu)\right]d\nu. \tag{13.31}$$

Поскольку излучение широкополосное, функция $\rho(\nu)$ меняется медленно по сравнению с острым пиком сечения перехода $\sigma(\nu)$. Следовательно, можно заменить под интегралом $\rho(\nu)/h\nu$ на $\rho(\nu_0)/h\nu_0$, что приводит к выражению

$$W_i = \frac{\rho(\nu_0)}{h\nu_0} c\int_0^\infty \sigma(\nu)d\nu = \frac{\rho(\nu_0)}{h\nu_0} cS. \tag{13.32}$$

Используя (13.25), получаем

$$W_i = \frac{\lambda^3}{8\pi h t_{\text{сп}}}\rho(\nu_0), \tag{13.33}$$

где $\lambda = c/\nu_0$ — длина волны в среде для излучения на центральной частоте ν_0. Вводя величину

$$\bar{n} = \frac{\lambda^3}{8\pi h}\rho(\nu_0), \tag{13.34}$$

представляющую собой среднее число фотонов на одну моду, мы можем записать (13.33) в удобном виде

$$\boxed{W_i = \frac{\bar{n}}{t_{\text{сп}}}.} \tag{13.35}$$

Интерпретация $\bar{n}$ как среднего числа фотонов, приходящегося на одну моду, следует из вида отношения

$$\frac{W_i}{P_{\text{сп}}} = \frac{\lambda^3\rho(\nu_0)}{8\pi h t_{\text{сп}}}\frac{1}{M(\nu_0)cS} = \frac{\rho(\nu_0)}{h\nu_0 M(\nu_0)}. \tag{13.36}$$

Величина $\rho(\nu_0)/h\nu_0$ представляет собой среднее число фотонов на единицу объема вблизи частоты ν_0, а $M(\nu_0)$ — число мод на единицу объема вблизи той же частоты ν_0. Таким образом, плотность вероятности W_i в $\bar{n}$ раз больше, чем для спонтанного излучения, поскольку каждая мода содержит в среднем $\bar{n}$ фотонов.

Коэффициенты Эйнштейна

Хотя Эйнштейн не знал формулы (13.33), он выполнил важный анализ обмена энергией между атомами и излучением, который позволил ему вывести общие выражения для плотностей вероятности спонтанных и вынужденных переходов. Предполагая, что атомы взаимодействуют с широкополос-

ным излучением со спектральной плотностью $\rho(\nu)$ в условиях теплового равновесия, он получил следующие выражения:

$$\boxed{P_{\text{сп}} = \mathbb{A};} \tag{13.37}$$

$$\boxed{W_i = \mathbb{B}\rho(\nu_0).} \tag{13.38}$$

Постулаты Эйнштейна

Постоянные $\mathbb{A}$ и $\mathbb{B}$ известны как **коэффициенты Эйнштейна**.

Сравнение с (13.24) и (13.33) показывает, что

$$\boxed{\mathbb{A} = \frac{1}{t_{\text{сп}}};} \tag{13.39}$$

$$\boxed{\mathbb{B} = \frac{\lambda^3}{8\pi h t_{\text{сп}}},} \tag{13.40}$$

что связано со спонтанными и вынужденными переходами соответственно. Отношение коэффициентов равно

$$\frac{\mathbb{A}}{\mathbb{B}} = \frac{\lambda^3}{8\pi h}. \tag{13.41}$$

Связь между $\mathbb{A}$ и $\mathbb{B}$ есть результат микроскопических (а не макроскопических) вероятностных законов взаимодействия между атомом и фотонами каждой моды. Мы приведем анализ, аналогичный предложенному Эйнштейном, в разд. 13.4.

Пример 13.1

Сравнение скоростей спонтанного и вынужденного излучения

Если скорость спонтанного излучения атома в возбужденном состоянии постоянна и равна $\mathbb{A} = 1/t_{\text{сп}}$, то скорость вынужденного излучения в присутствии широкополосного света, $\mathbb{B}\rho(\nu_0)$, пропорциональна спектральной плотности энергии света $\rho(\nu_0)$. Эти две скорости равны, когда

$$\rho(\nu_0) = \frac{\mathbb{A}}{\mathbb{B}} = \frac{8\pi h}{\lambda^2};$$

для больших значений спектральной плотности энергии скорость вынужденного излучения превосходит скорость спонтанного излучения. Если, например, $\lambda = 1$ мкм, то $\mathbb{A}/\mathbb{B} = 1{,}66\ \ 10^{-14}$ Дж/(м^3 Гц). Это соответствует спектральной плотности интенсивности $c\rho(\nu_0) \approx 5 \cdot 10^{-6}$ Вт/(м$^2 \cdot$ Гц) в свободном пространстве. Таким образом, при ширине линии $\Delta\nu = 10^7$ Гц оптическая интенсивность, при которой скорости спонтанного и вынужденного излучения равны, составляет 50 Вт/м^2, или 5 мВт/см^2.

Резюме

Атомный переход можно описывать резонансной частотой $\nu_0 = (E_2 - E_1)/h$, спонтанным временем жизни $t_{сп}$ и функцией формы линии $g(\nu)$, имеющей ширину $\Delta\nu$. Среднее сечение перехода

$$\bar{\sigma}(\nu) = \bar{S}g(\nu) = \frac{\lambda^2}{8\pi t_{сп}} g(\nu). \tag{13.26}$$

Спонтанное излучение

• Если атом находится на верхнем уровне и помещен в резонатор объема V, то плотность вероятности (в секунду) спонтанного испускания фотона в единственную *заданную* моду частоты ν равна

$$p_{сп} = \frac{c}{V}\sigma(\nu). \tag{13.12}$$

• Плотность вероятности спонтанного излучения в любую из доступных мод есть

$$P_{сп} = \frac{8\pi\bar{S}}{\lambda^2} = \frac{1}{t_{сп}}. \tag{13.24}$$

• Плотность вероятности испускания в моды, лежащие в полосе частот между ν и $\nu + d\nu$, равна

$$P_{сп}d\nu = \frac{1}{t_{сп}} g(\nu)d\nu.$$

Вынужденное излучение и поглощение

• Если атом в резонаторе находится на верхнем энергетическом уровне, а мода излучения содержит n фотонов частоты ν, то плотность вероятности излучения фотона в эту моду равна

$$W_i = n\frac{c}{V}\sigma(\nu). \tag{13.17}$$

Если вместо этого атом находится в нижнем состоянии, а мода содержит n фотонов, то вероятность поглощения фотона из этой моды также дается формулой (13.17).

• Если атом не находится в резонаторе, а освещается монохроматическим пучком света частоты ν со средней плотностью потока фотонов ϕ (число фотонов, проходящих в секунду через единицу площади), то плот-

ность вероятности вынужденного излучения (если атом находится на верхнем уровне) или поглощения (если на нижнем):

$$W_i = \phi\sigma(\nu). \tag{13.30}$$

• Если свет, падающий на атом, полихроматический, но узкополосный по сравнению с атомной шириной линии перехода, и имеет среднюю спектральную плотность потока фотонов ϕ_ν (число фотонов на единицу частоты, проходящих за единицу времени через единицу площади), то плотность вероятности вынужденного излучения (поглощения) равна

$$W_i = \int \phi_\nu \sigma(\nu) \mathrm{d}\nu. \tag{13.42}$$

• Если свет, освещающий атом, имеет спектральную плотность энергии $\rho(\nu)$, ширина которой намного больше атомной ширины линии, то плотность вероятности вынужденного излучения/поглощения равна

$$W_i = \mathbb{B}\rho(\nu_0), \tag{13.38}$$

где $\mathbb{B} = \lambda^3/8\pi h t_{\text{сп}}$ — коэффициент Эйнштейна.

Во всех этих формулах $c = c_0/n$ — скорость света в среде; $\lambda = \lambda_0/n$ — длина волны света в среде; n — показатель преломления.

13.3.4. Уширение линий

Поскольку функция формы линии $g(\nu)$ играет важную роль во взаимодействии атомов с фотонами, мы посвятим этот раздел краткому обсуждению ее природы. Одна и та же функция формы линии применима к спонтанному излучению, поглощению и вынужденному излучению.

Уширение за счет конечного времени жизни

Атомы могут переходить с одного энергетического уровня на другой в ходе как излучательных, так и безызлучательных процессов. Излучательные переходы сопровождаются поглощением или испусканием фотонов. При безызлучательных переходах передача энергии происходит через колебания кристаллической решетки, неупругие столкновения с другими атомами или стенками сосуда. Каждый атомный уровень имеет время жизни τ, обратное скорости спада его заселенности в результате излучательных и безызлучательных переходов на нижележащие уровни.

Время жизни τ_2 энергетического уровня 2, показанного на рис. 13.15, представляет собой величину, обратную скорости, с которой убывает заселенность этого уровня за счет переходов на уровень 1 и все нижележащие уровни энергии, не показанные на рисунке. Поскольку $1/t_{\text{сп}}$ — скорость радиационного распада с переходом с уровня 2 на уровень 1, полная скорость распада $1/\tau_2$

состояния 2 должна быть больше, т. е. $1/\tau_2 > 1/t_{сп}$, а значит $\tau_2 < t_{сп}$. Время жизни τ_1 уровня 1 определяется аналогично. Очевидно, что если 1 — самый нижний из возможных уровней энергии (основное состояние), то $\tau_1 = \infty$.

По существу, уширение из-за конечного времени жизни — следствие преобразования Фурье. Время жизни τ энергетического уровня связано с неопределенностью времени заселения этого уровня. Как показано в разд. А.1 приложения А, Фурье-образ экспоненциально затухающей гармонической функции времени $\exp(-t/2\tau)\exp(j2\pi\nu_0 t)$, энергия которой затухает как $\exp(-t/\tau)$ с постоянной времени τ, пропорционален $1/[1 + j4\pi(\nu - \nu_0)t]$. Полная ширина на высоте половины максимума квадрата модуля этой лоренцевой функции частоты равна $\Delta\nu = \pi\tau/2$. Эта спектральная неопределенность соответствует неопределенности энергии $\Delta E = h\Delta\nu = h/2\pi\tau$. Следовательно, энергетический уровень с временем жизни τ имеет разброс энергии $\Delta E = h/2\pi\tau$ при условии, что процесс распада можно моделировать простой экспонентой. В такой картине спонтанному излучению соответствует гармонический осциллятор с затуханием, генерирующий экспоненциально затухающую гармоническую функцию, как в осцилляторной модели Лоренца, представленной в подразд. 5.5.3.

Таким образом, если энергетические ширины уровней 1 и 2 равны $\Delta E_1 = h/2\pi\tau_1$ и $\Delta E_2 = h/2\pi\tau_2$ соответственно, то ширина энергетической разности, соответствующей переходу между двумя уровнями:

$$\Delta E = \Delta E_1 + \Delta E_2 = \frac{h}{2\pi}\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right) = \frac{h}{2\pi}\frac{1}{\tau}, \tag{13.43}$$

где τ — время жизни перехода; $\tau^{-1} = \tau_1^{-1} + \tau_2^{-1}$ Соответствующий разброс частоты перехода, который называется шириной линии из-за конечного времени жизни, выражается как

$$\boxed{\Delta\nu = \frac{1}{2\pi}\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right).} \tag{13.44}$$

Ширина линии из-за конечного времени жизни

Центр линии находится на частоте $\nu_0 = (E_2 - E_1)/h$, а функция формы линии имеет лоренцев профиль

$$\boxed{g(\nu) = \frac{\Delta\nu/2\pi}{(\nu - \nu_0)^2 + (\Delta\nu/2)^2}.} \tag{13.45}$$

Функция формы линии Лоренца

Более общая модель уширения из-за конечного времени жизни для атома или коллектива атомов выглядит следующим образом. Каждый фотон, испущенный при переходе, представляет собой волновой пакет с центральной частотой ν_0 (резонансная частота перехода), который имеет экспоненциально затухающую огибающую с временем затухания 2τ, что для энергии дает время затухания τ, равное времени жизни перехода. Как показано на рис. 13.21, излучаемый свет — последовательность таких волновых пакетов, испускаемых в

случайные моменты времени. Как обсуждается в примере 11.1, это соответствует случайному (частично когерентному) свету со спектральной интенсивностью, которая в точности описывается функцией Лоренца (13.51) с $\Delta\nu = 1/2\pi\tau$.

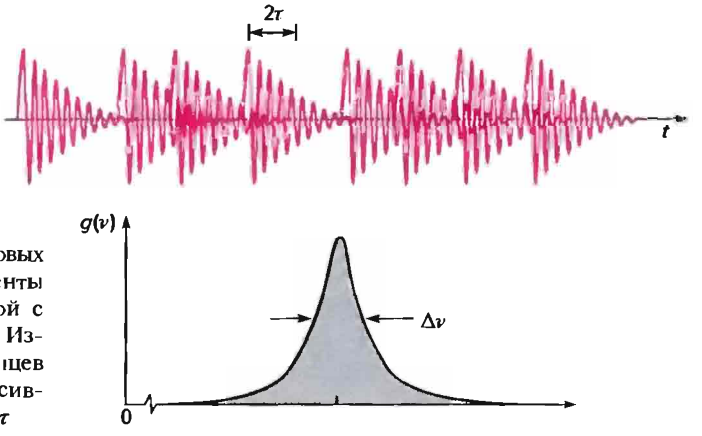


Рис. 13.21. Испускание волновых пакетов в случайные моменты времени атомной системой с временем жизни перехода τ. Излучаемый свет имеет лоренцев контур спектральной интенсивности с шириной $\Delta\nu = 1/2\pi\tau$

Величина лоренцевой функции формы линии на центральной частоте ν_0 составляет $g(\nu_0) = 2/\pi\Delta\nu$, так что максимальное сечение перехода, даваемое выражением (13.27), принимает значение

$$\boxed{\bar{\sigma}_0 = \frac{\lambda^2}{2\pi}\frac{1}{2\pi t_{\text{сп}}\Delta\nu}.} \tag{13.46}$$

Наибольшее сечение перехода получается при идеальных условиях, когда распад чисто радиационный, так что $\tau_2 = t_{\text{сп}}$ и $1/\tau_1 = 0$ (нижний уровень — основное состояние). Тогда $\Delta\nu = 1/2\pi t_{\text{сп}}$, так что

$$\bar{\sigma}_0 = \frac{\lambda^2}{2\pi}. \tag{13.47}$$

Это значит, что максимальное сечение имеет порядок квадрата длины волны. Если уровень 1 — не основное состояние или если существенны безызлучательные переходы, то $\Delta\nu$ может быть много больше $1/t_{\text{сп}}$, и в этом случае $\bar{\sigma}_0$ может быть существенно меньше, чем $\lambda^2/2\pi$. Например, для оптических переходов в диапазоне от $\lambda = 0{,}1$ до 10 мкм величина $\lambda^2/2\pi$ составляет от 10^{-11} до 10^{-7} см^2, в то время как типичные наблюдаемые значения попадают в интервал от 10^{-20} до 10^{-11} см^2 (см. табл. 14.1).

Столкновительное уширение

Столкновения с обменом энергией называются *неупругими столкновениями* и приводят к переходам между атомными уровнями энергии. Это влияет на скорости распада и времена жизни всех уровней и изменяет ширину линии излучения, рассмотренную выше.

Столкновения, не меняющие энергию, называются *упругими столкновениями* и также меняют ширину линии излучаемого поля. Упругие столкновения вносят вклад в случайные изменения фазы волновой функции, связанной с энергетически уровнем. В свою очередь, это приводит к случайному сдвигу фазы излучаемого поля при каждом столкновении. Как показано на рис. 13.22, синусоидальная волна, фаза которой претерпевает случайный сдвиг в случайные моменты времени (моменты столкновений), претерпевает спектральное уширение. Спектр такой случайно дефазируемой функции можно определить с помощью теории случайных процессов. В результате опять получается лоренцев контур, но с шириной $\Delta\nu = f_{ст}/\pi$, где $f_{ст}$ — частота столкновений (среднее число столкновений в секунду)[1]. Лоренцева функция формы линии с учетом вкладов конечного времени жизни и столкновений имеет полную ширину, являющуюся суммой отдельных ширин линий

$$\boxed{\Delta\nu = \frac{1}{2\pi}\left(\frac{1}{\tau_1} + \frac{1}{\tau_2} + 2f_{ст}\right).} \tag{13.48}$$

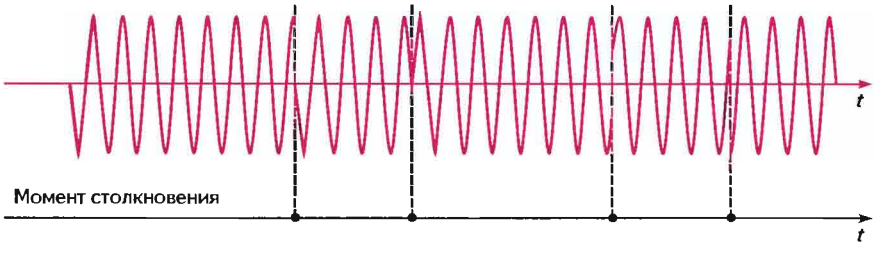


Рис. 13.22. Синусоидальная волна, прерываемая случайными сдвигами фазы с частотой $f_{ст}$, имеет лоренцев спектр с шириной $\Delta\nu = f_{ст}/\pi$

Неоднородное уширение

Уширение из-за времени жизни и ударное уширение — примеры **однородного уширения**, при котором все атомы среды можно считать одинаковыми и имеющими идентичные функции формы линии. При некоторых условиях, однако, различные атомы среды могут иметь различные функции формы линии или различные центральные частоты перехода. В этом случае можно определить среднюю функцию формы линии

$$\bar{g}(\nu) = \langle g_\beta(\nu) \rangle, \tag{13.49}$$

где $\langle \cdot \rangle$ обозначает усреднение по переменной β, различающей функции формы линии атомов $g_\beta(\nu)$. Средняя функция формы линии получается усреднением $g_\beta(\nu)$ с помощью весовой функции — доли атомов, обладающих свойством β, как показано на рис. 13.23.

---

[1] См., например, *Siegman A.E.* Lasers. Разд. 3.2. University Science, 1986.

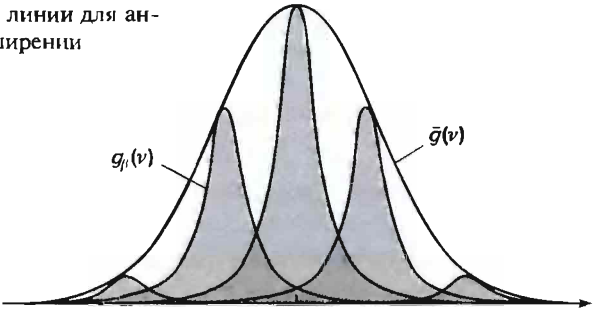


Рис. 13.23. Средняя функция формы линии для ансамбля атомов при неоднородном уширении

Частным случаем неоднородного уширения является доплеровское уширение. В результате **эффекта Доплера** атом, движущийся со скоростью v вдоль заданного направления, в неподвижной системе координат имеет спектр, сдвинутый на частоту $\pm(v/c)\nu_0$ при наблюдении в этом направлении, где ν_0 — центральная частота. Сдвиг происходит в сторону больших частот (знак «+»), если атом движется к наблюдателю, и в сторону меньших частот (знак «−»), если атом движется от наблюдателя. При произвольном направлении наблюдения сдвиг частоты составляет $\pm(v_{\parallel}/c)\nu_0$, где $v_{\parallel}$ — компонента скорости, параллельная направлению наблюдения. Поскольку ансамбль атомов в газе характеризуется распределением по скоростям, как показано на рис. 13.24, свет, испускаемый атомами, имеет распределение по частотам, приводящее к **доплеровскому уширению**.

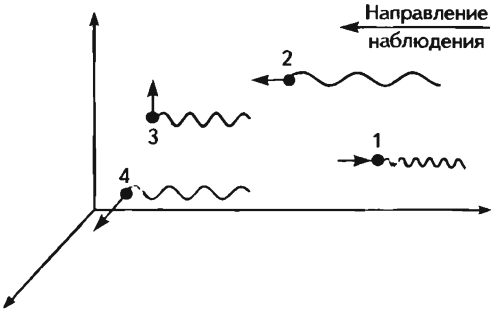


Рис. 13.24. Частота, испускаемая атомом, зависит от направления движения атома относительно направления наблюдения. Излучение атома 1 имеет большую частоту, чем у атомов 3 и 4. Излучение атома 2 имеет меньшую частоту

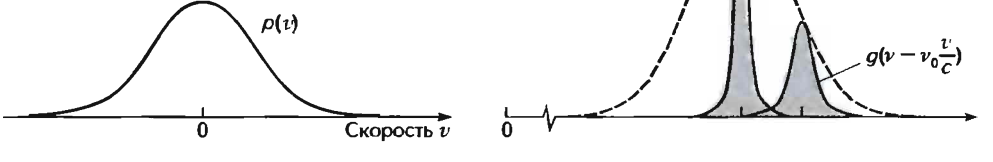


Рис. 13.25. Распределение по скоростям и средняя функция формы линии для атомной системы с доплеровским уширением

Таким образом, при доплеровском уширении роль параметра β играет скорость атома v и $\bar{g}(\nu) = \langle g_v(\nu) \rangle$. Как показано на рис. 13.25, если $p(v)\mathrm{d}v$ — вероятность того, что скорость данного атома лежит между v и $v + \mathrm{d}v$, то полная неоднородно уширенная эффектом Доплера функция формы линии равна

$$\bar{g}(\nu) = \int_{-\infty}^{\infty} g\left(\nu - \nu_0 \frac{v}{c}\right) p(v)\,\mathrm{d}v. \tag{13.50}$$

Упражнение 13.2

Функция формы линии при доплеровском уширении

а. Доля атомов газа, имеющих компоненту скорости вдоль некоторого направления, определяется гауссовой функцией плотности вероятности

$$p(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left(-\frac{v^2}{2\sigma_v^2}\right), \tag{13.51}$$

где

$$\sigma_v^2 = \frac{k_B T}{M};$$

M — атомная масса. При условии, что каждый атом имеет лоренцеву функцию формы линии с шириной $\Delta\nu$ и центральной частотой ν_0, получите выражение для средней функции формы линии $\bar{g}(\nu)$.

б. Покажите, что если $\Delta\nu \ll \nu_0\sigma_v/c$, то $\bar{g}(\nu)$ может быть аппроксимирована гауссовой функцией

$$\bar{g}(\nu) = \frac{1}{\sqrt{2\pi}\sigma_D} \exp\left[-\frac{(\nu - \nu_0)^2}{2\sigma_D^2}\right], \tag{13.52}$$

где

$$\sigma_D = \nu_0 \frac{\sigma_v}{c} = \frac{1}{\lambda}\sqrt{\frac{k_B T}{M}}. \tag{13.53}$$

Тогда для полной доплеровской ширины на уровне половины максимального значения $\Delta\nu_D$ получаем

$$\Delta\nu_D = \sqrt{8 \ln 2}\,\sigma_D \approx 2{,}35\sigma_D. \tag{13.54}$$

в. Вычислите доплеровскую ширину для перехода с $\lambda_0 = 632{,}8$ нм в Не и для перехода с $\lambda_0 = 10{,}6$ мкм в CO_2 при комнатной температуре, предполагая, что $\Delta\nu \ll \nu_0\sigma_v/c$. Эти переходы используются в газовых лазерах на Не—Ne и CO_2, соответственно.

г. Покажите, что максимальное значение сечения перехода для гауссовой функции формы линии в (13.52)

$$\sigma_0 = \frac{\lambda^2}{8\pi}\sqrt{\frac{4\ln 2}{\pi}}\frac{1}{t_{\text{сп}}\Delta\nu_D} \approx 0{,}94\frac{\lambda^2}{8\pi}\frac{1}{t_{\text{сп}}\Delta\nu_D}. \tag{13.55}$$

Сравните с (13.52) для лоренцевой функции формы линии.

Многие взаимодействия атомов с фотонами характеризуются уширением, которое является промежуточным между чисто неоднородным и чисто однородным. Такое смешанное уширение можно моделировать с помощью промежуточных функций формы линии, таких как профиль Фойгта.

*13.3.5. Усиленное спонтанное излучение

Все результаты, полученные до сих пор в разд. 13.3, основаны на предположении, что $\Delta\nu \gg \delta\nu$, т. е. что атомная ширина линии $\Delta\nu$ намного больше, чем ширина моды электромагнитного поля $\delta\nu$. Это условие выполняется обычно, но не всегда. В противоположном пределе, когда атомная ширина много меньше, чем ширина моды (рис. 13.26), может достигаться увеличение плотности вероятности спонтанного излучения, особенно в микрорезонаторах с высокой добротностью, что будет показано ниже. Усиление спонтанного излучения желательно для работы некоторых источников света, что обсуждается в разд. 17.4.

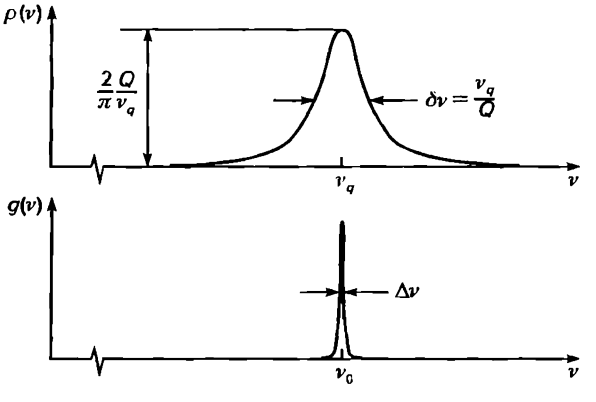


Рис. 13.26. Спонтанное излучение атома с нормированной функцией формы линии $g(\nu)$ в моду с более широкой нормированной лоренцевой функцией $\rho(\nu)$. Центральные частоты атомного перехода и моды резонатора обозначены ν_0 и ν_q соответственно, а их ширины — $\Delta\nu$ и $\delta\nu$. Мы рассматриваем случай, когда $\nu_0 = \nu_q$ и $\Delta\nu \ll \delta\nu$

Рассмотрим спонтанное излучение атома с резонансной частотой ν_0 в электромагнитную моду с центральной частотой $\nu_q = \nu_0$ в режиме $\Delta\nu \ll \delta\nu$, как изображено на рис. 13.26. В соответствии с (13.25), когда дипольный момент атома направлен вдоль электрического поля моды, плотность вероятности спонтанного излучения в одну моду резонатора $\rho(\nu)$ выражается как

$$P_{\text{сп}}^{\max} = \int_0^\infty \frac{c}{V}\sigma_{\max}(\nu)\rho(\nu)\,\mathrm{d}\nu \approx \frac{c}{V}\frac{3\lambda^2}{8\pi t_{\text{сп}}}\rho(\nu_0)\int_0^\infty g(\nu)\,\mathrm{d}\nu, \tag{13.56}$$

поскольку

$$\sigma_{\max}(\nu) = 3\bar{\sigma}(\nu); \quad \bar{\sigma}(\nu) = \frac{\lambda^2 g(\nu)}{8\pi t_{\text{сп}}},$$

как следует из (13.24) и (13.26), соответственно. Так как функция формы линии $g(\nu)$ нормирована, а высота нормированной функции Лоренца для моды резонатора равна $2Q/\pi\nu_q$, где $Q = \nu_q/\delta\nu$, получаем

$$P_{\text{сп}}^{\max} = \frac{1}{t_{\text{сп}}}\frac{3c\lambda^2}{8\pi V}\frac{2Q}{\pi\nu_q} = \frac{1}{t_{\text{сп}}}\frac{3}{4\pi^2}\frac{\lambda^3}{V}Q. \tag{13.57}$$

Конечный результат состоит в увеличении плотности вероятности спонтанного излучения по сравнению со свободным пространством на множитель, известный как **фактор Перселла**

$$\boxed{\frac{P_{\text{сп}}^{\max}}{P_{\text{сп}}} = \frac{3}{4\pi^2}\frac{\lambda^3}{V}Q.} \tag{13.58}$$

Фактор Перселла

Этот фактор имеет следующие особенности.

- Множитель 3 — результат учета взаимной ориентации дипольного момента атома и электрического поля моды.
- Величина λ^3/V, представляющая собой отношение куба длины волны к объему резонатора, существенно увеличивается в микрорезонаторах.
- Высокое значение добротности Q, т. е. узкая линия моды резонатора, увеличивает фактор Перселла, однако с ростом Q величина $\delta\nu = \nu_q/Q$ уменьшается, поэтому в конечном счете нарушается условие $\Delta\nu \ll \delta\nu$.

При отклонении ν_0 от ν_q высота контура моды на частоте ν_0 становится меньше, и усиление спонтанного излучения, в конце концов, превращается в его подавление.

*13.3.6. Лазерное охлаждение атомов и лазерные ловушки

Часто бывает желательно замедлить нейтральные атомы (**лазерное охлаждение**) и удерживать их в ограниченном объеме пространства (**атомная ловушка**). Ультрахолодные атомы обеспечивают не имеющую аналогов точность атомных часов. Лазерное охлаждение и удержание атомов можно реализовать путем расположения лазерных пучков таким образом, чтобы обеспечить селективную передачу импульса от фотонов к пучку атомов с хорошо регулируемыми скоростями (см. подразд. 12.1.4). Охлажденные и удерживаемые атомы — существенные компоненты **атомной оптики** — области исследований, связанные с манипулированием **волнами материи**. Структурированные оптические волны часто используются в качестве элементной базы для атомной оптики; при этом, как в обычной волновой оптике, на этих элементах наблюдаются отражение, преломление, дифракция, интерференция и рассеяние волн веще-

ства. Ожидается, что интерферометрия волн вещества сделает возможными исключительно чувствительные измерения локальных гравитационных аномалий. Охлажденные и удерживаемые атомы важны также для получения **конденсатов Бозе—Эйнштейна** (БЭК) — коллективов медленных и тесно расположенных атомов, волновые функции которых перекрываются.

Одна из простейших схем лазерного охлаждения основана на использовании фотонов лазерного пучка с малой шириной линии и центральной частотой, настроенной немного ниже центра атомной линии. Фотоны взаимодействуют с пучком атомов, летящих навстречу. После поглощения фотонов атомами, у которых за счет эффекта Доплера центральная частота перехода совпадает с частотой фотонов, эти атомы возвращаются в основное состояние в результате спонтанного или вынужденного испускания. Если испускание вынужденное, то импульс излученного фотона такой же, как у поглощенного, и импульс атома не изменяется. Если испускание спонтанное, то направление испущенного фотона произвольное, поэтому повторяющиеся акты поглощения-испускания приводят к общему уменьшению составляющей импульса атома в направлении навстречу лазерному пучку. В результате скорость атомов в данном направлении снижается, как схематически показано на рис. 13.27. В конце концов изменение импульса (и, следовательно, скорости) атомов приводит к тому, что атомы уходят из резонанса с лазерным пучком, что можно компенсировать периодическим изменением его частоты.

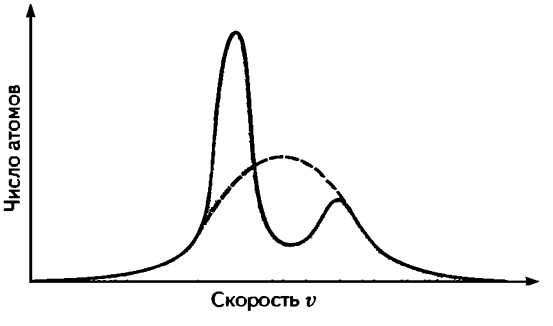


Рис. 13.27. Распределение по скоростям в пучке атомов (штриховая линия) и распределение после лазерного охлаждения (сплошная линия)

Несколько лазерных пучков можно использовать для создания оптической ловушки, в которой большое число нейтральных атомов можно удерживать в малом объеме пространства. Удерживаемые атомы можно быстро перемещать простым изменением направления лазерных пучков. Для того чтобы атомы удерживались, кинетическая энергия ансамбля атомов должна быть достаточно мала, чтобы атомы не «выпрыгивали» из ловушки. Использование методики лазерного охлаждения и удержания позволяет достигать температур в несколько микрокельвинов для нейтральных атомов (что соответствует скоростям порядка см/с). Это на много порядков ниже температур порядка милликельвинов, достигаемых обычным криогенным охлаждением. Еще более низких температур можно достичь «испарительным охлаждением», в котором глубина ловушки понижается, при этом атомы с большей энергией покидают ловушку, а остаются

наиболее медленные. Другой метод называется «охлаждением ниже отдачи» («subrecoil cooling»). В нем неопределенность импульса атома делается меньшей, чем импульс отдачи от одного фотона, и поддерживается на этом уровне в течение длительного времени за счет статистики Леви случайных блужданий импульса.

13.4. ТЕПЛОВОЕ ИЗЛУЧЕНИЕ

В условиях теплового равновесия и отсутствия других внешних источников энергии черными телами (телами, поглощающими все падающее на них излучение) испускается универсальный тип излучения, известный как **тепловое излучение**. В данном разделе мы определим свойства теплового излучения путем исследования взаимодействия ансамбля фотонов и атомов при тепловом равновесии, т. е. процессов поглощения, спонтанного и вынужденного излучения. Мы также покажем, как тепловое излучение предмета можно использовать для получения его изображения.

13.4.1. Тепловое равновесие между фотонами и атомами

Макроскопический подход, основанный на уравнениях баланса с учетом процессов спонтанного излучения, поглощения и вынужденного излучения, при условии теплового равновесия приводит к выражению для спектральной интенсивности теплового излучения. Отправной точкой нашего анализа являются формулы (13.24) и (13.35), описывающие спонтанное излучение и вынужденные переходы в присутствии широкополосного света. Рассмотрим полость единичного объема, стенки которой состоят из большого числа атомов с двумя энергетическими уровнями 1 и 2, разность энергий которых равна $h\nu$. В полости, имеющей температуру T, поддерживается широкополосное излучение. Пусть $N_1(t)$ и $N_2(t)$ — число атомов на единицу объема, заселяющих уровни 1 и 2 в момент времени t. Поскольку некоторые атомы первоначально находятся на уровне 2, что гарантируется конечной ненулевой температурой, спонтанные переходы вниз порождают поле излучения в полости. Это излучение, в свою очередь, может индуцировать поглощение и вынужденное испускание. Все три процесса сосуществуют. Рассмотрим установившийся (равновесный) режим. Допустим, что среднее число фотонов $\bar{n}$ находится в *каждой* из мод излучения, частоты которых лежат в пределах ширины линии атомного перехода в соответствии с (13.35).

Вначале рассмотрим одно спонтанное излучение. Вероятность того, что один атом, находящийся на верхнем уровне, спонтанно перейдет на нижний уровень с излучением фотона в любую моду за время между t и $t + \Delta t$, равна $P_{\text{сп}}\Delta t = \Delta t/t_{\text{сп}}$. Имеется $N_2(t)$ таких атомов, что среднее число испущенных за время Δt фотонов равно $N_2(t)\Delta t/t_{\text{сп}}$. Эта величина определяет также и число атомов, покинувших уровень 2 за время Δt. Поэтому (отрицательная) скорость увеличения

$N_2(t)$ из-за спонтанного излучения получается из дифференциального уравнения

$$\frac{dN_2}{dt} = -\frac{N_2}{t_{сп}}. \tag{13.59}$$

Решением является экспоненциально спадающая функция времени

$$N_2(t) = N_2(0)\exp\left(-\frac{t}{t_{сп}}\right),$$

график которой показан на рис. 13.28. Число атомов на верхнем уровне N_2 стремится к нулю с постоянной времени $t_{сп}$. Энергия уносится спонтанно излучаемыми фотонами.

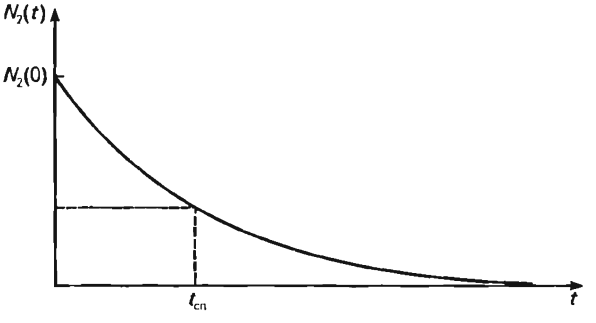


Рис. 13.28. Распад заселенности верхнего уровня за счет одного лишь спонтанного излучения

Включим теперь в рассмотрение процессы поглощения и вынужденного излучения, вносящие вклад в изменение заселенностей. Поскольку имеется N_1 атомов, способных поглощать фотоны, скорость возрастания числа атомов на верхнем уровне за счет поглощения можно выразить с использованием (13.35)

$$\frac{dN_2}{dt} = N_1 W_i = \frac{\bar{n}N_1}{t_{сп}}. \tag{13.60}$$

Аналогично, вынужденное излучение дает (отрицательный) вклад в скорость нарастания заселенности верхнего уровня:

$$\frac{dN_2}{dt} = -N_2 W_i = -\frac{\bar{n}N_2}{t_{сп}}. \tag{13.61}$$

Скорости атомного поглощения и вынужденного излучения пропорциональны среднему числу фотонов в каждой моде.

Объединяя (13.59), (13.60) и (13.61), чтобы учесть одновременное протекание процессов спонтанного излучения, поглощения и вынужденного излучения, получим скоростное уравнение

$$\boxed{\frac{dN_2}{dt} = -\frac{N_2}{t_{сп}} + \frac{\bar{n}N_1}{t_{сп}} - \frac{\bar{n}N_2}{t_{сп}}}. \tag{13.62}$$

Скоростное уравнение

Это уравнение игнорирует переходы на уровень 2 или с него за счет других эффектов, таких как взаимодействие с другими уровнями энергии, безызлучательные переходы и внешние источники возбуждения. Условие стационарности $dN_2/dt = 0$ приводит к выражению

$$\frac{N_2}{N_1} = \frac{\bar{n}}{1+\bar{n}}. \tag{13.63}$$

Очевидно, $N_2/N_1 \leq 1$. Теперь используем тот факт, что атомы находятся в состоянии теплового равновесия, а значит их заселенности подчиняются распределению Больцмана (13.8)

$$\frac{N_2}{N_1} = \exp\left(-\frac{E_2 - E_1}{k_B T}\right) = \exp\left(-\frac{h\nu}{k_B T}\right). \tag{13.64}$$

Подставляя (13.64) в (13.63), получаем среднее число фотонов на одну моду вблизи частоты ν в виде

$$\bar{n} = \frac{1}{\exp(h\nu/k_B T) - 1}. \tag{13.65}$$

Наш вывод основан на рассмотрении двух уровней, связанных поглощением, спонтанным и вынужденным излучением на частотах вблизи. Применимость (13.65), однако, намного шире. Это можно понять, если рассмотреть полость, стенки которой изготовлены из твердотельного материала, имеющего непрерывный спектр энергий, так что расстояние между уровнями и, соответственно, частота ν может принимать любое значение. Атомы стенок спонтанно излучают в полость. Испущенный свет затем взаимодействует с атомами, вызывая поглощение и вынужденное излучение. Если стенки поддерживаются при температуре T, комбинированная система атомов и излучения достигает теплового равновесия.

Выражение (13.65) совпадает с (12.43) — выражением для среднего числа фотонов в моде теплового излучения, для которой заселенности модовых энергетических уровней следуют распределению $p(n) \propto \exp(-nh\nu/k_B T)$. Это свидетельствует о самосогласованности нашего анализа. Фотоны, взаимодействующие с атомами при тепловом равновесии с температурой T, сами находятся в состоянии теплового равновесия с той же температурой T (см. подразд. 12.2.3). Ансамбль таких фотонов часто называют «фотонным газом».

13.4.2. Спектр излучения черного тела

На основе проведенного в подразд. 13.4.1 обсуждения средняя энергия $\bar{E}$ моды излучения равна просто $\bar{n}h\nu$, где $\bar{n}$ дается формулой (13.65), так что

$$\boxed{\bar{E} = \frac{h\nu}{\exp(h\nu/k_B T) - 1}}. \tag{13.66}$$

Средняя энергия моды при тепловом равновесии

Зависимость $\bar{E}$ от ν, совпадающая с (12.46), изображена на рис 13.29. Умножая среднюю энергию $\bar{E}$, приходящуюся на одну моду, на плотность мод $M(\nu) = 8\pi\nu^2/c^3$, определяемую формулой (10.82), получаем спектральную плотность энергии $\rho(\nu) = M(\nu)\bar{E}$ (энергию на единицу частоты и на единицу объема полости) в виде

$$\rho(\nu) = \frac{8\pi h\nu^3}{c^3}\frac{1}{\exp(h\nu/k_B T)-1}. \qquad (13.67)$$

Спектральная плотность энергии излучения черного тела

Эта формула известна как **спектр излучения черного тела**. График зависимости спектральной плотности энергии от частоты показан на рис. 13.30. Зависимость этой величины от температуры показана на рис. 13.31 (см. цв. вклейку). Полная мощность излучения черного тела быстро растет с температурой пропорционально T^4 **(закон Стефана—Больцмана)**.

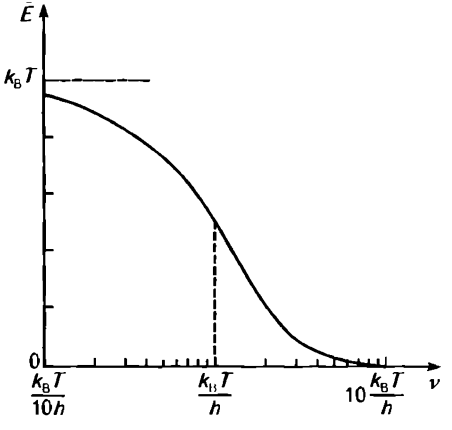


Рис. 13.29. Полулогарифмический график средней энергии $\bar{E}$ электромагнитной моды в состоянии теплового равновесия при температуре T как функции модовой частоты ν. При $T = 300$ °K $k_B T/h = 6{,}25$ ТГц, что соответствует длине волны 48 мкм

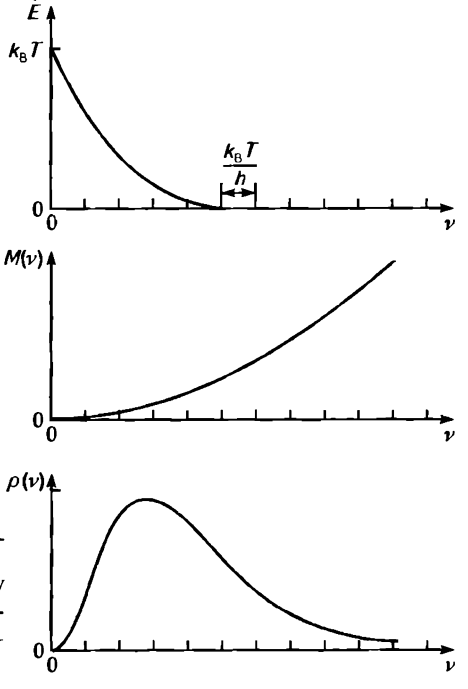


Рис. 13.30. Частотная зависимость энергии на одну моду $\bar{E}$, плотности мод $M(\nu)$ и спектральной плотности энергии $\rho(\nu) = M(\nu)\bar{E}$ в линейных координатах

Спектр излучения черного тела сыграл важную роль в открытии квантовой (фотонной) природы света (см. гл. 12). На основе классической электромагнитной оптики было известно, что плотность мод трехмерного резонатора является квадратичной функцией частоты, а именно $M(\nu) = 8\pi\nu^2/c^3$ (см. подразд. 10.3.3).

Однако закон равнораспределения энергии в классической статистической механике утверждал, что средняя энергия, приходящаяся на моду, должна быть постоянной при $\bar{E} = k_B T$, независимо от частоты моды.

Это приводило к выражению для $\rho(\nu)$, известному как формула Рэлея—Джинса для излучения черного тела, противоречащему экспериментальным данным. Более того, интеграл от него расходился. В 1900 г. Макс Планк обнаружил, что можно получить теоретическое выражение для спектра черного тела, согласующееся с экспериментом, путем квантования энергии каждой моды. Рассуждения Планка привели к выражению (13.66) для $\bar{E}$. Что касается формулы Рэлея—Джинса, то она приближенно справедлива в предельном случае малых энергий фотона, $h\nu \ll k_B T$. В этом случае

$$\exp\left(\frac{h\nu}{k_B T}\right) \approx 1 + \frac{h\nu}{k_B T},$$

в результате чего (13.66) превращается в классическую формулу равнораспределения $\bar{E} \approx k_B T$, а значит

$$\rho(\nu) \approx 8\pi\nu^2 \frac{k_B T}{c^3}.$$

Упражнение 13.3

Частота максимума спектральной плотности энергии

Используя закон излучения черного тела $\rho(\nu)$, покажите, что частота $\nu_п$, на которой спектральная плотность энергии максимальна, удовлетворяет уравнению $3(1 - e^{-x}) = x$, где $x = h\nu_п/k_B T$. Приближенно найдите x и определите $\nu_п$ при $T = 300$ °K.

Термография

Формула для спектральной плотности энергии излучения черного тела (13.67) полезна для получения карт (изображений) распределения температуры тепловых объектов. Это достигается с помощью камеры, чувствительной в диапазоне длин волн теплового излучения объекта (см. рис. 14.24). Горячие объекты, например Солнце, наиболее сильно излучают в видимой области, в то время как объекты с умеренной температурой, такие как Земля и человек, в типичном случае излучают в средней инфракрасной области. Холодные объекты излучают в дальней инфракрасной области. Получение изображений объектов за счет их собственного излучения называется **термографией**. Термографические камеры содержат матрицы фотоприемников, чувствительных в определенной области спектра (см. разд. 18.5). Метод часто используется в области длин волн $0{,}7 \le \lambda_0 \le 300$ мкм, что соответствует $12 \le T \le 5200$ °K. Хотя для термографии благоприятны более высокие температуры из-за того, что полная мощность излучения пропорциональна T^4, примеры изображений на рис. 13.32 (см. цв. вклейку) иллюстрируют широкий диапазон температур, которого можно достичь.

Термография используется для сбора информации об объектах и событиях, которым свойственны вариации температуры. Различные локальные температуры обычно отображаются искусственно генерируемом на компьютере цветом. Этот метод находит промышленные применения, такие как мониторирование перегрева монтажных плат и эволюции нефтяных пятен. Он помогает при поиске и спасении людей и животных, даже если они скрыты ночью в густой листве. Термография также применяется в клинической медицине, поскольку температура кожной поверхности является параметром диагностики нарушений кровотока и опухолей. В области охраны окружающей среды применения включают борьбу с пожарами и лесное хозяйство. Метод незаменим в астрономии и космологии, поскольку он позволяет получать изображения таких астрономических объектов, как относительно холодные красные звезды и красные гиганты в ближнем инфракрасном диапазоне, планеты, кометы и астероиды — в среднем инфракрасном и, наконец, центральные области галактик и излучение межзвездной пыли — в дальнем инфракрасном диапазоне.

13.5. ЛЮМИНЕСЦЕНЦИЯ И РАССЕЯНИЕ СВЕТА

Тепловое возбуждение — не единственный внешний источник энергии, который способен перевести атом или молекулу на более высокие энергетические уровни с последующим испусканием света. Другие источники возбуждения, например электронный удар или звуковая волна, также могут вызывать излучение света при возвращении системы в основное состояние. Возбуждение в виде одного или нескольких фотонов также может приводить к излучению света при фотолюминесценции. Нетепловые излучатели называются **люминесцентными излучателями**, а сам процесс такого излучения — **люминесценцией**.

Помимо люминесценции, при которой сначала происходит поглощение, а затем испускание фотонов, может происходить рассеяние света атомными или молекулярными системами резонансного или нерезонансного характера. Различные виды линейного и нелинейного рассеяния, такие как рэлеевское и комбинационное (рамановское) рассеяние, играют важную роль в фотонике.

13.5.1. Виды люминесценции

Виды люминесценции классифицируются по источнику возбуждения, как отмечено в приводимых примерах (рис. 13.33, см. цв. вклейку).

Катодолюминесценция

Катодолюминесценция — свечение материала в результате возбуждения электронами большой энергии. Примерами являются изображения на экране электронно-лучевой трубки или усилителя изображения, индуцированные путем возбуждения электронами покрывающего экран фосфора. Катодолюми-

несценция часто используется для определения состава материала, поскольку глубина проникновения может регулироваться за счет изменения энергии налетающих электронов, а различные компоненты материала дают свечение на разных длинах волн.

Сонолюминесценция

Сонолюминесценция — это испускание света жидкостью в результате акустической кавитации — образования, роста и коллапса пузырьков в жидкости, облучаемой высокоинтенсивным звуком или ультразвуком. Свет состоит из вспышек пикосекундной длительности, испускаемых в момент достижения коллапсирующим пузырьком минимального размера. Сонолюминесценция наблюдается от скоплений пузырьков, а при некоторых условиях — и от отдельных пузырьков. Существует возможность генерировать вспышки сонолюминесценции отдельных пузырьков со стабильным периодом и локализацией.

Хемилюминесценция

Так называется излучение света при химической реакции. Она наблюдается при тех сравнительно редких условиях, когда энергия, высвобождающаяся при химической реакции между двумя или более веществами, достаточна для заселения возбужденных состояний продукта реакции. Светящиеся палочки, например, светятся после того, как перегородка между двумя отделениями разрушается и содержащиеся в них реагенты смешиваются. Цвет свечения определяется красителем, входящим в химическую смесь. Такие источники света используются для освещения под водой и в военных условиях.

Биолюминесценция

Это хемилюминесценция живых организмов, таких как летающие и ползающие светляки. Свечение является способом связи и некоторые организмы, например летающие светляки, синхронизируют вспышки света. Многие глубоководные морские организмы естественно излучают свет, часто в синей части спектра, где морская вода более прозрачна. Биологи часто присоединяют биолюминесцентные белки медузы к генам других видов для оптического отслеживания генетических процессов.

Электролюминесценция

При электролюминесценции свет испускается в результате приложения к материалу электрического поля. Важный пример — инжекционная электролюминесценция, которая возникает, когда электрический ток инжектируется в прямосмещенный переход между полупроводниками, такой как в светоизлучающих диодах (СИД), которые будут обсуждаться в гл. 17. Рекомбинация инжектируемых электронов из зоны проводимости с дырками валентной зоны приводит к испусканию фотонов.

Фотолюминесценция

Фотолюминесценция — излучение света образцом после поглощения оптических фотонов. Примером является свечение некоторых материалов при облучении ультрафиолетовым светом. Фотолюминесценция, которая подробнее обсуждается в следующем разделе, является важным инструментом для исследования свойств полупроводниковых материалов. Свечение называется **радиолюминесценцией**, если возбуждение производится рентгеновскими или гамма-квантами.

Если излучательные переходы разрешены по спину, т. е. они происходят между двумя состояниями одинаковой мультиплетности (например, синглет—синглет или триплет—триплет, см. рис. 13.5), то процесс люминесценции называется **флуоресценцией**. Люминесценция на переходах, запрещенных по спину (например, триплет—синглет), называется **фосфоресценцией**. Времена жизни флуоресценции обычно сравнительно малы (от 0,1 до 20 нс), поэтому фотоны флуоресценции испускаются сразу после возбуждения. Это отличает ее от фосфоресценции, при которой «запрещенный» характер перехода влечет за собой большие времена жизни (обычно от 1 мс до 10 с) и, следовательно, существенный промежуток времени между возбуждением и испусканием. Усиление флуоресценции вынужденным испусканием составляет основу действия лазеров (см. гл. 14 и 15).

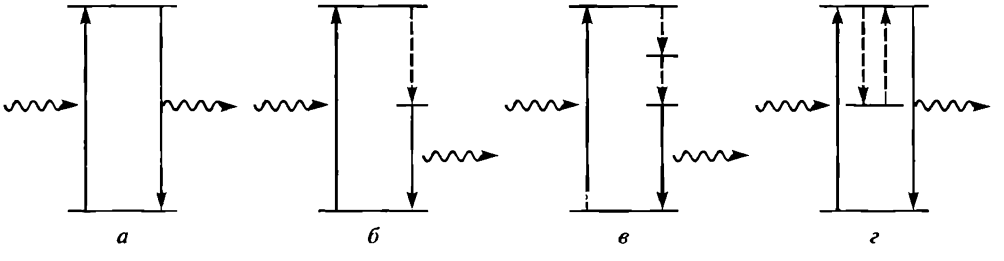


Рис. 13.34. Различные типы однофотонной фотолюминесценции

13.5.2. Фотолюминесценция

Фотолюминесценция происходит, когда система поглощает фотон и возбуждается на более высокий уровень энергии, после чего спонтанно излучает фотон и переходит на более низкий энергетический уровень. По закону сохранения энергии испускаемый фотон не может иметь большую энергию, чем возбуждающий. Несколько типов переходов, приводящих к фотолюминесценции, схематически показано на рис. 13.34. Частью процесса могут быть безызлучательные переходы вниз, показанные штриховыми линиями на рис. 13.34, *б* и *в*. За счет этого механизма ультрафиолетовый свет может быть преобразован в видимый. Возбужденный электрон может в течение продолжительного времени находиться в промежуточном долгоживущем состоянии (например, ловушке), результатом чего является задержанная люминесценция. Могут происходить промежуточные безызлучательные переходы вниз,

а затем снова вверх, как показано на рис. 13.34, *г*. В другой вариации на данную тему, называемой «квантовая резка», за поглощением ультрафиолетового фотона следует испускание двух видимых фотонов. Фотолюминесценция наблюдается во всех видах вещества.

Многофотонная фотолюминесценция

Фотолюминесценция может происходить, когда система возбуждается на более высокий уровень энергии путем поглощения более чем одного фотона с последующим переходом на нижележащие уровни, сопровождающимся испусканием фотона. Возбуждающие фотоны могут иметь одинаковые или разные энергии, а испускаемый фотон может иметь большую энергию, чем каждый из поглощаемых фотонов.

Многофотонная флуоресценция

Два или более фотонов одинаковой энергии могут объединиться, чтобы возбудить систему на более высокий уровень энергии, где она претерпевает фотолюминесценцию (флуоресценцию), как схематически показано на рис. 13.35, *а* и *б*. Двухфотонная флуоресценция, показанная на рис. 13.35, *а*, составляет основу метода визуализации, который называется **двухфотонной лазерной сканирующей микроскопией** (ДФЛСМ). Флуоресцентный зонд (**флуорофор**), присоединенный к объекту в специальном месте, поглощает пару фотонов, появляющихся в его окрестности, каждый с энергией $h\nu_1$, и излучает один фотон флуоресценции с энергией $h\nu_2$ ($>h\nu_1$), который регистрируется. Как показано в подразд. 12.2.3, вероятность наблюдения двух одновременно приходящих фотонов в данной точке в данный момент времени равна квадрату вероятности наблюдения одного такого фотона. Таким образом, благодаря (12.14), скорость двухфотонного поглощения в точке $\boldsymbol{r}$ в момент времени t, а значит и скорость испускания фотонов флуоресценции, ведет себя как квадрат интенсивности падающего света, т. е. пропорциональна $I^2(\boldsymbol{r}, t)$.

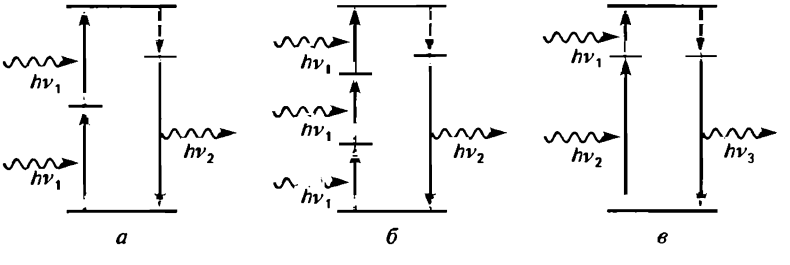


Рис. 13.35. Двухфотонная флуоресценция (*а*); трехфотонная флуоресценция (*б*); флуоресцентная ап-конверсия (*в*). Предполагается участие безызлучательных переходов во всех процессах распада. Возможны и другие сценарии

Достоинства ДФЛСМ в значительной мере вытекают из этой квадратичной зависимости: сфокусированный возбуждающий пучок локально поглощается в

непосредственной близости фокуса, поскольку двухфотонное поглощение происходит преимущественно там, где интенсивность наибольшая. По сравнению с обычной (однофотонной) микроскопией область, от которой наблюдается флуоресценция, уменьшается, что повышает разрешение. Кроме того, происходит подавление шумового фона от флуоресценции вне фокуса. Еще одним преимуществом ДФЛСМ в области биологии является удвоенная длина волны возбуждения, поскольку более длинные волны глубже проникают в биологические ткани. Чтобы гарантировать достаточно высокую пиковую интенсивность для двухфотонного поглощения при достаточно низкой средней интенсивности, чтобы избежать повреждения нежной биоткани, возбуждение часто производится лазером с синхронизацией мод, который генерирует ультракороткие (фемтосекундные) оптические импульсы с очень высокой пиковой мощностью, но низкой средней мощностью. **Многофотонная лазерная сканирующая микроскопия** действует по тому же принципу, но с поглощением не двух, а k независимых фотонов при каждом акте возбуждения, так что скорость испускания фотонов меняется как $I^k(\boldsymbol{r}, t)$.

Трехмерная многофотонная микролитография

Аналогичный подход используется для изготовления микрообъектов. Линза направляет оптические импульсы высокой мощности в некоторую точку специально разработанного прозрачного полимерного материала. Свет имеет достаточную интенсивность, чтобы вызвать многофотонную полимеризацию только в малой окрестности фокуса, причем достигает он этой окрестности, не оказывая влияния на материал. Двигая точку фокусировки линзы, можно записать любую заданную трехмерную микроструктуру. На практике сильное пороговое поведение полимеризационной нелинейности еще более увеличивает разрешение метода.

Ап-конверсия (преобразование частоты вверх) при флуоресценции

Многофотонная флуоресценция может происходить также при различных энергиях фотонов, участвующих в многофотонном переходе, как показано на рис. 13.35, *в*. Такая схема полезна для преобразования инфракрасных фотонов в видимые. Инфракрасный фотон с низкой энергией ($h\nu_1$) вместе с вспомогательным фотоном ($h\nu_2$) возбуждают систему, например, отдельный ион, который затем испускает фотон люминесценции с суммарной энергией ($h\nu_3 = h\nu_1 + h\nu_2$).

Ап-конверсия через последовательные акты поглощения наиболее часто наблюдается в материалах, содержащих ловушки, которые могут хранить возбужденные первым фотоном электроны в течение времени, достаточного для того, чтобы второй фотон появился и перевел систему в вышележащее флуоресцирующее состояние. Часто используются фосфоры, легированные редкоземельными ионами, такими как Eu^{3+} В некоторых материалах ловушки могут накапливать электроны в течение нескольких минут при освещении дневным

или ультрафиолетовым светом, обеспечивающим вспомогательные фотоны $h\nu_2$. Затем инфракрасный сигнальный фотон $h\nu_1$ высвобождает электрон из ловушки, результатом чего является видимый фотон люминесценции с энергией $h(\nu_1 + \nu_2)$. Ап-конверсия может происходить и по более сложному сценарию, например, в результате коллективного испускания двумя близко расположенными ионами, оба из которых были возбуждены.

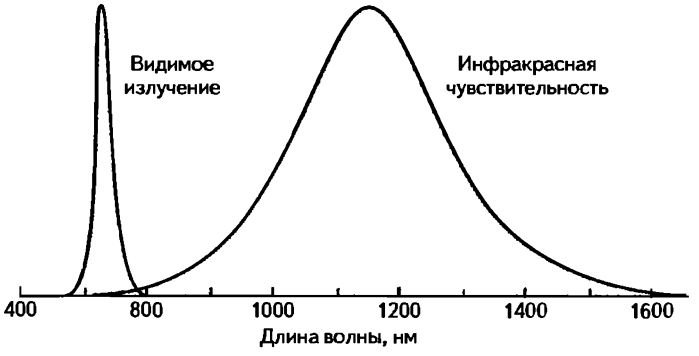


Рис. 13.36. Инфракрасная спектральная чувствительность и относительная спектральная интенсивность испускания при ап-конверсии для серийно выпускаемой карточки ИК-визуализации

Практические устройства часто имеют вид небольших отражающих или пропускающих карточек с активной площадью около 5 × 5 см — **карточек ИК-визуализации**. Порошок, обеспечивающий ап-конверсию, ламинируется между парой жестких листов прозрачного пластика, образуя карточку. Этот же порошок может быть диспергирован в объеме полимерного блока для обеспечения трехмерной визуализации. Хотя эффективность конверсии обычно очень мала, эти устройства, тем не менее, очень полезны для визуального наблюдения пространственного распределения инфракрасного пучка, например, генерируемого инфракрасным лазером. Относительная спектральная чувствительность и спектральная интенсивность испускания серийно выпускаемой карточки показаны на рис. 13.36.

13.5.3. Рассеяние света

Фотолюминесценция, рассмотренная в подразд. 13.5.2, включает резонансное поглощение фотона с переходом из основного состояния в реальное возбужденное состояние. Следующая за этим релаксация возбужденного состояния включает испускание фотона люминесценции. Поглощение и последующее излучение из реального возбужденного состояния — определяющие характеристики люминесценции, флуоресценции и фосфоресценции.

Процессы **рассеяния света** могут включать переходы через виртуальные состояния. Поскольку существуют нерезонансные взаимодействия, свет может рассеиваться в широком диапазоне частот. Мы рассмотрим по очереди три процесса рассеяния, важные для фотоники: рассеяние Рэлея, рассеяние Рамана (комбинационное рассеяние) и рассеяние Мандельштама—Бриллюэна (рис. 13.37).

Во многих случаях рассеяние неизбежно и не всегда желательно, однако, оно может быть полезным источником информации о характеристиках материала и применяться для создания источников света.

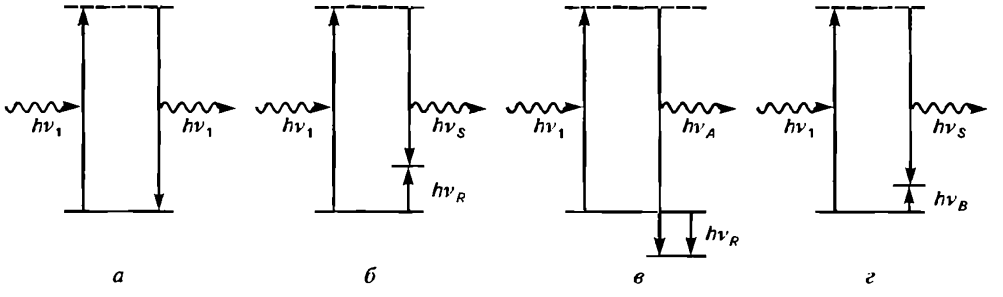


Рис. 13.37. Несколько видов рассеяния света:

а — рэлеевское; *б* — комбинационное (рамановское) стоксово; *в* — комбинационное (рамановское) антистоксово; *г* — рассеяние Мандельштама–Бриллюэна. Штриховыми горизонтальными линиями показаны виртуальные состояния, соответствующие нерезонансному рассеянию

Рэлеевское рассеяние

При рэлеевском рассеянии материал заставляет падающий фотон изменить направление. Это происходит в результате упругого взаимодействия, так что рассеянный и падающий фотоны имеют одинаковую энергию, как схематически показано на рис. 13.37. Рассеяние Рэлея происходит в газах, жидкостях и твердых телах. Оно обусловлено вариациями средств среды, например, случайными неоднородностями показателя преломления стекол (см. подразд. 9.3.1) или присутствием частиц, размеры которых много меньше длины волны света, таких как электроны, атомы, молекулы и наночастицы. Интенсивность рассеянного света пропорциональна ν^4 и, следовательно, $1/\lambda_0^4$, где ν и λ_0 — частота и длина волны освещения соответственно. Таким образом, короткие волны рассеиваются сильнее длинных волн. Рэлеевское рассеяние отвечает за голубой цвет неба. Рассеяние на сферических частицах большего размера, чем $\approx\lambda_0/10$, известно как **рассеяние Ми**, этот процесс слабо зависит от длины волны освещения и отвечает за белый ореол вокруг огней при наличии дымки или тумана.

Рамановское (комбинационное) рассеяние

Рассеяние Рамана (в русскоязычной литературе называемое **комбинационным рассеянием**) — это двухквантовый процесс, в котором в процессе взаимодействия с веществом фотон частоты ν_I исчезает и рождается новый фотон с более низкой частотой ν_S (стоксово рассеяние, $\nu_S = \nu_I - \nu_R$) или более высокой частотой ν_A (антистоксово рассеяние, $\nu_A = \nu_I + \nu_R$). Схемы этих процессов показаны на рис. 13.37, *б* и *в* соответственно. Комбинационное рассеяние может происходить в газах, жидкостях и твердых телах. В отличие от рэлеевского, комбинационное рассеяние — неупругий процесс. Изменение энергии фотона

происходит за счет обмена энергией $h\nu_R$ с колебательной и/или вращательной модой молекулы или твердого вещества. При **стоксовом рассеянии** фотон делится своей энергией с веществом, при **антистоксовом** — заимствует энергию у вещества. Таким образом, спектр рассеянного света в общем случае содержит несмещенную компоненту рэлеевского рассеяния на частоте падающего света, а также смещенные в красную и синюю сторону боковые компоненты стоксова и антистоксова комбинационного рассеяния соответственно. Хотя интенсивность боковых компонент для типичных нерезонансных взаимодействий мала и составляет около 10^{-7} от интенсивности падающего света, комбинационное рассеяние широко применяется для характеристики материалов. В кристаллических материалах колебательный спектр, как правило, дискретный, и рамановские линии узкие. Стекла, напротив, имеют широкие колебательные спектры, которые порождают широкие рамановские спектры. **Рассеяние Мандельштама—Бриллюэна**, схема которого изображена на рис. 13.37, *г*, аналогично рамановскому, однако обмен энергией $h\nu_B$ происходит не с колебательными модами молекул, а с акустическими модами сплошной среды.

Вынужденное комбинационное рассеяние

Вынужденное комбинационное рассеяние (ВКР) происходит, когда сигнальный фотон входит в нелинейную оптическую среду вместе с фотоном накачки, имеющим более высокую частоту (см. вставку на рис. 14.20). Сигнальный фотон стимулирует испускание второго такого же фотона, который получается в результате стоксова сдвига энергии фотона накачки. Избыточная энергия фотона отдается колебательным модам молекул вещества. Процесс имеет сходство с вынужденным излучением, однако, комбинационное взаимодействие является параметрическим нелинейным оптическим процессом третьего порядка (см. подразд. 21.3.2).

Вынужденное комбинационное рассеяние полезно для создания оптических усилителей (см. подразд. 14.3.4) и лазеров (см. подразд. 15.3.1). Определенное достоинство рамановского усиления и генерации заключается в том, что ширина полосы, в пределах которой они могут осуществляться, определяется колебательным спектром материала, а не шириной линии перехода для вынужденного излучения. Колебательный спектр стекла особенно широк, поэтому участок оптического волокна может служить оптическим усилителем или лазером в полосе длин волн в сотни нанометров шириной. Рамановские оптические усилители и рамановские волоконные лазеры находят все более широкое применение в системах оптической волоконной связи с разделением по длинам волн (см. подразд. 24.3.3).

ВКР является также полезным спектроскопическим инструментом, с помощью которого можно выявить колебательные характеристики материала. Чувствительность спектроскопии комбинационного рассеяния можно повысить при использовании **когерентного антистоксова рассеяния света (КАРС)**, в котором используется два лазера накачки, разность частот которых попадает в

резонанс с колебательной частотой исследуемого материала, что увеличивает эффективность смешения волн.

В другом важном приложении оптические волокна можно использовать для генерации широкополосного света с помощью процессов комбинационного рассеяния. Накачка вызывает спонтанное комбинационное рассеяние, получившееся излучение затем усиливается за счет ВКР по мере распространения света вдоль волокна. При достаточной длине волокна и интенсивности накачки получающиеся компоненты испытывают новое и новое комбинационное рассеяние, в конце концов образуя широкополосное излучение (суперконтинуум, см. подразд. 22.5.3). Стабилизация процесса возможна с применением резонатора. Вынужденное рассеяние Мандельштама—Бриллюэна протекает аналогично, за исключением того, что в нем участвуют акустические, а не молекулярные колебания.

Рекомендуемая литература

КНИГИ ПО ФИЗИКЕ АТОМА, МОЛЕКУЛЫ И ТВЕРДОГО ТЕЛА

См. также списки рекомендованной литературы к гл. 14, 15 и 16.

Haroche S., Raimond J.-M. Exploring the Quantum. Oxford University Press, 2006.

Plischke M., Bergersen B. Equilibrium Statistical Physics. World Scientific, 3rd ed. 2006.

Demtröder W. Molecular Physics: Theoretical Principles and Experimental Methods. Wiley-VCH, 2005.

Tang C.L. Fundamentals of Quantum Mechanics: For Solid State Electronics and Optics. Cambridge University Press, 2005.

Liu G., Jacquier B., eds. Spectroscopic Properties of Rare Earths in Optical Materials. Springer-Verlag, 2005.

Feynman R.P., Leighton R.B., Sands M. The Feynman Lectures on Physics. Vol. 3: Quantum Mechanics, 1965; Vol. 1: Mainly Mechanics, Radiation and Heat, 1963. Addison-Wesley, 2nd ed. 2005.

Harrison W.A. Elementary Electronic Structure. World Scientific, 2004.

Cohen- Tannoudji C. Atoms in Electromagnetic Fields. World Scientific, 2nd ed. 2004.

Foot C.J. Atomic Physics. Oxford University Press, 2004.

Kittel C. Elementary Statistical Physics. Wiley, 1958; Dover, reissued, 2004.

Lorentz H.A. The Theory of Electrons and its Applications to the Phenomena of Light and Radiant Heat. Teubner, 1906; Dover, reissued, 2004.

Henderson B., Bartram R.H. Crystal-Field Engineering of Solid-State Laser Materials. Cambridge University Press, 2000.

Levine I.N. Quantum Chemistry. Prentice Hall, 5th ed. 1999.

Roundhill D.M., Fackler Jr. J.P., eds. Optoelectronic Properties of Inorganic Compounds. Plenum, 1999.

Reif F. Complete Statistical Physics. Vol. 5. Berkeley Physics Course, McGraw-Hill, 1998.

Powell R.C. Physics of Solid-State Laser Materials. Springer-Verlag, 1998.

Cohen- Tannoudji C., Dupont-Roc J., Grynberg G. Atom-Photon Interactions: Basic Processes and Applications. Wiley, 1992; paperback ed. 1998.

Suter D. The Physics of Laser-Atom Interactions. Cambridge University Press, 1997.

Kan F.-H., Gan F. Laser Materials. World Scientific, 1995.

Yokoyama H., Ujihara K., eds. Spontaneous Emission and Laser Oscillation in Microcavities. CRC Press, 1995.

Lipson S.G., Lipson H., Tannhauser D.S. Optical Physics. Cambridge University Press, paperback ed., 1995.

Morrison C.A. Crystal Fields for Transition-Metal Ions in Laser Host Materials. Springer-Verlag, 1992.

Harris D.C., Bertolucci M.D. Symmetry and Spectroscopy: An Introduction to Vibrational and Electronic Spectroscopy. Oxford University Press, 1978; Dover, reissued 1989.

Born M. Atomic Physics. Blackie & Son, 1935; 8th ed. 1969; Dover, reissued 1989.

Letokhov V.S., ed. Laser Spectroscopy of Highly Vibrationally Excited Molecules. Adam Hilger, 1989.

Allen L., Eberly J.H. Optical Resonance and Two-Level Atoms. Wiley, 1975; Dover, reissued 1987.

Eisberg R., Resnick R. Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles. Wiley, 2nd ed. 1985.

Breene Jr. R.G. Theories of Spectral Line Shape. Wiley, 1981.

Kittel C., Kroemer H. Thermal Physics. Freeman, 2nd ed. 1980.

ter Haar D. The Old Quantum Theory. Pergamon, 1967. (Contains English translations of key early papers by Planck, Einstein, Rutherford, and Bohr.)

Herzberg G. Electronic Spectra and Electronic Structure of Polyatomic Molecules. Van Nostrand Reinhold, 1966.

Livesey D.L. Atomic and Nuclear Physics. Blaisdell, 1966.

Slater J.C. Quantum Theory of Atomic Structure. Vol. 1. McGraw-Hill, 1960.

Dirac P.A.M. The Principles of Quantum Mechanics. Oxford University Press, 4th ed. 1958.

Landau L.D., Lifshitz E.M. Quantum Mechanics. Addison-Wesley, 1958.

Herzberg G. Molecular Spectra and Molecular Structure. Vol. 1: Spectra of Diatomic Molecules. Van Nostrand Reinhold, 2nd ed. 1950.

Herzberg G. Atomic Spectra and Atomic Structure. Prentice Hall, 1937; Dover, reissued 1944.

Condon E.U., Shortley G.H. The Theory of Atomic Spectra. Cambridge University Press, 1935.

КНИГИ ПО ЛАЗЕРНОМУ ОХЛАЖДЕНИЮ И УДЕРЖАНИЮ АТОМОВ

Letokhov V. Laser Control of Atoms and Molecules. Oxford University Press, 2007.

Ashkin A. Optical Trapping and Manipulation of Neutral Particles Using Lasers: A Reprint Volume with Commentaries. World Scientific, 2006.

Bardou F., Bouchaud J.-P., Aspect A., Cohen- Tannoudji C. Levy Statistics and Laser Cooling: How Rare Events Bring Atoms to Rest. Cambridge University Press, 2002.

Metcalf H.J., van der Straten P. Laser Cooling and Trapping. Springer-Verlag, 1999.

КНИГИ ПО ТЕРМОГРАФИИ, ЛЮМИНЕСЦЕНЦИИ И РАССЕЯНИЮ

Lakowicz J.R. Principles of Fluorescence Spectroscopy. Springer, 3rd ed. 2006.

Young F.R. Sonoluminescence. CRC Press, 2005.

Tsuji A., Maeda M., Matsumoto M., Kricka L.J., Stanley P.E., eds., Bioluminescence and Chemiluminescence: Progress and Perspectives (Proceedings of the 13th International Symposium). World Scientific, 2005.

Breitenstein O., Langenkamp M. Lock-in Thermography: Basics and Use for Functional Diagnostics of Electronic Components. Springer-Verlag, 2003.

Krupa J.-C., Kulagin N.A., eds. Physics of Laser Crystals. Kluwer, 2003.

Long D.A. The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules. Wiley, 2002.

Damzen M.J., Vlad V.I., Babin V., Mocofanescu A. Stilnulated Brillouin Scattering: Fundamentals and Applications. Institute of Physics, 2002.

Wheelon A.D. Electromagnetic Scintillation. Cambridge University Press, 2001.

Berne B.J., Pecora R. Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics. Wiley, 1976; Dover, reissued 2000.

Bohren C.F., Huffman D.R. Absorption and Scattering of Light by Small Particles. Wiley, 1983; paperback ed. 1998.

Johnson Jr. C.S., Gabriel D.A. Laser Light Scattering. Chapter 5 of Spectroscopy in Biochemistry (Vol. II). T.E. Bell, ed. CRC Press, 1981; Dover, reissued 1995.

СТАТЬИ

Wise F.W., ed. Selected Papers 011 Semiconductor Quantunl Dots. SPIE Optical Engineering Press (Milestone Series. Vol. 180), 2005.

Baldacchini T., LaFratta C.N., Farrer R.A., Teich M.C., Saleh B.E.A., Naughton M.J., Fourkas J.T Acrylic-Based Resin with Favorable Properties for Three-Dimensional Two-Photon Polymerization. Journal of Applied Physics. Vol. 95, 2004. P. 6072—6076.

Masters B.R., ed. Selected Papers on Multiphoton Excitation Microscopy. SPIE Optical Engineering Press (Milestone Series. Vol. 175), 2003.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Weber M.J., ed. Selected Papers on Phosphors, Light Emitting Diodes, and Scintillators: Applications of Photoluminescence, Cathodolulninescence, Electrolunlinescence and Radioluminescence. SPIE Optical Engineering Press (Milestone Series. Vol. 151), 1998.

Tans S.J., Devoret M.H., Dai H., Thess A., Smalley R.E., Geerligs L.J., Dekker C. Individual Single-Wall Carbon Nanotubes as Quantum Wires. Nature. Vol. 386, 1997. P. 474—477.

Ashoori R.C. Electrons In Artifical Atoms. Nature. Vol. 379, 1996. P. 413—419.

Putterman S.J. Sonoluminescence: Sound Into Light, Scientific Anlerican. Vol. 272, № 2, 1995. P. 46—51.

Kerker M., ed. Selected Papers on Light Scattering. SPIE Optical Engineering Press (Milestone Series. Vol. 4), 1988.

van Vleck J.H., Huber D.L. Absorption, Emission, and Linebreadths: A Semihistorical Perspective. Reviews of Modern Physics. Vol. 49, 1977. P. 939—959.

Weisskopf V.F. How Light Interacts with Matter. Scientific American. Vol. 219, № 3, 1968. P. 60—71.

Purcell E.M. Spontaneous Emission Probabilities at Radio Frequencies, Proceedings of the American Physical Society. Cambridge, MA. April 25—27, 1946, Abstract B 10, (Physical Review. Vol. 69, p. 681, June 1946).

Göppert-Mayer M. Über Elementarakte mit zwei Quantenspriingen. Annalen der Physik. Vol. 9, 1931. P. 273—294.

Einstein A. Zur Quantentheorie der Strahlung. Physikalische Zeitschrift. Vol. 18, 1917. P. 121—128. [Translation: On the Quantum Theory of Radiation, in D. ter Haar, The Old Quantum Theory. Pergamon, 1967.]

Задачи

К РАЗДЕЛУ 13.3

1. **Сравнение вынужденного и спонтанного излучения.** Атом с двумя энергетическими уровнями, переход между которыми обладает следующими характеристиками: $\lambda_0 = 0{,}7$ мкм, $t_{сп} = 3$ мс, $\Delta\nu = 50$ ГГц при лоренцевой форме линии,

помещен в резонатор объема $V = 100$ см^3 с показателем преломления $n = 1$. Возбуждены две моды, одна на центральной частоте ν_0, другая на частоте $\nu_0 + \Delta\nu$, в каждой из мод имеется по 1000 фотонов. Определите плотность вероятности вынужденного излучения (поглощения). Пусть N_2 — число атомов на возбужденном уровне 2. Определите время распада N_2 за счет спонтанного и вынужденного излучения. Сколько должно быть фотонов, чтобы скорость распада из-за вынужденного излучения была такой же, как и из-за спонтанного излучения?

2. **Спонтанное излучение в заданные моды.**

а. Рассмотрим кубический резонатор объемом 1 мкм^3, заполненный средой с показателем преломления $n = 1$. Каковы модовые числа (q_1, q_2, q_3) для двух мод с наименьшими частотами (см. гл. 10)? Покажите, что эти частоты равны 260 и 367 ТГц.

б. Рассмотрим единичный атом в резонаторе, не содержащем фотонов. Пусть $p_{\text{сп}_1}$ — плотность вероятности (с$^{-1}$) испускания фотона атомом в моду (2, 1, 1), а $p_{\text{сп}_2}$ — плотность вероятности спонтанного испускания фотона с частотой 367 ТГц. Определите отношение $p_{\text{сп}_2}/p_{\text{сп}_1}$.

К РАЗДЕЛУ 13.4

1. **Скоростные уравнения для широкополосного излучения.** Резонатор единичного объема содержит атомы с двумя энергетическими уровнями 1 и 2, переход между которыми имеет частоту ν_0 и ширину линии $\Delta\nu$. На нижнем и верхнем уровнях находится N_1 и N_2 атомов соответственно, а в каждой из мод, частоты которых лежат в широкой полосе вокруг ν_0, имеется среднее число фотонов $\bar{n}$. Потери фотонов за счет неидеального отражения от стенок резонатора характеризуются скоростью $1/\tau_p$. Предполагая, что безызлучательные переходы между уровнями 2 и 1 отсутствуют, напишите скоростные уравнения для N_2 и $\bar{n}$.

2. **Подавленное спонтанное излучение.** Рассмотрим гипотетический двумерный излучатель — черное тело (например, квадратную пластинку площадью A) в состоянии теплового равновесия при температуре T.

а. Определите плотность мод $M(\nu)$ и спектральную плотность энергии (т. е. энергию в частотном интервале от ν до $\nu + d\nu$ на единицу площади) испускаемого излучения $\rho(\nu)$ (см. разд. 10.3).

б. Найдите плотность вероятности спонтанного излучения $P_{\text{сп}}$ для атома, расположенного в резонаторе, допускающем испускание только в двух измерениях. Такой резонатор можно изготовить, например, с использованием фотонно-кристаллических всенаправленных отражателей, расположенных с обеих сторон слоя.

3. **Сравнение вкладов вынужденного и спонтанного испускания в излучение черного тела.** Найдите температуру полости при тепловом равновесии, спектральная плотность излучения которой $\rho(\nu)$, когда скорости спонтанного и вынужденного излучения атомов, находящихся в стенках полости, равны при $\lambda_0 = 1$ мкм.

4. **Закон Вина.** Выведите выражение для спектральной плотности энергии $\rho_\lambda(\lambda)$ [энергия на единицу объема в интервале длин волн от λ до $\lambda + d\lambda$ равна $\rho_\lambda(\lambda)d\lambda$]. Покажите, что длина волны $\lambda_\text{п}$, на которой спектральная плотность энергии $\rho_\lambda(\lambda)$

максимальна, удовлетворяет уравнению $5(1 - e^{-y}) = y$, где $y = hc/\lambda_{\text{п}} k_{\text{B}} T$. Последнее означает, что $\lambda_{\text{п}} T = \text{const}$ (закон Вина). Приближенно определите $\lambda_{\text{п}} T$ Покажите, что $\lambda_{\text{п}} \neq c/\nu_{\text{п}}$, где $\nu_{\text{п}}$ — частота, на которой максимум имеет функция $\rho(\nu)$ (см. упражнение 13.3). Объясните результат.

5. **Спектральная плотность энергии излучения одномерного черного тела.** Рассмотрим гипотетический одномерный излучатель — черное тело длиной L в состоянии теплового равновесия при температуре T.

а. Определите плотность мод $M(\nu)$ (число мод на единицу частоты и на единицу длины) в одном измерении.

б. Используя среднюю плотность $\bar{E}$ моды с частотой ν, определите спектральную плотность энергии $\rho(\nu)$ излучения черного тела (т. е. энергию в спектральном интервале от ν до $\nu + \mathrm{d}\nu$ на единицу длины). Постройте примерный график зависимости $\rho(\nu)$ от ν.

6. **Закон Стефана—Больцмана.** Используя спектральную плотность энергии излучения черного тела (13.67), убедитесь, что полная мощность, излучаемая черным телом, пропорциональна T^4 в соответствии с законом Стефана—Больцмана. Определите коэффициент пропорциональности.

Указание:

$$\int_0^\infty \frac{x^3 \mathrm{d}x}{e^x - 1} = \frac{\pi^4}{15}.$$

К РАЗДЕЛУ 13.5

1. **Статистика света при катодолюминесценции.** Рассмотрим пучок электронов, падающий на фосфо́р (люминесцирующее вещество, не путать с химическим элементом фосфором — *примеч. пер.*) электронно-лучевой трубки. Пусть $\bar{m}$ — среднее число электронов, ударяющихся о единичную площадку фосфора в единицу времени. Считая, что число m электронов, падающих за фиксированное время, случайно и подчиняется распределению Пуассона, а число излучаемых фотонов, приходящееся на один электрон, также имеет пуассоновское распределение, но со средним значением $\bar{G}$, определите полное распределение $p(n)$ излучаемых при катодолюминесценции фотонов. Результат известен как распределение Неймана А-типа. Найдите выражения для среднего $\bar{n}$ и дисперсии σ_n^2.

Указание. Используйте условную вероятность.

ГЛАВА 14 ЛАЗЕРНЫЕ УСИЛИТЕЛИ

Чарльз Таунс
(род. 1915)

Николай Басов
(1922—2001)

Александр Прохоров
(1916—2002)

Таунс, Басов и Прохоров разработали принцип усиления света за счет вынужденного испускания излучения (Light Amplification by Stimulated Emission of Radiation — LASER). За эту работу в 1964 г. они получили Нобелевскую премию

Когерентный оптический усилитель — прибор, увеличивающий амплитуду оптического поля с сохранением его фазы. Если оптическое поле на входе в такой усилитель монохроматическое, то на выходе оно также будет монохроматическим с той же частотой. Амплитуда на выходе увеличивается по отношению к входу, а фаза остается неизменной или сдвигается на фиксированную величину. В отличие от этого **некогерентный оптический усилитель** увеличивает интенсивность оптической волны без сохранения ее фазы.

Когерентные оптические усилители имеют ряд важных приложений, таких как усиление слабых оптических импульсов, например, прошедших большой путь по оптическому волокну, или формирование сверхмощных оптических импульсов, например, для лазерного термоядерного синтеза. Кроме того, важно понять принципы, лежащие в основе действия лазерных усилителей, в качестве необходимой вводной части к анализу оптических генераторов (лазеров) в гл. 15.

Основной принцип достижения когерентного усиления света заключается в усилении света за счет вынужденного испускания излучения. Английское

сокращенное название этого процесса (Light Amplification by Stimulated Emission of Radiation — LASER) породило ныне общепринятый во всем мире термин «лазер». При вынужденном излучении (см. разд. 13.3) фотон из некоторой моды вынуждает атом, находящийся в возбужденном состоянии, совершить переход на более низкий энергетический уровень и при этом испустить точную копию (клон) исходного фотона той же моды. Теперь уже два одинаковых фотона могут стимулировать испускание еще двух своих точных копий и так далее. Результатом является когерентное оптическое усиление. Поскольку вынужденное излучение происходит, только когда энергия фотона примерно равна разности энергий атомных уровней, процесс ограничен полосой частот, определяемой шириной линии атомного перехода.

Лазерное усиление по многим признакам отличается от электронного. Электронные усилители основаны на приборах, в которых малые изменения инжектируемого электрического тока или приложенного напряжения приводят к большим изменениям скорости или потока носителей заряда, таких как электроны и дырки в полупроводниковых полевых транзисторах или биполярных плоскостных транзисторах. Настраиваемые электронные усилители имеют резонансные колебательные контуры (например, из катушки индуктивности и конденсатора) или резонаторы (металлические полости), ограничивающие область усиления нужной полосой частот. В отличие от них атомные, молекулярные и твердотельные лазерные усилители осуществляют селекцию частоты за счет разности энергий рабочих уровней энергии. Атомные переходы служат естественными резонансными элементами, определяющими рабочую частоту и ширину полосы прибора. Для дополнительной более тонкой настройки часто используются оптические резонаторы.

Свет, проходящий через вещество в состоянии теплового равновесия, ослабляется. Это происходит потому, что поглощение света преобладающей частью атомов, занимающей нижние уровни энергии, превалирует над вынужденным испусканием, в котором участвует незначительная доля атомов, заселяющих верхние энергетические уровни. Важным условием достижения лазерного усиления является наличие большего числа атомов в верхнем энергетическом состоянии по сравнению с нижним. Эта ситуация является неравновесной, как ясно из разд. 13.2. Достижение такой инверсии заселенностей требует внешнего источника энергии, чтобы возбудить («накачать») верхний уровень, как показано на рис. 14.1. Хотя в данной главе мы будем пользоваться терминами «атомы», «атомные уровни», их следует понимать в более широком смысле как «активная среда» и «лазерные уровни энергии».

Свойства идеального оптического, или электронного когерентного, усилителя схематически показаны на рис. 14.2, *а*. Это линейная система, увеличивающая амплитуду входного сигнала на фиксированный множитель — коэффициент усиления. Синусоидальный входной сигнал порождает синусоидальный выходной сигнал той же частоты, но большей амплитуды. Коэффициент усиления идеального усилителя постоянен во всей полосе усиления. Усилитель может вносить фазовый сдвиг, линейно меняющийся с частотой, что соответствует

временной задержке выходного сигнала относительно входного сигнала (см. разд. Б.1 Приложения Б).

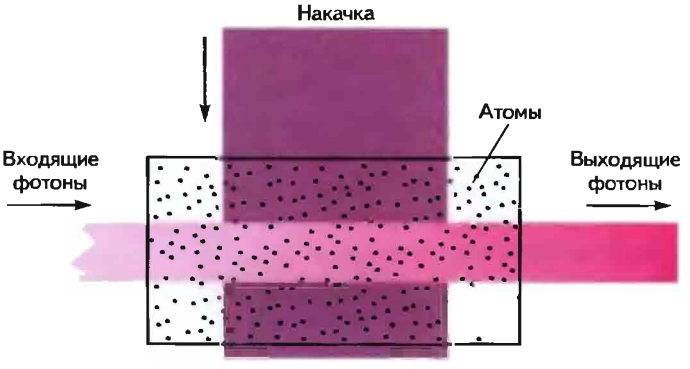


Рис. 14.1. Лазерный усилитель. Внешний источник энергии (накачка) возбуждает активную среду (показанную как ансамбль атомов), создавая инверсию заселенностей. Фотоны взаимодействуют с атомами. Когда вынужденное излучение начинает преобладать над поглощением, среда действует как когерентный усилитель

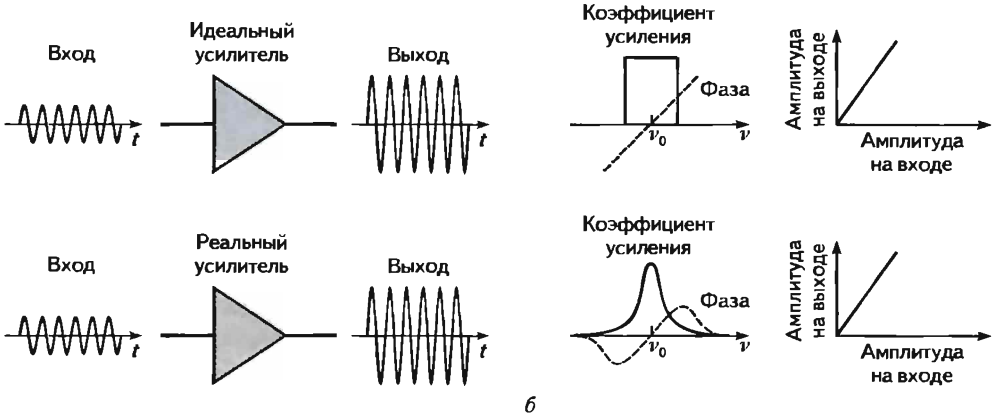


Рис. 14.2. Идеальный усилитель линеен. Он увеличивает амплитуду сигнала (частота которого попадает в полосу усиления) в постоянное число раз, возможно, внося линейный фазовый сдвиг (*а*). Реальный усилитель, как правило, имеет коэффициент усиления и фазовый сдвиг, зависящие от частоты, как показано на рисунке (*б*). Для больших входных сигналов выходной сигнал насыщается и усилитель проявляет нелинейность

Реальные когерентные усилители имеют коэффициент усиления и фазовый сдвиг, зависящие от частоты. Типичный вид такой зависимости показан на рис. 14.2, *б*. Коэффициент усиления и фазовый сдвиг определяют передаточную функцию усилителя. Для достаточно больших амплитуд входного сигнала реальные усилители, как правило, испытывают насыщение — форму нелинейного поведения, при котором амплитуда на выходе не возрастает пропорцио-

нально амплитуде на входе. За счет насыщения в выходном сигнале появляются гармоники, если ширина полосы усилителя достаточно велика, чтобы пропустить их. Реальные усилители также вносят шум, так что на выходе присутствует случайная флуктуирующая компонента, не зависящая от входного сигнала.

Таким образом, усилитель можно характеризовать следующими показателями:

- коэффициент усиления;
- ширина полосы;
- фазовый сдвиг;
- источник энергии;
- нелинейность и насыщение усиления;
- шум.

О данной главе

В данной главе мы обсудим перечисленные выше показатели в порядке очередности. В разд. 14.1 развита теория лазерного усиления, приводящая к выражениям для коэффициента усиления, ширины полосы и фазового сдвига. Механизмы накачки активной среды и создания инверсии заселенностей исследованы в разд. 14.2. Важные примеры лазерных усилителей рассмотрены в разд. 14.3. Разд. 14.4 и 14.5 посвящены нелинейности и шумам в процессе усиления. Данная глава опирается на материал гл. 13, особенно разд. 13.3.

14.1. ТЕОРИЯ ЛАЗЕРНОГО УСИЛЕНИЯ

Монохроматическая оптическая плоская волна, бегущая вдоль оси z, с частотой ν, электрическим полем $\mathcal{E}(z) = \mathrm{Re}\{E(z)\exp(j2\pi\nu t)\}$, комплексной амплитудой $E(z)$, интенсивностью

$$I(z) = \frac{|E(z)|^2}{2\eta}$$

и плотностью потока фотонов

$$\phi(z) = \frac{I(z)}{h\nu}$$

(число фотонов, проходящих в единицу времени через единичную площадку), будет взаимодействовать с атомной средой при условии, что атомы среды имеют два энергетических уровня, разность энергий между которыми примерно совпадает с энергией фотона $h\nu$. Числа атомов в единице объема на верхнем и нижнем энергетических уровнях обозначим N_2 и N_1 соответственно. Волна усиливается с коэффициентом усиления $\gamma(\nu)$ (на единицу длины) и испытывает фазовый сдвиг $\varphi(\nu)$ (на единицу длины). Перейдем к выводу выражений для $\gamma(\nu)$ и $\varphi(\nu)$. Положительный коэффициент $\gamma(\nu)$ соответствует усилению, отрицательный — ослаблению.

14.1.1. Коэффициент и ширина полосы усиления

Имеют место три вида взаимодействия фотонов с атомами (см. разд. 13.3). Если атом находится на нижнем энергетическом уровне, фотон может быть поглощен, а если на верхнем — может произойти клонирование фотона за счет вынужденного излучения. Эти два вида взаимодействия приводят к ослаблению и усилению соответственно. Третий вид взаимодействия — спонтанное излучение, при котором атом, находящийся на верхнем уровне, испускает фотон независимо от присутствия других фотонов, отвечает за шумы усилителя.

Плотность вероятности ($с^{-1}$) того, что невозбужденный атом поглотит один фотон, согласно (13.30) и (13.26), равна

$$W_i = \phi\sigma(\nu), \tag{14.1}$$

где $\sigma(\nu)$ — сечение перехода на частоте ν:

$$\sigma(\nu) = \frac{\lambda^2}{8\pi t_{сп}} g(\nu),$$

где $g(\nu)$ — нормированная функция формы линии; $t_{сп}$ — время жизни для спонтанного излучения; λ — длина волны света в среде. Плотность вероятности для вынужденного излучения такая же, как для поглощения.

Коэффициент усиления

Средняя плотность поглощенных фотонов (число фотонов на единицу времени и на единицу объема) равна $N_1 W_i$. Аналогично, средняя плотность клонированных фотонов, созданных в процессе вынужденного излучения, равна $N_2 W_i$. Тогда полное число фотонов, появившихся в единице объема в единицу времени, равно NW_i, где $N = N_2 - N_1$ — плотность разности заселенностей. Для удобства N часто называют просто разностью заселенностей. Если N положительна, то существует инверсия заселенностей, в случае которой среда может действовать как усилитель, и поток фотонов будет нарастать при прохождении через нее. Если N отрицательна, среда действует как ослабитель, и поток фотонов будет затухать. При $N = 0$ среда прозрачна.

Поскольку падающие фотоны распространяются вдоль оси z, вдоль той же оси будут распространяться и фотоны вынужденного излучения (рис. 14.3). Тогда внешняя накачка, создающая инверсию заселенностей ($N > 0$), вызывает рост плотности потока фотонов $\phi(z)$ с возрастанием z. Поскольку испущенные фотоны стимулируют дальнейшие акты испускания, рост их числа в любом месте z пропорционален их числу в этом месте, так что $\phi(z)$ растет экспоненциально.

Для явной демонстрации этого процесса рассмотрим бесконечно малый цилиндр длиной dz с единичной площадью основания, показанный на рис. 14.3. Если $\phi(z)$ и $\phi(z) + d\phi(z)$ — плотности потока фотонов, входящих в цилиндр и выходящих из него, то $d\phi(z)$ — плотность потока фотонов, испущенных внутри

цилиндра. Это приращение числа фотонов на единицу площади в единицу времени, $d\phi(z)$, — число фотонов, созданных в единичном объеме в единицу времени, NW_i, умноженное на высоту цилиндра dz:

$$d\phi = NW_i dz. \tag{14.2}$$

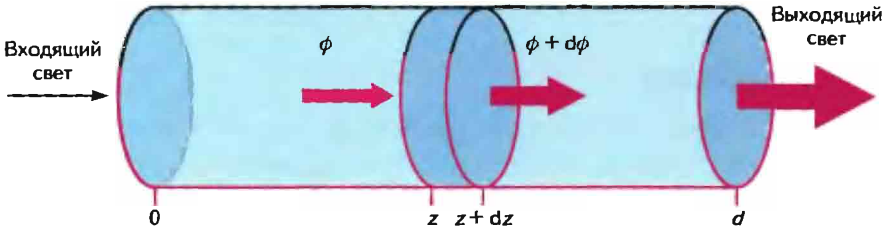


Рис. 14.3. Плотность потока фотонов ϕ [число фотонов/(см^2 · с)], входящих в бесконечно малый цилиндр, содержащий возбужденные атомы, увеличивается до $\phi + d\phi$ после прохождения расстояния dz

С помощью (14.1) можно записать (14.2) в виде дифференциального уравнения

$$\frac{d\phi(z)}{dz} = \gamma(\nu)\phi(z), \tag{14.3}$$

где

$$\boxed{\gamma(\nu) = N\sigma(\nu) = N\frac{\lambda^2}{8\pi t_{\text{сп}}}g(\nu).} \tag{14.4}$$

Коэффициент усиления

Коэффициент $\gamma(\nu)$ представляет полное усиление плотности потока фотонов на единицу длины среды. Решение (14.3) — экспоненциально нарастающая функция

$$\phi(z) = \phi(0)\exp[\gamma(\nu)z]. \tag{14.5}$$

Поскольку оптическая интенсивность равна $I(z) = h\nu\phi(z)$, (14.5) можно также записать для нее:

$$I(z) = I(0)\exp[\gamma(\nu)z]. \tag{14.6}$$

Таким образом, $\gamma(\nu)$ представляет собой также коэффициент усиления интенсивности на единицу длины среды.

Видно, что коэффициент усиления $\gamma(\nu)$ усилителя пропорционален разности заселенностей $N = N_2 - N_1$. Хотя в приведенном примере разность N была взята положительной, вывод не зависит от знака N. В отсутствие инверсии заселенностей N отрицательна ($N_2 < N_1$), отрицателен и коэффициент усиле-

ния. Среда будет ослаблять, а не усиливать свет, распространяющийся вдоль оси z, в соответствии с законом экспоненциального затухания

$$\phi(z) = \phi(0) \exp[-\alpha(\nu)z],$$

где коэффициент ослабления $\alpha(\nu) = -\gamma(\nu) = -N\sigma(\nu)$. Следовательно, среда, находящаяся в состоянии теплового равновесия, не может обеспечить лазерного усиления.

Усиление

Для области взаимодействия, полная длина которой d (см. рис. 14.3), полное усиление лазерного усилителя $G(\nu)$ определяется как отношение плотности потока фотонов на выходе к плотности потока фотонов на входе:

$$G(\nu) = \frac{\phi(d)}{\phi(0)},$$

так что

$$\boxed{G(\nu) = \exp[\gamma(\nu)d].} \qquad (14.7)$$

Усиление

Ширина полосы

Зависимость коэффициента усиления $\gamma(\nu)$ от частоты падающего света ν содержится в функции формы линии $g(\nu)$, как видно из (14.4). Последняя имеет вид пика шириной $\Delta\nu$ с центром на частоте атомного резонанса

$$\nu_0 = \frac{E_2 - E_1}{h},$$

где E_2 и E_1 — энергия атомных состояний. Следовательно, лазерный усилитель — резонансное устройство, резонансная частота и ширина полосы которого определяются функцией формы линии атомного перехода. Это происходит потому, что поглощение и вынужденное испускание определяются свойствами атомного перехода. Ширина линии $\Delta\nu$ измеряется либо в шкале частот (Гц), либо в шкале длин волн (нм). Эта ширина линии связана соотношением

$$\Delta\lambda = \left|\Delta\frac{c_0}{\nu}\right| = +\frac{c_0}{\nu^2}\Delta\nu = \frac{\lambda_0^2}{c_0}\Delta\nu.$$

Так, ширина линии $\Delta\nu = 1$ ТГц при $\lambda_0 = 0{,}6$ мкм соответствует $\Delta\lambda = 1{,}2$ нм.

Если, например, функция формы линии лоренцева, то (13.45) дает

$$g(\nu) = \frac{\Delta\nu/2\pi}{(\nu - \nu_0)^2 + (\Delta\nu/2)^2}. \qquad (14.8)$$

Тогда коэффициент усиления также имеет лоренцеву форму с той же шириной, т. е.

$$\gamma(\nu) = \gamma(\nu_0)\frac{(\Delta\nu/2)^2}{(\nu-\nu_0)^2+(\Delta\nu/2)^2}, \tag{14.9}$$

показанную на рис. 14.4, где $\gamma(\nu_0)$ — коэффициент усиления на центральной частоте ν_0:

$$\gamma(\nu_0) = N\frac{\lambda^2}{4\pi^2 t_{\text{сп}}\Delta\nu}.$$

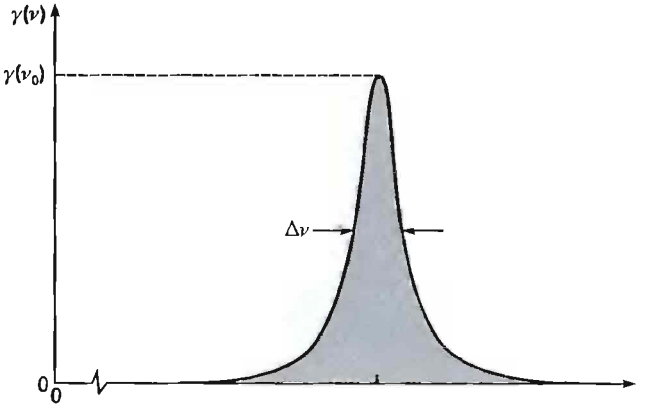


Рис. 14.4. Коэффициент усиления $\gamma(\nu)$ резонансного лазерного усилителя с лоренцевой формой линии

Упражнение 14.1

Ослабление и усиление в лазерном усилителе на рубине

а. Рассмотрим кристалл рубина с двумя энергетическими уровнями, разделенными разностью энергии, соответствующей длине волны $\lambda_0 = 694{,}3$ нм в свободном пространстве, с лоренцевой формой линии, имеющей ширину $\Delta\nu = 330$ ГГц. Время жизни спонтанного излучения $t_{\text{сп}} = 3$ мс, показатель преломления рубина $n = 1{,}76$. При $N_2 = N_a = 10^{22}$ см$^{-3}$ определите разность заселенностей $N = N_2 - N_1$ и коэффициент ослабления $\alpha(\nu_0)$ в центре линии в условиях теплового равновесия при $T = 300$ °K (справедливо распределение Больцмана, обсуждавшееся в разд. 13.2).

б. Какое значение должна принять разность заселенностей N, чтобы достичь коэффициента усиления $\gamma(\nu_0) = 0{,}5$ см$^{-1}$ на центральной частоте?

в. Какой длины должен быть кристалл, чтобы полное усиление на центральной частоте равнялось 4, когда $\gamma(\nu_0) = 0{,}5$ см$^{-1}$?

14.1.2. Фазовый сдвиг

Поскольку усиление резонансной среды зависит от частоты, среда обладает дисперсией (см. разд. 5.5) и с ее усилением должен быть связан зависящий от частоты фазовый сдвиг. Фазовый сдвиг, вносимый лазерным усилителем, можно определить, рассматривая взаимодействие света с веществом с помощью электрического поля, а не плотности потока фотонов или интенсивности, как мы делали до сих пор.

Перейдем к альтернативному подходу, в котором математические свойства системы, вытекающие из принципа причинности используются для определения фазового сдвига. Для однородно-уширенных сред коэффициент фазового сдвига $\varphi(\nu)$ (сдвиг фазы на единицу длины усиливающей среды) связан с коэффициентом усиления $\gamma(\nu)$ преобразованием Гильберта (см. разд. Б.1 приложения Б), так что знание $\gamma(\nu)$ на всех частотах однозначно определяет $\varphi(\nu)$.

Оптическая интенсивность и комплексная амплитуда поля связаны формулой $I(z) = |E(z)|^2/2\eta$. Поскольку $I(z) = I(0)\exp[\gamma(\nu)z]$ в соответствии с (14.6), комплексная амплитуда поля удовлетворяет соотношению

$$E(z) = E(0)\exp\left[\frac{1}{2}\gamma(\nu)z\right]\exp\left[-j\varphi(\nu)z\right], \tag{14.10}$$

где $\varphi(\nu)$ — коэффициент фазового сдвига. Следовательно, комплексная амплитуда поля в точке $z + \Delta z$ равна

$$\begin{aligned} E(z+\Delta z) &= E(0)\exp\left[\frac{1}{2}\gamma(\nu)(z+\Delta z)\right]\exp\left[-j\varphi(\nu)(z+\Delta z)\right] \\ &= E(z)\exp\left[\frac{1}{2}\gamma(\nu)\Delta z\right]\exp\left[-j\varphi(\nu)\Delta z\right] \approx E(z)\left[1+\frac{1}{2}\gamma(\nu)\Delta z - j\varphi\right. \end{aligned} \tag{14.11}$$

где экспоненциальные функции разложены в ряды Тейлора. Следовательно, приращение электрического поля $\Delta E(z) = E(z + \Delta z) - E(z)$ удовлетворяет уравнению

$$\frac{\Delta E(z)}{\Delta z} = E(z)\left[\frac{1}{2}\gamma(\nu) - j\varphi(\nu)\right]. \tag{14.12}$$

Такой бесконечно тонкий усилитель можно рассматривать как линейную систему с входным и выходным сигналами $E(z)$ и $\Delta E(z)/\Delta z$ соответственно и передаточной функцией

$$H(\nu) = \frac{1}{2}\gamma(\nu) - j\varphi(\nu). \tag{14.13}$$

Поскольку бесконечно тонкий усилитель — физическая система, он должен подчиняться принципу причинности. Но действительная и мнимая части передаточной функции линейной причинной системы связаны преобразованием Гильберта (см. разд. Б.1 приложения Б). Отсюда следует, что $-\varphi(\nu)$ является

результатом преобразования Гильберта функции $(1/2)\gamma(\nu)$, так что функция сдвига фазы усилителя определяется коэффициентом усиления.

Простым примером является атомная функция формы линии Лоренца малой ширины $\Delta\nu \ll \nu_0$, для которой коэффициент усиления $\gamma(\nu)$ дается формулой (14.9). Соответствующий коэффициент сдвига фазы $\varphi(\nu)$ дается формулой (Б.13) разд. Б.1:

$$\boxed{\varphi(\nu) = \frac{\nu - \nu_0}{\Delta\nu}\gamma(\nu).} \qquad (14.14)$$

Коэффициент фазового сдвига (лоренцева форма линии)

Графики лоренцевых коэффициентов усиления и фазового сдвига как функций частоты представлены на рис 14.5; при резонансе коэффициент усиления максимален, а коэффициент фазового сдвига равен нулю. Коэффициент фазового сдвига отрицателен на частотах ниже резонанса и положителен на частотах выше резонанса.

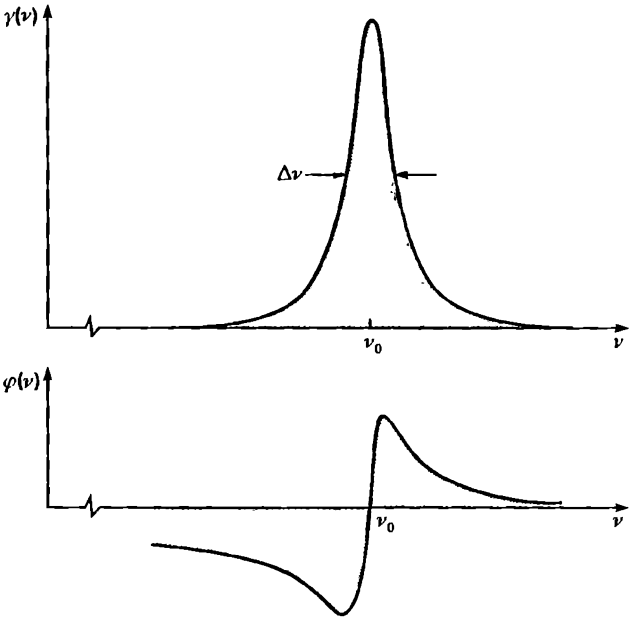


Рис. 14.5. Коэффициенты усиления $\gamma(\nu)$ и фазового сдвига $\varphi(\nu)$ для лазерного усилителя с лоренцевой функцией формы линии

14.2. НАКАЧКА УСИЛИТЕЛЯ

Как и все усилители, лазерный усилитель требует внешнего источника энергии, чтобы обеспечить приращение энергии входного сигнала. Накачка поставляет эту энергию посредством механизмов, обеспечивающих возбуждение электронов в атомах и заставляющих их совершать переходы с нижних энергетических состояний в верхние состояния. Чтобы обеспечить усиление, накачка должна таким образом создать инверсию заселенностей в рабочем переходе. Однако механизмы накачки часто включают использование дополни-

тельных энергетических уровней. Например, накачка атомов с уровня 1 на уровень 2 с целью достичь усиления на переходе 2 → 1 наиболее легко осуществляется посредством возбуждения атомов на уровень 3, откуда затем происходит заселение нижележащего уровня 2 в результате естественных процессов распада.

Накачка может быть оптической (например, от лампы-вспышки или лазера), электрической (например, от газового разряда, электронного пучка, ионного пучка, инжекции носителей заряда) или химической (например, от пламени или химической реакции, в которой конечный продукт оказывается в возбужденном состоянии). Для **непрерывного действия** скорости возбуждения и распада различных энергетических уровней, участвующих в процессе, должны быть сбалансированы для поддержания стационарной инверсии заселенностей на переходе 2 → 1.

14.2.1. Скоростные уравнения

Уравнения, описывающие скорости изменения плотностей заселенности N_1 и N_2 в результате накачки, а также радиационных и безызлучательных переходов, называются **скоростными уравнениями**, или уравнениями баланса. Они похожи на уравнения разд. 13.4, однако теперь частью процесса является селективная внешняя накачка, поэтому условия теплового равновесия не выполняются.

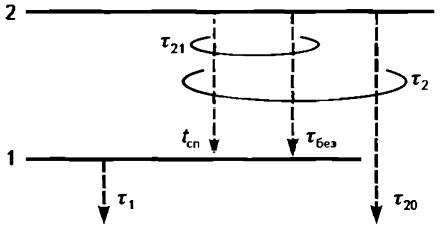


Рис. 14.6. Энергетические уровни 1 и 2 и их времена распада

Рассмотрим схему энергетических уровней на рис. 14.6. Сосредоточим внимание на уровнях 1 и 2, полные времена жизни которых, связанные со спонтанными переходами на более низкие уровни, равны τ_1 и τ_2 соответственно. Время жизни уровня 2 состоит из двух вкладов, один из которых вносят переходы на уровень 1 (τ_{21}), а другой (τ_{20}) — на все остальные уровни. Когда возможно несколько каналов распада, полная скорость перехода определяется как сумма составляющих скоростей. Поскольку скорости обратно пропорциональны временам жизни, имеем

$$\tau_2^{-1} = \tau_{21}^{-1} + \tau_{20}^{-1}. \tag{14.15}$$

Увеличение числа способов распада сокращает полное время жизни (так как распад становится более быстрым). Кроме вклада спонтанного радиационного затухания (с постоянной времени $t_{сп}$) в τ_{21} может присутствовать вклад безызлучательных процессов $\tau_{без}$, например столкновений атома со стенками сосуда, что приводит к депопуляции уровня 2, так что

$$\tau_{21}^{-1} = t_{сп}^{-1} + \tau_{без}^{-1}. \tag{14.16}$$

Если системе без накачки, как на рис. 14.6, предоставить возможность достичь стационарного состояния, то заселенности N_1 и N_2 исчезнут, так как, в конце концов, электроны перейдут на более низкие энергетические уровни.

Стационарные заселенности уровней 1 и 2 можно поддерживать, если непрерывно заселять энергетические уровни, лежащие над уровнем 2, распад которых будет приводить к заселению уровня 2, как показано на более реалистичной схеме уровней (рис. 14.7). За счет переходов с участием других уровней накачка обедняет уровень 1 и заселяет уровень 2 со скоростями R_1 и R_2 (на единицу объема в секунду) соответственно (рис. 14.8). В результате уровни 1 и 2 приобретают ненулевые стационарные заселенности.

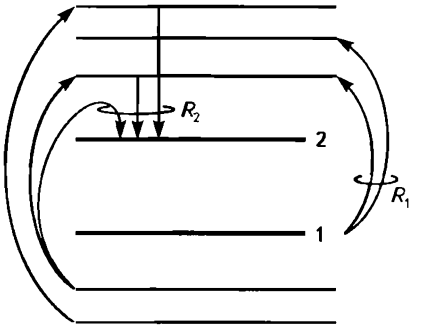


Рис. 14.7. Энергетические уровни 1 и 2 вместе с окружающими более высокими и более низкими уровнями энергии в присутствии накачки

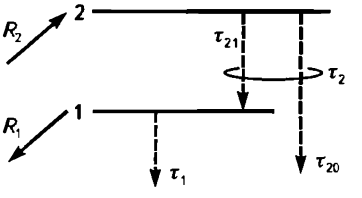


Рис. 14.8. Энергетические уровни 1 и 2 и их времена распада. Посредством накачки плотность заселенности уровня 2 увеличивается со скоростью R_2, а плотность заселенности уровня уменьшается со скоростью R_1

Перейдем к записи скоростных уравнений для этой системы как в отсутствие, так и в присутствии усиливаемого излучения (излучения, попадающего в резонанс с переходом $2 \rightarrow 1$).

Скоростные уравнения в отсутствие усиливаемого излучения

Скорости возрастания плотностей заселенности уровней 1 и 2 под действием накачки и распада равны, соответственно:

$$\frac{dN_2}{dt} = R_2 - \frac{N_2}{\tau_2}; \tag{14.17}$$

$$\frac{dN_1}{dt} = -R_1 - \frac{N_1}{\tau_1} + \frac{N_2}{\tau_{21}}, \tag{14.18}$$

В стационарных условиях ($dN_1/dt = dN_2/dt = 0$) можно решить уравнения (14.17) и (14.18) относительно N_1 и N_2 и определить разность заселенностей $N = N_2 - N_1$.

Результат имеет вид

$$N_0 = R_2\tau_2\left(1 - \frac{\tau_1}{\tau_{21}}\right) + R_1\tau_1, \qquad (14.19)$$

Стационарная разность заселенностей (усиливаемое излучение отсутствует)

где символ N_0 обозначает стационарную разность заселенностей N в отсутствие усиливаемого излучения.

В соответствии с (14.4) для большого коэффициента усиления требуется большая разность заселенностей, т. е. большое положительное значение N_0. Формула (14.19) показывает, что этого можно достичь за счет:

- больших R_1 и R_2;
- большого τ_2 (при этом $t_{сп}$, вносящее вклад в τ_2 через τ_{21}, должно быть достаточно малым, чтобы скорость радиационного перехода была большой, как будет видно ниже);
- малого τ_1, если $R_1 < (\tau_2/\tau_{21})R_2$.

Физические соображения, лежащие в основе этих условий, понятны. Верхний уровень должен сильно заселяться и медленно распадаться, чтобы поддерживать свою заселенность. Нижний уровень должен эффективно опустошаться и быстро терять заселенность. В идеале желательно иметь $\tau_{21} \approx t_{сп} \ll \tau_{20}$, так чтобы $\tau_2 \approx t_{сп}$ и $\tau_1 \ll t_{сп}$. При этих условиях (14.19) упрощается к виду

$$N_0 \approx R_2 t_{сп} + R_1\tau_1. \qquad (14.20)$$

В отсутствие откачки атомов с уровня 1 ($R_1 = 0$) или при $R_1 \ll (t_{сп}/\tau_1)R_2$ этот результат еще более упрощается:

$$N_0 \approx R_2 t_{сп}. \qquad (14.21)$$

Упражнение 14.2

Оптическая накачка

Предположим, что $R_1 = 0$ и R_2 реализуется путем возбуждения атомов из основного состояния $E = 0$ на уровень 2 с использованием фотонов частоты E_2/h, поглощаемых с вероятностью перехода W. Допустим, что $\tau_2 \approx t_{сп}$ и $\tau_1 \ll t_{сп}$, так что в стационарном состоянии $N_1 \approx 0$ и $N_0 \approx R_2 t_{сп}$. Если N_a — полная заселенность уровней 0, 1 и 2, покажите, что $R_2 \approx (N_a - 2N_0)W$, так что разность заселенностей

$$N_0 \approx \frac{N_a t_{сп} W}{1 + 2t_{сп} W}.$$

Скоростные уравнения при наличии усиливаемого излучения

Присутствие излучения вблизи резонансной частоты ν_0 делает возможными переходы между уровнями 2 и 1 с вынужденным излучением, а также поглощением фотонов. Эти процессы характеризуются плотностью вероятнос-

ти $W_i = \phi\sigma(\nu)$, см. формулу (14.1), и показаны на рис. 14.9. Скоростные уравнения (14.17) и (14.18) нужно расширить, чтобы включить в рассмотрение этот источник изменений заселенности обоих уровней

$$\frac{dN_2}{dt} = R_2 - \frac{N_2}{\tau_2} - N_2 W_i + N_1 W_i; \tag{14.22}$$

$$\frac{dN_1}{dt} = -R_1 - \frac{N_1}{\tau_1} + \frac{N_2}{\tau_{21}} + N_2 W_i - N_1 W_i. \tag{14.23}$$

Плотность заселенности уровня 2 уменьшается за счет вынужденного излучения с переходом $2 \to 1$ и увеличивается за счет поглощения с переходом $1 \to 2$. Спонтанное излучение учтено в τ_{21}.

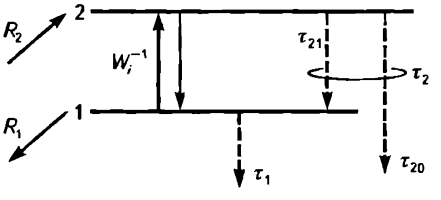


Рис. 14.9. Плотности заселенностей N_1 и N_2 ($см^{-3} \cdot с^{-1}$) энергетических уровней атомов среды 1 и 2 определяются тремя процессами: распадом (со скоростями $1/\tau_1$ и $1/\tau_2$, соответственно, включающими эффект спонтанного излучения), накачкой и откачкой (со скоростями R_2 и R_1 соответственно) и, наконец, поглощением и вынужденным излучением (со скоростью W_i и соответствующей постоянной времени $1/W_i$)

В стационарном режиме ($dN_1/dt = dN_2/dt = 0$) уравнения (14.22) и (14.23) легко решаются относительно N_1 и N_2. Для разности заселенностей $N = N_2 - N_1$ получаем

$$\boxed{N = \frac{N_0}{1 + \tau_s W_i},} \tag{14.24}$$

Стационарная разность заселенностей (при наличии усиливаемого излучения)

где N_0 — стационарная разность заселенностей в отсутствие усиливаемого излучения, определяемая выражением (14.19). Характерное время τ_s, которое всегда положительно, так как $\tau_2 \le \tau_{21}$, определяется выражением

$$\boxed{\tau_s = \tau_2 + \tau_1\left(1 - \frac{\tau_2}{\tau_{21}}\right).} \tag{14.25}$$

Время насыщения

В отсутствие усиливаемого излучения $W_i = 0$, так что (14.24) дает $N = N_0$, как и ожидалось. Поскольку τ_s положительно, стационарная разность заселенностей при наличии усиливаемого излучения всегда имеет меньшее значение, чем в его отсутствие, т. е. $|N| \le |N_0|$. Если излучение достаточно слабо, так что $\tau_s W_i \ll 1$ (**приближение слабого сигнала**), то можно положить $N \approx N_0$. По мере того, как усиливаемое излучение становится сильнее, W_i растет и в конечном

счете $N \to 0$ независимо от первоначального знака N_0, как показано на рис. 14.10. Это происходит потому, что вынужденное излучение и поглощение преобладают при очень больших W_i и имеют одинаковые плотности вероятности. Очевидно, что даже очень сильное излучение не может превратить отрицательную разность заселенностей в положительную и наоборот. Величина τ_s играет роль **времени насыщения**, что очевидно из рис 14.10.

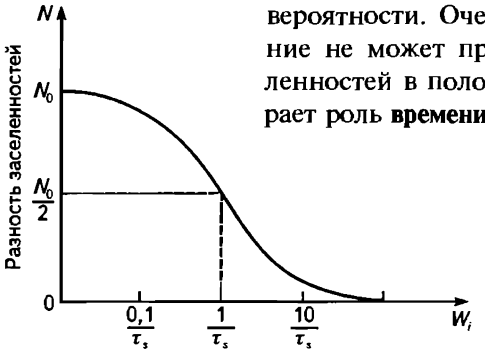


Рис. 14.10. Уменьшение стационарной разности заселенностей $N = N_2 - N_1$ по мере роста скорости поглощения и вынужденного излучения W_i. Когда $W_i = 1/\tau_s$, N снижается в два раза по сравнению со своим значением при $W_i = 0$

Упражнение 14.3

Время насыщения

Покажите, что если $t_{сп} \ll \tau_{без}$ (безызлучательная часть времени жизни τ_{21} перехода $2 \to 1$), $t_{сп} \ll \tau_{20}$ и $t_{сп} \gg \tau_1$, то $\tau_s \approx t_{сп}$.

14.2.2. Схемы накачки

Перейдем к изучению конкретных (трех- и четырехуровневых) схем накачки, применяемых на практике для получения инверсии заселенностей. Целью этих схем является использование такого процесса возбуждения, который увеличивает число атомов, заселяющих уровень 2, и уменьшает число атомов на уровне 1.

Четырехуровневая накачка

В схеме, показанной на рис. 14.11, уровень 1 лежит выше уровня 0 основного состояния. В состоянии теплового равновесия уровень 1 почти не заселен при условии, что $E_1 \gg k_B T$, — ситуация, несомненно, желательная. Накачка осуществляется на уровень (или несколько уровней), лежащих выше уровня 2; они обозначены как уровень 3. Переход $3 \to 2$ имеет малое время жизни (распад происходит быстро), так что накопление атомов на уровне 3 невелико. По причинам, которые выясняются в задаче 1 к разд. 14.2, двухуровневая схема накачки невозможна, поэтому уровень 2 заселяется через уровень 3, а не напрямую. Уровень 2 — долгоживущий, поэтому на нем накапливаются атомы, а уровень 1 — короткоживущий и быстро опустошается. Так устанавливается инверсия заселенностей между уровнями 2 и 1. В общей сложности используются четыре уровня, однако нужное для усиления оптическое взаимодействие имеет место лишь между уровнями 2 и 1.

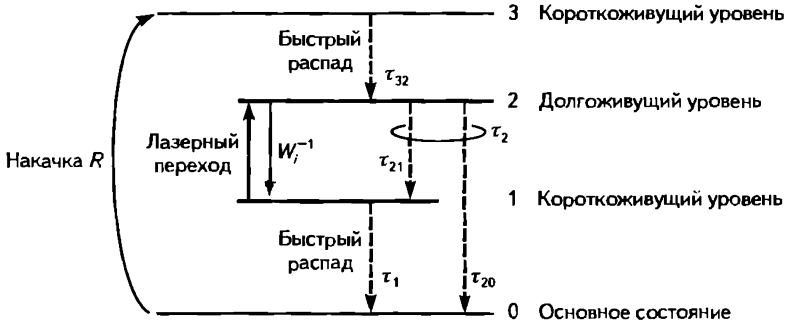


Рис. 14.11. Уровни энергии и скорости распада для четырехуровневой схемы. Четыре уровня выделены из множества уровней энергии, не показанных на рисунке. Предполагается, что скорость накачки на уровень 3 и откачки с уровня 0 одна и та же

Внешний источник энергии (например, фотоны с энергией E_3/h) накачивает атомы с уровня 0 на уровень 3 со скоростью R. Если распад в канале 3—2 происходит сравнительно быстро, то его можно считать мгновенным, тогда накачка уровня 3 эквивалентна накачке уровня 2 со скоростью $R_2 = R$. Тогда ситуация та же, что и показанная на рис. 14.9, и выражения (14.24) и (14.25) можно применять. Однако в этой конфигурации накачка не действует на атомы, находящиеся на уровне 1, поэтому $R_1 = 0$. Таким образом, в отсутствие усиливаемого излучения ($W_i = \phi = 0$) стационарная разность заселенностей определяется формулой (14.19) с $R_1 = 0$, т. е.

$$N_0 = R\tau_2\left(1 - \frac{\tau_1}{\tau_{21}}\right). \tag{14.26}$$

В большинстве четырехуровневых систем безызлучательным распадом в канале $2 \to 1$ можно пренебречь ($t_{\text{сп}} \ll \tau_{\text{без}}$) и $\tau_{20} \gg t_{\text{сп}} \gg \tau_1$ (см. упражнение 14.16), так что

$$N_0 \approx Rt_{\text{сп}}; \tag{14.27}$$

$$\tau_s \approx t_{\text{сп}}, \tag{14.28}$$

и, следовательно,

$$N \approx \frac{Rt_{\text{сп}}}{1 + t_{\text{сп}}W_i}. \tag{14.29}$$

В проведенном выводе явно использовалось предположение о независимости скорости накачки R от разности заселенностей $N = N_2 - N_1$. Это не всегда так, потому что плотности заселенностей основного состояния N_g и уровня 3, N_3, связаны с N_1 и N_2 соотношением

$$N_g + N_1 + N_2 + N_3 = N_a, \tag{14.30}$$

где полная плотность числа атомов N_a постоянна. Если процесс накачки включает переход между основным состоянием и уровнем 3 с вероятностью перехода W, то $R = (N_g - N_3)W$. Если уровни 3 и 1 короткоживущие, то $N_1 \approx N_3 \approx 0$, откуда $N_g + N_2 \approx N_a$, так что $N_g \approx N_a - N_2 \approx N_a - N$.

При этих условиях скорость накачки можно приближенно выразить как

$$R \approx (N_a - N)W, \tag{14.31}$$

откуда видно, что скорость накачки — линейно возрастающая функция разности заселенностей N и, очевидно, не является независимой от нее. Это происходит потому, что инверсия заселенностей, устанавливающаяся между уровнями 2 и 1, уменьшает число атомов, доступных накачке. Подставляя (14.31) в (14.28) и преобразуя, находим

$$N \approx \frac{t_{\text{сп}} N_a W}{1 + t_{\text{сп}} W + t_{\text{сп}} W_i}. \tag{14.32}$$

Наконец, разность заселенностей можно записать в виде (14.24),

$$N = \frac{N_0}{1 + \tau_s W_i}, \tag{14.33}$$

где, однако, N_0 и τ_s выражаются не формулами (14.27) и (14.28), а следующим образом:

$$\boxed{N_0 \approx \frac{t_{\text{сп}} N_a W}{1 + t_{\text{сп}} W}}; \tag{14.34}$$

$$\boxed{\tau_s \approx \frac{t_{\text{сп}}}{1 + t_{\text{сп}} W}}. \tag{14.35}$$

Для слабой накачки ($W \ll 1/t_{\text{сп}}$) $N_0 \approx t_{\text{сп}} N_a W$ пропорционально плотности вероятности перехода накачки W и $\tau_s \approx t_{\text{сп}}$, так что соотношения (14.27) и (14.28) справедливы. Однако по мере роста накачки N_0 уменьшается и в конце концов насыщается, в то время как τ_s уменьшается.

Трехуровневая накачка

В трехуровневой схеме накачки основное состояние ($E_1 = 0$) используется в качестве нижнего лазерного уровня 1, как показано на рис. 14.12. В схему опять входит вспомогательный третий уровень (обозначенный цифрой 3), причем распад в канале $3 \rightarrow 2$ происходит быстро и атомы не накапливаются на уровне 3. В канале $3 \rightarrow 1$, напротив, распад медленный ($\tau_{32} \ll \tau_{31}$), так что накачка служит для заселения верхнего лазерного уровня 2, который является долгоживущим и на котором накапливаются атомы. Накачка атомов на уровень 3 (например, путем поглощения фотонов с частотой E_3/h) происходит со скоростью R, быстрый безызлучательный распад эффективно заселяет уровень 2 со скоростью $R_2 = R$. Тепловая заселенность уровня 2 считается пренебрежимо малой.

Нетрудно видеть, что при быстром распаде 3 → 2 трехуровневая схема, представленная на рис. 14.12, является частным случаем схемы, показанной на рис. 14.9 (при условии, что R не зависит от N) с параметрами:

$$R_1 = R_2 = R; \quad \tau_1 = \infty; \quad \tau_2 = \tau_{21}. \tag{14.36}$$

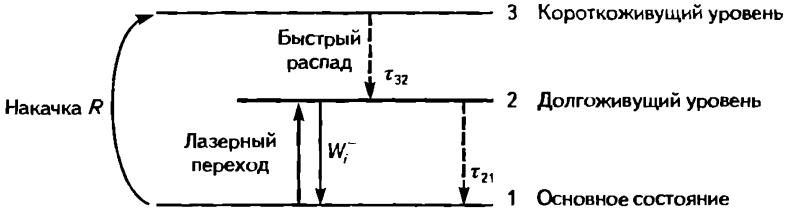


Рис. 14.12. Уровни энергии и скорости распада для трехуровневой схемы. Существует еще множество уровней, но они не относятся к проводимому рассмотрению. Предполагается, что скорость накачки на уровень 3 и откачки с уровня 1 одна и та же

Чтобы избежать алгебраических проблем, связанных с бесконечностью τ_1, вместо подстановки этих значений в (14.24) и (14.25) вернемся к исходным скоростным уравнениям (14.22) и (14.23). В стационарном состоянии оба эти уравнения дают одинаковый результат

$$0 = R - \frac{N_2}{\tau_{21}} - N_2 W_i + N_1 W \tag{14.37}$$

Невозможно определить и N_1, и N_2 из одного уравнения, связывающего их. Однако знание полной плотности числа атомов N_a в системе (на уровнях 1, 2 и 3) дает дополнительное условие, позволяющее определить N_1 и N_2. Поскольку время τ_{32} очень мало, уровень 3 сохраняет пренебрежимо малую стационарную заселенность; все атомы, попадающие на него, немедленно переходят на уровень 2. Таким образом,

$$N_1 + N_2 = N_a, \tag{14.38}$$

что дает возможность решить (14.37) относительно N_1 и N_2 и определить разность заселенностей $N = N_2 - N_1$ и время насыщения τ_s. Результат можно представить в обычном виде (14.24),

$$N = \frac{N_0}{1 + \tau_s W_i},$$

где на этот раз

$$N_0 = 2R\tau_{21} - N_a; \tag{14.39}$$

$$\tau_s = 2\tau_{21}. \tag{14.40}$$

Если безызлучательным распадом в канале $2 \to 1$ можно пренебречь ($t_{сп} \ll \tau_{без}$), то τ_{21} можно заменить на $t_{сп}$, после чего

$$N_0 \approx 2Rt_{сп} - N_a; \tag{14.41}$$

$$\tau_s \approx 2t_{сп}. \tag{14.42}$$

Интересно сравнить эти результаты с аналогичными выражениями (14.27) и (14.28) для четырехуровневой схемы. Достижение инверсии заселенностей ($N > 0$ и, следовательно, $N_0 > 0$) в трехуровневой схеме требует скорости накачки $R > N_a/2t_{сп}$. Таким образом, даже для того, чтобы сделать заселенности N_1 и N_2 равными (т. е. $N = 0$), требуется значительная мощность накачки $E_3 N_a/2t_{сп}$. Большая заселенность основного состояния (которое является нижним лазерным уровнем) является препятствием к достижению инверсии заселенностей, внутренне присущим трехуровневой схеме, но отсутствующим в четырехуровневой схеме, где нижний уровень лазерного перехода в нормальном состоянии пуст, так как время τ_1 мало. Время насыщения $\tau_s \approx t_{сп}$ для четырехуровневой схемы равно половине времени насыщения для трехуровневой схемы. И снова, как показано в задаче 1 к разд. 14.2, стационарная инверсия заселенностей не может быть получена прямой оптической накачкой перехода между уровнями 1 и 2.

Зависимость скорости накачки R от разности заселенностей N можно учесть при анализе трехуровневой схемы, если записать

$$R = (N_1 - N_3)W, \quad N_3 \approx 0;$$

$$N_1 = \frac{1}{2}(N_a - N),$$

откуда

$$R = \frac{1}{2}(N_a - N)W$$

Подставляя это в основное уравнение

$$N = \frac{2Rt_{сп} - N_a}{1 + 2t_{сп}W_i}$$

и производя преобразования, мы получаем разность заселенностей в обычном виде

$$N = \frac{N_0}{1 + \tau_s W_i}, \tag{14.43}$$

где, однако, теперь

$$\boxed{N_0 \approx \frac{N_a(t_{сп}W - 1)}{1 + t_{сп}W}}; \tag{14.44}$$

$$\boxed{\tau_s \approx \frac{2t_{сп}}{1 + t_{сп}W}}. \tag{14.45}$$

Как и в четырехуровневой схеме, N_0 и τ_s насыщаются с увеличением вероятности перехода накачки W.

Упражнение 14.4

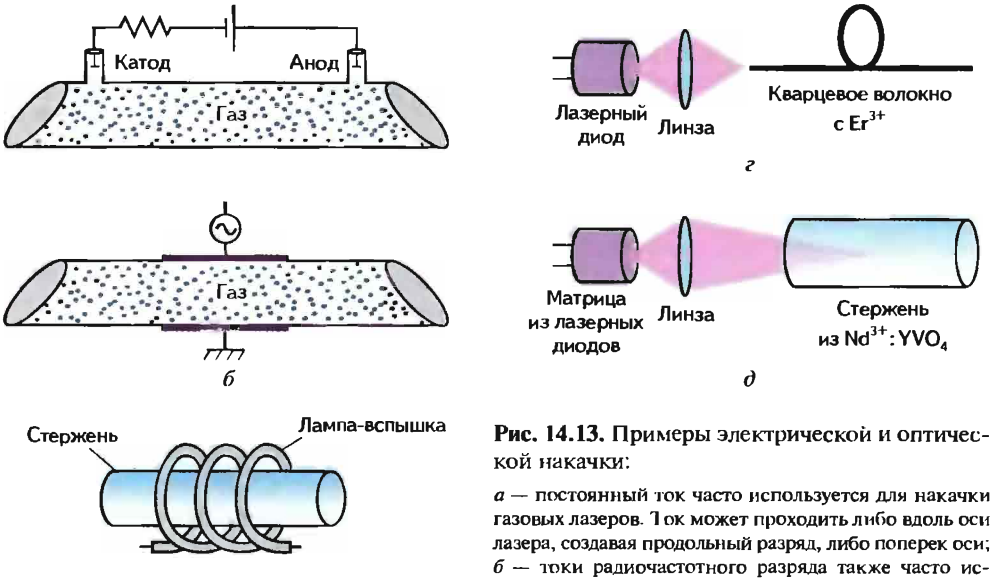
Мощность накачки в трех- и четырехуровневых схемах

а. Определите вероятность перехода накачки W, требуемую для достижения нулевой разности заселенностей в трех- и четырехуровневых лазерных усилителях.

б. Покажите, что если вероятность перехода накачки в трехуровневой системе равна $W = 2/t_{сп}$, а в четырехуровневой $W = 1/2t_{сп}$, то $N_0 = N_a/3$. Сравните мощности накачки, требуемые для достижения такой разности заселенностей.

Методы накачки

Как отмечалось ранее, накачку можно осуществить множеством методов, включая электрические, оптические и химические средства (см. обсуждение видов люминесценции в подразд. 13.5.1). Ряд часто применяемых электрических и оптических способов накачки схематически представлен на рис. 14.13. Ядерная накачка осуществляется пучком частиц высокой энергии или гамма-лучей от ядерного реактора или радиоактивного изотопа.



б *г* *д*

Рис. 14.13. Примеры электрической и оптической накачки:

а — постоянный ток часто используется для накачки газовых лазеров. Ток может проходить либо вдоль оси лазера, создавая продольный разряд, либо поперек оси; *б* — токи радиочастотного разряда также часто используются для накачки газовых лазеров; *в* — ксеноновые лампы-вспышки или криптоновые дуговые лампы непрерывного действия используют для оптической накачки рубиновых и редкоземельных твердотельных лазеров; *г* — полупроводниковые лазерные диоды часто применяются для накачки усилителей из кварцевого волокна с Er^{3+}; *д* — матрица из лазерных диодов применяется для оптической накачки лазеров на $Nd^{3+}:YVO_4$

Важно заметить, что R_1 и R_2 представляют собой число атомов в единицу времени и в единице объема, для которых достигается эффективная накачка. Процесс накачки может быть очень неэффективным. При оптической накачке, например, множество фотонов от лампы накачки теряется и не возбуждает активную среду.

14.3. РАСПРОСТРАНЕННЫЕ ЛАЗЕРНЫЕ УСИЛИТЕЛИ

Лазерное усиление может происходить в множестве материалов. Схемы уровней для ряда атомов, ионов, молекул и твердых тел, проявляющих лазерные свойства, показаны в разд. 13.1. На практике лазерные системы имеют множество взаимодействующих уровней энергии, которые влияют на заселенности уровней рабочего перехода N_1 и N_2, как показано на рис. 14.7. Тем не менее основные принципы действия лазерного усилителя можно описать на основе трех- и четырехуровневых схем.

Для иллюстрации сказанного рассмотрим три лазерных усилителя: трехуровневый на рубине, четырехуровневый на стекле с неодимом и трехуровневый на кварцевом волокне с эрбием. Большинство лазеров работает на основе четырехуровневой схемы накачки, рубин и кварцевое волокно с эрбием являются исключениями. Мы рассмотрим также важный усилитель, работающий на основе вынужденного комбинационного рассеяния. Все обсуждаемые здесь лазерные усилители также работают как генераторы (см. подразд. 15.3.1).

Большинство лазерных усилителей используется как **усилители мощности**. Этот тип усилителей предназначен для увеличения мощности высококачественного, но слабого лазерного излучения. Однако некоторые лазерные усилители, такие как усилитель на кварцевом волокне с эрбием, используются в оптических системах связи как **магистральные усилители** (оптические повторители) и как **оптические предусилители**, предназначенные для усиления сигнала перед фотоприемником (см. подразд. 24.1.3). Полупроводниковые оптические усилители описаны в разд. 17.2. Лазерные усилители часто используются в режиме насыщения (см. разд. 14.4).

14.3.1. Рубин

Рубин ($Cr^{3+}:Al_2O_3$) представляет собой сапфир (Al_2O_3), в котором ионы хрома (Cr^{3+}) замещают небольшой процент ионов алюминия (см. подразд. 13.1.3). Рубин — первый материал, в котором был обнаружен лазерный эффект (см. гл. 15). Он служит дидактическим примером, поскольку с наши дни этот лазерный усилитель используется редко. Как и в большинстве материалов, лазерное действие возможно на множестве переходов. Уровни энергии, отвечающие за хорошо известное красное излучение рубина, показаны на рис. 14.14 вместе со своими теоретико-групповыми символами. Это трехуровневая схема. Уровень 1 — основное состояние. Уровень 2 состоит из двух близ-

ко расположенных дискретных уровней; эти уровни не разрешены на рис. 14.14. Нижний из них, обозначаемый R_1, соответствует знаменитому красному лазерному переходу на $\lambda_0 = 694{,}3$ нм. Уровень 3 включает две широкие энергетические зоны с центрами около 550 нм (зеленая область) и 400 нм (фиолетовая область). Соответствующие полосы поглощения отвечают за красноватый цвет материала в проходящем свете.

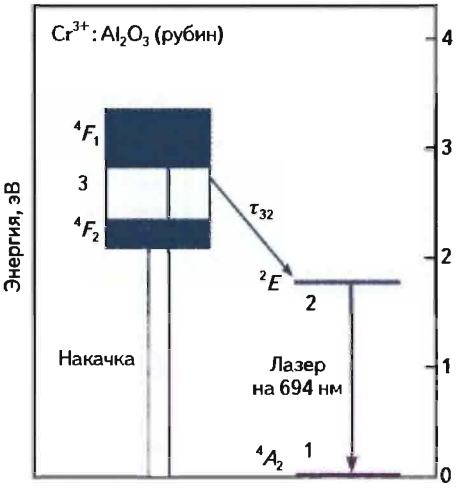


Рис. 14.14. Энергетические уровни, имеющие отношение к красному излучению рубинового лазера с длиной волны 694,3 нм на переходе $^2E \rightarrow {}^4A_2$

Рис. 14.15. Конструкции лазерных усилителей на рубине:

а — геометрия, использованная в первом лазерном генераторе на рубине, построенном Мейманом в 1960 г. (см гл. 15); *б* — высокоэффективная геометрия накачки с использованием прямолинейной лампы-вспышки в отражающем эллиптическом цилиндре

Как показано на рис. 14.15, оптическую накачку рубинового стержня с уровня 1 на уровень 3 можно осуществить, окружив его спиральной лампой-вспышкой или заключив вместе с лампой-вспышкой прямолинейной формы в отражающий цилиндр эллиптического сечения (см. рис. 1.6). Лампа-вспышка излучает белый свет, часть которого поглощается с переходом на очень широкий уровень 3 ионов Cr^{3+} Возбужденные ионы Cr^{3+} быстро переходят с уровня 3 на уровень 2 (с τ_{32} порядка пикосекунд). Они остаются на уровне 2 продолжительное время, поскольку время спонтанного радиационного перехода $2 \rightarrow 1$ достаточно велико ($t_{сп} \approx 3$ мс), что соответствует трехуровневой схеме накачки, показанной на рис. 14.12. Безызлучательный распад пренебрежимо мал ($\tau_{21} \approx t_{сп}$). Переход име-

ет однородно уширенную линию с $\Delta\nu \approx 330$ ГГц. Основной вклад в уширение дают упругие столкновения с фононами решетки. Характеристики лазерного перехода в рубине и рубинового лазера приведены в табл. 14.1 и 15.1 соответственно.

14.3.2. Стекло с неодимом

Усилители на стекле с Nd^{3+} могут быть изготовлены очень больших размеров и поэтому могут использоваться для генерации оптических импульсов экстремальной мощности, хотя и с малой частотой повторения из-за ограниченной теплопроводности стекла. Стекло может быть изготовлено с высоким оптическим качеством и сохраняет оптическую полировку. Кроме того, оно изотропно и легко легируется однородным образом.

Четырехуровневый лазерный усилитель на стекле с неодимом играет центральную роль в Национальном комплексе зажигания (National Ignition Facility, NIF) — научном комплексе для осуществления инерциального термоядерного синтеза, расположенном в Ливерморской национальной лаборатории им. Лоуренса в г. Ливермор, Калифорния (см. пример 14.5). Участвующие в накачке лазерного перехода на длине волны $\lambda_0 = 1{,}053$ мкм энергетические уровни для используемого в NIF фосфатного лазерного стекла показаны на рис. 14.16 (см. цв. вклейку). На этом рисунке также показан спектр ксеноновой лампы-вспышки, используемой для накачки этого усилителя.

Уровень 1 имеет энергию на 0,24 эВ выше основного состояния. Это значительно больше тепловой энергии при комнатной температуре 0,026 эВ, поэтому тепловая заселенность уровня 1 пренебрежимо мала. Уровень 3 представляет группу из четырех полос поглощения, каждая шириной примерно 30 нм, с центрами на 805, 745, 585 и 520 нм; эти полосы отвечают за пурпурный цвет материала при рассматривании в проходящем свете. Возбужденные ионы быстро релаксируют с уровня 3 на уровень 2, где остаются на продолжительное время ($t_{сп} = 375$ мкс). Поскольку τ_1 очень мало ($\approx$300 пс), схема энергетических уровней неодимового стекла соответствует четырехуровневой лазерной схеме, изображенной на рис. 14.11. Переход $2 \rightarrow 1$ неоднородно уширен вследствие аморфной структуры стекла, в которой каждый ион находится в различном окружении. Поэтому при комнатной температуре этот материал обладает значительной шириной линии $\Delta\nu \approx 7$ ТГц. Другие характеристики перехода $2 \rightarrow 1$ в этом материале детализированы в табл. 14.1. Усиление существенно выше, чем в рубине, по причине четырехуровневого характера схемы. Характеристики небольшого лазера на стекле с неодимом даны в табл. 15.1.

Пример 14.1

Лазерные усилители на стекле с неодимом в Национальном комплексе зажигания (NIF)

Усилители на стекле с неодимом широко используются в экспериментах по возбуждению управляемого термоядерного синтеза, которые, например, проводятся на NIF, а также в проекте Laser MegaJoule (LMJ) во Франции. Лазер-

ные системы в обоих проектах включают четыре кластера лазерных усилителей, каждый из которых состоит из шести стоп усилителей. Каждая стопа, в свою очередь, содержит 8 усиливающих пластин из лазерного стекла, расположенных в полости, засвечиваемой лампой-вспышкой, как показано на рис. 14.17 (см. цв. вклейку). Использование пучков квадратного сечения позволяет плотно упаковать усиливающие стеклянные пластины в компактную конфигурацию, что уменьшает размеры и стоимость системы. Хотя лазерная апертура квадратная, пластины из лазерного стекла прямоугольные, поскольку они расположены под углом Брюстера к направлению распространения пучка. Это минимизирует потери за счет френелевского отражения на поверхности пластин и усиливает взаимодействие света накачки от лампы-вспышки с их активным веществом. Конструкция усилителя NIF рассчитана на генерацию оптических импульсов с энергией 1,8 МДж длительностью 3,5 нс, что соответствует пиковой мощности в 500 ТВт. Чтобы дать стеклу время остыть между выстрелами, за 24 ч их можно произвести не более пяти. Затравочный оптический импульс создается лазером на волокне, легированном Yb^{3+} (см. подразд. 15.3.1).

Для инициирования процесса термоядерного синтеза частота усиленного излучения увеличивается в три раза, так что длина волны становится равной 351 нм (ультрафиолет). Это достигается в результате каскадного процесса, на первом шаге которого длина волны уменьшается до 526 нм за счет удвоения частоты (генерации второй гармоники) в нелинейно-оптическом кристалле дигидрофосфата калия (KDP). На втором этапе происходит параметрическое сложение частот первой и второй гармоник в кристалле дейтерированного дигидрофосфата калия (DKDP) (см гл. 21). Пучки утроенной частоты затем фокусируются в пятна миллиметровых размеров и направляются на мишень в виде таблетки, содержащей дейтерий и тритий. Комплекс занимает здание размером с футбольный стадион.

14.3.3. Кварцевое волокно, легированное эрбием

Оптические волоконные усилители (ОВУ) — средства усиления, необходимые для реализации преимуществ одномодовой волоконной оптики (см. гл. 9). В волоконных усилителях с редкоземельными ионами сигнал и накачка вводятся в легированное волокно, где сигнал усиливается за счет вынужденного излучения при переходах электронов с верхних уровней на нижние в подвергаемых накачке ионах. Обычно при этом используются Er^{3+}, Pr^{3+}, Tm^{3+}, Nd^{3+}, Yb^{3+} и Ho^{3+} Они действуют в широком диапазоне частот, главным образом, в ближнем ультрафиолете, однако, за исключением Er^{3+}, все легирующие добавки эффективно работают в обычной для телекоммуникации полосе частот только в сочетании с базовым стеклом, отличным от кварца.

Поскольку на практике используется именно кварц, **волоконные усилители с эрбием** (erbium-doped fiber amplifier — **EDFA**) широко применяются в системах волоконно-оптической связи, часто в виде последовательной цепочки (см. гл. 24). Наряду с другими достоинствами, они обеспечивают поляризационно-

независимое усиление, низкие вносимые потери и широкий переход вблизи $\lambda = 1550$ нм (соответствующий длине волны минимальных потерь в кварцевых оптических волокнах, как показано на рис. 9.24).

Накачка достигается продольным введением света в усиливающую среду от диодного лазера на InGaAs с напряженными квантовыми ямами, работающего на длине волны $\lambda_0 = 980$ нм. Как показано на рис. 14.18, свет накачки может вводиться в том же направлении, что и сигнал (прямое направление), в направлении, противоположном сигналу (обратное направление), или в обоих направлениях (двунаправленная схема). Несмотря на большую сложность, последняя конфигурация предпочтительна, поскольку она обеспечивает относительно однородное распределение мощности накачки по длине усиливающего волокна, что способствует повышению эффективности и снижению шумов. **Волокно с двойной оболочкой** может применяться для предотвращения нелинейно-оптических эффектов в сердцевине при высоких мощностях накачки, типичных для волоконных лазерных генераторов (см. рис. 15.27). Иттербий и тулий часто применяются для дополнительного легирования вместе с эрбием. Дополнительное легирование дает различные преимущества, такие, как расширение рабочего диапазона длин волн и достижение более высоких мощностей за счет механизмов переноса энергии и предотвращения кластеризации ионов эрбия.

Рис. 14.18. Продольная накачка волоконного лазерного усилителя. Накачка может осуществляться в прямом (*а*), обратном (*б*) и в обоих направлениях (*в*). Накачка усилителей из кварцевого волокна с эрбием часто производится диодным лазером на InGaAs с длиной волны $\lambda_0 = 980$ нм. ВКР-усилители на кварцевом волокне, обсуждаемые в подразд. 14.3.4, также можно накачивать диодным лазером на InGaAs, работающим на длине волны примерно на 100 нм ниже длины волны усиливаемого излучения

Как видно из рис. 14.19, волоконная система из кварца с Er^{3+} при накачке на $\lambda_0 = 980$ нм работает в окрестности $\lambda = 1550$ нм на линии флуоресцентного перехода ${}^4I_{13/2} \rightarrow {}^4I_{15/2}$. Она действует по трехуровневой схеме при $T = 300$ °K и по четырехуровневой схеме при охлаждении до $T = 77$ °K. Поскольку ион Er^{3+} относится к лантанидам, материал, в который погружены ионы, играет минимальную роль в формировании энергетического спектра. Уширение имеет смешанный тип с участием однородных (за счет фононов) и неоднородных (за счет вариаций локального поля в стекле) механизмов. Это приводит к формированию спектрального профиля усиления, который для некоторых приложений требует выравнивания. Благоприятными свойствами лазерного перехода

являются большое время спонтанной релаксации возбужденного состояния, отсутствие промежуточных уровней энергии между основным и возбужденным состояниями, а также отсутствие поглощения из возбужденного состояния.

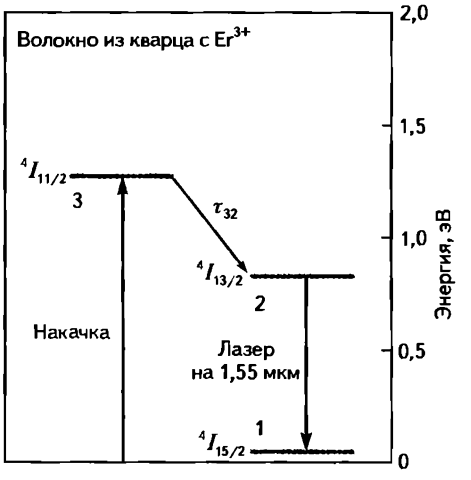


Рис. 14.19. Схема уровней энергии для возбуждения лазерного перехода $^4I_{13/2} \rightarrow {}^4I_{15/2}$ в окрестности $\lambda = 1550$ нм в кварцевом волокне, легированном ионами Er^{3+}. Этот усилитель работает по трехуровневой схеме при комнатной температуре ($T = 300$ °K). Три взаимодействующих уровня помечены цифрами. Систему можно заставить работать по четырехуровневой схеме в окрестности 2,9 мкм на переходе $^4I_{11/2} \rightarrow {}^4I_{13/2}$

Указанные свойства EDFA позволяют получать усиление свыше 50 дБ при мощностях накачки в десятки милливатт. В частности, усиление ≈30 дБ получается при введении ≈5 мВт накачки на 980 нм в 50-метровое волокно, содержащее ≈300 промилле Er_2O_3. Наилучшая эффективность усиления ≈10 дБ/мВт. Более того, можно генерировать выходную мощность свыше 100 Вт, поскольку выходная мощность растет пропорционально мощности накачки. Доступная ширина полосы, равная $\Delta\lambda \approx 40$ нм (что соответствует $\Delta\nu \approx 5{,}3$ ТГц), включает полосу связи *C* (обычная связь), простирающуюся от 1530 до 1565 нм (см. подразд. 24.1.1). Полоса *L* (длинноволновая) от 1565 до 1625 нм может быть легко включена в диапазон усиления, хотя оптимальные параметры EDFA для двух полос различны. Большая величина произведения усиления на ширину полосы, присущая этим усилителям, делает их весьма подходящими для использования в системах уплотнения с разделением сигналов по длине волны (см. разд. 24.3). Характеристики лазерного усилителя на основе кварцевого волокна с эрбием приведены в табл. 14.1. Введение обратной связи легко превращает усилитель в генератор, что обсуждается в подразд. 15.3.1.

Лазерный переход $^4I_{13/2} \rightarrow {}^4I_{15/2}$ допускает также прямую накачку светом диодного InGaAsP лазера на 1,48 мкм. Такая квазидвухуровневая схема менее эффективна, чем трехуровневая, реализуемая на 980 нм, поскольку усиление на единицу мощности накачки получается ниже, а шум выше. Однако ширина линии перехода накачки и мощность насыщающего сигнала выше на 1,48 мкм, поэтому накачка на этой длине волны иногда используется для усилителей более высокой мощности. В некоторых конфигурациях EDFA одновременно накачивают на обеих длинах волн.

Что касается других волоконных усилителей, легированных редкоземельными ионами, хорошие рабочие характеристики имеют волокна из фторидного или многокомпонентного силикатного стекла, легированного Tm^{3+}, действующие в области *S*-полосы 1460—1530 нм, волокна, легированные Pr^{3+}, в области 1300 нм и волокна из кварца с Yb^{3+} в области 1050—1120 нм. Другие классы широко используемых оптических усилителей — это волоконные ВКР-усилители, обсуждаемые в подразд. 14.3.4, и полупроводниковые оптические усилители, исследуемые в разд. 17.2. Относительные достоинства этих трех классов усилителей изучаются в подразд. 17.2.4 и 24.1.3.

Резюме

Волоконный усилитель из кварца с Er^{3+} широко применяется, благодаря следующим своим достоинствам:

- большое усиление;
- высокая выходная мощность;
- большая ширина полосы;
- поляризационная нечувствительность;
- высокая эффективность;
- низкие вносимые потери;
- низкие шумы.

14.3.4. Волоконные ВКР-усилители

Волоконные усилители, легированные эрбием и редкоземельными ионами, не единственный тип волоконных усилителей. Оптоволоконные усилители (ОВУ) могут основываться на принципах, отличных от вынужденного излучения. Важным вариантом ОВУ является волоконный ВКР-усилитель, основанный на вынужденном комбинационном рассеянии (ВКР).

Как обсуждалось в подразд. 13.5.3, вынужденное комбинационное рассеяние происходит, когда фотон накачки с энергией $h\nu_p$ вместе с сигнальным фотоном более низкой энергии $h\nu_s$ входит в нелинейную оптическую среду, например оптическое волокно. Нерезонансный вариант процесса схематически показан на врезке рис. 14.20, штриховая горизонтальная линия изображает виртуальное состояние. Сигнальный фотон стимулирует испускание своей точной копии, которая получается за счет стоксова сдвига энергии фотона накачки на величину $h\nu_R$, так что энергия вновь испущенного фотона в точности равна энергии падающего сигнального фотона. Избыток энергии фотона накачки передается колебательным модам стекла. Получающееся оптическое усиление аналогично вынужденному излучению в волоконном усилителе из кварца с эрбием, однако полоса усиления ВКР определяется колебательным спектром основного стекла, а не шириной перехода примесного иона. Сила эффекта, выражаемая коэффициентом усиления ВКР γ_R, зависит от нелинейных свойств

стеклянного волокна и пропорциональна интенсивности накачки $I_p = P/A$, где P — мощность накачки; A — площадь взаимодействия [см. (21.55)].

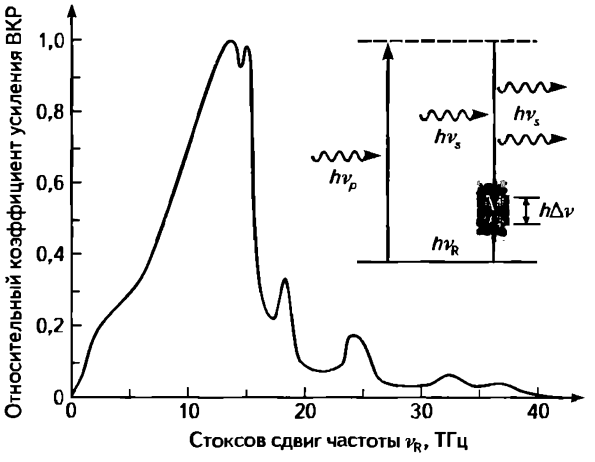


Рис. 14.20. Вынужденное комбинационное рассеяние (ВКР) схематически показано на врезке. ВКР-усиление возможно в широком диапазоне стоксовых частот, определяемом колебательными характеристиками материала. Кварц, германий, фосфор и боратные стекла имеют сильно различающиеся по форме и величине спектральные распределения. В кварцевом волокне, легированном германием, частота пика коэффициента усиления ВКР на 13 ТГц ниже пика накачки при ширине полосы 12,5 ТГц. (Кривая усиления заимствована из статьи *Stolen R.H., Lee C., Jain R.K.* Development of the Stimulated Raman Spectrum in Single-Mode Silica Fibers. Journal of the Optical Society of America B. Vol. 1, 1984. P. 652—657. Fig. 5.)

ВКР-усилители могут быть как распределенными, так и сосредоточенными. В **распределенном волоконном ВКР-усилителе** накачка вместе с сигнальным излучением распространяется по передающему волокну, одновременно служащему усиливающей средой. В **сосредоточенном волоконном ВКР-усилителе** усиление происходит в специальном коротком отрезке волокна с очень высокой нелинейностью. Сердцевина обычно делается малой, чтобы увеличить интенсивность накачки I_p и, следовательно, сократить необходимую длину волокна, которая может быть значительной. Как и в EDFA, излучение накачки может распространяться в прямом или обратном направлении по отношению к сигналу, а также в обоих направлениях (см. рис. 14.18). Чаще всего используется накачка в обратном направлении, так как при этом уменьшаются шумы, вносимые ею в сигнал.

Волоконные ВКР-усилители могут обеспечивать существенно большую ширину полосы, чем EDFA. Как видно из рис. 14.20, основной пик коэффициента усиления ВКР для обычного кварцевого волокна, легированного германием, сдвинут в стоксову сторону относительно частоты накачки приблизительно на $\nu_R = 13$ ТГц, что соответствует около 100 нм при $\lambda_0 = 1550$ нм. Однако

фосфосиликатные стеклянные волокна обладают значительно большими стоксовыми сдвигами (см. пример 15.3). Накачка волоконного ВКР-усилителя может осуществляться с помощью поляризационно-разнесенных лазерных диодов, волоконных лазеров, либо волоконных ВКР-лазеров, работающих на длине волны примерно на 100 нм короче, чем длина волны усиливаемого сигнала, если речь идет о кварцевом волокне, легированном германием. В этом материале ширина полосы, в которой возможно существенное усиление ВКР, примерно той же величины, что и сдвиг, $\Delta\nu \approx 12{,}5$ ТГц, что снова соответствует $\Delta\lambda \approx 100$ нм на 1550 нм. Однако, комбинируя различные накачки разных частот, полосы усиления можно намного расширить, поскольку стоксова частота привязана к частоте накачки. В принципе ВКР-усиление можно осуществить во всем диапазоне прозрачности волокна.

Волоконные ВКР-усилители обеспечивают усиление, достигающее 20 дБ. Эффективность усиления в кварцевом волокне, легированном германием, около 0,02 дБ/мВт; для сравнения у EDFA имеем 10 дБ/мВт. Таким образом, типичная мощность накачки для достижения приемлемых уровней ВКР-усиления в таком распределенном усилителе составляет сотни милливатт — намного больше, чем требуется для EDFA. В сосредоточенных ВКР-усилителях, где риск в плане безопасности не является предметом беспокойства, могут использоваться мощности накачки, превышающие 1 Вт. В отличие от EDFA требуется накачка с определенной поляризацией, поскольку усиление ВКР достигает максимума, когда поляризация накачки и сигнала одинакова. Эффективность волоконного ВКР-усилителя существенно меньше, чем EDFA, но ее можно заметно поднять, применяя волокно с компенсацией дисперсии. EDFA и волоконные ВКР-усилители иногда применяются в сочетании. Использование компенсации дисперсии вместе с доступностью матриц из диодных лазеров высокой мощности делает волоконные ВКР-усилители конкурентоспособными в волоконных линиях связи, особенно из-за требований к ширине полосы, которые продолжают повышаться по мере продвижения к передаче многих сотен терагерц на дальние расстояния.

Резюме

Волоконные ВКР-усилители имеют как достоинства, так и недостатки по сравнению с усилителями на волокне, легированном эрбием.

Достоинства волоконных ВКР-усилителей по сравнению с EDFA заключаются в следующем:

- большая ширина полосы;
- возможность расширения полосы с помощью накачки на нескольких частотах;
- широкий диапазон рабочих длин волн;
- произвольное базовое вещество волокна;
- совместимость с существующими каналами связи.

Недостатки волоконных ВКР-усилителей по сравнению с EDFA заключаются в следующем:

- меньшее усиление;
- большая мощность накачки и меньшая эффективность;
- большая длина волокна;
- чувствительность к поляризации сигнала.

Сравнение действия оптических волоконных усилителей с полупроводниковыми приведено в подразд. 17.2.4 и 24.1.3.

14.3.5. Таблица избранных лазерных переходов

Наиболее часто используемые лазерные усилители — те, что обсуждались в подразд. 14.3.2 и 14.3.3. Однако лазерное усиление возможно также в газах, органических красителях, системах со свободными электронами и полупроводниках. В табл. 14.1 приведены длины волн, сечения, спонтанные времена жизни, ширины линий и показатели преломления для ряда типичных лазерных переходов. Принцип мазера может быть реализован от микроволнового до рентгеновского диапазона, однако в табл. 14.1 выделены только переходы видимого и инфракрасного диапазона. Значения σ_0, $t_{сп}$ и $\Delta\nu$ меняются в широких пределах.

14.4. НЕЛИНЕЙНОСТЬ УСИЛИТЕЛЯ

14.4.1. Насыщение усиления в однородно-уширенной среде

Коэффициент усиления

Было установлено, что коэффициент усиления $\gamma(\nu)$ лазерной среды зависит от разности заселенностей N [см. (14.4)], которая, в свою очередь, определяется уровнем накачки [см. (14.29)], что N зависит также от скорости перехода W_i [см. (14.24)] и что W_i, в свою очередь, зависит от плотности потока фотонов ϕ [см. (14.1)]. Отсюда следует, что коэффициент усиления лазерной среды зависит от плотности потока фотонов, которая подлежит усилению. В этом состоит причина насыщения усиления и нелинейности лазерного усилителя, что мы сейчас продемонстрируем.

Подставляя (14.1) в (14.24), получаем

$$N = \frac{N_0}{1 + \phi/\phi_s(\nu)}, \tag{14.46}$$

где

$$\boxed{\frac{1}{\phi_s(\nu)} = \tau_s \sigma(\nu) = \frac{\lambda^2}{8\pi}\frac{\tau_s}{t_{сп}} g(\nu).} \tag{14.47}$$

Насыщающая плотность потока фотонов

Таблица 14.1. **Характеристики типичных лазерных переходов**

Лазерная среда	Длина волны перехода* λ_0, нм	Сечение перехода σ_0, см^2	Спонтанное время жизни $t_{сп}$	Ширина линии перехода** $\Delta\nu$		Показатель преломления n
C^{5+}	18,2	$5 \cdot 10^{-16}$	12 пс	1 ТГц	Н	≈1
Эксимер ArF	193	$3 \cdot 10^{-16}$	10 нс	10 ТГц	Н	≈1
Ar^+	515	$3 \cdot 10^{-12}$	10 нс	3,5 ГГц	Н	≈1
Краситель родамин 6Ж	560—640	$2 \cdot 10^{-16}$	5 нс	40 ТГц	О/Н	1,40
He—Ne	633	$3 \cdot 10^{-13}$	150 нс	1,5 ГГц	Н	≈1
$Cr^{3+}:Al_2O_3$	694	$2 \cdot 10^{-20}$	3 мс	330 ГГц	О	1,76
$Cr^{3+}:BeAl_2O_4$	700—820	$1\ 10^{-20}$	260 мкс	25 ТГц	О	1,74
$Ti^{3+}:Al_2O_3$	700—1050	$3 \cdot 10^{-19}$	3,9 мкс	100 ТГц	О	1,76
Yb^{3+} : YAG	1030	$2 \cdot 10^{-20}$	1 мс	1 ТГц	О	1,82
Nd^{3+} : фосфатное стекло	1053	$4 \cdot 10^{-20}$	370 мкс	7 ТГц	Н	1,50
Nd^{3+} : YAG	1064	$3 \cdot 10^{-19}$	230 мкс	150 ГГц	О	1,82
$Nd^{3+}:YVO_4$	1064	$8 \cdot 10^{-19}$	100 мкс	210 ГГц	О	2,0
InGaAsP***	1300—1600	$2 \cdot 10^{-16}$	2,5 нс	10 ТГц	О	3,54
Er^{3+} : кварцевое волокно	1550	$6 \cdot 10^{-21}$	10 мс	5 ТГц	О/Н	1,46
CO_2	10600	$3 \cdot 10^{-18}$	3 с	60 МГц	Н	≈1

* Длина волны в свободном пространстве, показанная в таблице, соответствует наиболее часто используемому переходу в каждой лазерной среде. Например, He—Ne лазерные системы чаще всего используются на красно-оранжевой линии 0,633 мкм, однако работают и часто применяются также на длинах волн 0,543; 1,15 и 3,39 мкм (а также имеют лазерные переходы на сотнях других длин волн).

** Значения, показанные для газов, таких как CO_2, типичны для низких давлений (атомная ширина линии в газах зависит от давления из-за уширения столкновениями, которое является однородным). О и Н обозначают преобладание однородного и неоднородного уширения соответственно.

*** Значения приведены для $In_{0,72}Ga_{0,28}As_{0,6}P_{0,4}$ в предположении, что концентрация инжектируемых носителей заряда $\Delta n = 1{,}8 \cdot 10^{18}$ см$^{-3}$ (см. примеры 17.3—17.5).

Так выражается зависимость разности заселенностей N от плотности потока фотонов ϕ. Теперь, подставляя (14.46) в выражение для коэффициента усиления (14.4), получаем **насыщенный коэффициент усиления** для однородно-уширенной среды:

$$\boxed{\gamma(\nu) = \frac{\gamma_0(\nu)}{1 + \phi/\phi_s(\nu)},} \qquad (14.48)$$

Насыщенный коэффициент усиления

где

$$\boxed{\gamma_0(\nu) = N_0\sigma(\nu) = N_0\frac{\lambda^2}{8\pi t_{\text{сп}}}g(\nu).} \qquad (14.49)$$

Коэффициент усиления слабого сигнала

Коэффициент усиления — убывающая функция плотности потока фотонов ϕ, графически показанная на рис. 14.21. Величина $\phi_s(\nu) = 1/\tau_s\sigma(\nu)$ представляет собой плотность потока фотонов, при которой коэффициент усиления снижается до половины своего максимального значения, поэтому она называется **насыщающей плотностью потока фотонов**. Когда $\tau_s \approx t_{\text{сп}}$, интерпретация $\phi_s(\nu)$ проста: грубо говоря, один фотон может быть испущен за каждое время спонтанного излучения в каждую площадь сечения перехода $[\sigma(\nu)\phi_s(\nu)t_{\text{сп}} = 1]$.

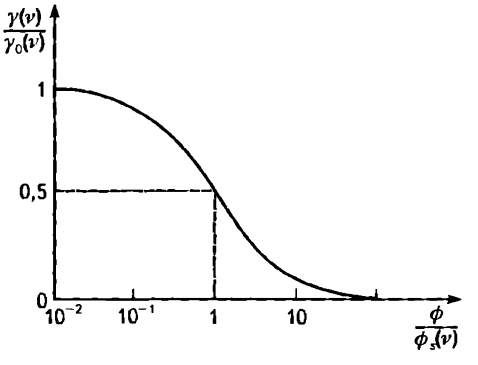


Рис. 14.21. Зависимость нормированного насыщенного коэффициента усиления $\gamma(\nu)/\gamma_0(\nu)$ от нормированной плотности потока фотонов $\phi/\phi_s(\nu)$. Когда ϕ равно своему насыщающему значению $\phi_s(\nu)$, коэффициент усиления снижается до половины своего ненасыщенного значения

Упражнение 14.5

Насыщающая плотность потока фотонов для рубина

Определите насыщающую плотность потока фотонов и соответствующую интенсивность насыщения для лазерного перехода в рубине на $\lambda_0 = 694{,}3$ нм при $\nu = \nu_0$. Используйте параметры, приведенные в табл. 14.1. Считайте, что $\tau_s = 2t_{\text{сп}}$ в соответствии с (14.42).

Упражнение 14.6

Спектральное уширение насыщенного усилителя

Рассмотрим усиливающую среду с однородным уширением и лоренцевой формой линии ширины $\Delta\nu$ [см. (14.8)]. Покажите, что при плотности потока фотонов ϕ коэффициент усиления $\gamma(\nu)$ принимает лоренцеву форму с шириной

$$\Delta\nu_s = \Delta\nu\sqrt{1+\frac{\phi}{\phi_s(\nu_0)}}. \qquad (14.50)$$

Ширина линии насыщенного усилителя

Видно, что насыщение усиления сопровождается увеличением ширины полосы, т. е. снижением избирательности по частоте (рис. 14.22).

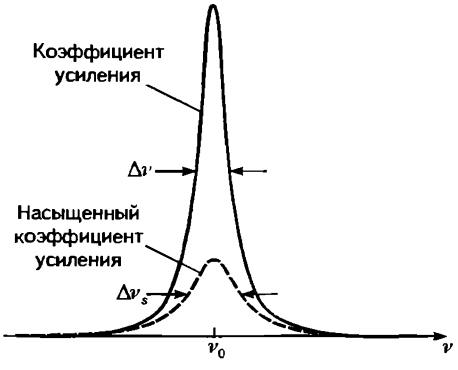


Рис. 14.22. Уменьшение коэффициента усиления и увеличение ширины полосы в результате насыщения при $\phi = 2\phi_s(\nu)$

Усиление

Определив влияние насыщения на *коэффициент* усиления (усиление на единицу длины), перейдем к определению поведения полного **насыщенного усиления** лазерного усилителя длиной d с однородным уширением (рис. 14.23, *а*). Для простоты опустим частотную зависимость в обозначениях $\gamma(\nu)$, $\phi_s(\nu)$ и будем писать просто γ, ϕ_s.

Если плотность потока фотонов в точке z — $\phi(z)$, то в соответствии с (14.48) коэффициент усиления в этой точке также является функцией z. Из (14.3) мы знаем, что приращение плотности потока фотонов в точке z есть $d\phi = \gamma\phi dz$, что приводит к дифференциальному уравнению

$$\frac{d\phi}{dz} = \frac{\gamma_0\phi}{1+\phi/\phi_s}. \qquad (14.51)$$

Переписывая это уравнение как

$$\left(\frac{1}{\phi}+\frac{1}{\phi_s}\right)d\phi = \gamma_0 dz$$

и интегрируя, получаем

$$\ln\frac{\phi(z)}{\phi(0)}+\frac{\phi(z)-\phi(0)}{\phi_s}=\gamma_0 z. \tag{14.52}$$

Следовательно, соотношение между плотностями потока фотонов на входе $\phi(0)$ и выходе $\phi(d)$ имеет вид

$$\left[\ln(Y)+Y\right]=\left[\ln(X)+X\right]+\gamma_0 d, \tag{14.53}$$

где X и Y — входное и выходное значения плотности потока фотонов, нормированные на насыщающую плотность потока фотонов, соответственно,

$$X=\frac{\phi(0)}{\phi_s};\quad Y=\frac{\phi(d)}{\phi_s}.$$

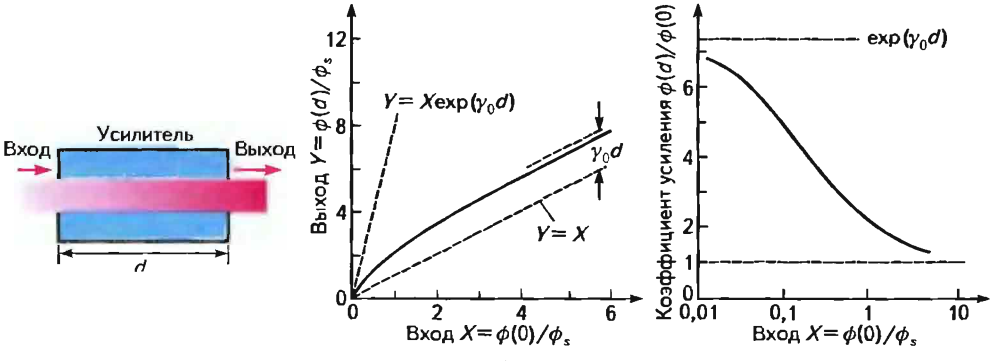


Рис. 14.23. Нелинейный (насыщенный) усилитель (*а*). Связь между нормированной выходной плотностью потока фотонов $Y=\phi(d)/\phi_s$ и нормированной входной плотностью потока фотонов $X=\phi(0)/\phi_s$ (*б*). Для $X\ll 1$ усиление $Y/X\approx\exp(\gamma_0 d)$. Для $X\gg 1$ получаем $Y\approx X+\gamma_0 d$. Численное решение уравнения (14.48) изображено сплошной линией. Усиление как функция нормированной плотности потока фотонов на входе X в усилителе длиной d при $\gamma_0 d=2$ (*в*)

Полезно исследовать усиление

$$G=\frac{\phi(d)}{\phi(0)}=\frac{Y}{X}$$

в двух предельных случаях.

1. Если как X, так и Y много меньше единицы (т. е. плотности потока фотонов намного меньше, чем насыщающая плотность потока фотонов), то X и Y пренебрежимо малы по сравнению с $\ln(X)$ и $\ln(Y)$, в результате чего получается приближенное соотношение $\ln(Y)\approx\ln(X)+\gamma_0 d$, откуда

$$Y\approx X\exp(\gamma_0 d). \tag{14.54}$$

В этом случае соотношение между Y и X линейно, а усиление равно

$$G = \frac{Y}{X} \approx \exp(\gamma_0 d)$$

(самая левая штриховая линия на рис. 14.23, *б*). Это согласуется с выражением (14.7), полученным в приближении слабого сигнала, справедливым, когда коэффициент усиления не зависит от плотности потока фотонов, т. е. $\gamma \approx \gamma_0$.

2. Когда $X \gg 1$, можно пренебречь $\ln(X)$ по сравнению с X, а $\ln(Y)$ — по сравнению с Y, в результате чего

$$Y \approx X + \gamma_0 d \tag{14.55}$$

или

$$\phi(d) \approx \phi(0) + \gamma_0 \phi_s d \approx \phi(0) + \frac{N_0 d}{\tau_s}. \tag{14.56}$$

В этих условиях сильного насыщения все атомы среды «заняты», испуская фотоны с постоянной плотностью потока $N_0 d/\tau_s$. Приходящие на вход фотоны просто проходят сквозь усилитель, при этом к ним добавляются фотоны с постоянной плотностью потока, которая не зависит от входного сигнала.

Для промежуточных значений X и Y уравнение (14.53) можно решить численно. График решения показан сплошной линией на рис. 14.23, *б*. Линейная связь входа и выхода, полученная в случае $X \ll 1$, и связь при очень сильном насыщении, $X \gg 1$, очевидно, являются предельными случаями численного решения. График усиления $G = Y/X$ для $\gamma_0 d = 2$ показан на рис. 14.23, *в*. Усиление имеет максимальное значение при малых плотностях потока фотонов на входе ($X \ll 1$) и уменьшается до единицы при $X \to \infty$.

Насыщаемые поглотители

Если коэффициент усиления γ_0 отрицателен, т. е. если распределение заселенностей нормальное, а не инвертированное ($N_0 < 0$), то среда вносит не усиление, а ослабление. Коэффициент ослабления $\alpha(\nu) = -\gamma(\nu)$ также испытывает насыщение в соответствии с соотношением

$$\alpha(\nu) = \frac{\alpha_0(\nu)}{1 + \phi/\phi_s(\nu)}.$$

Это значит, что при больших плотностях потока фотонов поглощение снижается. Материал, обладающий такими свойствами, называется **насыщаемым поглотителем.**

Связь между плотностями потока фотонов на входе $\phi(0)$ и выходе $\phi(d)$ в поглотителе длиной d описывается формулой (14.53) с отрицательным γ_0. Полный коэффициент пропускания поглотителя

$$\frac{Y}{X} = \frac{\phi(d)}{\phi(0)}$$

как функция $X = \phi(0)/\phi_s$ показан сплошной линией на рис. 14.24. Пропускание увеличивается с ростом $\phi(0)$, достигая предельного значения единица. Этот эффект объясняется тем, что $N \to 0$, так что поглощение исчезает.

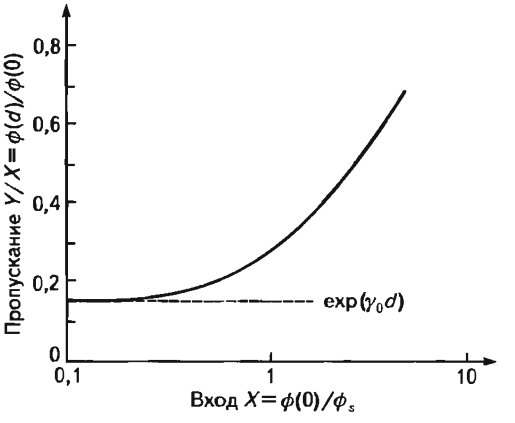


Рис. 14.24. Пропускание насыщаемого поглотителя $Y/X = \phi(d)/\phi(0)$ как функция нормированной плотности потока фотонов $X = \phi(0)/\phi_s$ при $\gamma_0 d = -2$. Пропускание растет с увеличением плотности потока фотонов на входе

*14.4.2. Насыщение усиления в неоднородно-уширенной среде

Среда с неоднородным уширением представляет собой собрание атомов с различающимися свойствами. Как обсуждалось в подразд. 13.3.4, каждая группа атомов, помеченная индексом β, имеет свою однородно-уширенную функцию формы линии $g_\beta(\nu)$. Функция формы линии всего ансамбля атомов $\bar{g}(\nu)$ получается усреднением по параметру β: $\bar{g}(\nu) = \langle g_\beta(\nu) \rangle$.

Поскольку согласно (14.49) коэффициент усиления слабого сигнала $\gamma_0(\nu)$ пропорционален $g(\nu)$, различные группы атомов β имеют различные коэффициенты усиления $\gamma_{0\beta}(\nu)$. Средний коэффициент усиления слабого сигнала тогда выражается формулой

$$\bar{\gamma}_0(\nu) = N_0 \frac{\lambda^2}{8\pi t_{\text{сп}}} \bar{g}(\nu). \tag{14.57}$$

Решение для насыщенного коэффициента усиления более сложно, поскольку насыщающая плотность потока фотонов $\phi_s(\nu)$, будучи пропорциональна $g(\nu)$, как следует из (14.47), сама зависит от β. Средний коэффициент усиления можно определить с помощью (14.48) и (14.47)

$$\bar{\gamma}(\nu) = \langle \gamma_\beta(\nu) \rangle, \tag{14.58}$$

где

$$\gamma_\beta(\nu) = \frac{\gamma_{0\beta}(\nu)}{1 + \phi/\phi_{s\beta}(\nu)} = b \frac{g_\beta(\nu)}{1 + \phi a^2 g_\beta(\nu)}; \tag{14.59}$$

$$b = N_0 \frac{\lambda^2}{8\pi t_{\text{сп}}}; \quad a^2 = \frac{\lambda^2}{8\pi} \frac{\tau_s}{t_{\text{сп}}}.$$

Вычисление среднего от (14.59) требует внимания, так как среднее от отношения не равно отношению средних.

Среда с доплеровским уширением

Хотя при **доплеровском уширении** все атомы среды имеют одинаковую форму линии $g(\nu)$, центральная частота каждой группы атомов β сдвинута на величину ν_β, пропорциональную скорости атомов v_β. Если функция $g(\nu)$ лоренцева с шириной $\Delta\nu$, то формула (14.8) дает

$$g(\nu) = \frac{\Delta\nu/2\pi}{(\nu-\nu_0)^2 + (\Delta\nu/2)^2}; \quad g_\beta(\nu) = g(\nu - \nu_\beta).$$

Подставляя $g_\beta(\nu)$ в (14.59), находим

$$\gamma_\beta(\nu) = \frac{b(\Delta\nu/2\pi)}{(\nu - \nu_\beta - \nu_0)^2 + (\Delta\nu_s/2)^2}, \tag{14.60}$$

где

$$\Delta\nu_s = \Delta\nu\sqrt{1 + \frac{\phi}{\phi_s(\nu_0)}}; \tag{14.61}$$

$$\phi_s^{-1}(\nu_0) = \frac{2a^2}{\pi\Delta\nu} = \frac{\lambda^2}{8\pi}\frac{\tau_s}{t_{\text{сп}}}\frac{2}{\pi\Delta\nu} = \frac{\lambda^2}{8\pi}\frac{\tau_s}{t_{\text{сп}}}g(\nu_0). \tag{14.62}$$

Выражение (14.61) было получено для насыщения усилителя с однородным уширением в упражнении 14.7 [см. (14.50)]. Очевидно, что группа атомов со скоростью v_β имеет насыщенный коэффициент усиления $\gamma_\beta(\nu)$ с лоренцевой формой ширины $\Delta\nu_s$, которая растет с увеличением плотности потока фотонов.

Среднее от $\gamma_\beta(\nu)$ в (14.58) легко получить, поскольку сдвиги ν_β характеризуются гауссовой функцией распределения вероятностей

$$p(\nu_\beta) = (2\pi\sigma_D^2)^{-1/2}\exp\left(-\frac{\nu_\beta^2}{2\sigma_D^2}\right)$$

со стандартным отклонением σ_D (см. упражнение 13.13). Таким образом, $\bar{\gamma}(\nu) = \langle\gamma_\beta(\nu)\rangle$ выражается как

$$\bar{\gamma}(\nu) = \int_{-\infty}^{\infty}\gamma_\beta(\nu)p(\nu_\beta)\,\mathrm{d}\nu_\beta. \tag{14.63}$$

Если $p(\nu_\beta)$ намного шире, чем $\gamma_\beta(\nu)$, т. е. доплеровское уширение намного больше $\Delta\nu_s$, можно рассматривать медленно меняющуюся функцию $p(\nu_\beta)$ как

постоянную и вынести ее из-под знака интеграла при вычислении $\bar{\gamma}(\nu_0)$. Полагая $\nu = \nu_0$ и $\nu_\beta = 0$ в показателе экспоненты, находим

$$\bar{\gamma}(\nu_0) = \frac{bp(0)}{\sqrt{1 + 2\phi a^2/\pi\Delta\nu}} = \frac{\bar{\gamma}_0}{\sqrt{1 + \phi/\phi_s(\nu_0)}}, \tag{14.64}$$

где средний коэффициент усиления слабого сигнала равен

$$\bar{\gamma}_0 = N_0 \frac{\lambda^2}{8\pi t_{\text{сп}}} \frac{1}{\sqrt{2\pi\sigma_D^2}}. \tag{14.65}$$

Формула (14.6) выражает средний коэффициент усиления среды с доплеровским уширением при насыщении на центральной частоте ν_0 как функцию плотности потока фотонов ϕ при $\nu = \nu_0$. С ростом ϕ коэффициент усиления испытывает насыщение по закону квадратного корня. Таким образом, при неоднородном уширении коэффициент усиления насыщается медленнее, чем при однородном [см. (14.48)], что иллюстрируется рис. 14.25.

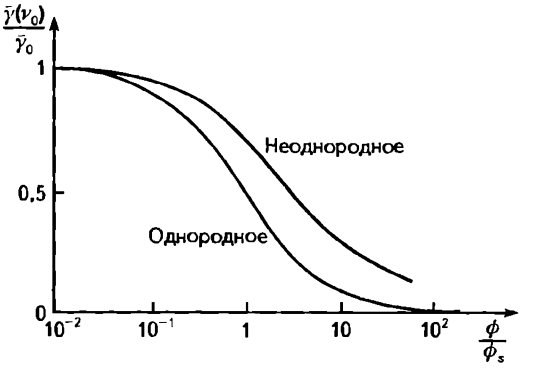


Рис. 14.25. Сравнение насыщения усиления при однородном и неоднородном уширении

Выжигание провала

При воздействии потока большой плотности монохроматических фотонов с частотой ν_1 на среду с неоднородным уширением насыщение коэффициента усиления происходит только у тех атомов, чьи функции формы линии перекрывают ν_1. Другие атомы просто не взаимодействуют с фотонами и остаются ненасыщенными. Если такую насыщенную среду зондировать с помощью слабого источника монохроматического света переменной частоты ν, то в профиле коэффициента усиления будет наблюдаться провал с центром на частоте ν_1, как показано на рис. 14.26. Это явление известно как **выжигание провала**.

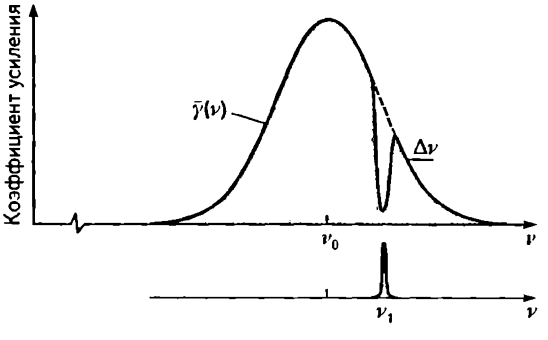


Рис. 14.26. Коэффициент усиления среды с неоднородным уширением локально насыщается потоком фотонов высокой плотности на частоте ν_1

Поскольку коэффициент усиления $\gamma_\beta(\nu)$ группы атомов со скоростью v_β имеет лоренцеву форму с шириной $\Delta\nu_s$, определяемой выражением (14.61), ширина провала также оказывается равной $\Delta\nu_s$. С ростом плотности потока насыщающих фотонов на частоте ν_1 как глубина, так и ширина провала увеличиваются.

*14.5. ШУМ УСИЛИТЕЛЯ

Резонансная среда, в которой происходит усиление посредством вынужденного излучения, генерирует также спонтанное излучение. Спонтанно испускаемый свет не зависит от сигнала на входе усилителя и является фундаментальным источником шума. В то время как усиливаемый сигнал имеет определенную частоту, направление и поляризацию, шум, связанный с **усиленным спонтанным излучением** (УСИ), является широкополосным, разнонаправленным и неполяризованным. Поэтому можно отфильтровать часть этого шума, расположив за усилителем узкополосный оптический фильтр, собирающую диафрагму и поляризатор.

Плотность вероятности (в секунду) того, что атом, находящийся на верхнем лазерном уровне, спонтанно испустит фотон с частотой между ν и $\nu + d\nu$ (см. упражнение 13.1), равна

$$P_{\text{сп}}(\nu)\,\mathrm{d}\nu = \frac{1}{t_{\text{сп}}}\,g(\nu)\,\mathrm{d}\nu. \tag{14.66}$$

Плотность вероятности испустить фотон произвольной частоты, конечно, равна $P_{\text{сп}} = 1/t_{\text{сп}}$. Если N_2 — плотность числа атомов на верхнем уровне энергии, то средняя плотность числа испущенных фотонов равна $N_2 P_{\text{сп}}(\nu)$. Следовательно, средняя мощность спонтанного излучения на единицу объема и на единицу частоты есть $h\nu N_2 P_{\text{сп}}(\nu)$. Эта плотность мощности излучается равномерно во всех направлениях и поровну делится между двумя поляризациями. Если выходное излучение усилителя собирается в телесном угле $\mathrm{d}\Omega$, как показано на рис. 14.27, и только для одной поляризации, то оно содержит только долю $(1/2)\mathrm{d}\Omega/4\pi$ от всей мощности спонтанного излучения. Более того, если использовать фильтр, полоса пропускания которого с центром на частоте усиливаемого сигнала ν имеет малую ширину B, то число фотонов, вносимых спонтанным излучением от бесконечно малого объема с единичной площадью и длиной $\mathrm{d}z$, будет равно $\xi_{\text{сп}}(\nu)\mathrm{d}z$, где

$$\xi_{\text{сп}}(\nu) = N_2 \frac{1}{t_{\text{сп}}}\,g(\nu)\,B\,\frac{\mathrm{d}\Omega}{8\pi} \tag{14.67}$$

представляет собой плотность потока шумовых фотонов на единицу длины усилителя.

При определении плотности потока шумовых фотонов, вносимых усилителем, плотность потока на единицу длины $\xi_{\text{сп}}(\nu)$ нельзя просто умножить на

длину усилителя. Дело в том, что шум спонтанного излучения сам усиливается средой, поэтому спонтанное излучение, генерируемое вблизи входа усилителя, вносит больший вклад в общий шум, чем генерируемое вблизи выхода. Один из способов учета шума спонтанного излучения заключается в замене дифференциального уравнения (14.3), определяющего нарастание плотности потока фотонов, на следующее:

$$\frac{d\phi}{dz} = \gamma(\nu)\phi + \xi_{сп}(\nu). \tag{14.68}$$

Уравнение (14.68) включает плотность потока фотонов как усиливаемого сигнала, так и шумового спонтанного излучения.

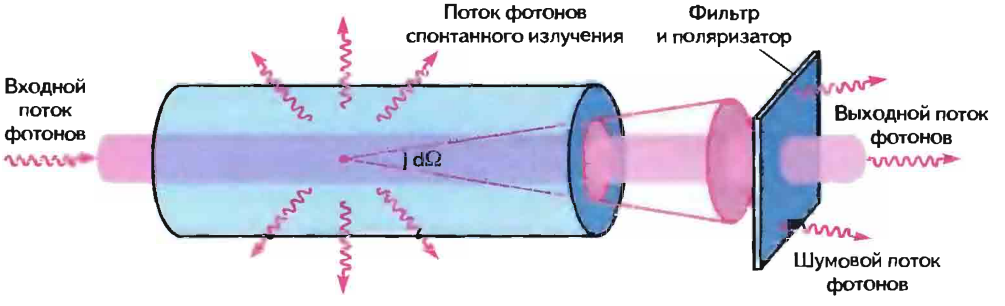


Рис. 14.27. Спонтанное излучение является источником шума усилителя. Оно широкополосное, испускается во всех направлениях и неполяризованно. С помощью оптики на выходе усилителя можно ограничить спонтанный шум узкой спектральной полосой, телесным $d\Omega$ углом и одной поляризацией

Упражнение 14.7

Усиленное спонтанное излучение (УСИ)

а. Пользуясь (14.68), покажите, что в отсутствие входного сигнала спонтанное излучение вызывает появление на выходе ненасыщенного усилителя [$\gamma(\nu) \approx \gamma_0(\nu)$] длиной d фотонов, плотность потока которых выражается как $\phi(d) = \phi_{сп}\{\exp[\gamma_0(\nu)d] - 1\}$, где $\phi_{сп} = \xi_{сп}(\nu)/\gamma_0(\nu)$.

б. Поскольку и $\xi_{сп}(\nu)$, и $\gamma_0(\nu)$ пропорциональны $g(\nu)$, $\phi_{сп}$ не зависит от $g(\nu)$, так что частотная зависимость $\phi(d)$ определяется множителем $\{\exp[\gamma_0(\nu)d] - 1\}$. Если $\gamma_0(\nu)$ имеет лоренцев контур с шириной $\Delta\nu$, т. е.

$$\gamma_0(\nu) = \gamma_0(\nu_0)\frac{(\Delta\nu/2)^2}{(\nu - \nu_0)^2 + (\Delta\nu/2)^2},$$

покажите, что спектральная ширина множителя $\{\exp[\gamma_0(\nu)d] - 1\}$ меньше, чем $\Delta\nu$, т. е. усиление спонтанного излучения сопровождается сужением его спектра.

В процессе усиления статистика числа фотонов (см. подразд. 12.2.3) входящего света меняется. Когерентный сигнал, подаваемый на вход усилителя, характеризуется пуассоновской статистикой числа фотонов с дисперсией σ_s^2, равной среднему числу сигнальных фотонов $\bar{n}_s$. Фотоны УСИ, напротив, характеризуются статистикой Бозе—Эйнштейна с $\sigma_{\text{УСИ}}^2 = \bar{n}_{\text{УСИ}} + \bar{n}_{\text{УСИ}}^2$ и, следовательно, обладают значительно большим шумом по сравнению с пуассоновскими. Статистика числа фотонов света после усиления, содержащего вклады как усиленного сигнала, так и УСИ, описывается промежуточным распределением между двумя упомянутыми. Если время счета мало, а выходящий свет линейно поляризован, то эта статистика в хорошем приближении может быть описана **полиномиальным распределением Лагерра** (см. задачу 2 к разд. 14.5), дисперсия которого

$$\sigma_n^2 = \bar{n}_s + \left(\bar{n}_{\text{УСИ}} + \bar{n}_{\text{УСИ}}^2\right) + 2\bar{n}_s\bar{n}_{\text{УСИ}}. \tag{14.69}$$

Эти флуктуации числа фотонов содержат отдельные вклады от сигнала и спонтанного излучения, а также перекрестный член.

Рекомендуемая литература

КНИГИ

См. также литературу к гл. 15.

Headley C., Agrawal G.P., eds. Raman Amplification in Fiber Optical Communication Systems. Elsevier, 2005.

Islam M.N., ed. Raman Amplifiers for Telecommunications. Vol. 1: Physical Principles. Springer-Verlag, 2004.

Islam M.N., ed. Raman Amplifiers for Telecommunications. Vol. 2: Sub-Systems and Systems. Springer-Verlag, 2004.

Desurvire E., Bayart D., Desthieux B., Bigo S. Erbium-Doped Fiber Amplifiers: Device and System Developments. Wiley, 2002.

Digonnet M.J.F., ed. Rare-Earth-Doped Fiber Lasers and Amplifiers. Marcel Dekker, 2nd ed. 2001.

Becker P.C., Olsson N.A., Simpson J.R. Erbium-Doped Fiber Amplifiers: Fundamentals and Technology. Academic Press, 1999.

Sudo S., Ohishi Y., Fujiura K., Kanamori T., Yamada M., Shimizu M. Optical Fiber Amplifiers: Materials, Devices, and Application Technologies. Artech, 1997.

Desurvire E. Erbium-Doped Fiber Amplifiers: Principles and Applications. Wiley, 1994.

Shimada S., Ishio H., eds. Optical Amplifiers and Their Applications. Wiley, 1994.

Bjarklev A. Optical Fiber Amplifiers: Design and System Applications. Artech, 1993.

СТАТЬИ

Bibeau C., Rhodes M.A., Atherton L.J. Innovative Technology Enables a New Architecture for the World's Largest Laser, Photonics Spectra. Vol. 40, № 6, 2006. P. 50—60.

Issue on fiber amplifiers and lasers. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 7, № 1, 2001.

Millennium issue. IEEE Journal of Selected Topics in Quantum Electronics. Vol. 6, № 6, 2000.

Campbell J.H., Suratwala T.I. Nd-Doped Phosphate Glasses for High-Energy/High-Peak-Power Lasers. Journal of Non-Crystalline Solids. Vol. 263/264, 2000. P. 318—341.

Li T., Teich M.C. Photon Point Process for Traveling-Wave Laser Amplifiers. IEEE Journal of Quantum Electronics. Vol. 29, 1993. P. 2568–2578.

Digonnet M.J., ed. Selected Papers on Rare Earth-Doped Fiber Laser Sources and Amplifiers. SPIE Optical Engineering Press (Milestone Series. Vol. 37), 1992.

Задачи

К РАЗДЕЛУ 14.1

1. **Усиление и длина стержня усилителя.** Промышленный лазерный усилитель на рубине с длиной стержня 15 см имеет усиление 12 при слабом сигнале. Каково усиление слабого сигнала при длине стержня 20 см? Эффектом насыщения пренебречь.

2. **Усиление и разность заселенностей.** Стержень из неодимового стекла длиной 15 см при использовании в качестве усилителя обеспечивает полное усиление 10 слабого сигнала на $\lambda_0 = 10{,}6$ мкм. С помощью данных табл. 14.1 определите разность заселенностей N (число ионов Nd^{3+} на $см^3$), необходимую для достижения такого усиления.

3. **Усиление широкополосного сигнала.** Переход между двумя энергетическими уровнями характеризуется лоренцевой формой линии с центральной частотой $\nu_0 = 5 \cdot 10^{14}$ Гц и шириной $\Delta\nu = 1$ ТГц. Инверсия заселенностей обеспечивает максимальный коэффициент усиления $\gamma(\nu_0) = 0{,}1$ $см^{-1}$. Среда имеет дополнительный источник потерь $\alpha_s = 0{,}05$ $см^{-1}$, не зависящий от ν. Оценить потери или усиление, приобретаемые светом на длине 1 см, если он имеет равномерное спектральное распределение мощности с центральной частотой ν_0 и шириной $2\Delta\nu$.

К РАЗДЕЛУ 14.2

1. **Двухуровневая схема накачки.** Запишите скоростные уравнения для двухуровневой системы и покажите, что стационарная инверсия заселенностей не может быть получена прямой оптической накачкой перехода $1 \rightarrow 2$.

2. **Две лазерные линии.** Рассмотрим атомную систему с четырьмя уровнями 0 (основное состояние) 1, 2 и 3. Производится накачка двух переходов: между основным состоянием и уровнем 3 со скоростью R_3 и между основным состоянием и уровнем 2 со скоростью R_2. Возможна инверсия заселенностей между уровнями 3 и 1 и/или между уровнями 2 и 1 (как в четырехуровневом лазере). Предполагая, что распад за счет переходов с уровня 3 на уровень 2 невозможен, а спонтанными переходами с уровней 3 и 2 в основное состояние можно пренебречь, запишите скоростные уравнения для заселенностей уровней 1, 2 и 3 с временами жизни τ_1, τ_{31} и τ_{21}. Определите стационарные заселенности N_1, N_2 и N_3 и исследуйте возможность одновременной инверсии для переходов $3 \rightarrow 1$ и $2 \rightarrow 1$. Покажите, что присутствие излучения в канале $2 \rightarrow 1$ уменьшает разность заселенностей в канале $3 \rightarrow 1$.

К РАЗДЕЛУ 14.4

1. **Значение насыщающей плотности потока фотонов.** В общей двухуровневой схеме атома (см. рис. 14.8) τ_2 представляет собой время жизни уровня 2 в отсут-

ствие вынужденного испускания. При наличии вынужденного испускания скорость распада уровня 2 увеличивается, а эффективное время жизни уменьшается. Найдите плотность потока фотонов ϕ, при которой время жизни снижается в два раза. Как эта величина связана с насыщающей плотностью потока фотонов ϕ_s?

2. **Интенсивность насыщения.** Определите насыщающую плотность потока фотонов $\phi_s(\nu_0)$ и соответствующую интенсивность насыщения $I_s(\nu_0)$ для однородно-уширенных лазерных переходов в рубине и $Nd^{3+}:YAG$, приведенных в табл. 14.1.

3. **Рост плотности потока фотонов в усилителе с насыщением.** Рост плотности потока фотонов $\phi(z)$ в лазерном усилителе описывается уравнением (14.52). Постройте график зависимости $\phi(z)/\phi_s$ от $\gamma_0 z$ для $\phi(0)/\phi_s = 0{,}05$. Где начинает проявляться насыщение в этом усилителе?

4. **Резонансное поглощение в среде при тепловом равновесии.** Среда объемом 1 см^3 с показателем преломления 1 содержит $N_a = 10^{23}$ атомов в состоянии теплового равновесия. Основное состояние — уровень 1, уровень 2 — лежит на 2,48 эВ выше ($\lambda_0 = 0{,}5$ мкм). Переход между этими двумя уровнями характеризуется спонтанным временем жизни $t_{сп} = 1$ мс и лоренцевой формой линии с шириной $\Delta\nu = 1$ ГГц. Рассмотрим два значения температуры T_1 и T_2, таких что $k_B T_1 = 0{,}026$ эВ и $k_B T_2 = 0{,}26$ эВ.

а. Определите заселенности N_1 и N_2.

б. Определите число фотонов, спонтанно испускаемых каждую секунду.

в. Определите коэффициент поглощения в этой среде на $\lambda_0 = 0{,}5$ мкм, считая падающий поток фотонов слабым.

г. Нарисуйте график зависимости коэффициента поглощения от частоты и отметьте на нем важные параметры.

д. Найдите значение плотности потока фотонов, при котором коэффициент поглощения снижается в 2 раза (т. е. насыщающую плотность потока фотонов).

е. Нарисуйте график зависимости прошедшей плотности потока фотонов $\phi(d)$ от падающей $\phi(0)$ для $\nu = \nu_0$ и $\nu = \nu_0 + \Delta\nu$, если $\phi(0)/\phi_s \ll 1$.

5. **Усиление в насыщаемой усиливающей среде.** Рассмотрим лазерную усиливающую среду с однородным уширением длиной $d = 10$ см с насыщающей плотностью потока фотонов $\phi_s = 4 \cdot 10^{18}$ фотонов/(см$^2 \cdot$ с). Известно, что если на входе плотность потока фотонов $\phi(0) = 4 \cdot 10^{15}$ фотонов/(см$^2 \cdot$ с), то на выходе она составляет $\phi(d) = 4 \cdot 10^{16}$ фотонов/(см$^2 \cdot$ с).

а. Определите величину усиления слабого сигнала G_0 для этого усилителя.

б. Определите коэффициент усиления слабого сигнала γ_0.

в. Какова плотность потока фотонов, при которой коэффициент усиления снижается в пять раз?

г. Определите коэффициент усиления для входной плотности потока фотонов $\phi(0) = 4 \cdot 10^{19}$ фотонов/(см$^2 \cdot$ с). Больше, меньше или равно будет полное усиление G при этих условиях в сравнении с усилением слабого сигнала G_0 в пункте а?

ЗАДАЧИ К РАЗД. 14.5

1. **Отношение мощностей сигнала и шума УСИ.** Ненасыщенный усилитель длиной d с коэффициентом усиления $\gamma_0(\nu)$ усиливает входной сигнал $\phi_s(0)$ частоты ν и вносит шум за счет усиленного спонтанного излучения (УСИ) со скоростью $\xi_{сп}$ (на единицу длины). Плотность потока фотонов усиленного сигнала обозначим $\phi_s(d)$, а УСИ на выходе — $\phi_{УСИ}$. Изобразите графически зависимость отношения $\phi_s(d)/\phi_{УСИ}$ от произведения коэффициента усиления на длину усилителя $\gamma_0(\nu)d$.

2. **Распределение числа фотонов когерентного света после усиления.** Линейно поляризованная суперпозиция интерферирующих теплового и когерентного света служит подходящей моделью света, выходящего из лазерного усилителя. Известно, что такая суперпозиция характеризуется случайными флуктуациями энергии w, которые подчиняются нецентральному хи-квадрат-распределению вероятностей

$$p(w) = \frac{1}{\bar{w}_{УСИ}} \exp\left(-\frac{w + w_s}{\bar{w}_{УСИ}}\right) I_0\left[\frac{2\sqrt{w_s w}}{\bar{w}_{УСИ}}\right],$$

при условии, что время измерения достаточно мало[1]. Здесь I_0 обозначает модифицированную функцию Бесселя нулевого порядка; $\bar{w}_{УСИ}$ — средняя энергия УСИ; w_s — постоянная энергия усиливаемого когерентного сигнала.

а. Вычислите среднее значение и дисперсию w.

б. С помощью (12.49) и (12.50) определите среднее число фотонов $\bar{n}$ и дисперсию σ_n^2 и подтвердите справедливость (14.69).

в. С помощью (12.48) покажите, что распределение числа фотонов имеет вид

$$p(n) = \frac{\bar{n}_{УСИ}^n}{(1 + \bar{n}_{УСИ})^{n+1}} \exp\left(-\frac{\bar{n}_s}{1 + \bar{n}_{УСИ}}\right) L_n\left[\frac{\bar{n}_s/\bar{n}_{УСИ}}{1 + \bar{n}_{УСИ}}\right],$$

где L_n — полином Лагерра

$$L_n(-x) = \sum_{k=0}^{n} \binom{n}{k} \frac{x^k}{k!};$$

$\bar{n}_s$ и $\bar{n}_{УСИ}$ — средние числа фотонов сигнала и УСИ, соответственно.

г. Постройте график $p(n)$ для $\bar{n}_s/\bar{n} = 0$; 0,5; 0,8 и 1 при $\bar{n} = 5$ и покажите, что он сводится к распределению Бозе—Эйнштейна при $\bar{n}_s/\bar{n} = 0$ и к распределению Пуассона при $\bar{n}_s/\bar{n} = 1$.

---

[1] См., например, *Saleh B.E.A.* Photoelectron Statistics. Springer-Verlag, 1978.

Учебное издание

Заявки на книги присылайте по адресам:
zakaz@id-intellect.ru
solo@id-intellect.ru
id-intellect@mail.ru

тел. (495) 579-96-45
факс (495) 617-41-88

В заявке обязательно указывайте
свои реквизиты (для организаций) и почтовый адрес!

Подробная информация о книгах на сайте
http://www.id-intellect.ru

Бахаа Е.А. Салех, Малвин Карл Тейх

ОПТИКА И ФОТОНИКА. ПРИНЦИПЫ И ПРИМЕНЕНИЯ

Том 1

Компьютерная верстка – Н.А. Попова
Корректор – Г.Н. Петрова
Художник – С.Ю. Биричев
Ответственный за выпуск – Л.Ф. Соловейчик

Формат 70x100/16. Печать офсетная.
Гарнитура Ньютон.
Печ. 47,5 л. Зак. М-1029.
Бумага офсетная №1, плотность 80 г/м²

Издательский Дом «Интеллект»
141700, Московская обл., г. Долгопрудный,
Промышленный пр-д, д. 14,
тел. (495) 617-41-85

Отпечатано в типографии филиала ОАО «ТАТМЕДИА»
«ПИК «Идел Пресс».
420066, г. Казань, ул. Декабристов, д. 2

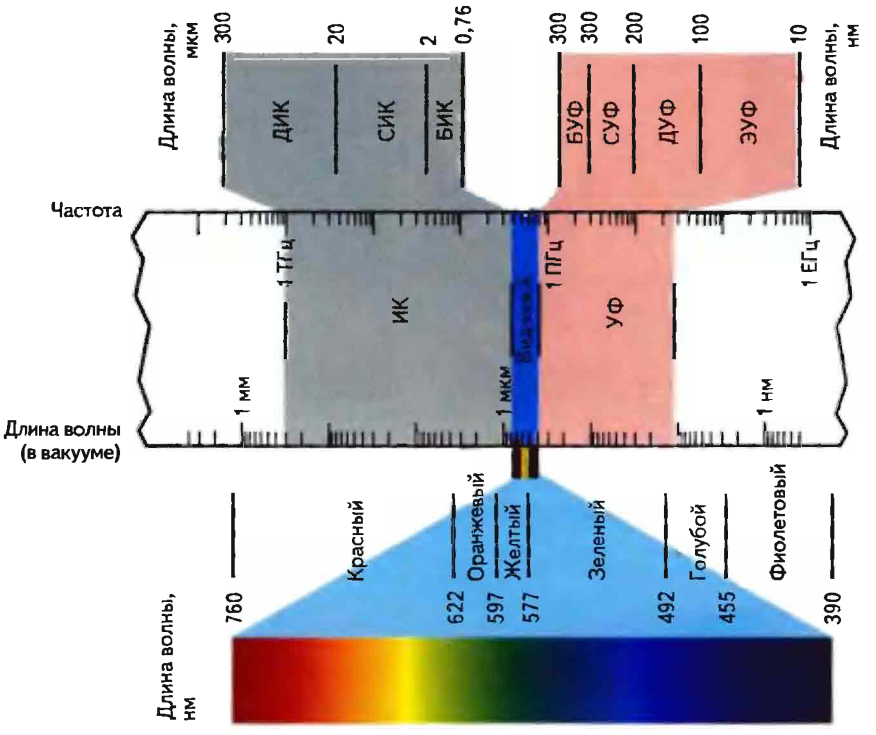


Рис. 2.1. Оптические частоты и длины волн. Инфракрасный (**ИК**) диапазон спектра включает ближний инфракрасный (**БИК**), средний инфракрасный (**СИК**) и дальний инфракрасный (**ДИК**) поддиапазоны. Ультрафиолетовый (**УФ**) диапазон включает ближний ультрафиолетовый (**БУФ**), средний ультрафиолетовый (**СУФ**), дальний ультрафиолетовый (**ДУФ**) и экстремальный ультрафиолетовый (**ЭУФ**) поддиапазоны. Экстремальное ультрафиолетовое излучение известно также как мягкие рентгеновские (**МР**) лучи. Вакуумное ультрафиолетовое излучение (**ВУФ**) включает ДУФ- и ЭУФ-поддиапазоны. Инфракрасный, видимый и ультрафиолетовый диапазоны называются оптическими, поскольку в них используются одинаковые элементы, такие как линзы и зеркала

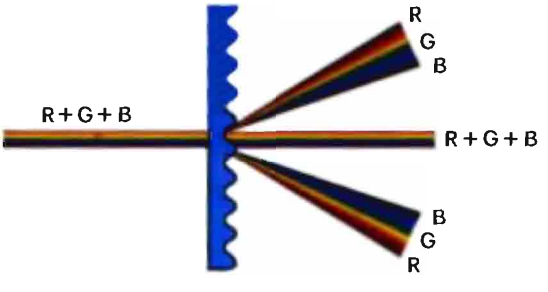


Рис. 2.20. Дифракционная решетка направляет две волны с различными длинами волн λ_1 и λ_2 под двумя различными углами θ_1 и θ_2. Следовательно, она действует как анализатор спектра, или спектрометр

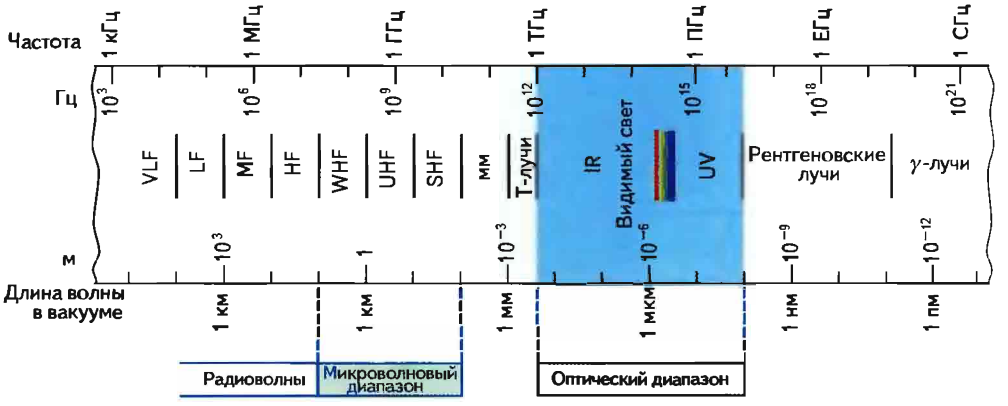


Рис. 5.1. Спектр электромагнитных излучений от низких частот (длинных волн) для высоких частот (коротких волн). Оптический диапазон, выделенный штриховкой, показан более подробно на рис. 2.1

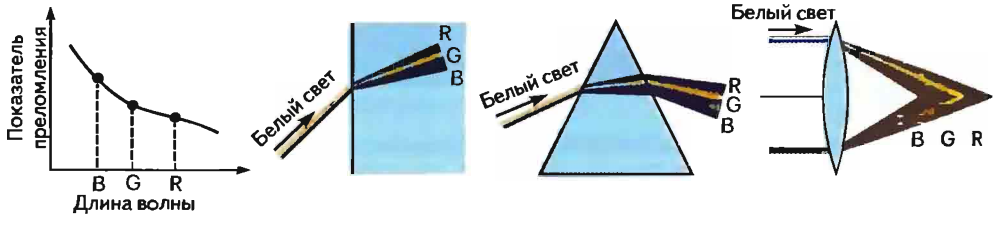


Рис. 5.15. Оптические элементы, изготовленные из диспергирующего материала, преломляют волны с различной длиной волны на различные углы (B — синий, G — зеленый, R — красный цвета)

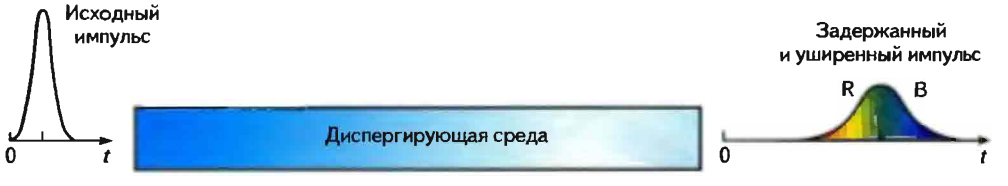


Рис. 5.16. Диспергирующая среда уширяет импульс, так как различные частотные компоненты, составляющие его, распространяются с различными скоростями. На данной иллюстрации низкочастотные компоненты (с большей длиной волны, обозначенные R) распространяются быстрее, чем высокочастотные (с меньшей длиной волны, обозначенные B), и поэтому достигают выхода раньше

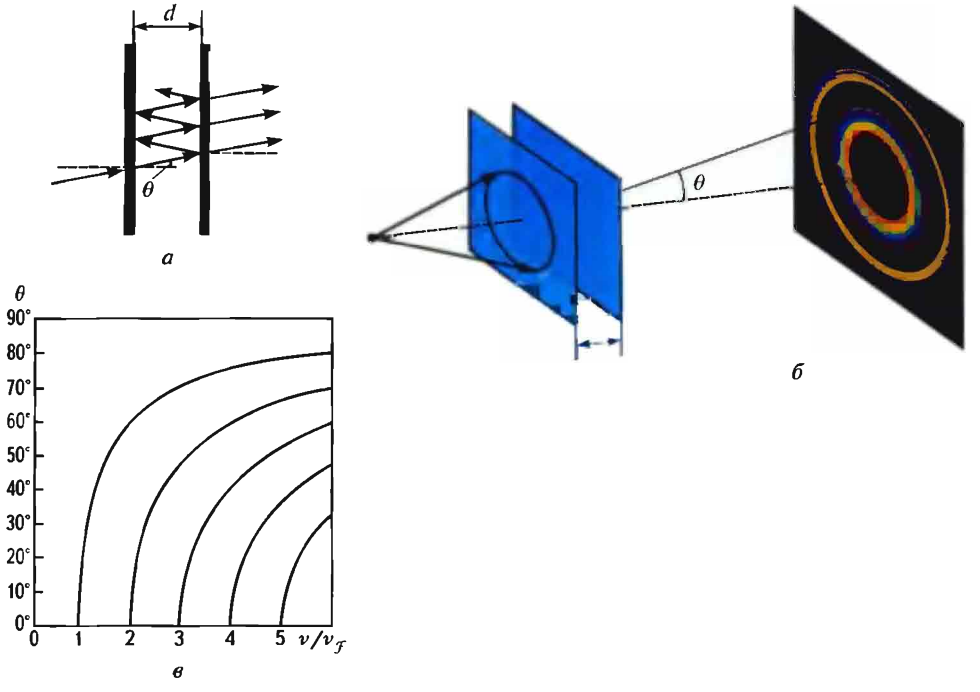


Рис. 7.8. Внеосевая волна, проходящая сквозь зеркальный эталон Фабри–Перо (*а*). Проходя через эталон, белый свет от точечного источника порождает набор концентрических колец различных частот (цветов) (*б*). Частоты и углы, удовлетворяющие условию максимального пропускания (7.43) (*в*)

Рис. 11.6. Зависимость спектральной плотности от длины волны в трех точках цветного изображения (Анри Матисс, «Георгины»)

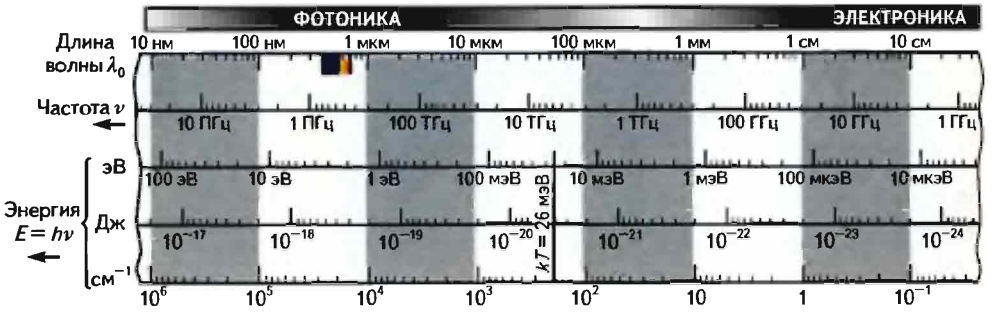


Рис. 12.3. Связь между длиной волны фотона λ_0, частотой ν и энергией E (выраженной в эВ, Дж и см$^{-1}$). Фотон с длиной волны в свободном пространстве $\lambda_0 = 1$ мкм имеет частоту $\nu = 300$ ТГц и энергию $E = 1{,}24$ эВ $= 1{,}99 \cdot 10^{-19}$ Дж $= 10^4$ см$^{-1}$. Отмечены области, относящиеся к фотонике и электронике

Рис. 13.12. Фотолюминесценция коллоидных квантовых точек CdSe (с молекулами олеиламина, покрывающими поверхность), взвешенных в *n*-гексане, под действием ультрафиолетового возбуждения на $\lambda_0 = 365$ нм. Зависимость энергии от размеров квантовой ямы позволяет менять цвет свечения, меняя размер частиц (предоставлено Донг-Кьюн Сео, Государственный университет штата Аризона)

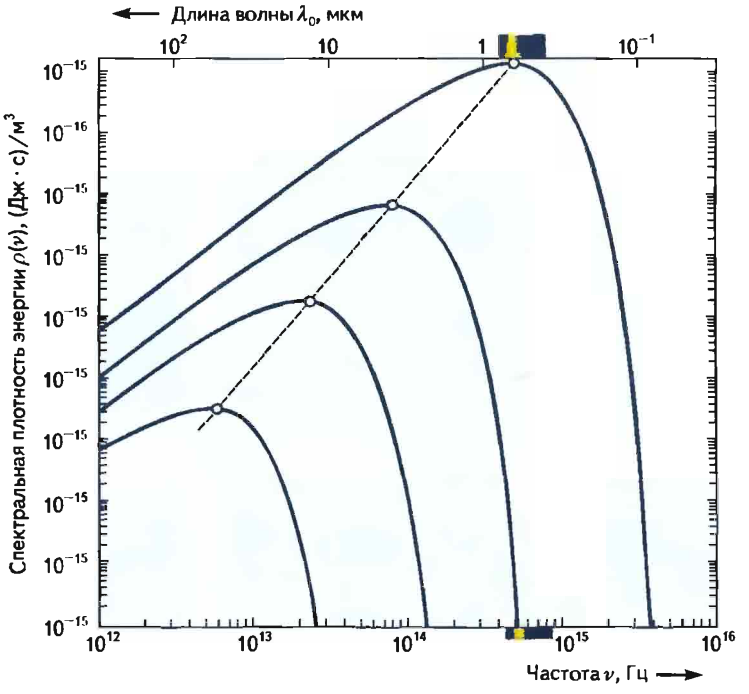


Рис. 13.31. Зависимость спектральной плотности энергии $\rho(\nu)$ от частоты, построенная в дважды логарифмических координатах для семи различных температур

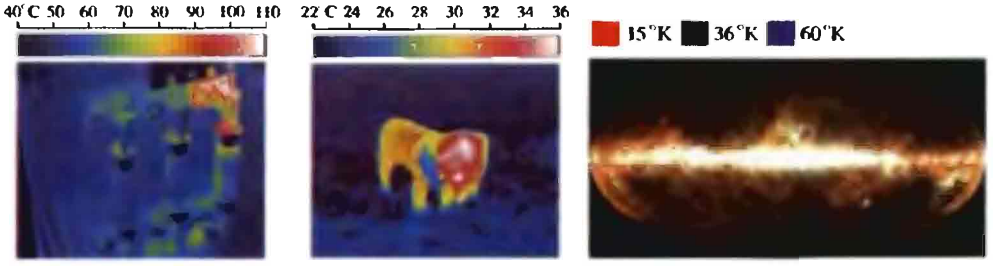


Рис. 13.32. Примеры термографических изображений в различных температурных областях: *а* — анализ промышленных систем; *б* — поисково-спасательные работы; *в* — космология

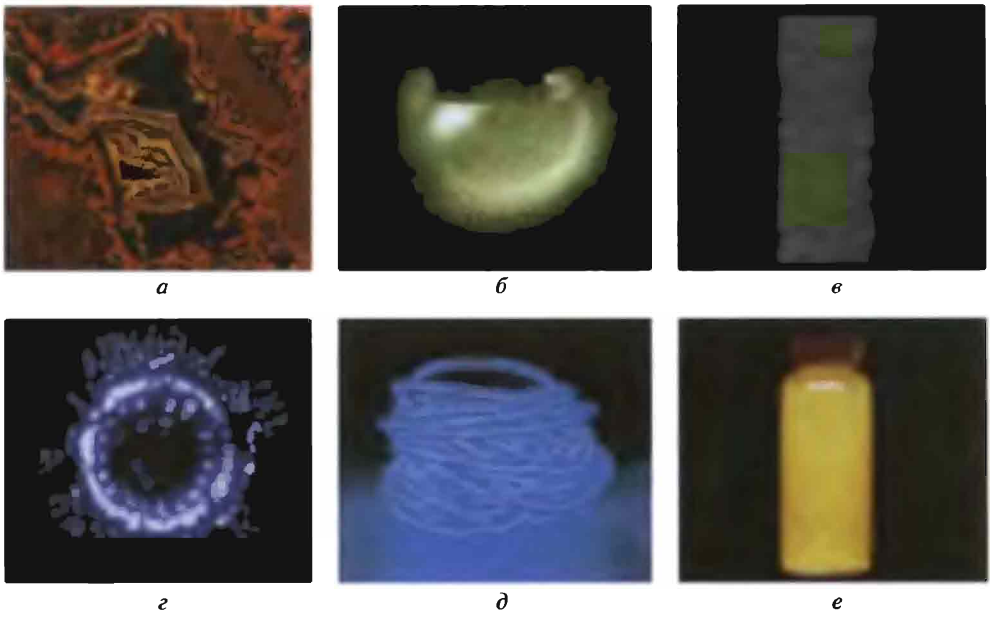


Рис. 13.33. Катодолюминесценция минерального образца позволяет выявить зоны содержания кальцита и доломита в ячеистой брекчии (*а*). Размер края 1,3 мм, энергия электронов 22 кэВ (с разрешения Чарльза Онасча, Государственный университет Боулинг Грин). Многопузырьковая сонолюминесценция, полученная при погружении источника ультразвука в жидкость (*б*) (с разрешения К.С. Суслика, Университет штата Иллинойс в Урбана-Шампейн). Хемилюминесценция световой палочки (*в*). Глубоководная чашеобразная медуза *Atolla vanhoeffeni* (диаметр около 3 см) распространена во всем мире, ее биолюминесценция является ответом на раздражения (*г*) (с разрешения Е.А. Уиддер, Ассоциация исследований и сохранения океана). Электрическое поле, приложенное поперек пары параллельных проводов при различных потенциалах, вызывает электролюминесценцию покрывающего их порошкового материала (*д*). Фотолюминесценция коллоидных квантовых точек CdSe в гексане после освещения ультрафиолетовым светом (*е*) (см. рис. 13.12) (с разрешения Д.-К. Сео, Государственный университет штата Аризона)

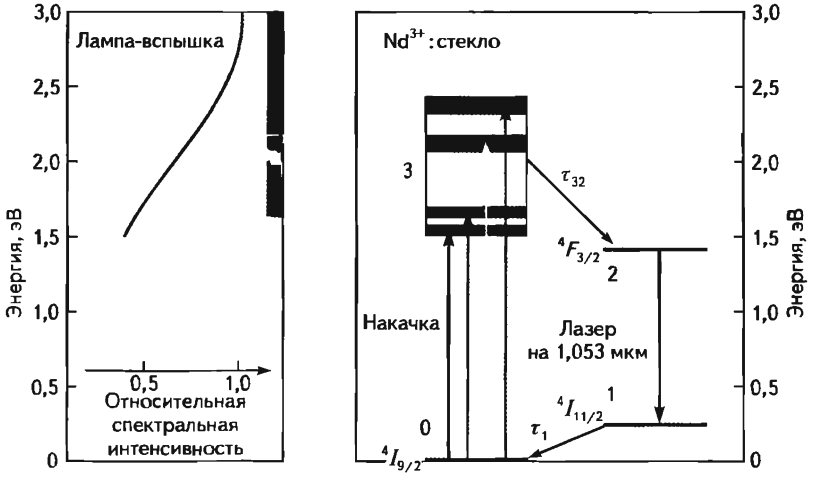


Рис. 14.16. *Слева*: спектральный профиль широкополосного излучения ксеноновой лампы-вспышки, используемой для накачки усилителей на стекле с неодимом в Национальном комплексе зажигания (NIF) *Справа*: схема накачки лазерного перехода ${}^4F_{3/2} \rightarrow {}^4I_{11/2}$ на длине волны $\lambda_0 = 1{,}053$ мкм в фосфатном стекле с неодимом Schott LG-770

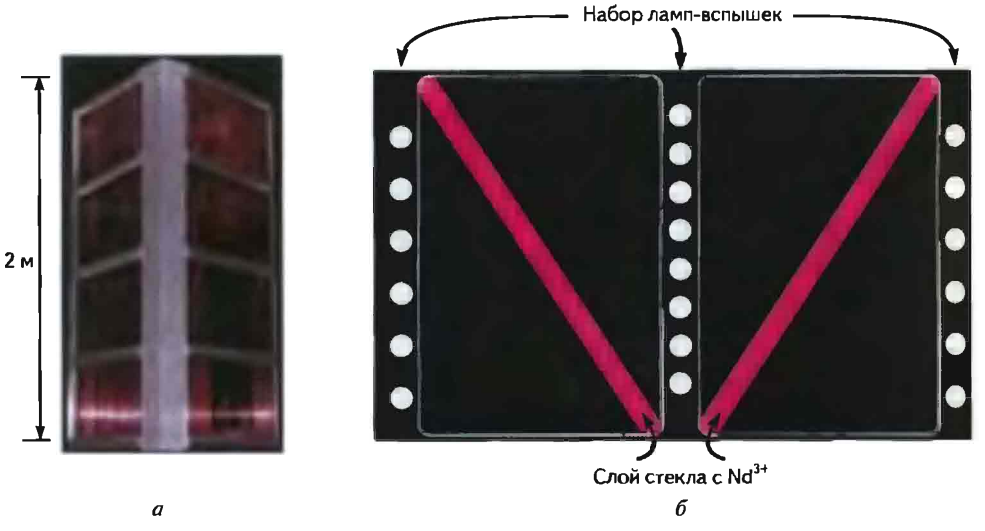


Рис. 14.17. Стопа усилителей в Национальном комплексе зажигания Ливерморской национальной лаборатории им. Лоуренса включает восемь пластин из лазерного стекла, размещенных в полости, засвечиваемой набором ламп-вспышек (*а*). Каждая пластина изготовлена из специального разработанного фосфатного лазерного стекла с уровнем легирования неодимом 2 моль % (Schott LG-770 или Hoya LHG-8) и имеет размеры 46 × 81 × 4 см. Высота стопы из восьми усилителей — примерно 2 м. Шесть таких стоп в NIF образуют кластер, а четыре кластера включают 192 отдельные пучковые линии. Каждая линия, в свою очередь, состоит из 16 этих пластины. Вид сверху на лампы-вспышки и усиливающие пластины из неодимового стекла в стопе (*б*)